

# Sarcasm Detection : Building a contextual hierarchy

**Taradheesh Bali**

Centre of Exact Humanities  
IIIT, Hyderabad  
Telangana, India

taradheesh.bali@research.iiit.ac.in

**Navjyoti Singh**

Centre of Exact Humanities  
IIIT, Hyderabad  
Telangana, India

navjyoti@iiit.ac.in

## Abstract

The conundrum of understanding and classifying sarcasm has been dealt with by the traditional theorists as an analysis of a sarcastic utterance and the ironic situation that surrounds it. The problem with such an approach is that it is too narrow, as it is unable to sufficiently utilize the two indispensable agents in making such an utterance, viz. the speaker and the listener. It undermines the necessary context required to comprehend a sarcastic utterance. In this paper, we propose a novel approach towards understanding sarcasm in terms of the existing knowledge hierarchy between the two participants, which forms the basis of the context that both agents share. The difference in relationship of the speaker of the sarcastic utterance and the disparate audience found on social media, such as Twitter, is also captured. We then apply our model on a corpus of tweets to achieve significant results and consequently, shed light on subjective nature of context, which is contingent on the relation between the speaker and the listener.

## 1 Introduction

Though deceptively simple for humans, it should come as no surprise that automatic recognition of sarcasm by machines is a highly complex task. Even more so when it is only through text, which is where even human annotators have struggled and have had differences of opinion as reported by Tsur et al. (2010). Computational detection of sarcasm, therefore, has been a more recent field of study. Davidov et al. (2010) made the first notable contribution through their semi-supervised classifier for tweets and Amazon product reviews. Since then a host of researches (González-Ibáñez et al., 2011; Riloff et al., 2013; Barbieri et al., 2014) have primarily focused on lexical features, such as word frequency (to detect the most commonly occurring words in a sarcastic statement), conjunctively occurring noun and verb phrases, sentiment analysis of such phrases etc. In addition to these, González-Ibáñez et al. (2011) included pragmatic features such as emoticons and in-reply-to user (though this only validated the interaction as a conversation). Similarly, Bamman and Smith (2015) also tried to make use of the two-fold relation between the author and the audience of a sarcastic utterance by maintaining a familiarity score between both. Joshi et al. (2015) lexical analysis a step further by harnessing the inter-sentential context incongruity. Sulis et al. (2016)

In this paper we provide a computational model built and structured on our proposed theoretical improvements to the framework for understanding sarcasm. Treating the sarcastic utterance as a multi-agent process between the speaker and the listener, we create a hierarchical knowledge structure for understanding their context. As a result, we are able to highlight and tackle the drawbacks of the past theories as well as computational attempts for understanding sarcasm. Recent studies have also tried to distinguish between verbal irony and sarcasm (Sulis et al., 2016) but for the scope of this paper we have treated them alike and both terms have been used interchangeably.

Rest of the paper is organized as follows. We first discuss related theoretical and computational work and their drawbacks in Section 2. Then Section 3 contains our proposed framework based on under-

---

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

standing context through knowledge hierarchy. Section 3.1 introduces the belief contradiction. Section 4 comprises of our methodology for data extraction, interpretation and classification. Sections 5 and 6 contain analysis and future work respectively.

## 2 Related work

### 2.1 Theoretical work

The traditional view of sarcasm/verbal irony has been that in such an utterance the speaker means the opposite of what he says. Many significant works in that direction are based on the maxims proposed by Grice (1970)<sup>1</sup>. However, when understood in such a manner it is easy to see why a typical Gricean explanation based on the violation of his conversational implicatures would suffice. But on closer examination it appears that while sarcasm and irony both tend to follow the trend of violation of Gricean Maxims in most (not all) cases, so do metaphors as well. This is because his account treats both the phenomenon as deviations from Gricean conventions. Also according to the Gricean account of meaning inversion, where the speaker intends to convey opposite of what he said is incomplete because in cases other than assertions or declarative statements it is not always possible to construe the opposite of what is said. Opposite in different utterances can mean different things like negation of proposition, negation of predicate, negation of implicature etc. This is because the speaker can selectively target from a specific part of the speech act to its whole to express his opinion sarcastically. Additionally, in cases such as hyperbole, rhetorical questions or over-polite requests where the no maxim is directly violated the line becomes even more blurred. Moreover, such a theory gives no significant reason as to why a speaker would choose to convey what he intends in such a way instead of simply stating what he wants and is therefore a better fit for incongruity resolution rather than for understanding sarcasm. But then again his theory of implicatures was not specifically meant to deal with such problems (at least not initially (Grice, 1978)) but was more about defining maxims for more conventional conversations.

