Towards a Contextual Pragmatic Model to Detect Irony in Tweets

Jihen Karoui, Farah Benamara, Véronique Moriceau, Nathalie Aussenac-Gilles, Lamia Hadrich Belguith

Towards a Contextual Pragmatic Model to Detect Irony in Tweets

Jihen Karoui
IRIT, MIRACL
Toulouse University, Sfax University
karoui@irit.fr

Farah Benamara Zitoune
IRIT, CNRS
Toulouse University
benamara@irit.fr

Véronique Moriceau
LIMSI-CNRS
Univ. Paris-Sud
moriceau@limsi.fr

Nathalie Aussenac-Gilles
IRIT, CNRS
Nathalie.Aussenac-Gilles@irit.fr

Lamia Hadrich Belguith
MIRACL
University of Sfax
l.belguith@fsegs.rnu.tn

Abstract

This paper proposes an approach to capture the pragmatic context needed to infer irony in tweets. We aim to test the validity of two main hypotheses: (1) the presence of negations, as an internal propriety of an utterance, can help to detect the disparity between the literal and the intended meaning of an utterance, (2) a tweet containing an asserted fact of the form $Not(P_1)$ is ironic if and only if one can assess the absurdity of P_1. Our first results are encouraging and show that deriving a pragmatic contextual model is feasible.

1 Motivation

Irony is a complex linguistic phenomenon widely studied in philosophy and linguistics (Grice et al., 1975; Sperber and Wilson, 1981; Utsumi, 1996). Despite theories differ on how to define irony, they all commonly agree that it involves an incongruity between the literal meaning of an utterance and what is expected about the speaker and/or the environment. For many researchers, irony overlaps with a variety of other figurative devices such as satire, parody, and sarcasm (Clark and Gerrig, 1984; Gibbs, 2000). In this paper, we use irony as an umbrella term that covers these devices focusing for the first time on the automatic detection of irony in French tweets.

According to (Grice et al., 1975; Searle, 1979; Attardo, 2000), the search for a non-literal meaning starts when the hearer realizes that the speaker's utterance is context-inappropriate, that is an utterance fails to make sense against the context. For example, the tweet: *"Congratulation #lesbleus for your great match!"* is ironic if the French soccer team has lost the match. An analysis of a corpus of French tweets shows that there are two ways to infer such a context: (a) rely exclusively on the lexical clues internal to the utterance, or (b) combine these clues with an additional pragmatic context external to the utterance. In (a), the speaker intentionally creates an explicit juxtaposition of incompatible actions or words that can either have opposite polarities, or can be semantically unrelated, as in "***The Voice*** *is more important than* ***Fukushima*** *tonight*". Explicit opposition can also arise from an explicit positive/negative contrast between a subjective proposition and a situation that describes an undesirable activity or state. For instance, in " ***I love*** *when my phone* ***turns the volume down automatically***" the writer assumes that every one expects its cell phone to ring loud enough to be heard. In (b), irony is due to an implicit opposition between a lexicalized proposition P describing an event or state and a pragmatic context external to the utterance in which P is false or is not likely to happen. In other words, the writer asserts or affirms P while he intends to convey P' such that $P' = Not(P)$ or $P' \neq P$. The irony occurs because the writer believes that his audience can detect the disparity between P and P' on the basis of contextual knowledge or common background shared with the writer. For example, in "*#Hollande is really a good diplomat #Algeria.*", the writer critics the foreign policy of the French president Hollande in Algeria, whereas in "*The #NSA wiretapped a whole country. No worries for #Belgium:* ***it is not a whole country.***", the irony occurs because the fact in bold font is not true.

