

Chinese Irony Corpus Construction and Ironic Structure Analysis

Yi-jie Tang and Hsin-Hsi Chen

Department of Computer Science and Information Engineering
National Taiwan University, Taipei, Taiwan
tangyj@nlg.csie.ntu.edu.tw, hhchen@ntu.edu.tw

Abstract

Non-literal expression recognition is a challenging task in natural language processing. An ironic expression implies the opposite of the literal meaning, causing problems in opinion mining and sentiment analysis. In this paper, ironic messages are collected from microblogs to form an irony corpus based on the use of emoticons, linguistic forms, and sentiment polarity. Five linguistic patterns are mined by using the proposed bootstrapping approach. We also analyze the linguistic structure and elements used to convey irony. Based on our observations, ironic words/phrases and contextual information are the necessary elements in irony, while the contextual information can be hidden in linguistic forms. A rhetorical element, which is optional in irony, can also be used to help strengthen the effects and understandability of an ironic expression. The ironic elements in each instance of our irony corpus are labelled based on this structure. This corpus can be used to study the usage of ironic expressions and the identification of ironic elements, and thus improve the performance of irony recognition.

1 Introduction

Dealing with non-literal meaning is a challenging task in natural language processing. Linguistic context and background knowledge are required to interpret non-literal utterances properly. An ironic expression, where the meaning is the opposite of what is literally expressed, is one of the indirect and non-literal linguistic forms that cannot be easily processed and detected. One cannot capture the real meanings of opinions and sentiments expressed in a document or conversation if irony is not taken into account.

The challenges of irony processing involve the following issues: (1) No comprehensive irony corpus is available. (2) Irony analysis is related to semantics, pragmatics and discourse studies, which are the most challenging in natural language processing. (3) Contextual information and background knowledge are necessary, but they are hard to obtain and process. (4) Non-linguistic or non-verbal factors, e.g., intonations, gestures and talking speed in speech, and spaces, punctuations and typography in writing, have to be considered.

This paper focuses on irony corpus construction, ironic pattern mining, and ironic structure analysis. Messages were collected from a microblogging platform based on emoticons, and ironic messages and patterns were extracted to build an irony corpus. The structure of ironic expressions and the clarification of the uses of ironic elements were also analyzed. Labels representing the ironic elements are added to each message in the irony corpus. To the best of our knowledge, this is the first Chinese irony corpus available for research.

This paper is organized as follows. Section 2 surveys the related work. Section 3 proposes a methodology to construct an irony corpus. Section 4 presents the patterns mined from the corpus. Section 5 discusses the results of ironic expressions collected from a different type of corpus. Section 6 makes the error analysis. Section 7 analyzes linguistic structure of Chinese irony. Section 8 concludes the remarks.

This work is licensed under a Creative Commons Attribution 4.0 International License. Page numbers and proceedings footer are added by the organizers. License details: <http://creativecommons.org/licenses/by/4.0/>

2 Related Work

Sarcasm and irony have been studied by linguistics and cognitive scientists (Giora and Fein, 1999; Gibbs and Colston, 2007) for years, but there has been no concrete claim on the linguistic structure of irony. Some studies have started focusing on the processing of sarcasm and irony recently, but it is still not clear whether sarcasm and irony differ significantly or represent the same concept.

The research of non-literal expression identification has drawn attention in recent years. Katz and Giesbrecht (2006) use meaning vectors for literal and non-literal expression classification. Li and Sporleder (2010) focus on distinguishing literal and non-literal usages of idioms.

Filatova (2012) uses crowdsourcing to generate an irony and sarcasm corpus. Veale and Hao (2010) construct a corpus of ironic similes using the wildcarded query “as * as a *” on a search engine. Davidov et al. (2010) collect messages from Twitter and product reviews from Amazon.com using the Mechanical Turk service. The #sarcasm hashtag is used as ground truth, and a k-nearest neighbor strategy is used for classification. González-Ibáñez et al. (2011) also make use of hashtags in Twitter as labels to build a sarcasm corpus. In their study, both human classification and automatic classification achieve low accuracy in sarcasm detection. Reyes et al. (2012) analyze humor and irony based on the user-generated tags, such as “#humor” and “#irony”, in twitter. Lukin and Walker (2013) use a bootstrapping method to improve the performance of the classifiers for identifying sarcastic and nasty utterances in online dialogues.