Since then there have two significant approaches towards explaining sarcasm, viz., the pretense based approach and the echoic mention approach, both disagree with the Gricean claim and treat sarcasm not merely as semantic inversion but also as an act of speaker's expression of his attitude. While people on each side have claimed and tried to justify their approach to be superior, the debate has subdued as recent hybrid approaches have emerged which incorporate some aspects from both sides have proven to be better suited to explaining the process.

The most basic claims of these theories are as follows -

- **The Echoic Theory** - According to the echoic mention theory, a sarcastic utterance is understood by the listener when he is able to detect that the speaker of such an utterance is mocking or expressing his attitude towards some previously stated proposition (Sperber, 1984). That proposition can either be explicitly expressed in the conversation before or be implied implicitly. But this definition of echo was deemed too narrow and was later expanded to include the allusion to established social norms as well (Kumon-Nakamura et al., 1995).
- **The Pretense Theory** - Grice, without giving any proper explanation, had suggested that the being ironic involves an act of pretense which the pretender intends the audience to catch. Fowler (1965) introduced the concept of two types of audience, where one on listening to a sarcastic utterance could not get past the literal meaning but the other who are a part of the inner circle. Clark and Gerrig (1984) build on Grice's and Fowler's approach towards sarcasm and postulate that when uttering a sarcastic statement U, the speaker S is pretending to be another person P to whom S ascribes the utterance to, and intends for the listener L to look past/understand this pretense and know S's opinion/emotional attitude towards U and how ridiculous U is and in turn tries to mock either both P and U. That person P can be either a real or an imaginary person who endorses the idea of U.

---

<sup>1</sup>Grice's four maxims of quality, quantity, relevance and manner

Kumon-Nakamura et al. (1995) and Camp (2012) have also provided two significant combined accounts which incorporate elements from both the pretense account as well as the echoic account.

### 2.1.1 Criticisms

The echo theorists mainly emphasise on the content of the sarcastic utterance and suggest that the pretense theory focuses more on the form of the utterance (Wilson, 2006). Conforming to the echo theory, an argument often made is that while most instances of sarcasm/verbal irony can be explained through pretense, it is not a necessary condition for all the occurrences. According to Wilson (2006), while pretense can successfully mimic the form of another speech act, the content which the speaker alludes to, a necessary component of any sarcastic utterance, can only be delivered through echo. As a result, a pretense account cannot capture the essential attributive element of the sarcastic utterance, and determine towards whom or what is the speaker directing his derogatory attitude. Currie (2006) while agreeing that attribution is an indispensable constituent of a sarcastic utterance he also points out that such attribution is not always because of echoic content. For example, in cases of parodic sarcasm, one need not necessarily imitate bring forward the content but rather it is the form in this case which generates the echo. Therefore, reminding effect that echoic theory banks upon can be explained through both reminiscence in content as well as in form.

A major flaw in the criticisms of the pretense accounts is that their definitions of pretense are very limited and often is the case that they use a definition of pretense which only suits their criticism of it. Pretense is a very powerful concept that can be applied to a variety of speech acts. If not applied precisely enough pretense theories such as Clark and Gerrig (1984) or Clark (1996) can lead to a very generic and noisy classification, where even though the points they make are correct they are neither necessary nor sufficient. For example, it is not always the case that I consider myself as a gullible audience first in order to see through the pretense and nor do I as an audience necessarily take part actively in the speaker's pretense as is suggested in their dual audience or joint-pretense theories respectively.

