With 1 Follower I Must Be AWESOME :P.
Exploring the Role of Irony Markers in Irony Recognition

Debanjan Ghosh, § Smaranda Muresan ‡
§ School of Communication & Information, Rutgers University, NJ, USA
‡ Data Science Institute, Columbia University, NY, USA
debanjang.ghosh@rutgers.edu, smara@columbia.edu

Abstract

Conversations in social media often contain the use of irony or sarcasm, when the users say the opposite of what they really mean. Irony markers are the meta-communicative clues that inform the reader that an utterance is ironic. We propose a thorough analysis of theoretically grounded irony markers in two social media platforms: Twitter and Reddit. Classification and frequency analysis show that for Twitter, typographic markers such as emoticons and emojis are the most discriminative markers to recognize ironic utterances, while for Reddit the morphological markers (e.g., interjections, tag questions) are the most discriminative.

Introduction

With the advent of social media, irony and sarcasm detection has become an active area of research in Natural Language Processing (NLP) (Joshi, Bhattacharyya, and Carman 2016; Riloff et al. 2013; Joshi, Sharma, and Bhattacharyya 2015; Ghosh, Fabbri, and Muresan 2017). Most computational studies have focused on building state-of-the-art models to detect whether an utterance or comment is ironic/sarcastic or not, sometimes without theoretical grounding. In linguistics and discourse studies, Attardo (2000) and later Burgers (2010) have studied two theoretical aspects of irony in the text: irony factors and irony markers. Irony factors are characteristics of ironic utterances that cannot be removed without destroying the irony. In contrast, irony markers are a meta-communicative clue that “alert the reader to the fact that a sentence is ironical” (Attardo 2000). They can be removed and the utterance is still ironic.

In this paper, we examine the role of irony markers in social media for irony recognition. Although punctuations, capitalization, and hyperboles are previously used as features in irony detection (Bamman and Smith 2015; Muresan et al. 2016), here we thoroughly analyze a set of theoretically-grounded types of irony markers, such as tropes (e.g., metaphors), morpho-syntactic indicators (e.g., tag questions), and typographic markers (e.g., emojis) and their use in ironic utterances. Consider the following two irony examples from Twitter and Reddit given in Table 1.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reddit</td>
<td>Are you telling me iPhone 5 is only marginally better than iPhone 4S? I thought we were reaching a golden age with this game-changing device. /s</td>
</tr>
<tr>
<td>Twitter</td>
<td>With 1 follower I must be AWESOME. :P # ironic</td>
</tr>
</tbody>
</table>

Table 1: Use of irony markers in two social media platforms

Both utterances are labeled as ironic by their authors (using hashtags in Twitter and the /s marker in Reddit). In the Reddit example, the author uses several irony markers such as rhetorical question (e.g., “are you telling” . . .) and metaphor (e.g., “golden age”). In the Twitter example, we notice the use of capitalization (“AWESOME”) and emoticons (“:*” (tongue out)) that the author uses to alert the readers that it is an ironic tweet.

We present three contributions in this paper. First, we provide a detailed investigation of a set of theoretically-grounded irony markers (e.g., tropes, morpho-syntactic, and typographic markers) in social media. We conduct the classification and frequency analysis based on their occurrence. Second, we analyze and compare the use of irony markers on two social media platforms (Reddit and Twitter). Third, we provide an analysis of markers on topically different social media content (e.g., technology vs. political subreddits).

Data

Twitter: We use a set of 350K tweets for our experiments. The ironic/sarcastic tweets are collected using hashtags, such as #irony, #sarcasm, and #sarcastic whereas the non-sarcastic tweets do not contain these hashtags, but they might include sentiment hashtags, such as #happy, #love, #sad, #hate (similar to (González-Ibáñez, Muresan, and Waecholder 2011; Ghosh, Guo, and Muresan 2015)). As preprocessing, we removed the retweets, spam, duplicates, and tweets written in languages other than English. Also, we deleted all tweets where the hashtags of interest were not located at the very end (i.e., we eliminated “#sarcasm is something that I love”). We lowercased the tweets, except the words where all the characters are uppercased.

Reddit: Khodak, Saunshi, and Vodrahalli (2018) introduced an extensive collection of sarcastic and non-sarcastic...
posts collected from different subreddits. In Reddit, authors mark their sarcastic intent of their posts by adding “/s” at the end of a post/comment. We collected 50K instances from the corpus for our experiments (denoted as Reddit), where the sarcastic and non-sarcastic replies are at least two sentences (i.e., we discard posts that are too short).

