

A Survey in Automatic Irony Processing: Linguistic, Cognitive, and Multi-X Perspectives

Qingcheng Zeng

Department of Linguistics
Northwestern University
qingchengzeng@outlook.com

An-Ran Li

Department of Chinese and Bilingual Studies
The Hong Kong Polytechnic University
an-ran.li@connect.polyu.hk

Abstract

Irony is a ubiquitous figurative language in daily communication. Previously, many researchers have approached irony from linguistic, cognitive science, and computational aspects. Recently, some progress have been witnessed in automatic irony processing due to the rapid development in deep neural models in natural language processing (NLP). In this paper, we will provide a comprehensive overview of computational irony, insights from linguistic theory and cognitive science, as well as its interactions with downstream NLP tasks and newly proposed multi-X irony processing perspectives.

1 Introduction

Irony, which generally refers to the expressions that have opposite literal meanings to real meanings, is a representative rhetoric device in human languages (Li and Huang, 2020). For example, in sentences *I just love when you test my patience!!*, and *Had no sleep and have got school now*, compared to their literal meanings, both sentences are conveying reverse meanings and emotions, which mean *not love* and *not happy*. People are using ironies to communicate their affective states more implicitly or explicitly, to strengthen the claims depending on the needs, which contribute to neologism across human languages. However, the inner incongruity makes ironic expressions harder for machines to understand. Consequently, accurate irony processing systems are essential and challenging for downstream tasks and natural language understanding (NLU) research.

Previously, many researchers approached irony processing from various perspectives. The last survey in computational irony was published seven years ago (Wallace, 2015) and mainly focused on irony detection and pragmatic context model. Therefore, a systematic and comprehensive survey on automatic irony processing remains ab-

sent, encouraging us to focus on a review of advances in irony processing, from traditional machine learning, recurrent neural networks (RNNs) methods, to deeper pretrained language models (PLMs) throughout. Besides, del Pilar Salas-Zárate et al. (2020) revealed the great imbalance in figurative language research, in which sarcasm was dominant and twice as much as irony research. With this review, We aim to encourage a balanced and equal research environment in figurative languages.

Moreover, the design and further improvement in present irony processing systems should be concatenated with both the theoretical accounts in irony theories and its role in communication, followed by downstream applications in NLP tasks. Generally, most researchers approached irony processing as a single irony detection branch and three main shared tasks (Ghanem et al., 2019; Van Hee et al., 2018a; Ortega-Bueno et al., 2019) all focused on irony detection, whereas other aspects are constantly under-explored. And theoretical-informed or cognitive-informed discussions are rarely seen in irony research compared to tiny amendments in neural networks' architectures.

In this paper, we aim to offer a comprehensive review in irony processing from machine learning, linguistic theory, cognitive science, and newly proposed multi-X perspectives and to evaluate irony processing in NLP applications. The remaining sections are organized as follows. In the second part, we will approach irony theories in linguistics and cognitive science research, and discuss the concrete differences between irony and sarcasm. Then, we will review irony datasets in world languages and discuss potential problems in annotation schemes throughout. In the fourth part we will retrospect research progress in automatic irony processing, including traditional feature engineering, neural network architectures, to PLMs and relatively under-explored research fields in irony processing. Finally, we will discuss irony's in-

teractions with downstream NLP tasks like sentiment analysis and opinion mining, and multi-X perspectives for further development in computational irony research.

2 Theoretical Research in Irony

2.1 Irony Theories

Various definitions have been given to irony. Early studies suggested that irony is the expression whose real meaning is contradictory to its literal meaning (Grice, 1975). The Merriam-Webster Dictionary, The Oxford English Dictionary, and The Collins English Dictionary all adopted this definition and used the words "opposite" or "contrary" to explain the relationship between the literal and contextual meanings of irony.