Irony detection is quite a hot topic in the research community also due to its importance for efficient sentiment analysis (Ghosh et al., 2015). Several approaches have been proposed to detect irony casting the problem into a binary classification task relying on a variety of features. Most of them are gleaned from the utterance internal context going from n-grams models, stylistic (punctuation, emoticons, quotations, etc.), to dictionary-based features (sentiment and affect dictionaries, slang languages, etc.). These features have shown to be useful to learn whether a text span is ironic/sarcastic or not (Burfoot and Baldwin, 2009; Davidov et al., 2010; Tsur et al., 2010; Gonzalez-Ibanez et al., 2011; Reyes et al., 2013; Barbieri and Saggion, 2014). However, many authors pointed out the necessity of additional pragmatic features: (Utsumi, 2004) showed that opposition, rhetorical questions and the politeness level are relevant. (Burfoot and Baldwin, 2009) focused on satire detection in newswire articles and introduced the notion of validity which models absurdity by identifying a conjunc-

tion of named entities present in a given document and queries the web for the conjunction of those entities. (Gonzalez-Ibanez et al., 2011) exploited the common ground between speaker and hearer by looking if a tweet is a reply to another tweet. (Reyes et al., 2013) employed opposition in time (adverbs of time such as *now* and *suddenly*) and context imbalance to estimate the semantic similarity of concepts in a text to each other. (Barbieri and Saggion, 2014) captured the gap between rare and common words as well as the use of common vs. rare synonyms. Finally, (Buschmeier et al., 2014) measured the imbalance between the overall polarity of words in a review and the star-rating. Most of these pragmatic features rely on linguistic aspects of the tweet by using only the text of the tweet. We aim here to go further by proposing a novel computational model able to capture the "outside of the utterance" context needed to infer irony in implicit oppositions.

2 Methodology

An analysis of a corpus of French ironic tweets randomly chosen from various topics shows that more than 62.75% of tweets contain explicit negation markers such as "ne...pas" (not) or negative polarity items like "jamais" (never) or "personne" (nobody). Negation seems thus to be an important clue in ironic statements, at least in French. This rises the following hypotheses: (H1) the presence of negations, as an internal propriety of an utterance, can help to detect the disparity between the literal and the intended meaning of an utterance, and (H2) a tweet containing an asserted fact of the form $Not(P)$ is ironic if and only if one can prove P on the basis of some external common knowledge to the utterance shared by the author and the reader.

To test the validity of the above hypotheses, we propose a novel three-step model involving three successive stages: (1) detect if a tweet is ironic or not relying exclusively on the information internal to the tweet. We use a supervised learning method relying on both state of the art features whose efficiency has been empirically proved and new groups of features. (2) Test this internal context against the "outside of the utterance" context. We design an algorithm that takes the classifier's outputs and corrects the misclassified ironic instances of the form $Not(P)$ by looking for P in reliable external sources of information on the Web, such as Wikipedia or online newspapers. We experiment when labels are given by gold standard annotations and when they are predicted by the classifier. (3) If the literal meaning fails to make sense, i.e. P is found, then the tweet is likely to convey a non-literal meaning.

To this end, we collected a corpus of 6,742 French tweets using the Tweeter API focusing on tweets relative to a set of topics discussed in the media during Spring 2014. Our intuition behind choosing such topics is that a media-friendly topic is more likely to be found in external sources of information. We chose 184 topics split into 9 categories (politics, sport, etc.). For each topic, we selected a set of keywords with and without hashtag: politics (*e.g.* Sarkozy, Hollande, UMP), health (*e.g.* cancer, flu), sport (*e.g.* #Zlatan, #FIFAworldcup), social media (*e.g.* #Facebook, Skype, MSN), artists (*e.g.* Rihanna, Beyoncé), TV shows (*e.g.* TheVoice, XFactor), countries or cities (*e.g.* NorthKorea, Brasil), the Arab Spring (*e.g.* Marzouki, Ben Ali) and some other generic topics (*e.g.* pollution, racism). Then we selected ironic tweets containing the topic keywords, the #*ironie* or #*sarcasme* hashtag and a negation word as well as ironic tweets containing only the topic keywords with #*ironie* or #*sarcasme* hashtag but no negation word. Finally, we selected non ironic tweets that contained either the topic keywords and a negation word, or only the topic keywords. We removed duplicates, retweets and tweets containing pictures which would need to be interpreted to understand the ironic content. Irony hashtags (#*ironie* or #*sarcasme*) are removed from the tweets for the following experiments. To guarantee that tweets with negation words contain true negations, we automatically identified negation usage of a given word using a French syntactic dependency parser[1]. We then designed dedicated rules to correct the parser's decisions if necessary. At the end, we got a total of 4,231 tweets with negation and 2,511 without negation, among them, 30.42% are ironic with negation and 72.36% are non ironic with negation. At the end, we got a total of 4,231 tweets with negation and 2,511 without negation: among them, 30.42% are ironic with negation and 72.36% are non ironic with negation. To capture the effect of negation on our task, we split these tweets in three corpora: tweets with negation only (*NegOnly)*, tweets with no negation (*NoNeg*), and a corpus that gathers all the tweets of the previous 2 corpora (*All*). Table 1 shows the repartition of tweets in our corpora.

