Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing

Elena Filatova

Computer and Information Science Department
Fordham University
filatova@cis.fordham.edu

Abstract

The ability to reliably identify sarcasm and irony in text can improve the performance of many Natural Language Processing (NLP) systems including summarization, sentiment analysis, etc. The existing sarcasm detection systems have focused on identifying sarcasm on a sentence level or for a specific phrase. However, often it is impossible to identify a sentence containing sarcasm without knowing the context. In this paper we describe a corpus generation experiment where we collect regular and sarcastic Amazon product reviews. We perform qualitative and quantitative analysis of the corpus. The resulting corpus can be used for identifying sarcasm on two levels: a document and a text utterance (where a text utterance can be as short as a sentence and as long as a whole document).

1. Introduction

The ability to identifying sarcasm and irony has got a lot of attention recently. The task of irony identification is not just interesting. Many systems, especially those that deal with opinion mining and sentiment analysis, can improve their performance given the correct identification of sarcastic utterances (Hu and Liu, 2004; Pang and Lee, 2008; Popescu and Etzioni, 2005; Sarmento et al., 2009; Wiebe et al., 2004).

One of the major issues within the task of irony identification is the absence of an agreement among researchers (linguists, psychologists, computer scientists) on how one can formally define irony or sarcasm and their structure. On the contrary, many theories that try to explain the phenomenon of irony and sarcasm agree that it is impossible to come up with a formal definition of these phenomena. Moreover, there exists a belief that these terms are not static but undergo changes (Nunberg, 2001) and that sarcasm even has regional variations (Dress et al., 2008). Thus, it is not possible to create a definition of irony or sarcasm for training annotators to identify ironic utterances following a set of formal criteria. However, despite the absence of a formal definition for the terms irony and sarcasm, often human subjects have a common understanding of what these terms mean and can reliably identify text utterances containing irony or sarcasm.

There exist systems that target the task of automatic sarcasm identification (Carvalho et al., 2009; Davidov et al., 2010; Tsur et al., 2010; González-Ibáñez et al., 2011). However, the focus of this research is on sarcasm detection rather than on corpus generation. Also, these systems focus on identifying sarcasm at the sentence level (Davidov et al., 2010; Tsur et al., 2010), or via analyzing a specific phrase (Tepperman et al., 2006), or "exploring certain oral or gestural clues in user comments, such as emoticons, onomatopoeic expressions for laughter, heavy punctuation marks, quotation marks and positive interjections" (Carvalho et al., 2009).

It is agreed that many cases of sarcastic text utterances can be understood only when placed within a certain situation or within a broader text context. Thus, we set up our data collection experiment so that the collected corpus can be used for in-depth study of different linguistic phenomena that make a text utterance sarcastic or ironic. We use product reviews posted on www.Amazon.com. In contrast to the existing corpora that are also collected using Amazon product reviews (Davidov et al., 2010; Tsur et al., 2010), our corpus consists of text documents rather than separate sentences. In some cases whole documents can be considered sarcastic, in other cases only specific parts of documents are sarcastic. We believe that providing a context is of tremendous use for learning patterns of text utterances containing sarcasm. To better understand the phenomenon of sarcasm we collect *pairs* of Amazon product reviews, where both reviews are written for the same product, and one of the reviews is identified as sarcastic and the other one as regular (without sarcasm).

The rest of the paper is structured as follows. In Section 2. we describe the related work on the topic of sarcasm and irony detection as well as several examples of corpora generated for sarcasm detection. While discussing the related work we also provide the reasoning and examples that motivate us to create a new corpus with examples of documents containing sarcasm. In Section 3. we describe our procedure for collecting a set of documents that can be used for irony and sarcasm detection on two levels (document and sentences level). In Section 4. we analyze the collected corpus. Finally, in Section 5. we draw conclusions based on the collected corpus analysis and discuss the ways this corpus can be used within the task of sarcasm and irony detection. The corpus is available for downloading.[1]

2. Related Work

Myers (1977) notices that "Irony is a little bit like the weather: we all know it's there, but so far nobody has done much about it." A substantial body of research has been done since then in the fields of Psychology, Linguistics,

---

[1] `http://storm.cis.fordham.edu/~filatova/`
`SarcasmCorpus.html`

Media Studies and Computer Science to explain the phenomenon of sarcasm and irony. Several attempts have been made to create computational models of irony (Littman and Mey, 1991; Utsumi, 1996).