Popa-Wyatt (2014) illustrates that, although Kumon-Nakamura et al. (1995) and Camp (2012), have tried to merge the two theories and related aspects together, they are unable to reasonably explain the relation between echo and pretense, and the interplay between both when identifying a sarcastic utterance. Camp (2012) suggests that for understanding a sarcastic utterance an inversion of a pre-supposed normative scale takes place but remains fairly unclear on the meaning of pre-supposition of the said normative scale. Popa-Wyatt (2014) while criticising Walton's (1990) account of pretense for identifying the right target of sarcasm also uses the argument that it can't be assumed that one disbelieves what one pretends to believe. While her objection is correctly placed, she does not progress the argument significantly. Her account follows *abductive reasoning*<sup>2</sup> when explaining that there is some connection between the pretend defective thought and the real/conceivable thought (which is the target), whereas a sufficient theory should be able to properly explain this connection that they both share and also be able to answer the questions of how one invokes the other and why does this effect take place.

## 2.2 Computational work and drawbacks

One of the key contributions of this paper is that instead of treating the problem of detecting sarcasm as a mere lexical exercise, we have tried to model a system that tries to build an understanding of the whole process as a human would do. González-Ibáñez et al. (2011) have shown that even after including the significant pragmatic features such as emoticons but leaving out of account the context associated with a particular tweet results in misclassification. They admitted that the brevity of a tweet sometimes makes it necessary to include additional information about the interaction between the tweeters such as common ground and world knowledge.

Bamman and Smith (2015) tried to tackle this problem by incorporating a familiarity score between the speaker and the audience, which was a measure for how much interaction both the parties have had. They also maintained a historical profile of the author and the audience of the sarcastic comment with features such as maintaining a score for their historical salient terms or his historical topics, which again

---

<sup>2</sup>Abductive reasoning, is a form of logical inference which goes from an observation to a theory which accounts for the observation, ideally seeking to find the simplest and most likely explanation.

is similar to the previous word frequency models albeit with the inclusion/acknowledgement of author-audience interaction for the first time. Although these features helped in increasing the accuracy, they fail to provide any significant improvement when both the users had low familiarity score or historical background. The efficiency of their proposition gets diminished even further when even though both the speaker and listener have some common context but the speaker has made a sarcastic utterance for the first time.

In their paper Joshi et al. (2015) tried to capture context incongruity, with their premise being - *sarcasm is a contrast between positive sentiment word and a negative situation and vice versa* - an approach similar to that of Riloff et al. (2013), but while Riloff (2013) focused only on contrast between words bearing positive sentiment and a negative situation; Joshi (2015) also laid equal emphasis on contrast between noun phrases with negative sentiment juxtaposed with a positive situation (comprised of the verb phrase). Though their research helped in shedding new light on the matter, the scope of the definition for context that they use is very limited and as reported in their paper, led to errors in cases where the context was highly subjective and could not be detected from only a single tweet.

### 3 Proposed Framework

Knowledge and uncertainty, especially the hierarchy that it generates in multi-agent systems has been a serious area of enquiry in game theory and philosophy. Inside a conversation, what the participants know about each other and what the participants know about the knowledge of other participants Lee (2001)<sup>3</sup>, all play a central role in how pretense is incorporated or acted out. Sarcastic pretense, as an act of speech, involves a fair degree of such higher order thinking, which can be understood to some degree by this vocabulary borrowed from game theory and epistemology. Also, one must note, that while the original treatment of these constructs deal with modelling agents in terms of bayesian rationality, we have no such unrealistic ambitions and are only using the broader framework to shed some light on the layers of perception that shape a speech act.

If the participants of a conversation are Speaker  $S$  and Listener  $L$ , then let us use  $B^s$  and  $B^l$  to denote the set and structure of beliefs held by them respectively.  $B_0^s$ , then, essentially represents the real identity of  $S$ , and similarly  $B_0^l$  for  $L$ . Given that, let us denote what  $S$  believes to be  $L$ 's beliefs by  $B_1^s$ , ie. what  $L$  believes in accordance to  $B_0^l$  (and similarly  $B_1^l$  denote what  $L$  believes to be  $S$ 's beliefs).