For brevity, we denote ironic utterances as I and non-ironic utterances as NI. Both Twitter and Reddit datasets are balanced between the I and NI classes. We use 80% of the datasets for training, 10% for development, and the remaining 10% for testing.

Irony Markers

Three types of markers — tropes, morpho-syntactic, and typographic are used as features.

Tropes:

Tropes are figurative use of expressions.

- Metaphors - Metaphors often facilitate ironic representation and are used as markers. We have drawn metaphors from different sources (e.g., 884 and 8,600 adjective/noun metaphors from (Tsvetkov et al. 2014) and (Gutiérrez et al. 2016), respectively, and used them as binary features. We also evaluated the metaphor detector (Rei et al. 2017) over Twitter and Reddit datasets. We considered metaphor candidates that have precision $\geq 0.75$ (see Rei et al. (2017)).

- Hyperbole - Hyperboles or intensifiers are commonly used in irony because speakers frequently overstate the magnitude of a situation or event. We use terms that are denoted as “strong subjective” (positive/negative) from the MPQA corpus (Wilson, Wiebe, and Hoffmann 2005) as hyperboles. Apart from using hyperboles directly as the binary feature we also use their sentiment as features.

- Rhetorical Questions - Rhetorical Questions (for brevity RQ) have the structure of a question but are not typical information seeking questions. We follow the hypothesis introduced by Oraby et al. (2017) that questions that are in the middle of a comment are more likely to be RQ since questions followed by text cannot be typical information seeking questions. Presence of RQ is used as a binary feature.

Morpho-syntactic (MS) irony markers:

This type of markers appear at the morphologic and syntactic levels of an utterance.

- Exclamation - Exclamation marks emphasize a sense of surprise on the literal evaluation that is reversed in the ironic reading (Burgers 2010). We use two binary features, single or multiple uses of the marker.

- Tag questions - We built a list of tag questions (e.g., “didn’t you?”, “aren’t we?”) from a grammar site and use them as binary indicators.2

- Hashtag - Particularly in social media. Figure 1 shows a tweet with two emojis (e.g., “unassumed” and “confounded” faces respectively) used as markers. We use an emoji library of 1,400 emojis to identify the particular emoji used in ironic utterances and use them as binary indicators.5

- Emoticon - Emoticons are frequently used to emphasize the ironic intent of the user. For example, in the ironic tweet “nice to wake up to cute text. #suck”, the hashtag “#suck” depicts the negative sentiment. We use binary sentiment feature (positive or negative) to identify the sentiment of the hashtag, while comparing against the MPQA sentiment lexicon. Often multiple words are combined in a hashtag without spacing (e.g., “fun” and “night” in #fun-night). We use an off-the-shelf tool to split words in such hashtags and then checked the sentiment of the words.3

- Interjections - Interjections seem to undermine a literal evaluation and occur frequently in ironic utterances (e.g., “‘yeah”, ‘wow”, “yay”, “ouch” etc.). Similar to tag questions we assembled interjections (a total of 250) from different grammar sites.

Typographic irony markers:

- Capitalization - Users often capitalize words to represent their ironic use (e.g., the use of “GREAT”, “SO”, and “WONDERFUL” in the ironic tweet “GREAT i’m SO happy shattered phone on this WONDERFUL day!!!”).

- Quotation mark - Users regularly put quotation marks to stress the ironic meaning (e.g., “great” instead of GREAT in the above example).

- Other punctuation marks - Punctuation marks such as “?”, “.”, “;” and their various uses (e.g., single/multiple/mix of two different punctuations) are used as features.

- Hashtag - Particularly in Twitter, hashtags often represent the sentiment of the author. For example, in the ironic tweet “nice to wake up to cute text. #suck”, the hashtag “#suck” depicts the negative sentiment. We use binary sentiment feature (positive or negative) to identify the sentiment of the hashtag, while comparing against the MPQA sentiment lexicon. Often multiple words are combined in a hashtag without spacing (e.g., “fun” and “night” in #fun-night). We use an off-the-shelf tool to split words in such hashtags and then checked the sentiment of the words.4

- Emoticon - Emoticons are frequently used to emphasize the ironic intent of the user. For example, in the example “I love the weather :) #irony”, the emoticon “:)” (wink) alerts the reader to a possible ironic interpretation of weather (i.e., bad weather). We collected a comprehensive list of emoticons (over one-hundred) from Wikipedia and also used standard regular expressions to identify emoticons in our datasets.4 Beside using the emoticons directly as binary features, we use their sentiment as features as well (e.g., “wink” is regarded as positive sentiment in MPQA).