Though irony has many forms and is considered to be elusive (Muecke, 1970), researchers agree that there two distinct types of irony: verbal irony and situation irony. Verbal irony is often called sarcasm. Our corpus contain cases of both verbal irony or sarcasm and situational irony.

2.1. Understanding Irony

Littman and Mey (1991) identify and characterize three types of ironic situations[2] using a computational model irony. This model can be used to distinguish between ironic and non ironic situations. For this model to work, the information about such basic elements, as Agents, Goals, Plans, Effects, etc. is required, and often learning this information is a non-trivial task by itself. According to Lucariello (1994) unexpectedness is a central property of ironic events. Utsumi (1996) presents a model that given a brief description of a situation that involves one or more people identifies whether an utterance rendered by one of the situation participants is ironic or not.

Researchers who deal with verbal irony or sarcasm often base their theories on the violation of the quality maximum (*Be Truthful*), one of the four cooperative principals or maxims of the pragmatic theory (Grice, 1975). Sperber and Wilson (1981) treat verbal irony as a type of echoic allusion to an attributed utterance or thought. The literal meaning of an ironic statement "echoes" an expectation which has been violated. Clark and Gerrig (1984) propose a pretence-based explanation of irony, where the speaker of an ironical utterance is not performing a genuine speech act but merely pretending to perform one, while expecting her audience to see through the pretence and recognize the skeptical, mocking or contemptuous attitude behind it. Kumon-Nakamura et al. (1995) introduce an "allusional pretence" explanation of irony, where an ironical utterance must not only be pragmatically insincere (that is, a case not of saying but of making as if to say) but also allude to a failed expectation or norm.

There exist other linguistic and psychological theories of both verbal and situational irony. All these theories have differences in the interpretation of the phenomenon of irony, and none of the above theories give an exact definition of an irony. However, despite the absence of a formal definition of irony, the researchers agree that people, including 5-6 year old children (Creusere, 1999; Nakassis and Snedeker, 202), are usually good in recognizing irony. Kreuz and Caucci (2007) ran an experiment with examples from fiction. They collected examples containing the phrase "*said sarcastically*," removed this phrase, and presented the updated utterances to annotators. Psychology undergraduate students were asked to identify sarcastic utterances in the absence of the definition of the term sarcasm, and they reliably differentiated between sarcastic and non-sarcastic utterances.

---

[2]O. Henry's short story, *The Gift of the Magi*, presents a classic example of situational irony.

Sarcasm is often treated as a special case of irony, "Ironic insults, where the positive literal meaning is subverted by the negative intended meaning, will be perceived to be more positive than direct insults, where the literal meaning is negative" (Dews and Winner, 1995). Within developmental research (Creusere, 1999) sarcastic utterances are the utterances with positive literal meanings, negative intended meanings, and clear victims.

2.2. Automatic Irony Identification

Tepperman et al. (2006) work on sarcasm detection in speech. The phrase "*yeah right*" is classified whether it is used sarcastically or not according to the "prosodic, spectral, and contextual cues." The contextual cues used in this work are: laughter, whether the phrase comes in the beginning or at the end of the speaker's turn, etc.

Carvalho et al. (2009) collect opinionated user posts from the web site of a popular Portuguese newspaper. On average, user comments have about four sentences. The goal of the experiment is to identify which pre-defined patterns can be good indicators of sarcastic sentences which, otherwise, are classified as positive. According to the manual evaluation of the corpus sentences, the most reliable features signaling the presence of irony are: (i) emoticons and onomatopoeic expressions for laughter, (ii) heavy punctuation marks, (iii) quotation marks and (vi) *positive* interjections.