Likewise,  $B_2^s$  is what  $S$  believes about the beliefs held by  $L$  about  $S$ 's beliefs, ie.  $B_2^s$  is the beliefs of  $S$  in accordance to  $S$ 's perception of  $B_1^l$ . And similarly, for any  $i$ ,  $B_i^s$  denotes what  $S$  believes about the beliefs about  $S$  held in  $B_{i-1}^l$ .

In these terms, sarcastic pretense can be viewed as an act of speech where the if  $S$  has spoken an utterance  $U$  pretending to be  $P$ , then

- $U$  is not entirely aligned to  $S$ 's beliefs, ie.  $B_0^s$
- $L$  is able to distinguish  $P$  from  $S$ , using  $U$ . ie.  $U$  is sufficient to distinguish between  $B_1^l$  and  $P$
- $S$  knows that  $L$  will be able to distinguish  $P$  from  $S$  using  $U$ , ie.  $U$  is sufficient to distinguish between  $B_2^s$  and  $P$ . Otherwise, if  $S$  is not certain that  $L$  will be able to distinguish, an explicit marker of pretense would be required to convey the intended message.

An important thing to note here is the speaker's intent behind the sarcastic utterance. The intent of the speaker behind an act of sarcastic pretense is for the audience to see-through the pretense, and hence, through our structure of belief hierarchy we are also able to differentiate between a lie and a sarcastic utterance, which an echoic account is unable to capture due to its focus being mainly on content.

<sup>3</sup>Mutual Knowledge - An event is mutual knowledge if all agents know that the event occurred. However, mutual knowledge by itself implies nothing about what agents know about other agents' knowledge: i.e. it is possible that an event is mutual knowledge but that each agent is unaware that the other agents know it has occurred.

Common Knowledge - There is common knowledge of  $p$  in a group of agents  $G$  when all the agents in  $G$  know  $p$ , they all know that they know  $p$ , they all know that they all know that they know  $p$ , and so on ad infinitum.

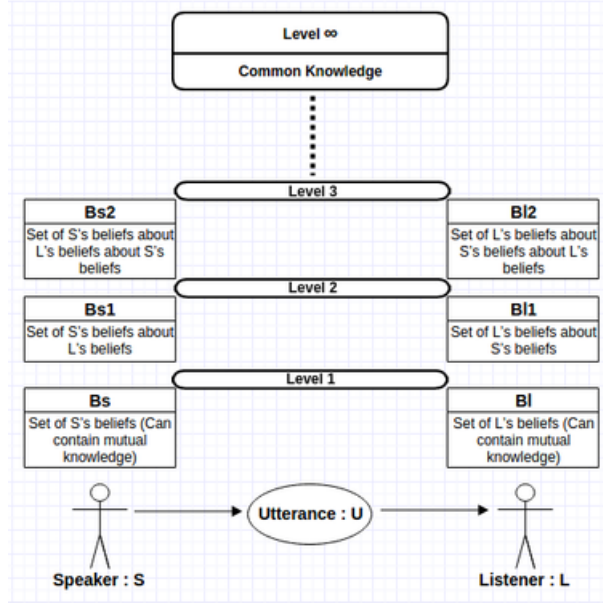


Figure 1: Knowledge Hierarchy

### 3.1 Belief Contradiction

One common drawback of the previous research on this topic is that it tries to isolate a sarcastic utterance from the conversation. Any conversation in general, not only on social media, has to have a speaker and an audience. The additional feature of conversations on social media is that much of it is public, especially on twitter. A tweet even when directed to a single user as a part of a conversation between two people who know each other is accessible to everyone who views it. Therefore, an uninitiated audience can also share their opinions on the matter with varying degrees of knowledge about the previous conversation. Such an unconventional structure can make the interpretation of a sarcastic utterance very difficult. As we have discussed above a joint account of pretense and echo is best suited for explaining the process of sarcasm. Besides, a better understanding of the process of pretense is required to explain how the speaker decides to carry out a pretense so that the audience that he is aiming for to see through the pretense understands it.