However, more research into various types of ironic examples revealed that the contextual meaning of irony does not have to be "opposite" or "contrary" to the literal one. According to Sperber and Wilson (1986); Wilson and Sperber (2012), some expressions have no "literal meaning" to be challenged because no "literal meaning" is mentioned in the context, based on which they raised **relevance theory** and the "echoic" concept. They considered irony as "an echoic use of language in which the speaker tacitly dissociates herself from an attributed utterance or thought" (Wilson, 2006). That is, if the "echoic use" is incongruous in some ways, the expression can be ironic. Based on this theory, Seto (1998) put forward that there are some "echo-markers" like *definitely*, *really*, and *indeed*.

Li and Huang (2020) provided instances to show that "incongruity" does not have to be between the literal and contextual meanings of irony in certain circumstances. They believed that irony's true nature is a psychological activity as much as a verbal representation. The speaker or listeners must finish the "reversal" process on a psychological level for it to be completed. When compared to the concepts of "echoic" and "incongruity," "reversal" is concerned not only with the results but also with the psychological processes that the speakers/listeners go through.

2.1.1 Types of Irony

Booth (1974) divided irony into tragic and comic irony by literary genre, as well as stable and unstable irony by determinacy. He also categorized irony into dramatic irony, situational irony, verbal irony, and rhetorical irony by the range of context it has to

refer to. Researchers in computational irony paid the most attention to situational irony and verbal irony. Situational irony is a circumstance in which the outcome differs from what was expected, or a situation that contains striking contrasts (Imagine a situation that a lifeguard is saved from drowning). Verbal irony is the expression in which the speaker's intended meaning significantly different from the literal one. Abrams and Harpham (2014) considered that verbal irony usually involves the explicit presentation of one attitude or evaluation, but with signals that the speaker wants to express a totally different attitude or opinion. It differs from situational irony in that it is purposely manufactured by speakers. For example, *Very well, keep insulting me!*

Verbal irony is the kind to which computational linguistics researchers pay the most attention. Li (2021) made a further classification of them. She put forward eight kinds of reversals which are rhetorical reversal, expectation reversal, evaluation reversal, reversal of sentiment, reversal of factuality, relationship reversal, reversal from opposite pair, and reversal from satiation. The paper considered that verbal ironies can be classified by the kind of reversal they generate.

2.2 Linguistic Features

2.2.1 Irony Markers and Constructions

Although most of the studies saw irony as a pragmatics phenomenon, people also considered that it can be reflected on the verbal, grammatical, or semantic level. For example, on the verbal level, people often use words like *thank*, *congratulate*, *welcome*, *happy*, and *interesting* to express ironic meanings. Laszlo (2017) found 15 core evaluative words which often show in ironic expressions. She generated patterns from these core evaluative words to extract ironic sentences from the corpus. For example, when the word *love* is in the pattern "NP + would/ 'd/ wouldn't + love", it is highly possible to be an ironic expression. On a grammatical level, people often use the subjunctive when they intend to be ironic. Besides that, semantic conflict is the most direct way to express ironic meaning. The incompatibility between the main words of the proposition leads to the ridiculousness of the proposition (e.g. It's very considerate of you to make such a loud noise while I was asleep). Besides, (Ghosh and Muresan, 2018) also categorized irony markers according to trope, morphosyntactic, and

typographic types.

Li (2021) considered that ironies are often expressed by specific “constructions”, especially in short discourses. Larger than “core evaluative words” in Laszlo (2017), the “constructions” mentioned in Li (2021) are mostly in the form of idioms or phrases. The crucial feature of them is the lack of predictability. Most of them do not have to rely on too much contextual information, they themselves can provoke the process of reversal for readers or listeners (e.g. 贵人多忘事 (honorable people frequently forget things)).

2.2.2 Irony in Communication

Researchers claim that by using ironies, people have several kinds of intentions.