The hashtag-based approaches are not always suitable for irony corpus construction for all the languages. As of March 9, 2014, only 113 messages are found to contain the hashtag #反諷 (#irony) in Weibo, the largest Chinese language microblogging platform. This paper differs from the previous work in that we employ negative emoticons and positive words as clues to capture the irony. The linguistic patterns mined from the irony corpus can be used to detect if a sentence is ironic.

3 Irony Corpus Generation from Microblogs

This section introduces a bootstrapping methodology to construct an irony corpus and mine irony patterns. While Lukin and Walker (2013) also used a bootstrapping method to improve sarcasm and nastiness classifiers, this paper, in contrast, focuses on irony pattern mining and corpus construction.

3.1 An Emotion-Tagged Corpus

The traditional definition of verbal irony is adopted, where the speaker says something that seems to be the opposite of what they mean (Gibbs and Colston, 2007). Under this definition, texts annotated with polarity information that expresses the actual meaning should be collected, and the literal meanings of words in the texts should be identified. If any disagreement exists between the actual meaning and literal meaning, then we say the text contains irony.

Nowadays, emoticons are used quite often in social media to express the feelings of the posters. The tagged emoticons specify their actual meanings in some sense. Based on this idea, messages were collected from Plurk¹, a microblogging platform similar to Twitter. It lets users post messages limited to 140 characters, and allows them to use graphical emoticons in their messages.

It was assumed that these emoticons can represent the poster’s sentiments, and, therefore, be regarded as sentiment labels of the messages. Among 35 emoticons, 23 are categorized into positive, and 12 are categorized into negative. Collected messages are dated from Jun 21, 2008 to Nov 7, 2009, and all of them are in Traditional Chinese.

On the other hand, the literal meanings of the posted messages need to be known. Many sentiment analysis algorithms (Liu, 2012) can be explored. A lexicon-based approach was adopted. The NTU Sentiment Dictionary, or NTUSD (Ku and Chen, 2007), was employed to determine the sentiment of a word. This dictionary provides 21,056 positive and 22,751 negative words. Most of these words are in Traditional Chinese.

¹ <http://www.plurk.com>

3.2 Candidates Extraction

Possible irony messages were extracted from the Plurk corpus by using NTUSD. Since the typical social function of irony is expressing negative meaning with positive words, as mentioned in Gibbs and Colston (2007), focus was directed on those messages with negative emoticons and positive words. A total of 3,178,372 messages was found containing at least one negative emoticon. Among them, 304,754 messages with at least one positive word are found and form an irony candidate dataset.

Discourse relation determines how two discourse units cohere to each other. Sentiment transition of two clausal arguments is identified based on their discourse relation (Zhou et al., 2011; Wang et al., 2012; Huang et al., 2013). In the sentence “he is nice but not attractive,” positive opinion in the beginning is transformed to a negative one by the discourse connective “but.” Both the positive word “nice” and the negative phrase “not attractive” are used literally. Thus, it was necessary to filter out messages containing such connectives.

Messages are removed only when the positive word occurs earlier than the discourse connectives with a comparison function, due to Chinese grammatical structure. The Chinese discourse connectives used here include “但”, “但是”, “可是”, “只是”, “不過” (all the above are equivalent to the English word *but*), “然而” (however), “卻” (comparatively), “可惜” (unfortunately), “偏偏” (contrarily), “反而” (oppositely), and “倒是” (on the contrary). A total of 254,836 messages remains after this process.

3.3 Pattern Mining

Although irony can be used without any customary linguistic patterns, some ironic expressions do exhibit specific forms of language use. Colston and O’Brien (2000) suggest that both irony and hyperbole create contrasts between expected and ensuing events. It was assumed that exaggerated expressions could be used with irony to strengthen the effects of the speech act. In the expression 我真是太幸運啦! (I am really and extremely lucky!), the adverbs *really* and *extremely* are used to strengthen the ironic effect. Thus, combinations of degree adverb phrases and a positive adjective are used as patterns to find possible irony expressions automatically in the candidate dataset.