For a successful understanding of the pretense and subsequently the sarcasm/sarcastic intent behind my utterance, the audience has to recurse to the second level and validate, whether what I said either -

1. Contradicts or expresses a contrasting attitude towards  $B_1^l$  or,
2. Contradicts or expresses a contrasting attitude towards the beliefs in  $B^l$  (this set of beliefs can contain  $L$ 's own bias as well as the established norms).

While for the listener to understand the sarcastic utterance he has to recurse to the second level, in order for the speaker of the sarcastic utterance to make sure that the listener understands the sarcastic utterance he has to recurse to the third level. Then if  $B_2^s$  is populated in respect to the attitude he wants to convey he utters the sentence. Otherwise with the additional functionality of written social media he can add a #sarcasm or #sarcastic in the end to make his attitude a common knowledge.

Consequently, the audience that has to understand the sarcastic utterance can broadly be divided into two classes :

#### Case 1 : $B_1^l$ is not empty and $U$ is in contrast

This means that both agents have either some established common knowledge before or some perceived/inferred knowledge that  $L$  possesses which was conveyed by  $S$ . In such a case,  $L$  compares  $U$  or

its literal implicature against elements in  $B_1^l$ , and if the attitude expressed by  $S$  through  $U$  is in contrast with the relevant beliefs in  $B_1^l$  then there is increased likelihood that  $L$  understands the utterance  $U$  to be sarcastic.

Figure 2 shows  $S$  conveying and thereby, populating  $B_1^l$  with his (negative) emotional attitude. A literal (positive) understanding of the belief conveyed through  $U$  (Figure 3) shows  $S$  contrasting his previously conveyed belief. Thus, informing  $L$  about the increased probability of the  $U$  being sarcastic.



Figure 2: Belief conveyed in prior conversation



Figure 3: Belief contradiction through sarcastic utterance  $U$

## Case 2 : $B_1^l$ is not empty but has no priors related to $U$ OR $B_1^l$ is empty

This means that either,  $L$  does not have any previous knowledge about  $U$  and related implicatures or he does not have any previous shared knowledge with  $S$ . In both the cases  $L$  has only  $B^l$  to rely upon in relation to  $U$ , therefore if the attitude expressed by  $S$  through  $U$  is now in contrast with relevant beliefs in  $B^l$  then there is increased likelihood that  $L$  understands the utterance  $U$  to be sarcastic.

Let us call both the audience  $L_1$  and  $L_2$  respectively.

## 4 Method

### 4.1 Data and its interpretation

To ensure that high precision tweets are chosen for the positive set we limit our scope only to those tweets which have #sarcasm or #sarcastic explicitly mentioned in them. We keep only English tweets with number of words  $\geq 5$ , filter out hyperlinks or retweets. Tweets were mined using tweepy, a python library for accessing twitter API. A datapoint in this dataset will be our  $U$  (sarcastic utterance). In order to ensure that there is a communicative context we only choose those tweets that are in reply to another parent tweet say  $p$ . Cases where  $p$  is protected (inaccessible via API due to restricted permissions) were also accounted for and such dangling pairs, along with their references were cleaned from our dataset as a part of pre-processing. The user who has replied with the sarcastic tweet is  $S$  and the user to whose tweet he is replying to is  $L$ . We store all the hashtags used in both  $p$  and  $U$ , in a set  $H$ . These hashtags function as indicators of the topics that both are talking about. We then crawl and retrieve the most recent tweets of both the authors, which contain an element of  $H$  in order to populate  $B^s$  and  $B^l$ . Now we check if  $L$  and  $S$  have had any previous communication in relation to an element of  $H$ .

If yes, then it is Case 1 (with audience type  $L_1$ ) and each such  $B_1^l$  is populated with a maximum of 10 tweets containing an element of  $H$  which  $S$  has tweeted to  $L$ .