Following the experiment set-up used in by Kreuz and Caucci (2007), several corpora of Twitter messages are created (Davidov et al., 2010; González-Ibáñez et al., 2011). The Twitter messages in these corpora are explicitly marked by their authors with hashtags #sarcasm, #sarcastic. González-Ibáñez et al. (2011) classify the collected Twitter messages into sarcastic and straight-forwardly positive or negative messages. While Davidov et al. (2010) use the created collections of Twitter messages and sentences from Amazon product reviews to distinguish between sarcastic and non-sarcastic messages.

Tsur et al. (2010) analyze sentences extracted from Amazon product reviews. To run their classification algorithm, they first apply a semi-automatic procedure for corpus generation. They start with a small set of seed sentences that are classified on the 1..5 scale where 5 means a definitely sarcastic sentence and 1 means a clear absence of sarcasm. These seed sentences are used to extract features and construct a model that is used to automatically retrieve sarcastic sentences in Amazon product reviews. Another assumption that is used in this work to enrich the collection of sarcastic messages is that those sentences that appear next to sarcastic sentences are also likely to be sarcastic. Thus, the semi-automatically collected training corpus contains 471 positive examples (sarcastic sentences) and 5020 negative examples (non-sarcastic sentences).

One of the major characteristics of the above text corpora (both Twitter collections, the Amazon collection, as well as the Portuguese newspaper comments) is that the text utterances in these corpora are short. Twitter messages do not exceed 140 characters. As far as the Amazon corpus is concerned, it contains stand alone sentences extracted from product reviews rather than complete reviews, and these sentences are used independently of the reviews from

where they were extracted. Thus, the sentences in the Amazon corpus are analyzed outside their broader textual context.

However, it has been noted that in many cases a stand alone text utterance (e.g., sentence) cannot be reliably judged as ironic/sarcastic or not without the surrounding context. For example, the sentence "*Where am I?*" can be marked as ironic only if an annotator knows that this sentence comes from a review of a GPS device.[3]

Also, in many cases, only analyzing several sentences together can reveal the presence of irony. In the below example, if the two sentences are analyzed separately, each of them is considered as non-ironic. The utterance becomes ironic only if the two sentences are analyzed together.

> *Gentlemen, you can't fight in here! This is the War Room.*
>> P. Sellers as President Muffley in *Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb*, 1964

In our work we describe the procedure for corpus generation that contains examples of documents with and without irony. As noted above, in the absence of a strict definition of irony, it is impossible to train experts who could reliably identify irony, at the same time, the task of irony detection seems to be quite intuitive. Thus, to create a corpus of reasonable size we use the Mechanical Turk service[4] that allows to "crowdsource" labor intensive tasks and is now being used as a source of subjects for data collection and annotation in many fields (Paolacci et al., 2010), including NLP (Callison-Burch and Dredze, 2010). We use the studies conducted to assess the reliability of MTurk annotators (Ipeirotis et al., 2010) to ensure the quality of the collected data.

3. Corpus Generation

As noted, often a stand alone sentence cannot be reliably judged as either ironic or non-ironic without the surrounding context. Thus, we design our experiment to collect whole documents rather than sentences.

We use the reviews published on www.Amazon.com as the source of documents for our corpus. Amazon reviews are frequently used as the source of data for sentiment analysis (Ghose et al., 2007) and also have already been used for sarcasm detection (Tsur et al., 2010).

As discussed in the previous section, the current corpora dealing with sarcasm identification, consist of short text utterances (Tweets or sentences from Amazon reviews). To better understand the phenomenon of sarcasm we collect document (Amazon review) pairs describing the same product, where one review is judged as containing sarcasm and the other review is judged as a regular review (i.e., without sarcasm).

We collect a corpus that can be used for identifying sarcasm on macro (document) and micro (text utterance) level. Together with the product reviews we collect additional information that can be used by irony detection systems including the link to the product, the number of stars assigned

by the reviewer to the product under analysis, etc. Thus, this corpus can also be used to confirm the hypothesis that is frequently used in the definitions of irony, namely, that irony is often used to describe something negative.