Not all degree adverbs in Chinese are used because some of them are mostly used in formal texts and not frequently present in microblogs. The degree adverb phrases used here include the combinations of the adverbs “還” (*hái*), “也” (*yě*), “未免” (*wèimiǎn*), “可” (*kě*) and “實在” (truly) and the degree adverbs “真” (really), “太” (extremely) and “非常” (very).

The following bootstrapping procedure was used to find more patterns.

(1) Which patterns should be used is decided. At the very beginning of the bootstrapping procedure, the [degree adverb + positive adjective] pattern mentioned above is used.

(2) Messages containing the patterns in step (1) are automatically retrieved from the candidates. NTUSD is used to determine sentiment polarity, and CKIP parser is used to get parts of speech².

(3) Messages retrieved in step (2) were reviewed by the annotator to decide which of them are actually ironic.

(4) If the annotator finds new irony patterns in the reviewed messages, then the procedure starts again from step (1) and uses the patterns to repeat the process.

This process was repeated for four times. After the fourth iteration, no more new patterns were found by the annotator. Finally, 2,825 messages are found to have any of the patterns, and 1,005 of them are confirmed to be ironic and make up the NTU Irony Corpus.³ Examples of these patterns and ironic messages are shown in Section 4.

4 Irony Patterns

All the patterns mined by the approach used in Section 3 are categorized into the following five groups.

4.1 Degree Adverbs + Positive Adjective

In this pattern, the following two components must exist:

² <http://ckipsvr.iis.sinica.edu.tw>.

³ The NTU Irony Corpus is available at http://nlg.csie.ntu.edu.tw/nlpresource/irony_corpus/.

- (a) Degree adverb phrase + positive adjective phrase
- (b) Negative context

The negative context can occur either before or after the component (a). For example, the following expression is used when someone has to wait for a long time to start ordering in a restaurant. In total, 13.03% of all the messages in the corpus contain this pattern.

- (s1) 點餐都要等半小時，服務還真是好阿
I have to wait for half an hour to order. The service is definitely really good.

The underlined expression is the contextual information described in (b), and the double-underlined expression is the linguistic form described in (a). In the second clause the adverbs “還” (*hái*) and “真” (really) are combined to form a degree adverb phrase for intensification or hyperbole. Although the positive word *good* is used, the speaker means the opposite. The first clause indicates why they think the service is not good, and, therefore, provides the contextual information.

4.2 The Use of Positive Adjective with High Intensity

In this pattern, the following two components must exist:

- (a) Positive adjective with high intensity
- (b) Negative context

Specific positive adjectives with high intensity are used to form ironic expressions with or without other rhetorical elements. Since the context is negative, the positive adjective is used to express non-literal meanings. The adjectives we found in the corpus include “偉大” (*great*), “了不起” (*remarkable*) and “天才” (*genius*). Only 2.09% of the messages in the corpus contain this pattern. For example, the word *great* is used in the following message:

- (s2) 我的 plurk 「又」發生不明錯誤了...這真是這世紀最偉大的發明啊
My Plurk account encountered an unknown error 'again'... This is indeed the greatest invention in the century.

4.3 The Use of Positive Noun with High Intensity

In this pattern, the following two components must exist:

- (a) Positive noun with high intensity
- (b) Negative context

Specific nouns that represent highly positive meanings are also used to express irony. These nouns include “巨星” (*superstar*), “大禮” (*big gift*) and “境界” (*wonderful state*). When they are used with a negative context, an ironic expression is formed. This pattern is not found frequently in the corpus. Only 2.00% of the messages in the corpus contain this pattern. An example is listed below:

- (s3) 中秋節收到的大禮是.....長了一堆肉
The big gift I received in the Mid Autumn Festival was a lot of fat in my body.

4.4 The Use of “很好” (very good)

In this pattern, the following two components must exist:

- (a) Sentence boundary + 很好 + punctuation
- (b) Negative context

A sentence boundary occurs before the word “很好” (very good) because there is no subject. Multiple punctuations, and particularly exclamation marks and ellipses, can be used after “很好” to increase the intensity. In the following example, exclamation marks are used:

(s4)感冒... 很好!! 我的假期飛了
I caught a cold... Very good!! My vacation is gone.

