An Empirical, Quantitative Analysis
of the Differences Between Sarcasm and Irony

Jennifer Ling and Roman Klinger[(✉)]

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart,
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
{jennifer.ling,roman.klinger}@ims.uni-stuttgart.de

Abstract. A variety of classification approaches for the detection of ironic or sarcastic messages has been proposed in the last decade to improve sentiment classification. However, despite the availability of psychologically and linguistically motivated theories regarding the difference between irony and sarcasm, these typically do not carry over to a use in predictive models; one reason might be that these concepts are often considered very similar. In this paper, we contribute an empirical analysis of Tweets and how authors label them as irony or sarcasm. We use this distantly labeled corpus to estimate a model to distinguish between39 both classes of figurative language with the aim to, ultimately, improve the semantically correct interpretation of opinionated statements. Our model separates irony from sarcasm with 79 % accuracy on a balanced set. This result suggests that the task is harder than separating irony or sarcasm from regular texts with 89 % and 90 % accuracy, respectively. A feature analysis shows that ironic Tweets have on average a lower number of sentences than sarcastic Tweets. Sarcastic Tweets contain more positive words than ironic Tweets. Sarcastic Tweets are more often messages to a specific recipient than ironic Tweets. The analysis of bag-of-words features suggests that the comparably high classification performance to distinguish irony from sarcasm is supported by specific, reoccurring topics.

1 Introduction

Irony and sarcasm are rhetoric devices present in everyday life. With the advent of social media platforms, the interest increases to differentiate ironic and sarcastic textual contributions from regular ones. This is particularly necessary to correctly link opinions to semantic concepts. For instance the review fragment [1]

> "i would recomend this book to friends who have insomnia or those who i absolutely despise."

contains the positive phrase "recommend this book". However, the irony markers "have insomnia" and "who i absolutely despise" suggest a non-literal meaning. The detection of these semantic differences between ironic, sarcastic and regular language is a prerequisite for a subsequent aggregation of harvested information.

Many previous approaches to detect irony and sarcasm assume that both concepts are sufficiently similar to not make an explicit differentiation when distinguishing them from regular utterances. In this paper, we challenge this assumption: We do not focus on the difference between figurative (*e.g.*, ironic and sarcastic, amongst other) and non-figurative utterances but on an empirical evaluation and analysis of the *difference between sarcasm and irony.* We develop a data-driven model to distinguish sarcasm and irony based on authors' labels of their own Tweets. We assume that such an approach can potentially help to improve polarity detection in sentiment analysis settings [2], as sarcasm is often considered a more drastic form of irony which is used to attack something or someone [3]. We provide a data-driven interpretation of the difference between sarcasm and irony based on the author's point of view. Similarly to previous work in sentiment classification on Twitter [4], we use information attached to each Tweet to distantly label it. Such approach might lead to noisy labels, but this is not an issue in our setting: We are interested in what authors consider to be ironic and sarcastic, though these users might not be aware of formal definitions of these concepts.

Therefore, our main contributions are:

(1) We retrieve and publish a corpus of 99000 English Tweets, 33000 of which contain the hashtag #irony or #ironic and 33000 contain #sarcasm or #sarcastic. The remaining ones contain none of those.
(2) We compare typical features used for sentiment analysis and figurative language detection in a classification setting to differentiate between irony and sarcasm.
(3) We perform a feature analysis with the aim to discover the latent structure of ironic and sarcastic Tweets.

The remainder of the paper is organized as follows: We briefly review previous work in Sect. 2. In Sect. 3, we explain our corpus collection and introduce the features and the classifiers. In Sect. 4, we discuss our findings and conclude in Sect. 5 with a summary and present the key results of our experiments.

2   Background and Related Work

Irony and sarcasm are important devices in communication that are used to convey an attitude or evaluation towards the content of a message [5]. Between the age of six and eight years, children gain the ability to recognize ironic utterances [6–8]. The principle of inferability [9] states that figurative language is used if the speaker is confident that the addressee will interpret the utterance and infer the communicative intention of the speaker/author correctly. Irony is ubiquitous, with up to 8 % of utterances exchanged being ironic [10].