Be polite: According to Brown et al. (1987), when unfavorable attitudes such as resistance, criticism, and complaints are stated with irony, the threat to the listener’s reputation is reduced. The irony, as stated in Giora (1995), is an indirect negation. People prefer to utilize indirect negation to be polite to their listeners because direct negation can generate great unhappiness;

Ease criticisms: As reported by Dews and Winner (1995), irony helps to ease the expression’s evaluative function. They believe that the incompatibility between literal meaning and contextual meaning can make it difficult to articulate negative feelings. However, Toplak and Katz (2000) argued that, while irony literally avoids conflict, it is more aggressive from the perspective of the speaker’s goal;

Self-protection: Sperber and Wilson (1986) proposed the “echoic” idea, which stated that irony is a detached utterance that is simply an echo of another people’s thought. It’s a self-protection tactic, especially when the speakers are members of marginalized groups. According to Gibbs (2000), the irony is an “off-record” statement that allows speakers to deny their true intentions and avoid being challenged;

Be amusing: Gibbs (2000) reported that when young people intend to be humorous, 50% of their communication is ironic. It can assist people in creating a dialogue platform on which speakers and listeners can agree and communicate more easily.

2.3 Irony and Sarcasm

Most of the studies saw sarcasm as a subset of irony (Tannen et al., 2005; Barbe, 1995; Kumon-Nakamura et al., 1995; Leggitt and Gibbs, 2000;

Bowes and Katz, 2011). Sarcasm is often recognized as “a nasty, mean-spirited, or just relatively negative form of verbal irony, used on occasion to enhance the negativity expressed relative to direct, non-figurative criticism” (Colston, 2017).

One of the peculiarities of sarcasm is whether or not the speakers intend to offend the listeners. Kumon-Nakamura et al. (1995), for example, believe that sarcastic irony always conveys a negative attitude and is intended to harm the object being discussed. The non-sarcastic irony, on the other hand, can communicate either a good or negative attitude, and it is rarely meant to be hurtful. Barbe (1995) concurred that the core difference was “hurtful”. She claimed that irony is a face-saving strategy while sarcasm is a face-threatening action. Ridicule is another feature of sarcasm. According to Lee and Katz (1998), sarcasm is closer to ridicule than irony. Their experiment revealed that sarcasm is directed at a single person, but irony is directed toward a large group of people. Haiman et al. (1998) claimed that one of the most distinguishing characteristics of sarcasm is that the literal meaning of its words is always positive. However, he did not convey his thoughts on irony. Whereas Littman and Mey (1991) viewed this topic from another angle. While there are many various forms of ironies, they believe that there is only one type of sarcasm because “sarcasm cannot exist independently of the communication setting.”

Cognitive scientists approached the difference in experimental studies. Previous research in child language acquisition (Glenwright and Pexman, 2010) reported that children understood sarcastic criticism later than they could understand the non-literal meanings of irony and sarcasm, implying different pragmatic purposes of irony and sarcasm. Filik et al. (2019) utilized fMRI and found out sarcasm is associated with wider activation of semantic network in human brains compared to irony.

However, most computational linguistics researchers used irony and sarcasm interchangeably, since the boundary between these two concepts is too vague for even human beings, let alone for machines. Joshi et al. (2016) and Sulis et al. (2016) verified this claim from both human annotators and computational perspectives. Although in this paper we will mainly focus on the literal **irony processing** and discuss the inspirations from recent research output in sarcasm processing, we aim to encourage **unify sarcasm under the framework**

of irony via fine-grained annotation schemes.

3 Irony Datasets Perspectives

3.1 Irony Textual Datasets

Some main target databases for irony processing research include social media platforms like Twitter and online shopping websites like Amazon. For example, Reyes and Rosso (2012) collected a 11,861-document irony dataset based on customer reviews from several websites. Some other preliminary attempts to build irony benchmark datasets is Reyes et al. (2012) and Reyes et al. (2013), in which they used self-generated hashtag *#irony* as the gold standard and constructed 40,000-tweet and 50000-tweet datasets from Twitter respectively, each including 10,000 ironic tweets and remaining non-ironic ones. The irony benchmark dataset that is now widely used is from Van Hee et al. (2018a), consisting of 4,792 tweets and half of them were ironic. This dataset was also constructed via searching hashtags including *#irony*, *#sarcasm*, and *#not*.