To collect Amazon product review pairs we use the Mechanical Turk service which is now being widely used as a source of subjects for NLP data collection (Callison-Burch and Dredze, 2010). Our data collection experiment consists of two steps: a product review pairs collection step and a step that combines quality control and data analysis.

3.1. Step 1: Data Collection

First, we asked MTurkers to find pairs of Amazon reviews for the same product where one review contains sarcasm and the other - does not. Here are the instructions that we provided for this task:

- First review should be ironic or sarcastic. Together with this review you should
    1. cut-and-paste the text snippet(s) from the review that makes this review ironic/sarcastic
    2. select the review type: ironic, sarcastic or both (ironic and sarcastic)

- The other review should be a regular review (neither ironic nor sarcastic)

This task explanation was followed by a detailed outline of the review submission procedure. Thus, for each review that contains irony we obtained:

1. a permalink that can be used to retrieve the text of the product review together with many other useful information including, the number of stars assigned to the review;

2. ironic/sarcastic/both labeling that can be used to test our hypothesis on whether people can reliably distinguish between irony and sarcasm.

We asked for 1000 pairs of Amazon product reviews. However, some of the submitted data points missed the requested information and thus, were excluded from further consideration.

On purpose, we did not provide any guidelines regarding the size of the text snippet that makes a review sarcastic. By omitting the size restrictions we test whether it is possible to pinpoint in text the irony part. This information can also be used to analyze whether it is possible to localize irony in text as it has been done for opinion concentration (Brooke and Hurst, 2009). The size of the submitted text snippets varies from a phrase to the whole document.

After Step 1 is completed we perform a data cleaning procedure are remove from our corpus duplicates as well as the documents that are submitted as ironic but are not assigned the corresponding label. By duplicate submission we mean identical reviews. If for a certain product several different reviews are submitted we keep all these reviews. Thus, we end up with 1905 documents, most of which are paired into ironic – non-ironic pairs. However, for some of the submitted reviews, the counterpart review is deleted as part of data cleaning procedure. Thus, not all reviews in the collected corpus are paired.

---

[3]This example is from (Tsur et al., 2010).

[4]https://www.mturk.com

According to Ipeirotis et al. (2010) and Paolacci et al. (2010), the data submitted by MTurkers can contain noise and spam. To ensure the quality of our corpus, eliminate questionable documents, and also test several hypothesis about the nature of irony, we ran a second experiment on MTurk (Step 2).

3.2. Step 2: Data Quality Control

During Step 2 we recruit a new set of MTurkers so that each document collected during Step 1 is annotated by five new annotators. This step is designed to serve several goals. First, it allows us to eliminate those product reviews that are submitted as sarcastic but are either submitted by spammers or not considered as clearly sarcastic by other annotators. Second, we check whether MTurkers can guess the number of stars assigned to the product by the review author.

Our goal is to identify those documents that are clearly sarcastic or clearly regular (i.e., non-sarcastic). In the absence of a formal definition of sarcasm we use inter-annotator agreement to identify those documents that are considered as sarcastic by some annotators and as non-sarcastic by other and thus, should be eliminated from further consideration. We use two quality control procedures: simple majority voting and the data quality control algorithm that is based on computing Krippendorff's alpha coefficient to distinguish between reliable annotators and unreliable annotators; the labels from reliable annotators get high weight in computing the final label for a data point (Ipeirotis et al., 2010).

First, we use simple majority voting and keep for further consideration only those documents whose initial label (sarcastic or regular) is supported by at least three out of five new annotators. The labels: ironic, sarcastic, and both are all considered to support the same type of documents (i.e., sarcastic). After computing majority votes and leaving only those documents whose initial type is supported by at least three new voters, we end up with 486 Amazon reviews containing irony, and 844 regular Amazon reviews. However, the labels submitted during Step 2 can contain noise as well. Thus, using simple majority voting might not be the ideal approach for identifying incorrectly submitted documents.