Utsumi proposed that an ironic utterance can only occur in an *ironic environment* [11]. The theories of the *echoic account* [12], the *pretense theory* [13] or the *allusional pretense theory* [14] have challenged the understanding that an

ironic utterance typically conveys the opposite of its literal propositional content. However, in spite of the fact that the attributive nature of irony is widely accepted (see [15]), no formal definition of irony is available as of today which is instantiated across operational systems and empirical evaluations.

Commonly, the difference between irony and sarcasm is not made explicit (for instance in the context of an ironic environment, which holds for sarcasm analogously [16]). [17] state that irony and sarcasm are often used interchangeably, other publications mention a high similarity between sarcasm, satire, and irony [18]. Sarcasm is often considered a specific case of irony [19], often more negative and "biting" [20,21]. In other approaches, it is defined as a synonym to verbal irony [9,22].

Systems for automatic recognition of irony and sarcasm typically focus on one of both classes (irony *or* sarcasm) or do not explicitly encode the difference (combine irony *and* sarcasm) [23–25].

There is only little work on automated systems which aim at learning the difference between sarcasm and irony, neither of data-driven analyzes. One approach we are aware of assumes in a corpus-based analysis that the hashtag #irony refers to situational irony and #sarcasm to the intended opposite of the literal meaning [2]. In a corpus of 257 Tweets with the hashtag #irony, only 2 refer to verbal irony. About 25 % of this corpus involve clear situational irony. However, this work focused on the impact of irony detection on opinion mining. Similarly, one goal of the SemEval 2015 Task 11 [26] was to evaluate sentiment analysis systems separately on sarcastic and ironic subcorpora. It has been shown that figurative language-specific methods improve the result [27].

Wang [28] focused on the analysis of irony vs. sarcasm – to our knowledge the first work with this goal. Her approach applies sentiment analysis and performs a manual, qualitative sub-corpus analysis. She finds that irony is used in two senses: One which is equivalent to the use of sarcasm and intends to attack something or someone, and one which is to describe an event; therefore refers to situational irony. However, in contrast to our work, she does not perform an automatic classification of irony and sarcasm, nor a detailed feature analysis. Our work, in contrast to her work, focuses more on quantitative aspects than qualitative aspects.

Very recently, Sulis et al. [29] analyzed the differences between the hashtags #irony, #sarcasm, and #not. Their approach is similar to our experiments. However, the feature sets described in their work and in our work complement each other.

3   Methods

3.1   Corpus

Our assumption is that users on Twitter annotate their messages ("Tweets") to be ironic or sarcastic using the respective hashtags. These hashtags are therefore the *irony markers*. It cannot necessarily be expected that additional irony markers exist. Therefore, our distant supervision assumption is different from

the findings and conclusions in which the labels should approximate a ground truth. On the contrary, our research goal is to understand the specific use of the difference of the authors' labels and not to generalize to Tweets without such labels.

The basis for our analysis of irony and sarcasm is a corpus of 99 000 messages crawled between July and September 2015. The corpus consists of 33 000 Tweets (30 000 training and 3 000 test data) from each of the categories *irony*, *sarcasm*, and *regular*, selected with respective hashtags #irony/#ironic, #sarcasm/#sarcastic. Frequent other hashtags in these subcorpora are #drugs, #gopdebate, #late, #news and #peace which we use as search terms to compile the corpus of regular Tweets together with #education, #humor and #politics, which were used in previous work (*cf.* [23,24]).

Each Tweet is preprocessed as follows: We store the post date, the author name, the Tweet ID, the text and, if available, the geolocation (available in 1241 cases) and the provided location (5860 cases). Retweets (marked with "RT"), Tweets shorter than five tokens and near-duplicates (based on token-based Jaccard similarity with a threshold of 0.8) are discarded [30]. Tweets are assigned to the ironic subcorpus if they contain #irony/#ironic but not #sarcasm/#sarcastic and vice versa. If both classes of figurative language are present, the Tweet is discarded (0.24 % of the Tweets with at least one label also contain the other). The text is tokenized with a domain-specific approach [31] (with small adaptations to keep UTF8 emoticons, acronyms and URLs each in one token) and part-of-speech-tagged with the GATE Twitter PoS Tagger [32]. The corpus is available at http://www.romanklinger.de/ironysarcasm.

3.2   Features

In line with previous work with the aim to distinguish between figurative language and regular language [24,33,34], we employ a set of features to measure characteristics of each text, which we list and explain briefly in the following.