Corpus	**Ironic**	**Non ironic**	TOTAL
NegOnly	470	3,761	**4,231**
NoNeg	1,075	1,436	**2,511**
All	1,545	5,197	**6,742**

Table 1: Tweet repartition.

3 Binary classifier

We experiment with SMO under the Weka toolkit with standard parameters. We also evaluated other learning algorithms (naive bayes, decision trees, logistic regression) but the results were not as good as those obtained with SMO. We have built three classifiers, one for each corpus, namely C_{Neg}, C_{NoNeg}, and C_{All}. Since the number of ironic instances in the first corpus is relatively small, we learn C_{Neg} with 10-cross validation on a balanced subset of 940 tweets. For the second and the last classifiers, we used 80% of the corpus for training

[1]We have used Malt as a syntactic parser.

and 20% for test, with an equal distribution between the ironic (henceforth IR) and non ironic (henceforth NIR) instances[2]. The results presented in this paper have been obtained when training C_{NoNeg} on 1,720 and testing on 430 tweets. C_{All} has been trained on 2,472 tweets (1432 contain negation –404 IR and 1028 NIR) and tested on 618 tweets (360 contain negation – 66 IR and 294 NIR). For each classifier, we represent each tweet with a vector composed of six groups of features. Most of them are state of the art features, others, in italic font are new.

Surface features include tweet length in words (Tsur et al., 2010), the presence or absence of punctuation marks (Gonzalez-Ibanez et al., 2011), words in capital letters (Reyes et al., 2013), interjections (Gonzalez-Ibanez et al., 2011), emoticons (Buschmeier et al., 2014), quotations (Tsur et al., 2010), slang words (Burfoot and Baldwin, 2009), opposition words such as "but" and "although" (Utsumi, 2004), a sequence of exclamation or a sequence of question marks (Carvalho et al., 2009), a combination of both exclamation and question marks (Buschmeier et al., 2014) and finally, *the presence of discourse connectives that do not convey opposition* such as "hence, therefore, as a result" since we assume that non ironic tweets are likely to be more verbose. To implement these features, we rely on manually built French lexicons to deal with interjections, emoticons, slang language, and discourse connectives (Roze et al., 2012).

Sentiment features consist of features that check for the presence of positive/negative opinion words (Reyes and Rosso, 2012) and the number of positive and negative opinion words (Barbieri and Saggion, 2014). We add three new features: *the presence of words that express surprise or astonishment*, and *the presence and the number of neutral opinions*. To get these features we use two lexicons: CASOAR, a French opinion lexicon (Benamara et al., 2014) and EMOTAIX, a publicly available French emotion and affect lexicon.

Sentiment shifter features group checks if a given *tweet contains an opinion word which is in the scope of an intensifier adverb or a modality*.

Shifter features tests if a tweet contains an intensifier (Liebrecht et al., 2013), a negation word (Reyes et al., 2013), or *reporting speech verbs*.

Opposition features are new and check for the presence of specific lexico-syntactic patterns that verify whether a tweet contains a sentiment opposition or an explicit positive/negative contrast between a subjective proposition and an objective one. These features have been partly inspired from (Riloff et al., 2013) who proposed a bootstrapping algorithm to detect sarcastic tweets of the form $[P_+].[P'_{obj}]$ which corresponds to a contrast between positive sentiment and an objective negative situation. We extended this pattern to capture additional types of explicit oppositions. Some of our patterns include: $[Neg(P_+)].[P'_+]$, $[P_-].[P'_+]$, $[Neg(P_+)].[P'_{obj}]$, $[P'_{obj}].[P_-]$. We consider that an opinion expression is under the scope of a negation if it is separated by a maximum of two tokens.