Therefore, we compute a second metric to ensure the quality of the documents in our corpus. We apply the quality control algorithm designed specifically for quality management on Amazon Mechanical Turk and described in (Ipeirotis et al., 2010). This algorithm computes the reliability of every annotator and allows to rate the annotators based on the quality of their work. Using this algorithm we find those annotators that submit good quality labels and then keep only those documents that are labeled by these annotators and where the majority (out of five) reliable annotators agree that a document that was initially submitted as containing irony is indeed ironic, and a document that was initially submitted as regular indeed does not contain irony.

Thus, we use three-level quality control and keep in the final collection of documents only those whose initial type (sarcastic or regular) is supported by both majority voting using the labels obtained on Step 2 and the algorithm described in (Ipeirotis et al., 2010).

After the completion of the two-step data collection experiment and application of the data quality control management algorithm, we end up with 437 sarcastic reviews and 817 regular reviews. Out of these reviews we get 331 pairs of sarcastic and regular reviews submitted for the same Amazon product as well as 106 sarcastic and 486 regular unpaired reviews.

The number of regular reviews is almost twice as high as the number of sarcastic reviews. However, this characteristic of the corpus does not reflect the real distribution of ironic and regular reviews among Amazon product reviews. Rather, it reflects the fact that it is more difficult for people to agree on whether a review is ironic or not rather than on whether the review has or does not have irony.

According to the manual analysis of several cases that are initially submitted as ironic but judged as regular by the spammer elimination software, we can say that the applied criteria for leaving reviews in the corpus is very strict. For example, the below review is submitted as ironic during Step 1, however, during Step 2 one MTurker considered this review ironic, one - sarcastic, and three other MTurkers considered this review to be a regular review. Thus, despite the fact that the authors of this paper believe that this review is sarcastic and that the utterance submitted for this review provides enough proof of the sarcastic nature of this review, nonetheless, this review is not included into our corpus.[5]

- *Submitted review label:* ironic

- *Product:* Moon (2009)

- *Stars:* 1 out of 5

- *Review Title:* Boring and unoriginal. I liked "District 9" much more.

- *Review:* This movie didn't really do anything for me. It was just a variation of the theme of a person isolated from other people. It didn't even bring anything new about the theme to the table. This type of situation has been done so many times before (Castaway, Robinson Crusoe, I Am Legend, etc.)in ways that I thought were more interesting. Copying HAL from 2001 didn't help in the originality department, either. I have to admit I couldn't get into the movie from the beginning because I couldn't believe they would send ONE person into space for three years- ONE person! After that, I just couldn't get into it (so I probably missed/forgot some things). I also think that a clone of someone wouldn't have such a different personality from the original person if that clone had only been in that kind environment. If they can clone him, why don't they just have a bunch of clones up there? They'd get things done in a lot shorter time than three years. Also, why did it take him almost three years to go nearly insane? I probably would have lost it after three days.

---

[5]In all the examples of Amazon product reviews presented in this paper we preserve the authors' punctuation and spelling.

- *Utterance that makes this review ironic:* I have to admit I couldn't get into the movie from the beginning because I couldn't believe they would send ONE person into space for three years- ONE person!

Interestingly, there is some level of disagreement not only regarding whether a review is ironic, but also some of the reviews that were initially submitted as regular were assigned a higher probability of being ironic after Step 2. Most of such reviews are written for the type of Amazon products for which users write mainly sarcastic reviews such as uranium ore, Zenith watch that is sold at 40% discount price for mere $86,999.99, etc.

3.2.1. Guessing Star Rating

Finally, we check whether MTurkers can guess the number of stars assigned to a product based on its review. On Step 2, we provide annotators with the plain text of a review (without its URL or the number of stars assigned by the review's author to the product under analysis) and ask to submit the number of stars that this review is likely to give to the product. For each review we get five MTurkers guessing the number of stars assigned by this review to a product and compute the average of these five values. We then compute the correlation of this average with the initial number of stars assigned to the product based on the review under analysis. This correlation is quite high: 0.889 – for all reviews, 0.821 – for sarcastic reviews, and 0.841 – for regular reviews.