There were also tremendous attempts to construct benchmark datasets in other languages. For example, Tang and Chen (2014) firstly built the NTU Irony Corpus including 1,005 ironic messages from Plurk, a Twitter-like social media platform, by mining specific ironic patterns and manually checking extracted messages. A recent Chinese benchmark dataset for irony detection was constructed by Xiang et al. (2020), which includes 8,766 Weibo posts, labelled from *not ironic* to *strongly ironic* in a five-scale system. Besides, irony datasets in Spanish, Greek, and Italian are also widely available. A comprehensive overview of the irony datasets is listed in Table 1.

3.2 Data Source and Construction Methodology

Twitter, as one of the most trending social platforms, is the major source of irony benchmark datasets. Given Plurk and Weibo's similarity to Twitter, online short-text social medias are almost the only origin for present datasets, which might lead to several potential problems. For example, the limitation of 140-word introduced specific bias towards short-text classification and long texts remain a problem. Besides, the judgment of irony might be highly dependent on contextual information like previous comments or retweets and one single tweet could be meaningless itself. At last, topics on social media platforms might also be

highly biased towards political or sports topics. Interestingly, previous research (Ghosh and Muresan, 2018) reported that for Twitter and Reddit, different ironic markers played the most important roles, further emphasizing the needs of multiple sources for robust NLU.

As for the construction methodology, most datasets adopted "keywords and hashtags" filtering strategy, and some of them followed by human annotations. With annotation schemes put aside, Sykora et al. (2020) explored how self-generated tags could correspond to real labels correctly via a manual semantic analysis. At the worst case, only 16% of the tagged *#sarcasm* tweets are unambiguous sarcastic tweets according to well-trained linguists, which further emphasized the necessity of human annotation.

3.3 Annotation Schemes

As shown in Table 1, various datasets have been labelled differently. Most datasets were annotated as binary classes, ironic versus non-ironic. Both Chinese datasets adopted different strategies by annotating ironic elements and intensity respectively. Van Hee et al. (2018a) is the only one doing fine-grained labelling, into verbal irony by polarity contrast, other verbal irony, situational irony, and non-ironic.

Another annotation scheme was proposed by Cignarella et al. (2020b). They annotated irony activators at the morphosyntactic level and distinguished different types of irony activation. This annotation scheme dived into syntactic information and offered information for analyzing ironical constructions.

3.4 Future work in datasets

A high-quality benchmark dataset is crucial to measure NLU capability and advance future research. We are calling improvements for a future benchmark dataset from following perspectives.

- 1) Diverse data sources like literature, daily conversations, and news articles should be involved, and the distribution of text length should be relatively balanced.
- 2) A uniform annotation scheme and strategy is awaiting for construction. For example, situational irony is needed; ironic intensity should be labelled as reference rather than exclusive labels in Xiang et al. (2020); a scheme to unify sarcasm and irony could be expected.