Sometimes this pattern is followed by an exclamation word, such as “啊” (*a*), “呀” (*ya*), and “嘛” (*ma*). These exclamations, like punctuations, can help strengthen the level of the speaker’s feelings. In our irony corpus, this pattern is used in 50.84% of all ironic messages. Obviously, this is a common way when people want to express their negative feelings with an ironic expression.

4.5 “可以再...一點” (It’s okay to be worse)

In this pattern, the following expression must exist:

可以再 + negative adjective + 一點
(It is okay to be more + negative adjective)

This pattern literally states that it is okay for something to become worse and is a commonly used pattern to express irony in our corpus. It can be found in 33.53% of the messages in the corpus. In most cases, even when no proper contextual information is present, the listener can tell the literal meaning is not meant because it violates most people’s inclinations. Thus, the use of this pattern is usually non-literal and ironic. An example is shown below.

(s5)零下十一度... 你可以再冷一點
It's -11°C... It is okay to be colder

A message can contain more than one pattern, causing the sum of the percentages of the above five patterns to be greater than 100%. For example, both patterns 4.4 and 4.5 are used in the following message:

(s6)很好!!!! 我可以再笨一點 再笨一點阿...
Very Good!!!! It is okay for me to be more idiotic...

The patterns in Sections 4.4 and 4.5 are mainly based on their linguistic forms and frequently used in ironic expressions. We argue that these patterns are more static than the others, and we call them the *customary patterns*. On the other hand, the patterns in Sections 4.1, 4.2 and 4.3 are called *non-customary patterns*.

5 Collecting Ironic Expressions from Blogs

In order to understand how irony is conveyed in different types of media, we use the methodology and mined patterns described in Sections 3 and 4 to collect irony expressions from the Yahoo Kimo Blogs corpus.

5.1 The Yahoo Blog Corpus

The Yahoo Kimo Blog corpus, referred to the Yahoo corpus in the following sections, contains blog articles from November 1, 2005 to August 20, 2007 (Yang, Lin and Chen, 2009). Out of all the posts in the dataset, 2,764,202 posts have at least one emoticon. The articles posted in July 2006 are used here, and they are divided into 341,932 smaller units by the full stop symbol. All articles are in Traditional Chinese.

Since the Plurk platform can be used as an instant messaging system, and readers of the message are usually on the author’s friend list, these messages are usually conversational. On the other hand, Ya-

hoo blogs are not limited in length and a blog article itself is not part of the conversation. Thus, the blog articles are usually more formal compared to microblog messages.

Although the articles are separated by a full stop into shorter units, these units are not necessarily identical to sentences due to the conventional usage of the Chinese period symbol. They can consist of multiple sentences and thus contain a discourse structure, which makes them suitable for this corpus study.

5.2 Extract Ironic Expressions

A similar approach to the steps described in Section 3.3, is used to collect ironic expressions from the Yahoo corpus, but four patterns of irony found in Plurk are used to perform step (1). These patterns, as listed below, are adopted because they are the most frequently used ones in our Plurk irony corpus. They can also reflect the uses of customary and non-customary irony patterns as the first two patterns are customary, and the last two are non-customary. Pattern 1 and Pattern 2 are the same patterns as mentioned in Section 4.4 and Section 4.5, respectively. Pattern 3 and Pattern 4 are two forms from the pattern described in Section 4.1. Only step (1) to step (3) are performed, and step (4) is bypassed; that is, the process is not repeated.

Pattern 1:

- (a) Sentence boundary + 很好 + punctuation
- (b) Negative context

Pattern 2:

可以再 + negative adjective + 一點

Pattern 3:

- (a) 還真 + positive adjective
- (b) Negative context

Pattern 4:

- (a) 真是 + positive expression
- (b) Negative context

5.3 Results and Discussion

A total of 36 ironic texts is obtained. All the four irony patterns seen in Plurk can be found in Yahoo. The final results are shown in Table 1.

	Number of Ironic Expressions	Percentage
Pattern 1	14	38.89%
Pattern 2	10	27.78%
Pattern 3	5	13.89%
Pattern 4	7	19.44%

Table 1: Ironic texts found for the four Patterns in Yahoo.