Bag-of-Words/Unigrams, Bigrams. All words (unigrams) and the 50 000 most frequent bigrams are used as features individually. We refer to this feature as *BoW*.

Figurative Language. Each of the following features is instantiated in four versions, the count ($c \in \mathbb{N}$), Boolean (true iff $c > 0$), the count normalized by the number of tokens $|\mathbf{t}|$ in the Tweet ($r = \frac{c}{|\mathbf{t}|}$), and a series of Boolean features representing r by stacked binning in steps of 0.2.

The feature *Emoticon* counts the number of emoticons corresponding to facial expressions in the Tweet [18,24]. We use all UTF8 encoded emoticons and symbols from [35]. For instance, the Tweet "the joke's on me 🙆👍😁😜" has an *Emoticon* count of 2, the Boolean version is true, the normalized value is $\frac{2}{8}$, and stacked binning Boolean features are true for >0 and for >0.2. Analogously, we

use features for *Emoticon Positive* and *Emoticon Negative* and *Symbol* for non-facial symbols [36]. The feature *Emoticon-Sequence* is analogously defined as the length of the longest sequence of consecutive emoticons (the Boolean version is true if count is greater 1).

The feature *Capitalization* counts the number of capitalized words per Tweet [23]. The feature *Interjection* measures the count of occurrences of phrases like "aha", "brrr" and "oops" [37,38]. The feature *Additional Hashtags* measures the occurrence of hashtags in addition to the label hashtags.

The following features are implemented Boolean-only: *User* holds if a Tweet contains a string starting with "@", which typically refers to an addressee [18,36]. The feature *URL* captures the occurrences of a URL. *Ellipsis and Punctuation* holds if "..." is directly followed by a punctuation symbol [23,33]. In addition, we check if *Punctuation Marks* [23,36,37], a *Laughing Acronym* (`lol`, `lawl`, `luls`, `rofl`, `roflmao`, `lmao`, `lmfao`), a *Grin* (`*grin*`, `*gg*`, `*g*`), or *Laughing Onomatopoeia* (`[mu-|ba-]haha`, `hihi`, `hehe`) occur [33].

Sentiment. The expression of a sentiment or opinion is a central method to formulate emotions and attitudes. We calculate a dictionary-based sentiment score to classify Tweets into positive or negative [39]. The feature *Sentiment Score Positive* and *Sentiment Score Negative* are summed-up scores of positive words and negative words, respectively [28]. The *Sentiment Score Gap* is the absolute difference between the scores of the most positive and the most negative word [40]. Boolean features are *Valence Shift*, which holds if polarity switches within a four word window. *Pos/Neg Hyperbole* holds if a sequence of three positive or three negative words occurs, *Pos/Neg Quotes* if up to two consecutive adjectives, adverbs or nouns in quotation marks have a positive or negative polarity, *Pos/Neg &Punctuation* if a span of up to four words with at least one positive/negative word occurs but no negative/positive word ends with at least two exclamation marks or a sequence of a question mark and an exclamation mark. *Pos/Neg &Ellipsis* is defined analogously [33].

Syntactic. In the four versions mentioned above, we measure the number of *Stopwords*, *Nouns*, *Verbs*, *Adverbs*, *Adjectives* [37], *Pronomina* ("I/we", "my/our", "me/us", "mine/ours"). The *Tweet Length* as an integer is measured in addition and, as Boolean feature only, the occurrence of *Negations* ("n't", "not") [36]. *Sentence Length Gap* measures the difference between the shortest and the longest sentence. The Boolean version of this feature holds if the sentences of the Tweet have different lengths. The Boolean feature *Repeated Word* holds if one word is occurring more than once in a window of four [37].

4   Results

We aim at answering the following questions:

(1) Is it possible to predict if a Tweet has been labeled to be ironic or sarcastic by the author (without having access to the actual label)?
(2) Which features have an impact on making this prediction?
(3) Can we get qualitative insight with this approach what users consider to be sarcasm or irony?

We apply statistical models to distinguish between these classes to answer these questions. Further, we analyze how our features are distributed in the different subcorpora and qualitatively analyze words with high (pointwise) mutual information.