Study	Language	Data Source	Construction Methodology	Size	Annotation Scheme
Reyes and Rosso (2012)	English	online websites	filtering low-star reviews	11861	ironic / non-ironic
Reyes et al. (2013)	English	Twitter	#irony hashtag	40000	ironic / non-ironic
Wallace et al. (2014)	English	Reddit	Sub-reddit	3020	ironic / non-ironic
Van Hee et al. (2016)	English, Dutch	Twitter	#irony, #sarcasm, and #not hashtags	3000, 3179	three-scale irony annotation
Van Hee et al. (2018a)	English	Twitter	#irony, #sarcasm, and #not hashtags	4792	four-scale ironic types
Tang and Chen (2014)	Chinese	Plurk	pattern mining	1005	annotating ironic elements
Xiang et al. (2020)	Chinese	Weibo	pattern mining	8766	five-scale ironic intensity
Barbieri et al. (2016)	Italian	Twitter	keywords, hashtags, and etc.	9410	ironic / non-ironic
Cignarella et al. (2018)	Italian	Twitter	keywords, hashtags, and etc.	4849	ironic / non-ironic / sarcastic
Charalampakis et al. (2015)	Greek	Twitter	keywords	61427	ironic / non-ironic
Ortega-Bueno et al. (2019)	Spanish	Twitter	comments and tweets	9000	ironic / non-ironic
Ghanem et al. (2019)	Arabic	Twitter	keywords and hashtags	5030	ironic / non-ironic
Corrêa et al. (2021)	Portuguese	Twitter and news articles	keywords, hashtags, and news articles	34306	ironic / non-ironic
Vijay et al. (2018)	Hindi-English code-mixed	Twitter	keywords and hashtags	3055	ironic / non-ironic

Table 1: Irony benchmark datasets

3) Multi-X perspectives should be incorporated into the construction of datasets. For example, a multimodal and multilingual dataset could enhance irony identification in a grounded environment.

4 Irony Processing Systems

In this section, we will discuss the progress of irony processing systems comprehensively, organized along the development of machine learning and deep learning.

4.1 Irony Detection

4.1.1 Rule-based Detection Methods

Tang and Chen (2014) extracted ironic expressions based on five patterns summarized from mandarin Chinese. Li and Huang (2020) further expanded ironic constructions to more than twenty and proposed a systematic irony identification procedure (IIP). Besides Chinese, Frenda (2016) utilized sentiment lexicons, verb morphology, quotation marks, and etc. to design an Italian irony detection model and got competitive performance with machine learning models. However, rule-based models are too complex and hard to generalize for wider applications.

4.1.2 Supervised non-neural network era

Most research took irony detection as a simple classification problem. Before the popularity of deep learning, feature engineering is crucial for accurate irony detection. Generally, features could be divided into several levels.

Lexical features Lexical features are at the foundational level of NLP features, basically divided into bags of words (BOW) sets, word form sets, and conditional n-gram probabilities (Van Hee et al., 2018b). Representative BOW sets mainly include n-grams and character n-grams. Word form sets fo-

cus on number and frequency, such as punctuation numbers, emoticon frequencies, character repetitions, etc. Despite their easiness to get, lexical features were proved effective in much research.

Syntactic features Syntax is mainly quantified via parts-of-speech and named entities. After tagging, the number and frequency of both characteristics could act as features in classification models. Besides, hand-crafted syntactic features also included clash before verb tenses (Reyes et al., 2013) and dependency parsing (Cignarella et al., 2020a).

Semantic features Van Hee et al. (2018b) approached semantic features based on the presence or not in semantic clusters, which were trained on a irony Twitter corpus with the Word2Vec algorithm (Mikolov et al., 2013).

Linguistic-motivated features Irony processing is deeply associated with sentiments and emotions. Therefore, researchers have offered many characteristics to capture irony patterns. For example, Reyes et al. (2013) proposed the feature of *contextual imbalance*, which was quantified via measuring the semantic similarity pairwise. Generally, most features could be categorized into *ambiguity* (Reyes et al., 2012) and *incongruity* (Joshi et al., 2015). Take incongruity as an example, implicit incongruity was defined as a boolean feature checking containing implicit sentiment phrases or not; explicit incongruity was defined as number of times a polarity contrast appears. Theoretical research (Ghosh et al., 2020) is encouraging more semantic and pragmatic features to better capture ironies.

Features at various levels were concatenated with classifiers, including naive bayes, decision trees, support vector machines (SVMs), etc. (Van Hee et al., 2018b) to get final classification results.