Finally, **internal contextual** deals with the presence/absence of *personal pronouns, topic keywords* and *named entities*, as predicted by the parser's outputs.

For each classifier, we investigated how each group of features contributes to the learning process. We applied to each training set a feature selection algorithm (Chi2 and GainRatio), then trained the classifiers over all relevant features of each group[3]. In all experiments, we used all surface features as baseline. Table 2 presents the result in terms of precision (P), recall (R), macro-averaged F-score (MAF) and accuracy (A). We can see that C_{All} achieves better results. An analysis of the best features combination for each classifier suggests four main conclusions: (1) surface features are primordial for irony detection. This is more salient for *NoNeg*. (2) Negation is an important feature for our task. However, having it alone is not enough to find ironic instances. Indeed, among the 76 misclassified instances in C_{All}, 60% contain negation clues (37 IR and 9 NIR). (3) When negation is concerned, opposition features are among the most productive. (4) Explicit opinion words (i.e sentiment and sentiment shifter) are likely to be used in tweets with no negation. More importantly, these results empirically validate hypothesis (H1), i.e. negation is a good clue to detect irony.

	Ironic (IR)			**Not ironic (NIR)**		
	P	R	F	P	R	F
C_{Neg}	88.9	56.0	68.7	67.9	93.3	78.5
C_{NoNeg}	71.1	65.1	68.0	67.80	73.50	70.50
C_{All}	93.0	81.6	86.9	83.6	93.9	88.4

	Overall Results	
	MAF	A
C_{Neg}	73.6	74.5
C_{NoNeg}	69.2	69.3
C_{All}	87.6	87.7

Table 2: Results for the best features combination.

Error analysis shows that misclassification of ironic instances is mainly due to four factors: presence of similes (ironic comparison)[4], absence of context within the utterance (most frequent case), humor and satire[5], and wrong *#ironie* or *#sarcasme* tags. The absence of context can manifest itself in several ways: (1) there is no pointer that helps to identify the main topic of the tweet, as in "*I've been missing her, damn!*". Even if the topic is present, it is often lexicalized in several collapsed words or funny hashtags (*#baddays, #aprilfoll*),

[2]For C_{NoNeg} and C_{All}, we also tested 10-cross validation with a balanced distribution between the ironic and non-ironic instances but results were not conclusive.

[3]Results with all features are lower.

[4]e.g. "Benzema in the French team is like Sunday. He is of no use.. :D"

[5]*e.g.* "I propose that we send Hollande instead of the space probes on the next comet, it will save time and money ;) #HUMOUR"

which are hard to automatically analyze. (2) The irony is about specific situations (Shelley, 2001). (3) False assertions about hot topics, like in "*Don't worry. Senegal is the world champion soccer*". (4) Oppositions that involve a contradiction between two words that are semantically unrelated, a named entity and a given event (*e.g.* "Tchad and "democratic election"), etc. Case (4) is more frequent in the *NoNeg* corpus.

Knowing that tweets with negation represent 62.75% of our corpus, and given that irony can focus on the negation of a word or a proposition (Haverkate, 1990), we propose to improve the classification of these tweets by identifying the absurdity of their content, following Attardo's relevant inappropriateness model of irony (Attardo, 2000) in which a violation of contextual appropriateness signals ironical intent.

4 Deriving the pragmatic context

The proposed model included two parts: binary classifiers trained with tweet features, and an algorithm that corrects the outputs of the classifiers which are likely to be misclassified. These two phases can be applied successively or together. In this latter case, the algorithm outputs are integrated into the classifiers and the corrected instances are used in the training process of the binary classifier. In this paper, we only present results of the two phases applied successively because it achieved better results.