4.1   Classification of Irony, Sarcasm and Regular Tweets

In the experiments for classification of Tweets, subsets of features described in Sect. 3.2 and labels based on hashtags (the feature extraction does not have access to them) are used as input for different classification methods, namely support vector machines (SVM, in the implementation of liblinear [41]) [42], decision trees [43] (in the Weka J48 implementation [44]) and maximum entropy classifiers (MaxEnt) [45]. Consistent with previous experiences in the domain [33] and throughout our meta-parameter optimization and feature selection with 10-fold cross validation on the training set, the MaxEnt model outperforms the other approaches (drop of performance in decision trees of up to 14 % points, SVM up to 2 % points). Therefore, we only report on the MaxEnt model in the following.

Table 1 shows results for four different classification tasks: Throughout all feature settings (unigram, unigram+bigram, all, all−BoW (=only domain-specific features)) the differences between the classification tasks are comparable. Distinguishing sarcasm and regular Tweets leads to the highest performance with an accuracy of 0.90, closely followed by irony vs. regular Tweets with 0.89 accuracy. Separating figurative (without making the distinction between ironic and sarcastic) from regular text is slightly harder than the separate tasks with up to 0.88 accuracy. Distinguishing irony from sarcasm leads to a lower performance when compared to the other three tasks: The accuracy is $\approx$10 % points lower with 0.79 accuracy with all features. The performance in this task with domain-specific features only (All−BoW) of 0.64 reveals that the actual words in the text have a high impact on the classification performance. However, this result might be surprisingly high, given that both classes are often considered synonymous or at least similar. To achieve a better understanding of the structure of the task and the importance of the different features, we analyze their contribution in more detail in the following.

Table 1. Accuracy of Maximum Entropy Classifier on independent test set with all features described in Sect. 3.2 in comparison to unigram and bigram bag-of-words models. The data sets are balanced.

Feature set	Irony vs. Sarcasm	Irony vs. Regular	Sarcasm vs. Regular	Figurative vs. Regular
Baseline	0.50	0.50	0.50	0.50
All−BoW	0.64	0.81	0.83	0.82
BoW Unigram	0.76	0.87	0.89	0.87
BoW Uni+Bigram	0.78	0.88	0.90	0.88
All	0.79	0.89	0.90	0.88

4.2   Feature Analysis

We discuss the features with highest impact on one of the classification tasks in the following. A complete list of counts and values is depicted in Table 2.

The most important features to distinguish figurative from regular Tweets are *URL*, *Additional Hashtags*, *Verbs*, *Stopwords*, *Adverbs*, *User* and the Boolean *Sentiment Score Gap*, *Sentence Length Gap*, and the *Tweet Length* (these features are marked with FR in Table 2). The most important features to distinguish irony from sarcasm are *Nouns*, *Sentiment Score Positive*, and *Stopwords*, *Sentence Length Gap*, *Tweet Length*, *Verbs*, *Sentiment Score Negative*, *Interjection*, *Sentiment Score Gap*, (all marked in table with IS).

Emoticons are used more frequently in regular Tweets, and more frequent in their positive versions. Sarcastic Tweets use more of them than ironic Tweets. *Laughing Onomatopoeia* are used more often for figurative Tweets. Further, sarcastic Tweets are more often positive, and ironic tweets are more often negative. This suggests that both contexts are commonly used to express the opposite of the literal meaning, as sarcasm can be considered a more negative version of irony (*cf.* Sect. 2). Ironic Tweets tend to be longer than sarcastic and regular Tweets (*Tweet Length*). One reason might be that irony often refers to descriptions of situations, which are more complex to be described than verbal irony [2,28]. *User* is used more often in figurative Tweets (which is in line with the interpretation of [28]).

4.3   Bag-Of-Words Analysis

The bag-of-words model is a very strong baseline (*cf.* Sect. 4.1), and these features have a high impact in comparison to our domain-specific feature set. In the following, we discuss a selection of words in a ranked top-50 list of pointwise mutual information for the different classes in the whole corpus of 99000 Tweets. The words can be divided into two interesting subclasses: Those which express a sentiment and those which are corpus or topic-specific (presumably because of a limited time-frame of crawling).

Table 2. Counts of features and corpus statistics for irony, sarcasm and non-figurative Tweets. All numbers refer to the number of occurrences of the Boolean versions except where denoted with ∗. Top features for Figurative/Regular discrimination are marked with FR, those for Irony/Sarcasm discrimination with IS.