If no, then it is Case 2 (with audience type  $L_2$ ) and each such  $B^l$  is populated with a maximum of 10 tweets containing an element of  $H$  which either  $L$  has tweeted independently or have been tweeted by rest of the world. This is done because in this case either  $L$  could have explicitly tweeted his opinions on elements in  $H$  or we assume that his views even though not categorically expressed in any previous tweets are aligned with what the overall popular sentiment exhibits through the 10 most recent tweets (tweeted by the rest of the world) on topics in  $H$ .  $L$ 's independent tweets represent his biases and the tweets from the rest of the world represent the norms.

For negative data the above mentioned process was repeated but for replies without an explicit #sarcasm or #sarcastic marker. Finally, the negative data was manually cross-verified and tweets with implicit sarcastic mentions were removed. We maintain a Reference Table for reference IDs for each type of tweet for both positive as well as the negative data. This table helps in cross-checking the negative data such that it does not overlap via any user, tweet or retweet with the positive data. This yields us total of 2000 conversational instances with a positive set of 1000 instances, 500 for each case, and an equal number of negative instances.

Table 1: Feature Set

<b>Lexical</b>	
Word Unigrams and Bigrams	For tweets of both $S$ and $L$ , we used unigrams and bigrams found in the training corpus.
Brown cluster unigrams	Again for both tweets $S$ and $L$ , we used brown clusters which helped us in grouping words used in similar contexts into the same cluster.
Part of Speech	Output of the POS tagger of CMU <sup>4</sup> of each lexical item in the tweet.
<b>Pragmatic</b>	
Capitalization	Number of capitalized letters.
Emoticons and Expressions	Number of emoticons and expressions such as lol, haha, :D
Frequent Expressions in sarcastic utterances	For sarcastic tweets of both $S$ and $L$ , we kept a list of top 100 words according to their $tf-idf$ score. This feature indicated the presence of a such a word in the current tweets.
<b>Contextual</b>	
Belief Contradiction	Binary feature stating if a belief has been contradicted or not.

## 4.2 Classification

Similar to Bamman and Smith (2015), we adopted a binary logistic regression with  $l_2$  regularization using 10 fold cross validation to do the binary classification task of marking tweets as SARCASTIC and NOT SARCASTIC. 8 folds out of 10 were used for training, 1 for tuning and remaining last was used for testing. Our baseline model, which predicted the most common of the two labels, gave an accuracy of 48%.

Our feature set is described in Table 1. We used three sets of features, namely, Lexical, Pragmatic, and Contextual. The lexical and pragmatic features are self-explanatory. As our contextual feature we used the proposed belief contradiction method. The noisy non-standard tokens in the data were normalized using the English Normalizer implemented by Sharma et al. (2016). We then used Stanford's Sentiment Analyzer (Socher et al., 2013) to calculate the sentiment of a tweet which we used as the belief of  $L$  /  $S$  regarding the topic and for the current tweet as well. These belief scores were normalized between -1

<sup>4</sup>[www.cs.cmu.edu/~ark/TweetNLP/](http://www.cs.cmu.edu/~ark/TweetNLP/) (2011)

to 1, with -1 being very negative and 1 being very positive. The overall sentiment scores for  $B_1^l$  and  $B^l$  are calculated as the average of that set. A belief is deemed to be violated when it exudes a contrasting sentiment as compared to the sentiments of either  $B_1^l$  or  $B^l$ , depending on the type of audience.

Table 2: Results  
Results Table

Features	Accuracy
Lexical(Baseline)	71.2%
Lexical+Pragmatic	75.8%
Lexical+Pragmatic+Contextual	78.7%

## 5 Analysis

As Table 2 shows the impact of contextual features based on our belief contradiction method as it hugely improves the accuracy from baseline and therefore, is statistically significant. Also, we have addressed problems faced in previous researches, such as, Joshi et al. (2015), where samples containing highly subjective incongruity were misclassified. Our classifier correctly predicted most such samples, as it is able to capture the incongruence in cases where context is highly subjective and varies for a speaker in each interaction. For instances of Case 2, when the speaker and the audience have no previous context, Bamman and Smith (2015) model does not perform well. But our framework captures the violation of normative expectation through an overall recent sentiment score for such topics and hence, is able to correctly classify such instances as well.