The proportions of the four patterns in Plurk and Yahoo are also compared. The percentages are calculated by dividing the occurrence of each pattern by the occurrence of all four patterns in the same datasets. As can be seen in Figure 1, the proportions of patterns (1) and (2) in Plurk are significantly higher than in Yahoo, and the proportions of patterns (3) and (4) in Plurk are significantly lower than in Yahoo ($p < 0.05$ according to the t-test). This suggests that patterns (1) and (2) tend to be used in informal and conversational texts while patterns (3) and (4) tend to be used in formal articles to convey irony. Also, this may suggest that customary patterns are more likely to be used in conversations, and authors of formal articles prefer an indirect way to express irony, although more data are required for further studies in the future.

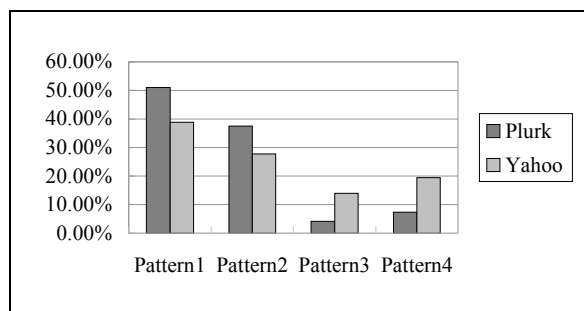


Figure 1: Comparison of the proportion of the four patterns in Plurk and Yahoo.

6 Error Analysis

In this section, we analyze why non-ironic messages were retrieved by the automatic processes. The 1,820 wrong messages specified in Section 3.3 are classified into the following two categories.

(1) Sentiment identification

Using the patterns to find possible ironic messages involves the correct sentiment identification. NTUSD does not cover some new words used on Internet informal conversations. The sentiment of a word can also be changed depending on its context. For example, “太強” (so strong) is listed as a positive term in NTUSD. However, it is used to indicate a negative condition in the example (s7).

(s7) 止痛藥的副作用也太強了吧，昏睡一整晚

The side effect of the pain reliever was so strong, making me sleep through the whole night.

(2) Opinion targets

In a Plurk message, even though the message poster is talking about the same topic, more than one entity with associated opinions can be present. For example:

(s8) 最近公司生意很好，好累ㄟ

The business of our company is running so well. I am so tired.

The poster expresses negative sentiment by using the word “tired.” Although the positive word “很好” (very good) is also used, it modifies the word “business” rather than the poster’s condition. That is, the opinion targets of the two words are different, and this causes problems when automatically retrieving ironic messages.

7 Linguistic Structure of Irony

In this section, the linguistic structure of irony is analyzed based on our observations on the corpus.

7.1 Ironic Word

As described, the literal meaning of an ironic word or phrase is opposite to the actual meaning. An ironic word/phrase is necessary to separate irony from regular utterances. If the ironic word of an utterance is reverted, the speaker’s actual sentiment or intention is reconstructed.

However, it is not easy to identify the ironic word in an utterance. Sometimes more than one word can be an ironic word. In our corpus, 94.93% of the ironic words are adjectives, while others are used as adverbs, verbs or nouns. The recognition of ironic word/phrase is a challenging task, but other ironic elements described in Sections 7.2 and 7.3 can be analyzed side by side to help improve the performance.

7.2 Contextual Information

Contextual information is usually provided as part of ironic utterances to help convey irony. For example, the underlined sentence in the following utterance is crucial for irony interpretation:

- (s9) 我掛彩了，真是太好運了
I was injured. I was really lucky.

Without the first sentence, it is hard to tell if *lucky* is actually meant. Although a speaker can still use ironic words/phrases without providing contextual information, this can be an ineffective way to communicate the actual meanings of irony. According to the cooperative principle proposed by Grice (1975), the speaker must give enough information in order to enable successful communication and implicatures. The four maxims of the cooperative principle include:

- (1) **Maxim of Quantity:** The speaker should make their contribution as informative as is required. Do not make the contribution more informative than is required.
- (2) **Maxim of Quality:** The speaker should not say what they believe to be false, and should not say that for which they lack adequate evidence.
- (3) **Maxim of Relation:** The speaker should be relevant.
- (4) **Maxim of Manner:** The speaker should avoid obscurity of expression, avoid ambiguity, be brief and be orderly.