4. Corpus Analysis

One of the key characteristics of sarcasm that is described by linguists and psychologists, and exploited by sarcasm identification system is the fact that a sarcastic reviews use positive words but express negative opinion. Negative opinion is captured on Amazon by low star ratings assigned by review authors. As expected, in the collected corpus, the majority of the sarcastic reviews are written by people who assign low scores to the reviewed products: 59.94% - one-star reviews. The majority of regular reviews (74%), on the other hand, assign 5 stars to the reviewed products. We explain this phenomenon by the fact that annotators who submit reviews for our corpus during Step 1 of the corpus collection experiment have a general understanding for the irony and sarcasm phenomena as figures of speech that change the positive literary meaning of a text utterance to negative. Thus, it is easier to find sarcastic reviews among those that assign low scores to products, and the chance of finding an ironic review among the reviews that assign high scores to a product is low. The distribution of the reviews according to how many stars are assigned to them is presented in Table 1

		Number of reviews with				
		1 ⋆	2 ⋆	3 ⋆	4 ⋆	5 ⋆
sarcastic	437	262	27	20	14	114
regular	817	64	17	35	96	605

Table 1: Distribution of stars assigned to reviews.

Many of the five-start sarcastic reviews are written for the products that the review authors consider strange/funny/odd and thus, these ironic reviews are mostly show cases of the review authors' wits. Examples of products whose ironic reviews assigned five stars to these products include uranium ore, a herpes plush doll, and even a T-shirt with a certain print can be a reason for people to exercise her wit (see Figure 1):

Figure 1: The Mountain Three Wolf Moon Short Sleeve Tee.

- *Submitted review label:* ironic

- *Product:* The Mountain Three Wolf Moon Short Sleeve Tee.

- *Stars:* 5 out of 5

- *Review Title:* Three wolves is just two wolves plus another wolf.

- *Review:* I had a two-wolf shirt for a while and I didn't think life could get any better. I was wrong. Life got 50% better, no lie.

4.1. What Makes an Sarcastic Review Sarcastic

During Step 1 of our data collection experiment we asked MTurkers to submit text utterances in support of the presence of irony in the product review. The length of such utterances varies. For example, the presence of irony in the review of the Canada Green Grass Seed is supported by the following sentence: "*Perhaps it grows in Canada but not Washington State!!.*" In some cases irony is supported by several text utterances that can be consecutive or can be extracted from different parts of the review. In other cases, a complete review text is submitted in support of its sarcastic nature.

5. Conclusion

In this paper we describe a corpus generation experiment whose goal is to obtain Amazon product review pairs that can be used to analyze the notion of sarcasm in text and for training sarcasm detection system. Our corpus can be used for understanding sarcasm on two levels: document and text utterance level. Using this corpus we test and confirm two hypothesis:

- Sarcasm is often used in the reviews that give a negative score to the product under analysis;

- The presence of irony in a product review does not affect the readers' understanding of the product quality: given the text of the review (irrespectively, whether it is a sarcastic or a regular review), people are good in understanding the attitude of the review author to the product under analysis and can reliably guess how many stars the review author assigned to the product.

6. References

Julian Brooke and Matthew Hurst. 2009. Patterns in the stream: Exploring the interaction of polarity, topic, and discourse in a large opinion corpus. In *Proceedings of the ACM Conference on Information and Knowledge Management, 1st International Workshop on Topic-Sentiment Analysis for Mass Opinion Measurement*, Hong Kong, November.

Chris Callison-Burch and Mark Dredze. 2010. Creating speech and language data with Amazon's Mechanical Turk. In *Proceeding of the NAACL workshop on Creating Speech and Language Data With Mechanical Turk (NAACL–HLT 2010)*, Los Angeles, CA, USA, June.

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 *Proceeding of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, TSA '09, pages 53–56, New York, NY, USA. ACM.

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

Marlena Creusere. 1999. Theories of adults understanding and use of irony and sarcasm: Applications to and evidence from research with children. *Developmental Review*, 19:213–262.

Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In *Proceeding of Computational Natural Language Learning (CoNLL 2010)*, Uppsala, Sweden, July.

Shelly Dews and Ellen Winner. 1995. Muting the meaning: A social function of irony. *Metaphor and Symbolic Activity*, 10(1):3–19.