Our approach is to query Google via its API to check the veracity of tweets with negation that have been classified as non ironic by the binary classifier in order to correct the misclassified tweets (if a tweet saying $Not(P)$ has been classified as non-ironic but P is found online, then we assume that the opposite content is checked so the tweet class is changed into ironic). Let $WordsT$ be the set of words excluding stop words that belong to a tweet t, and let kw be the topic keyword used to collect t. Let $N \subset WordsT$ be the set of negation words of t. The algorithm is as follows:

1. Segment t into a set of sentences S.

2. For each $s \in S$ such that $\exists neg \in N$ and $neg \in s$:

2.1 Remove # and @ symbols, emoticons, and neg, then extract the set of tokens $P \subset s$ that are on the scope of neg (in a distance of 2 tokens).

2.2 Generate a query $Q_1 = P \cup kw$ and submit it to Google which will return 20 results (title+snippet) or less.

2.3 Among the returned results, keep only the reliable ones (Wikipedia, online newspapers, web sites that do not contain "blog" or "twitter" in their URL). Then, for each result, if the query keywords are found in the title or in the snippet, then t is considered as ironic. STOP.

3. Generate a second query $Q_2 = (WordsT - N) \cup kw$ and submit it again to Google and follow the procedure in 2.3. If Q_2 is found, then t is considered as ironic. Otherwise, the class predicted by the classifier does not change.

Let us illustrate our algorithm with the topic *Valls* and the tweet: *#Valls has learnt that Sarkozy was wiretapped in newspapers. Fortunately he is not the interior minister.* The first step leads to two sentences s_1 (*#Valls has learnt that Sarkozy was wiretapped in newspapers.*) and s_2 (***Fortunately he is not the interior minister***). From s_2, we remove the negation word "not", isolate the negation scope $P = \{interior,\ minister\}$ and generate the query $Q_1 = \{Valls\ interior\ minister\}$. The step 2.3 allows to retrieve the result:

> <Title>*Manuel **Valls** - Wikipedia, the free encyclopedia*</Title>
> <Snippet>*... French politician. For the Spanish composer, see Manuel **Valls** (composer). .... **Valls** was appointed **Minister** of the **Interior** in the Ayrault Cabinet in May 2012.*</Snippet>.

All query keywords were found in this snippet (in bold font), we can then conclude that the tweet is ironic.

We made several experiments to evaluate how the query-based method improves tweet classification. For this purpose, we have applied the method on both corpora *All* and *Neg*: ① A first experiment evaluates the method on tweets with negation classified as `NIR` but which are ironic according to gold annotations. This experiment represents an ideal case which we try to achieve or improve through other ones. ②: A second experiment consists in applying the method on all tweets with negation that have been classified as `NIR` by the classifier, no matter if the predicted class is correct or not. Table 3 shows the results for both experiments.

	①		②	
`NIR` tweets for which:	*All*	*Neg*	*All*	*Neg*
Query applied	37	207	327	644
Results on Google	25	102	166	331
Class changed into `IR`	5	35	69	178
Classifier Accuracy	87.7	74.46	**87.7**	**74.46**
Query-based Accuracy	**88.51**	**78.19**	78.15	62.98

Table 3: Results for the query-based method.

All scores for the query-based method are statistically significant compared to the classifier's scores ($p\_value < 0,0001$ when calculated with the McNemar's test.). An error analysis shows that 65% of tweets that are still misclassified with this method are tweets for which finding their content online is almost impossible because they are personal tweets or lack internal context. A conclusion that can be drawn is that this method should not be applied on this type of tweets. For this purpose, we made the same experiments only on tweets with different combinations of relevant features. The best results are obtained when the method is applied only on `NIR` tweets with negation selected via the internal context features, more precisely on tweets which do not contain a personal pronoun and which contain named entities: these results are coherent with

the fact that tweets containing personal pronouns and no named entity are likely to relate personal content impossible to validate on the Web (*e.g. I've been missing her, damn! #ironie*). Table 4 shows the results for these experiments. All scores for the query-based method are also statistically significant compared to the classifier's scores.