Based on Grice's maxims, it is assumed *enough*, *correct*, *relevant*, and *understandable* contextual information should be provided with ironic expressions. However, the speaker sometimes assumes the listener already knows about the conditions where the irony takes place and has the required background knowledge; thus the contextual information is hidden in the ironic utterance.

Four types of context can be used to interpret irony:

- (1) **Linguistic context:** The linguistic context refers to the words that are expressed before and/or after the irony words in a sentence or discourse. It is easier to obtain and analyze than the other three types of context.
- (2) **Physical context:** Physical context refers to what is actually present and/or happening in the environment or circumstance where the conversation is taking place. It is also related to the timing. In online conversations, participants are not usually in the same location, but they can be aware of the same ongoing events and situations. It is not necessary for the speaker to provide physical context information if they assume the objects or situations are noticeable to the listeners.
- (3) **Epistemic context:** The background knowledge shared by the participants in a conversation can also be used to interpret the irony. This type of context does not change over time. For example, people know rocks are hard, so they can understand the expression *the bed is as soft as a rock* is not literal.
- (4) **Social context:** Social relationship can be important for expressing and interpreting irony, especially in online messages.

We argue that at least one type of contextual information must exist, but it can be hidden if the speaker thinks the listener is already aware of it. Physical, epistemic and social context can be hidden, while linguistic contextual information must be present.

7.3 Rhetoric

As shown in Section 4, degree adverbs, punctuations and exclamations can be used to convey irony. Some of them can even be repeated to intensify the effects. These elements increase contradiction and strengthen the degree of negative opinions. Unlike ironic words and context, rhetoric elements are not necessary to convey irony.

Liebrecht et al. (2013) call the words used to strengthen evaluative utterances *intensifiers*. In their experiments, non-hyperbolic sarcastic messages often contain an explicit marker on Twitter. They argue that sarcasm is often signaled by hyperbolic words, including intensifiers and exclamations, and sarcastic utterances with hyperbolic words are easier to identify by listeners/readers than sarcastic utterances without hyperbolic words. It can be seen that adverbs, adjectives, punctuations and exclamations with high intensity observed in our irony patterns have very similar effects.

Among the 113 messages containing the #反諷 (#irony) hashtag in Weibo, which was mentioned in Section 2, 83.19% do not exhibit hyperbole or uses of intensifiers. This observation is similar to the argument suggested in Liebrecht et al. (2013) and is one of the reasons why the hashtag is not suitable

for the irony pattern mining task in this study. In comparison, this methodology helps find more clues of irony that can be seen from their linguistic forms.

7.4 Corpus Labeling

To increase the usefulness of the corpus, ironic element tags are added to each message. An example is shown in Figure 2.

<context sentiment="pos">才剛買的書，竟然掉頁了，</context>這品質<rhetic>也太
</rhetic><ironic sentiment="neg">好</ ironic>了<rhetic>吧</rhetic>.

English translation:
<context sentiment="pos">The book I just bought has fallen apart.</context> The quality is <rheto-
ric>just extremely</rhetic> <ironic sentiment="neg">good</ ironic>le<rhetic>ba</rhetic>.

Figure 2: An example message with ironic element tags.

As can be seen in the example, “好” (good) is the word that is used in the opposite way, so it is marked with the ironic word/phrase label <ironic>. The preceding sentence states what actually happened, and is marked with the label <context>. The message poster also uses the degree adverb “太” (extremely) and used the exclamation “吧” (*ba*, a sentence-final partical). These two words are marked with the <rhetic> label. The sentiment polarity marks of the ironic word and contextual information, shown as either *pos* or *neg*, are also added.