	Feature	Irony	Sarcasm	Regular	
Stat.	*# Tweets*	*33000*	*33000*	*33000*	
	Average # Sentences	*1.82*	*2.06*	*1.43*	
Figurative-specific	Emoticon	1155	1453	1925	
	Emoticon Positive	670	754	1101	
	Emoticon Negative	316	547	504	
	Symbol	331	363	928	
	Emoticon-Sequence	663	796	1681	
	Capitalization	11798	12453	13324	
	Interjection	411	1152	97	IS
	Additional Hashtags	14224	14024	25939	FR
	User	11866	13138	5619	FR
	URL	8660	7704	26523	FR
	Ellipsis and Punctuation	129	165	25	
	Punctuation Marks	273	546	119	
	Laughing Acronym	622	568	1227	
	Grin	0	2	0	
	Laughing Onomatopoeia	179	194	91	
Sentiment	Sentiment Score Positive	8423	13156	6128	IS
	Sentiment Score Negative	9096	6481	4443	IS
	Sentiment Score Gap	18396	20473	10940	FR IS
	Valence Shift	1729	1905	719	
	Pos Hyperbole	7	22	5	
	Neg Hyperbole	17	11	10	
	Pos Quotes	110	93	10	
	Neg Quotes	157	61	11	
	Pos & Punctuation	85	336	66	
	Neg & Punctuation	107	186	43	
	Pos & Ellipsis	413	655	233	
	Neg & Ellipsis	427	412	225	
Syntactic	Negations	5044	5495	1905	
	Sentence Length Gap	14531	18222	9901	FR IS
	Repeated Word	7706	7948	6549	
	Stopwords*	3.27	2.69	1.74	FR IS
	Nouns*	3.78	3.03	3.91	IS
	Verbs*	2.77	2.46	1.54	FR IS
	Adverbs*	0.69	0.77	0.32	FR
	Adjectives*	1.05	1.08	0.87	
	Pronomina*	0.25	0.26	0.13	
	Tweet Length*	17.05	15.71	15.39	FR IS

High-PMI words which can be considered negatively connotated from the irony corpus are: *complains*, *claiming*, *criticizing*, *illegally*, *betrayed*, *burnt*, *aborted*, *accusing*, and *however*. Positive are *#karma* and *giggle*.

In the sarcasm subcorpus, words with a high PMI which express an opinion are *Yay*, *Duh*, *thrilled*, *goody*, *#kidding*, *#yay*, *C'mon*, *Soooo*, *AWESOME*, *Mondays*, *exciting*, *Yup*, *nicer*, *#not*, *shocked*, *woah*, and *shocker*. It is less straightforward to detect a dominating polarity in these words, but an investigation of Tweets reveal that these are used for instance in patterns like "[Statement about something negative] [Positive Word] [#sarcasm]". This suggests, similarly to the feature analysis of sentiment characteristics in Sect. 4.2, that irony and sarcasm are commonly used to express the opposite, but irony more often to express something positive with a negative literal meaning and sarcasm to express something negative with a positive literal meaning.

Examples for topic specific words with high PMI for sarcasm are #creativity (because of a popular news article about a relationship between sarcasm and creativity), #Royals (related to a sports team), @businessinsider (answering to different posts on a news aggregator site), playoff, Research, @Etsy, @LenKaspar. An interesting special case is the word "coworker", which does not seem to relate to a specific event nor can be considered to be of specific opinion. However, the respective Tweets are mostly negative.

Examples for specific topics in the irony subcorpus are *Independence* and *Labor* (both mentioning reasons for not celebrating the respective holiday), *Dismaland*, *#refugees*, *Syria*, *hypocrisy*, *extremists*. These topics are more clearly discussed in a negative context than those marked with #sarcasm.

5   Discussion and Future Work

In this paper, we have shown that a classification model can distinguish ironic and sarcastic Tweets with substantial accuracy. Without taking into account word-specific features, the performance is limited (0.64 accuracy). When words are taken into account in addition, the accuracy increases (0.79). Compared to distinguishing ironic and sarcastic Tweets from regular Tweets (0.88 accuracy), this result is comparably low. This suggests that this task is, as expected, more challenging than recognizing figurative Tweets without making this difference.