Megan Dress, Roger Kreuz, Kristen Link, and Gina Caucci. 2008. Regional variation in the use of sarcasm. *Journal of Language and Social Psychology*, 27(1):71–85.

Anindya Ghose, Panagiotis G. Ipeirotis, and Arun Sundararajan. 2007. Opinion mining using econometrics: A case study on reputation systems. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, Czech Republic, July.

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 ()*, pages 581–586, Portland, Oregon, USA, June. Association for Computational Linguistics.

Paul Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3: Speech Acts:41–58.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD 2004)*, pages 168–177, Seattle, WA, USA, August.

Panagiotis Ipeirotis, Foster Provost, and Jing Wang. 2010. Quality management on Amazon Mechanical Turk. In *Proceedings of the Second KDD Human Computation Workshop (KDD-HCOMP 2010)*, Washington DC, USA, July.

Roger Kreuz and Gina Caucci. 2007. Lexical influences on the perception of sarcasm. In *Proceedings of the NAACL workshop on Computational Approaches to Figurative Language (HLT–NAACL 2007)*, Rochester, NY, USA, April.

Sachi Kumon-Nakamura, Sam Glucksberg, and Mary Brown. 1995. How about another piece of pie: The allusional pretense theory of discourse irony. *Journal of Experimental Psychology: General*, 124(1):3–21.

David C. Littman and Jacob L. Mey. 1991. The nature of irony: Toward a computational model of irony. *Journal of Pragmatics*, 15(2):131–151.

Joan Lucariello. 1994. Situational irony: A concept of events gone awry. *Journal of Experimental Psychology: General*, 123(2):129–145.

Douglas Colin Muecke. 1970. Irony and the ironic. *The Critical Idiom*, 13.

Alice R. Myers. 1977. Toward a definition of irony. In Ralph W. Fasold and Roger W. Shuy, editors, *Studies in language variation: semantics, syntax, phonology, pragmatics, social situations, ethnographic approaches*, pages 171–183. Georgetown University Press.

Constantine Nakassis and Jesse Snedeker. 202. 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*, Boston, MA, USA, July.

Geoffrey Nunberg. 2001. *The Way we Talk Now: Commentaries on Language and Culture*. Boston: Houghton Mifflin.

Bo Pang and Lillian Lee. 2008. *Opinion Mining and Sentiment Analysis.*, volume 2. Foundations and Trends® in Information Retrieval.

Gabriele Paolacci, Jesse Chandler, and Panagiotis G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), August.

Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT–EMNLP 2005)*, pages 339–346, Vancouver, B.C., Canada, October. Morristown, NJ, USA: Association for Computational Linguistics.

Luís Sarmento, Paula Carvalho, Mário J. Silva, and Eugénio de Oliveira. 2009. Automatic creation of a reference corpus for political opinion mining in user-generated content.

Dan Sperber and Deirdre Wilson. 1981. Irony and the use-

mention distinction. *Radical pragmatics*, pages 295–318.

Joseph Tepperman, David Traum, and Shrikanth S. Narayanan. 2006. "Yeah right": Sarcasm recognition for spoken dialogue systems. In *Proceedings of InterSpeech*, pages 1838–1841, Pittsburgh, PA, USA, September.

Oren Tsur, Dmitry Davidov, and Ari Rappoport. 2010. ICWSM - A great catchy name: Semi-supervised recognition of sarcastic sentences in product reviews. In *Proceeding of AAAI Conference on Weblogs and Social Media (ICWSM–10)*, Washington, DC, USA, May.

Akira Utsumi. 1996. A unified theory of irony and its computational formalization. In *Proceedings of the 16th conference on Computational linguistics (COLING 1996)*, pages 962–967, Stroudsburg, PA, USA. Association for Computational Linguistics.

Janyce Wiebe, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin. 2004. Learning subjective language. *Computational Linguistics*, 30:277–308.

Deirdre Wilson and Dan Sperber. 2004. Relevance theory. In L. R. Horn and G. Ward, editors, *The Handbook of Pragmatics*, pages 607–632. Oxford: Blackwell.