- Emoji - Emojis are like emoticons, but they are actual pictures and recently have become very popular in social media. Figure 1 shows a tweet with two emojis (e.g., “unassumed” and “confounded” faces respectively) used as markers. We use an emoji library of 1,400 emojis to identify the particular emoji used in ironic utterances and use them as binary indicators.5

---

2http://www.perfect-english-grammar.com/tag-questions.html

3https://github.com/matchado/HashTagSplitter

4http://sentiment.christopherpotts.net/code-data/

5https://github.com/vdurmont/emoji-java
Classification Experiments and Results

We first conduct a binary classification task to decide whether an utterance (e.g., a tweet or a Reddit post) is ironic or non-ironic, exclusively based on the irony marker features. We use Support Vector Machines (SVM) classifier with linear kernel (Fan et al. 2008). Table 2 and Table 3 present the results of the ablation tests for Twitter and Reddit. We report Precision ($P$), Recall ($R$) and F1 scores of both I and NI categories.

Table 2 shows that for ironic utterances in Twitter, removing tropes have the maximum negative effect on Recall, with a reduction on F1 score by 15%. This is primarily due to the removal of hyperboles that frequently appear in ironic utterances in Twitter. Removing typographic markers (e.g., emojis, emoticons, etc.) have the maximum negative effect on the Precision for the irony I category, since particular emojis and emoticons appear regularly in ironic utterances (Table 4). For Reddit, Table 3 shows that removal of typographic markers such as emoticons does not affect the F1 scores, whereas the removal of morpho-syntactic markers, e.g., tag questions, interjections have a negative effect on the F1.

Table 4 and Table 5 represent the top most discriminative features for both categories based on the feature weights learned during the SVM training for Twitter and Reddit, respectively. Table 4 shows that for Twitter, typographic features such as emojis and emoticons have the highest feature weights for both categories. Interestingly, we observe that for ironic tweets users often express negative sentiment directly via emojis (e.g., angry face, rage) whereas for non-ironic utterances, emojis with positive sentiments (e.g., hearts, wedding) are more familiar. For Reddit (Table 5), we observe that instead of emojis, other markers such as exclamation marks, negative tag questions, and metaphors are discriminatory markers for the irony category. In contrary, for the non-irony category, positive tag questions and negative sentiment hyperboles are influential features.

Frequency analysis of markers

We also investigate the occurrence of markers in the two platforms via frequency analysis (Table 7). We report the mean of occurrence per utterance and the standard deviation (SD) of each marker. Table 7 shows that markers such as hyperbole, punctuations, and interjections are popular in both platforms. Emojis and emoticons, although the two most popular markers in Twitter are almost unused in Reddit. Exclamations and RQs are more common in the Reddit corpus. Next, we combine each marker with the type they belong to (i.e., either trope, morpho-syntactic and typographic) and compare the means between each pair of types via independent t-tests. We found that the difference of means is significant ($p \leq 0.005$) for all pair of types across the two platforms.

Irony markers across topical subreddits

Finally, we collected another set of irony posts from (Khedak, Saunshi, and Vodrahalli 2018), but this time we collected posts from specific topical subreddits. We collected irony posts about politics (e.g., subreddits: politics, hillary, the_donald), sports (e.g., subreddits: nba, football, soccer), religion (e.g., subreddits: religion) and technology (e.g., subreddits: technology). Table 6 presents the mean and SD for each genre. We observe that users use tropes such as hyperbole and RQ, morpho-syntactic markers such as exclamation and interjections and multiple-punctuations more in politics and religion than in technology and sports. This is expected since subreddits regarding politics and religion are often more controversial than technology and sports and the users might want to stress that they are ironic or sarcastic using the markers.

Conclusion

We provided a thorough investigation of irony markers across two social media platforms: Twitter and Reddit. Classification experiments and frequency analysis suggest that
RQ. 0.06 (0.23) 0.11 (0.32) 0.07 (0.44) 0.01 (0.32) 0.01 (0.32)

MS

Exclamation

Tag Question

Interjection

Capitalization

Typographic

Punctuations

Table 6: Frequency of irony markers in different genres (subreddits). The mean and the SD (in bracket) are reported.

Table 7: Frequency of irony markers in two platforms. The mean and the SD (in bracket) are reported.

typographic markers such as emojis and emoticons are most frequent for Twitter whereas tag questions, exclamation, metaphors are frequent for Reddit. We also provide an analysis across different topical subreddits. In future, we are planning to experiment with other markers (e.g., ironic echo, repetition, understatement).

References


Gutiérrez, E. D.; Shutova, E.; Marghetis, T.; and Bergen, B. 2016. Literal and metaphorical senses in compositional distributional semantic models. In *ACL (1).*