To get a better understanding of the specific task structure, we performed an analysis of the importance of features. In summary, the key results of our experiments are as follows:

Key Result 1: Lengths of Tweets. Regular Tweets have fewer sentences than ironic Tweets. However, most sentences occur in sarcastic Tweets. On the contrary to this observation, ironic Tweets contain more tokens than sarcastic Tweets, followed by regular posts (sentence counts: Sarcasm: 2.06 > Irony: 1.82 > Regular: 1.43; token counts: Regular: 15.39 < Sarcasm: 15.71 < Irony: 17.05).

One interpretation is that the irony hashtag is commonly used to describe and explain ironic situations; for instance as in *"Mum signed me up to so many*

job sites as a hint to get a job, all they do is spam my email &now I can't find an email from my job. #Irony" [46]. Whereas irony demands for many words to illustrate the situation's circumstances and convey the irony, sarcasm (as an instance of verbal irony) can be expressed using the interaction of different phrases or sentences to introduce the sarcastic meaning incrementally. An example is *"You can smile, you know. Whoa, really?? I had no fucking idea!! Please, tell me more of what I can do. I'm so interested #sarcasm"* [47].

Key Result 2: Polarity. Sarcastic Tweets use more positive words than ironic Tweets (13146 positive sarcastic Tweets vs. 8423 positive ironic Tweets). On the other side, irony is used more often with negative words (9096 vs. 6481). This is in line with the common understanding of irony to express the opposite of the literal meaning and sarcasm as a generally negative version of it. An example is *"Another wonderful day #blessed #sarcasm I love it when people are nice... #sarcasm"* [48]. This characteristic of formulating the opposite is also supported by the frequencies of negations: Whereas 5044 ironic and 5495 sarcastic Tweets contain negations, they are only used in 1905 regular Tweets (for instance *"Nah, I'm sure they aren't mad for being ignored. #Sarcasm #Never-Forget #Venezuela"* [49]).

Key Result 3: Hashtags, Usernames, URLs. Sarcastic Tweets are comparably often messages to somebody and therefore mention usernames more often than ironic Tweets (13138 sarcastic vs. 11866 ironic Tweets that contain usernames). Usernames are less frequently mentioned in regular Tweets (with 5619). The following example supports one possible interpretation of sarcastic Tweets being more often targeted to a specific entity: *"Thanks @sonicdrivein for giving me so many onion rings on my meal. #sarcasm #sonic #hungry"* [50].

Regular Tweets refer more often to other things: Hashtags are most frequent in them (25939 vs. 14223 in irony and 14024 in sarcasm) as well as are URLs (26523 vs. 8660 and 7704).

These results support existing theories of irony and sarcasm with task-specific features and help to understand the actual use of these devices by authors. However, our study also revealed that word-based features are of high impact for the automated classification task. This suggests that topic-specific background knowledge is helpful to detect these devices. Only a subset of these words support interpretations of [28] for irony and sarcasm and for figurative vs. regular language [36], namely that irony and sarcasm are commonly used to express the opposite of the literal meaning, in line with the common understanding of these devices.

On the contrary, many words refer to specific topics under discussion. This suggests that the high performance in the classification is at least partially a result of overfitting to the data. Another interpretation is that these words build an environment for irony or sarcasm such it can actually be understood. To investigate this further, it is important to focus on a further data-driven analysis of topics in the different subcorpora and aim at discovering latent patterns in them. This will support the analysis of concept drift: training a classifier and testing

it on Tweets from a different (and distant) time frame will reveal differences between features which generalize over specific events.