4.1.3 Supervised Neural Network Era

There have been irony detection tasks in various languages after 2015, among which most participants used convolutional neural networks (CNNs) and RNNs methods to detect ironical expressions. For example, in SemEval-2018 Task 3 (Van Hee et al., 2018a), Wu et al. (2018) classified irony tweets and their ironical types in a densely connected LSTM network, together with multitask learning (MTL) objectives in the optimization. In IroSvA (Ortega-Bueno et al., 2019), González et al. (2019) utilized transformer encoders only for detecting Spanish ironical tweets.

Besides, Ilić et al. (2018) firstly utilized contextualized word representations ELMo (Peters et al., 2018) with Bi-LSTM to detect ironies. Zhang et al. (2019) enhanced irony detection with sentiment corpora based on attention Bi-LSTM RNNs, and achieved state-of-the-art results on Reyes et al. (2013) dataset.

4.1.4 Pretraining and Fine-tuning Paradigm

Since BERT (Devlin et al., 2019), the “pretraining and fine-tuning paradigm” has become the mainstream in NLP research due to its extraordinary capacity in dealing with contextualized information and learning general linguistic knowledge. Recent development in irony detection also witnessed the usage of PLMs. Xiang et al. (2020) released several baseline results along with the dataset, among which BERT had highest accuracy (5% higher than Bi-LSTM methods). Potamias et al. (2020) proposed a recurrent convolutional neural network (RCNN)-RoBERTa (Liu et al., 2019) strategy and improved the results on the SemEval-2018 dataset by a large degree. Besides, Cignarella et al. (2020a) explored syntax-augmented irony detection with multilingual BERT (mBERT) in multilingual settings. To sum up, some representative studies in irony detection are detailed in Table 2.

4.2 Irony Generation

Irony generation is mostly an underexplored research field besides Zhu et al. (2019), in which they defined irony generation as a style transfer problem, and utilized a Seq2Seq framework (Sutskever et al., 2014) with reinforcement learning to generate ironical counterparts from a non-ironic sentence. Concretely, they designed the overall reward as a harmonic mean of irony reward and sentiment reward, which was trying to capture the sentiment incongruity. In terms of the evaluation, besides

traditional natural language generation metrics like BLEU, they also designed task-specific evaluation metrics, which should be further enhanced in irony and even figurative language research.

Future work in irony generation could be advanced in new PLMs and theoretical accounts. For example, no attempts were made to generate ironical expressions after generative PLMs like BART (Lewis et al., 2020). Controllable irony generation and its interaction with agents are interesting topics remaining for future exploration. Besides, irony theories could be further utilized. In recent research on unsupervised sarcasm generation (Mishra et al., 2019; Chakrabarty et al., 2020), context incongruity, valence reversal, and semantic incongruity were merged to enhance the generation.

5 Discussion

5.1 Irony for Downstream NLP Tasks

Irony is directly associated with downstream NLU tasks like sentiment analysis and opinion mining. For example, the sentence retrieved from Filatova (2012) *I would recommend this book to friends who have insomnia or those who I absolutely despise.* is classified as positive by fine-tuned sentiment analysis RoBERTa model (Heitmann et al., 2020), which is apparently opposite to human evaluation. Wrong sentiment judgments will potentially lead to contrary opinion mining. We suggest that irony could be further captured through introducing incongruity embedding or specific pattern matching. Joshi et al. (2015) designed linguistic-motivated features implicit and explicit incongruity, which are inspiring for enhancing irony understanding. Consider another example task, machine translation, in which wrong translation will potentially lead to totally opposite meanings. We encourage to model discourse features (Voigt and Jurafsky, 2012), such as ironic patterns and punctuation as embeddings for robust irony translation.

In addition, we are looking forward to the research of irony and sarcasm processing in NLP for social good (NLP4SG), especially considering the strong sentiments hidden in ironies. A recent work (Chia et al., 2021) explored cyberbullying detection and this was a starting point to handle online harmful ironical contents.