	①		②	
NIR *tweets for which:*	*All*	*Neg*	*All*	*Neg*
Query applied	0	18	40	18
Results on Google	-	12	17	12
Class changed into IR	-	4	7	4
Classifier Accuracy	87.7	74.46	**87.7**	74.46
Query-based Accuracy	**87.7**	**74.89**	86.57	**74.89**

Table 4: Results when applied on "non-personal" tweets.

For experiment ①, on *All*, the method is not applied because all misclassified tweets contain a personal pronoun and no named entity. The query-based method outperforms the classifier in all cases, except on *All* where results on Google were found for only 42.5% of queries whereas more than 50% of queries found results in all other experiments (maximum is 66.6% in *NegOnly*). Tweets for which no result is found are tweets with named entities but which do not relate an event or a statement (*e.g. AHAHAHAHAHA! NO RESPECT #Legorafi*, where "Legorafi" is a satirical newspaper). To evaluate the task difficulty, two annotators were also asked to label as ironic or not the 50 tweets (40+18) for which the method is applied. The inter-annotator score (Cohen's Kappa) between both annotators is only $\kappa = 0.41$. Among the 12 reclassifications into IR, both annotators disagree with each other for 5 of them. Even if this experiment is not strong enough to lead to a formal conclusion because of the small number of tweets, this tends to show that human beings would not do it better.

It is interesting to note that even if internal context features were not relevant for automatic tweet classification, our results show that they are useful for classification improvement. As shown by ①, the query-based method is more effective when applied on misclassified tweets. We can then consider that using internal contextual features (presence of personal pronouns and named entities) can be a way to automatically detect tweets that are likely to be misclassified.

5 Discussion and conclusions

This paper proposed a model to identify irony in implicit oppositions in French. As far as we know, this is the first work on irony detection in French on Twitter data. Comparing to other languages, our results are very encouraging. For example, sarcasm detection achieved 30% precision in Dutch tweets (Liebrecht et al., 2013) while irony detection in English data resulted in 79% precision (Reyes et al., 2013).

We treat French irony as an overall term that covers other figurative language devices such as sarcasm, humor, etc. This is a first step before moving to a more fine-grained automatic identification of figurative language in French. For interesting discussions on the distinction/similarity between irony and sarcasm hastags, see (Wang, 2013).

One of the main contribution of this study is that the proposed model does not rely only on the lexical clues of a tweet, but also on its pragmatic context. Our intuition is that a tweet containing an asserted fact of the form $Not(P_1)$ is ironic if and only if one can prove P_1 on the basis of some external information. This form of tweets is quite frequent in French (more than 62.75% of our data contain explicit negation words), which suggests two hypotheses: (H1) negation can be a good indicator to detect irony, and (H2) external context can help to detect the absurdity of ironic content.

To validate if negation helps, we built binary classifiers using both state of the art features and new features (explicit and implicit opposition, sentiment shifter, discourse connectives). Overall accuracies were good when the data contain both tweets with negation and no negation but lower when tweets contain only negation or no negation at all. Error analysis show that major errors come from the presence of implicit oppositions, particularly in C_{Neg} and C_{All}. These results empirically validate hypothesis (H1). Negation has been shown to be very helpful in many NLP tasks, such as sentiment analysis (Wiegand et al., 2010). It has also been used as a feature to detect irony (Reyes et al., 2013). However, no one has empirically measured how irony classification behaves in the presence or absence of negation in the data.

To test (H2), we proposed a query-based method that corrects the classifier's outputs in order to retrieve false assertions. Our experiments show that the classification after applying Google searches in reliable web sites significantly improves the classifier accuracy when tested on C_{Neg}. In addition, we show that internal context features are useful to improve classification. These results empirically validate (H2). However, even though the algorithm improves the classifier performance, the number of queries is small which suggests that a much larger dataset is needed. As for negation, querying external source of information has been shown to give an improvement over the basic features for many NLP tasks (for example, in question-answering (Moldovan et al., 2002)). However, as far as we know, this approach has not been used for irony classification.

This study is a first step towards improving irony detection relying on external context. We plan to study other ways to retrieve such a context like the conversation thread.