References

1. Amazon Review: "worst book i have ever read" (2010). http://www.amazon.com/review/R86RAMEBZSB11. Accessed 29 Feb 2016
2. Maynard, D., Greenwood, M.: Who cares about sarcastic tweets? investigating the impact of sarcasm on sentiment analysis. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), Reykjavik, Iceland, European Language Resources Association (ELRA). pp. 4238–4243. ACL Anthology Identifier: L14–1527, May 2014
3. Kreuz, R.J., Glucksberg, S.: How to be sarcastic: the echoic reminder theory of verbal irony. J. Exp. Psychol. Gen. **118**(4), 374 (1989)
4. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. Technical report, Stanford University (2009). http://www.stanford.edu/alecmgo/papers/TwitterDistantSupervision09.pdf
5. Abrams, M.H.: A Glossary of Literary Terms, 7th edn. Heinle & Heinle, Thomson Learning, Boston (1999)
6. Nakassis, C., Snedeker, J.: Beyond sarcasm: intonation and context as relational cues in children's recognition of irony. In: Proceedings of the Twenty-sixth Boston University Conference on Language Development. Cascadilla Press (2002)
7. Creusere, M.A.: A developmental test of theoretical perspective on the understanding of verbal irony: children's recognition of allusion and pragmatic insincerity. In: Raymond W Gibbs, J., Colston, H.L. (eds.) Irony in Language and Thought: A Cognitive Science Reader, 1st edn, pp. 409–424. Lawrence Erlbaum Associates, New York (2007)
8. Glenwright, M.H., Pexman, P.M.: Children's perceptions of the social functions of verbal irony. In: Raymond W Gibbs, J., Colston, H.L. (eds.) Irony in Language and Thought: A Cognitive Science Reader, 1st edn, pp. 447–464. Lawrence Erlbaum Associates, New York (2007)
9. Kreuz, R.J.: The use of verbal irony: cues and constraints. In: Mio, J.S., Katz, A.N. (eds.) Metaphor: Implications and Applications. Lawrence Erlbaum Associates, Mahwah (1996)
10. Gibbs, R.W.J.: Irony in talk among friends. In: Raymond W Gibbs, J., Colston, H.L. (eds.) Irony in Language and Thought: A Cognitive Science Reader, 1st edn. Lawrence Erlbaum Associates, New York (2007)
11. Utsumi, A.: A unified theory of irony and its computational formalization. In: Proceedings of the 16th Conference on Computational Linguistics - vol. 2. COLING 1996, pp. 962–967. Association for Computational Linguistics, Stroudsburg (1996)
12. Wilson, D., Sperber, D.: On verbal irony. Lingua **87**, 53–76 (1992)
13. Clark, H.H., Gerrig, R.J.: On the pretense theory of irony. J. Exp. Psychol. Gen. **113**(1), 121–126 (1984)
14. Kumon-Nakamura, S., Glucksberg, S., Brown, M.: How about another piece of pie: the allusional pretense theory of discourse irony. J. Exp. Psychol. Gen. **124**(1), 3–21 (1995)

15. Wilson, D., Sperber, D.: Explaining Irony. Cambridge University Press, Cambridge (2012)
16. Utsumi, A.: Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. J. Pragmatics **32**(12), 1777–1806 (2000)
17. Clift, R.: Irony in conversation. Lang. Soc. **28**(4), 523–553 (1999)
18. Ptáček, T., Habernal, I., Hong, J.: Sarcasm detection on czech and english twitter. In: Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, Dublin, Ireland, Dublin City University and Association for Computational Linguistics, pp. 213–223, August 2014
19. Filatova, E.: Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In: Chair, N.C.C., Choukri, K., Declerck, T., Doan, M.U., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey, European Language Resources Association (ELRA), May 2012
20. Reyes, A., Rosso, P., Veale, T.: A multidimensional approach for detecting irony in Twitter. Lang. Resour. Eval. **47**(1), 239–268 (2013)
21. Rajadesingan, A., Zafarani, R., Liu, H.: Sarcasm detection on twitter: a behavioral modeling approach. In: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM 2015, 97–106. ACM, New York (2015)
22. Tepperman, J., Traum, D., Narayanan, S.S.: "Yeah right": sarcasm recognition for spoken dialogue systems. In: Proceedings of InterSpeech, Pittsburgh, PA, pp. 1838–1841, September 2006
23. Tsur, O., Davidov, D., Rappoport, A.: ICWSM - a great catchy name: semi-supervised recognition of sarcastic sentences in online product reviews. In: International AAAI Conference on Web and Social Media (ICWSM), Washington D.C., USA (2010)
24. Barbieri, F., Saggion, H.: Modelling irony in twitter. In: Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, pp. 56–64. Association for Computational Linguistics, April 2014
25. Riloff, E., Qadir, A., Surve, P., De Silva, L., Gilbert, N., Huang, R.: Sarcasm as contrast between a positive sentiment and negative situation. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pp. 704–714. Association for Computational Linguistics, Seattle, October 2013
26. Ghosh, A., Li, G., Veale, T., Rosso, P., Shutova, E., Barnden, J., Reyes, A.: SemEval-2015 task 11: sentiment analysis of figurative language in twitter. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 470–478. Association for Computational Linguistics, Denver, June 2015
27. Van Hee, C., Lefever, E., Hoste, V.: LT3: sentiment analysis of figurative tweets: piece of cake #NotReally. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 684–688. Association for Computational Linguistics, Denver, June 2015
28. Wang, P.Y.A.: #Irony or #Sarcasm - a quantitative and qualitative study based on twitter. In: 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC), Taiwan, Taipei, pp. 349–356 (2013)
29. Sulis, E., Farías, D.I.H., Rosso, P., Patti, V., Ruffo, G.: Figurative messages and affect in Twitter: differences between #irony, #sarcasm and #not. Knowledge-Based Systems, May 2016. (in Press)
30. Jaccard, P.: Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull. Soc. Vaudoise des Sci. Nat. **37**, 547–579 (1901)