## 6 Future Work

The number of features for lexical and pragmatic analyses could be increased and diversified to enhance our system. Also, while we were analysing the overall sentiment of a tweet, breaking a tweet into parts according to each topic could be done to analyse the sentiment of the speaker towards each topic. This could provide us with a better understanding of his beliefs on that topic. It also allows us to capture the sentiment contrast at a sub-tweet level rather than the whole tweet. In the future, this experiment can be extended to the sarcastic tweets of collections that have been used previously, for instance the one used in the shared task 11 of SemEval-2015 (Ghosh et al., 2015). The scope of this paper stretches beyond sarcasm detection in tweets only. Our model can also be incorporated in conversational chat bots to detect sarcasm in a user’s reply and respond accordingly. The historical features that we propose can ameliorate the responses even for smart messaging services such as Google's Allo.

## References

- David Bamman and Noah A Smith. 2015. Contextualized sarcasm detection on twitter. In *Ninth International AAAI Conference on Web and Social Media*.
- Francesco Barbieri, Horacio Saggion, and Francesco Ronzano. 2014. Modelling sarcasm in twitter, a novel approach. *ACL 2014*, page 50.
- Elisabeth Camp. 2012. Sarcasm, pretense, and the semantics/pragmatics distinction. *Noûs*, 46(4):587–634.
- Herbert H Clark and Richard J Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1):121–126.
- HH Clark. 1996. Using language cambridge university press cambridge.
- Gregory Currie. 2006. Why irony is pretence. *The architecture of the imagination*, pages 111–33.
- Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in twitter and amazon. In *Proceedings of the fourteenth conference on computational natural language learning*, pages 107–116. Association for Computational Linguistics.



- Henry Watson Fowler. 1965. *Modern english usage* (revised by sir ernest gowers).
- Aniruddha Ghosh, Guofu Li, and Tony Veale. 2015. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter.
- Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A Smith. 2011. Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 42–47. Association for Computational Linguistics.
- Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. 2011. Identifying sarcasm in twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 581–586. Association for Computational Linguistics.
- Herbert P Grice. 1970. *Logic and conversation*. na.
- H Paul Grice. 1978. Further notes on logic and conversation. *1978*, 1:13–128.
- Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya. 2015. Harnessing context incongruity for sarcasm detection. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, volume 2, pages 757–762.
- Sachi Kumon-Nakamura, Sam Glucksberg, and Mary Brown. 1995. How about another piece of pie: The allusionary pretense theory of discourse irony. *Journal of Experimental Psychology: General*, 124(1):3.
- Benny PH Lee. 2001. Mutual knowledge, background knowledge and shared beliefs: Their roles in establishing common ground. *Journal of pragmatics*, 33(1):21–44.
- Mihaela Popa-Wyatt. 2014. Pretence and echo: towards an integrated account of verbal irony. *International Review of Pragmatics*, 6(1):127–168.
- Ellen Riloff, Ashequl Qadir, Prafulla Surve, Lalindra De Silva, Nathan Gilbert, and Ruihong Huang. 2013. Sarcasm as contrast between a positive sentiment and negative situation. In *EMNLP*, volume 13, pages 704–714.
- Arnav Sharma, Sakshi Gupta, Raveesh Motlani, Piyush Bansal, Manish Srivastava, Radhika Mamidi, and Dipti M Sharma. 2016. Shallow parsing pipeline for hindi-english code-mixed social media text. *arXiv preprint arXiv:1604.03136*.
- Richard Socher, Alex Perelygin, Jean Y Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer.
- Dan Sperber. 1984. Verbal irony: Pretense or echoic mention?
- Emilio Sulis, Delia Irazú Hernández Farías, Paolo Rosso, Viviana Patti, and Giancarlo Ruffo. 2016. Figurative messages and affect in twitter: differences between# irony,# sarcasm and# not. *Knowledge-Based Systems*.
- Oren Tsur, Dmitry Davidov, and Ari Rappoport. 2010. Icwsm-a great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *ICWSM*.
- Kendall L Walton. 1990. *Mimesis as make-believe: On the foundations of the representational arts*. Harvard University Press.
- Deirdre Wilson. 2006. The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116(10):1722–1743.