8 Conclusion

In this paper, five types of irony patterns are mined, and an irony corpus is constructed based on linguistic forms and sentiment classification. Four verbal forms in Plurk and Yahoo were further examined. The former platform restricts short text conversation, and the latter platform allows for the long text description. The experimental results show that the customary forms tend to be used in informal and conversational texts while the non-customary forms tend to be used in formal articles to convey irony. The three basic elements that form a successful ironic speech act were also analyzed. These elements, including the words/phrases with reversed meanings, contextual information and rhetorical words, should be identified first in order to properly process ironic expressions and perform linguistic analysis. In the mined patterns, it was found that hyperbole was frequently present. In future work, we will explore other opinion mining and sentiment analysis algorithms, and focus on automatic recognition of hyperbole and the ironic elements.

Acknowledgements

This research was partially supported by National Taiwan University under grant 103R890858. We are also very thankful to the anonymous reviewers for their helpful comments to revise this paper.

References

- Herbert L. Colston and Jennifer O'Brien. 2000. Contrast of Kind Versus Contrast of Magnitude: the Pragmatic Accomplishments of Irony and Hyperbole. *Discourse and Processes*, 30(3):179-199.
- 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-2010)*, pages 107-116, Uppsala, Sweden.
- Elena Filatova. 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pages 392-298, Istanbul, Turkey.
- Raymond W. Gibbs and Herbert L. Colston. 2007. *Irony in Language and Thought*. Lawrence Erlbaum Associates, New York.

- Rachel Giora and Ofer Fein. 1999. Irony: Context and Salience. *Metaphor and Symbol*, 14:241-257.
- 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: Short Papers*, pages 581-586, Portland, Oregon, USA.
- H. P. Grice. 1975. Logic and Conversation. In P. Cole and J. J. Morgan, eds. *Syntax and Semantics, 3: Speech Acts*. New York: Academic Press.
- Hen-Hsen Huang, Chi-Hsin Yu, Tai-Wei Chang, Cong-Kai Lin and Hsin-Hsi Chen. 2013. Analyses of the Association between Discourse Relation and Sentiment Polarity with a Chinese Human-Annotated Corpus. In *Proceedings of ACL 2013 7th Linguistic Annotation Workshop & Interoperability with Discourse*, pages 70-78, Sofia, Bulgaria.
- Graham Katz and Eugenie Giesbrecht. 2006. Automatic Identification of Non-compositional Multiword Expressions Using Latent Semantic Analysis. In *Proceedings of the ACL/COLING-06 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pages 12-19, Sydney, Australia.
- Lun-Wei Ku and Hsin-Hsi Chen. 2007. Mining Opinions from the Web: Beyond Relevance Retrieval. *Journal of American Society for Information Science and Technology, Special Issue on Mining Web Resources for Enhancing Information Retrieval*, 58(12):1838-1850.
- Linlin Li and Caroline Sporleder. 2010. Linguistic Cues for Distinguishing Literal and Non-Literal Usages. In *Proceedings of 23rd International Conference on Computational Linguistics (COLING 2010), Poster Volume*, pages 683-691, Beijing, China.
- Christine Liebrecht, Florian Kunneman, and Antal van den Bosch. 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, Atlanta, Georgia.
- Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies, Morgan & Claypool Publishers.
- Stephanie Lukin and Marilyn Walker. 2013. Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 30-40, Atlanta, Georgia.
- Antonio Reyes, Paolo Rosso, and Davide Buscaldi. 2012. From Humor Recognition to Irony Detection: The Figurative Language of Social Media. *Data & Knowledge Engineering*, 74:1-12.
- Tony Veale and Yanfen Hao. 2010. Detecting Ironic Intent in Creative Comparisons. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI 2010)*, pages 765-770, Lisbon, Portugal.
- Fei Wang, Yunfang Wu, and Likun Qiu. 2012. Exploiting Discourse Relations for Sentiment Analysis. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012): Posters*, pages 1311-1320, Mumbai, India.
- Changhua Yang, Kevin Lin, and Hsin-Hsi Chen 2009. Writer Meets Reader: Emotion Analysis of Social Media from both the Writer's and Reader's Perspectives. In *Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2009)*, pages 287-290, Milan, Italy.
- Lanjun Zhou, Binyang Li, Wei Gao, Zhongyu Wei, and Kam-Fai Wong. 2011. Unsupervised Discovery of Discourse Relations for Eliminating Intro-Sentence Polarity Ambiguities. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 162-171, Edinburgh, Scotland, UK.