31. Potts, C.: Twitter-aware tokenizer (2011). http://sentiment.christopherpotts.net/code-data/happyfuntokenizing.py. Accessed 03 Mar 2016
32. Derczynski, L., Ritter, A., Clark, S., Bontcheva, K.: Twitter part-of-speech tagging for all: Overcoming sparse and noisy data. In: Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013, Hissar, Bulgaria, pp. 198–206. Incoma Ltd., Shoumen, September 2013
33. Buschmeier, K., Cimiano, P., Klinger, R.: An impact analysis of features in a classification approach to irony detection in product reviews. In: Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pp. 42–49. Association for Computational Linguistics, Baltimore, June 2014
34. Davidov, D., Tsur, O., Rappoport, A.: Semi-Supervised Recognition of Sarcasm in Twitter and Amazon. In: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, pp. 107–116. Association for Computational Linguistics, Uppsala, July 2010
35. Whitlock, T.: Emoji unicode tables (2015). http://apps.timwhitlock.info/emoji/tables/unicode. Accessed 26 Feb 2016
36. González-Ibáñez, R., Muresan, S., Wacholder, N.: Identifying sarcasm in twitter: A closer look. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 581–586. Association for Computational Linguistics, Portland, June 2011
37. Kreuz, R., Caucci, G.: Lexical influences on the perception of sarcasm. In: Proceedings of the Workshop on Computational Approaches to Figurative Language, pp. 1–4. Association for Computational Linguistics, Rochester, April 2007
38. Holen, V.: Dictionary of interjections (2016). http://www.vidarholen.net/contents/interjections/. Accessed 03 Mar 2016
39. Nielsen, F.: Afinn, March 2011. http://www2.imm.dtu.dk/pubdb/p.php?6010. Accessed 18 Mar 2016
40. Barbieri, F., Saggion, H.: Modelling irony in twitter: feature analysis and evaluation. In: Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014), pp. 4258–4264. European Language Resources Association (ELRA), Reykjavik, May 2014
41. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: Liblinear: a library for large linear classification. J. Mach. Learn. Res. **9**, 1871–1874 (2008)
42. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods. Cambridge University Press, Cambridge (2000)
43. Utgoff, P.E.: Incremental induction of decision trees. Mach. Learn. **4**(2), 161–186 (1989)
44. Witten, I.H., Frank, E., Hall, M.H.: Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, San Francisco (2010)
45. Nigam, K., Lafferty, J., McCallum, A.: Using maximum entropy for text classification. In: IJCAI Workshop on Machine Learning for Information Filtering, Stockholm, Sweden (1999)
46. @Katie_Blair_: Mom signed me up... Twitter (2015). https://twitter.com/Katie_Blair_status/624580420668104704
47. @Bellastar12597: You can smile... Twitter (2015). https://twitter.com/Bellastar12597/status/630851771599069184

48. @MrHoffman9: Another wonderful day... Twitter (2015). https://twitter.com/Mr Hoffman9/status/635539576426262529
49. @CidHialeah: Another wonderful day... Twitter (2015). https://twitter.com/Cid Hialeah/status/623120702221103104
50. @Deadstitch: Another wonderful day... Twitter (2015). https://twitter.com/Dea dstitch/status/622931551064453120