Acknowledgements

This work was funded by the French National Research Agency (ASFALDA project ANR-12-CORD-023).

References

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

Francesco Barbieri and Horacio Saggion. 2014. Modelling Irony in Twitter: Feature Analysis and Evaluation. In *Proceedings of Language Resources and Evaluation Conference (LREC)*, pages 4258–4264.

Farah Benamara, Véronique Moriceau, and Yvette Yannick Mathieu. 2014. Fine-grained semantic categorization of opinion expressions for consensus detection (Catégorisation sémantique fine des expressions d'opinion pour la détection de consensus) [in French]. In *TALN-RECITAL 2014 Workshop DEFT 2014 : DÉfi Fouille de Textes (DEFT 2014 Workshop: Text Mining Challenge)*, pages 36–44, July.

Clint Burfoot and Clint 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.

Konstantin Buschmeier, Philipp Cimiano, and Roman Klinger. 2014. 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*, pages 42–49.

Paula Carvalho, Luís Sarmento, Mário J Silva, and Eugénio De Oliveira. 2009. Clues for detecting irony in user-generated contents: oh...!! it's so easy;-). In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 53–56. ACM.

Herbert H Clark and Richard J Gerrig. 1984. On the pretense theory of irony. *Journal of Experimental Psychology: General*, 113(1):121–126.

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*, CoNLL '10, pages 107–116.

Aniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, John Barnden, and Antonio Reyes. 2015. Semeval-2015 task 11: Sentiment Analysis of Figurative Language in Twitter. In *Proc. 9th Int. Workshop on Semantic Evaluation (SemEval 2015), Co-located with NAACL*, page 470478. Association for Computational Linguistics.

Raymond W Gibbs. 2000. Irony in talk among friends. *Metaphor and symbol*, 15(1-2):5–27.

Roberto Gonzalez-Ibanez, Smaranda Muresan, and Nina Wacholde. 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.

H Paul Grice, Peter Cole, and Jerry L Morgan. 1975. Syntax and semantics. *Logic and conversation*, 3:41–58.

Henk Haverkate. 1990. A speech act analysis of irony. *Journal of Pragmatics*, 14(1):77 – 109.

Christine Liebrecht, Florian Kunneman, and Bosch Antal van den. 2013. The perfect solution for detecting sarcasm in tweets# not. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 29–37. New Brunswick, NJ: ACL.

Dan I Moldovan, Sanda M Harabagiu, Roxana Girju, Paul Morarescu, V Finley Lacatusu, Adrian Novischi, Adriana Badulescu, and Orest Bolohan. 2002. LCC Tools for Question Answering. In *TREC*.

Antonio Reyes and Paolo Rosso. 2012. Making objective decisions from subjective data: Detecting irony in customer reviews. *Decision Support Systems*, 53(4):754–760.

Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, 47(1):239–268.

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*, pages 704–714.

Charlotte Roze, Laurence Danlos, and Philippe Muller. 2012. Lexconn: A French lexicon of discourse connectives. *Discours, Multidisciplinary Perspectives on Signalling Text Organisation*, 10:(on line).

J. Searle. 1979. *Expression and meaning: Studies in the theory of speech acts*. Cambridge University.

Cameron Shelley. 2001. The bicoherence theory of situational irony. *Cognitive Science*, 25(5):775–818.

Dan Sperber and Deirdre Wilson. 1981. Irony and the use-mention distinction. *Radical pragmatics*, 49:295–318.

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*.

Akira Utsumi. 1996. A unified theory of irony and its computational formalization. In *Proceedings of the 16th conference on Computational linguistics-Volume 2*, pages 962–967. Association for Computational Linguistics.

Akira Utsumi. 2004. Stylistic and contextual effects in irony processing. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society*, pages 1369–1374.

Po-Ya Angela Wang. 2013. #Irony or #Sarcasm-A Quantitative and Qualitative Study Based on Twitter.

Michael Wiegand, Alexandra Balahur, Benjamin Roth, Dietrich Klakow, and Andrés Montoyo. 2010. A Survey on the Role of Negation in Sentiment Analysis. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, pages 60–68. Association for Computational Linguistics.