5.2 Multi-X Perspectives

Recent developments in NLP and sarcasm processing encourage us to approach discourse processing

Study	Input Features	Architecture	Dataset	Performance (F1 Score)
Reyes et al. (2013)	hand-crafted high-level features	decision tree	Reyes et al. (2013)*	70 / 76 / 73
Barbieri and Saggion (2014)	frequency, intensity, sentiments, etc.	decision tree	Reyes et al. (2013)*	73 / 75 / 75
Nozza et al. (2016)	unsupervised	topic-irony model	Reyes et al. (2013)*	84.77 / 82.92 / 88.34
Zhang et al. (2019)	word embeddings	sentiment-transferred Bi-LSTM	Reyes et al. (2013)*	94.69 / 95.69 / 96.55
Van Hee et al. (2018b)	lexical, syntactic and semantic features	SVMs	Van Hee et al. (2016)**	70.11
Rohanian et al. (2018)	intensity, contrast, topics, etc.	ensemble voting classifier	Van Hee et al. (2018a)***	65.00 / 41.53
Wu et al. (2018)	word embeddings, POS tags, sentiments, etc.	LSTM + MTL	Van Hee et al. (2018a)***	70.54 / 49.47
Cignarella et al. (2020a)	mBERT output, autoencoders	LSTM	Van Hee et al. (2018a)***	70.6
Potamias et al. (2020)	RoBERTa output	RCNN	Van Hee et al. (2018a)***	80.0
Santilli et al. (2018)	word space vectors, BOW sets	SVMs	Cignarella et al. (2018)****	70.00 / 52.00
Cimino et al. (2018)	word embeddings	LSTM + MTL	Cignarella et al. (2018)****	73.60 / 53.00
González et al. (2019)	word embeddings	transformer encoders	Ortega-Bueno et al. (2019)	68.32

* Three binary classification systems were trained respectively for this dataset.

** This result was obtained on the 3000 English tweets subset.

*** This dataset had two sub-tasks, identifying ironic or not and classifying ironic types. The latter two only focused on the first sub-task.

**** This dataset had two sub-tasks, irony detection and further identifying sarcasm.

Table 2: Representative Irony Detection Systems

from multiple angles. In this part, we will review and suggest several multi-X perspectives for irony and figurative language processing.

5.2.1 Multimodal Irony Processing

Linguistic interactions are not solely consisted of texts. Besides, facial expressions and speech communications are crucial to convey emotions and feelings. For example, Skrelin et al. (2020) reported people could classify ironies based on phonetic characteristics only. Consequently, it is conceivable that multimodal methods could help with irony detection. Schifanella et al. (2016) made the first attempt in multimodal sarcasm detection, in which they extracted posts from three multimodal social media platforms based on hashtags. Then they used SVMs and neural networks to prove the validity of visual information in enhancing sarcasm detection.

Castro et al. (2019) made a great improvement in multimodal sarcasm detection by introducing audio features into the dataset. The experiments also verified the importance of more modalities in sarcasm processing.

Future work in multimodal irony processing should include a comprehensive multimodal irony dataset based on MUsTARD dataset (Castro et al., 2019) with more fine-grained annotation schemes. Additionally, most methods (Pan et al., 2020; Liu et al., 2021) explored sarcasm by introducing inter-modality and intra-modality attention in single-stream setting. How double-stream multimodal pretrained models (MPMs) will encode and interact in complex discourse settings remains an interesting problem to solve.

5.2.2 Multilingual Irony Processing

To understand irony in a multilingual context is even harder due to cultural gaps. Previously listed dataset includes a Hindi-English code-mixed irony dataset (Vijay et al., 2018), in which they offered an example:

- **Text:** The kahawat ‘old is gold’ purani hogae. Aaj kal ki nasal kehti hai ‘gold is old’, but the old kahawat only makes sense. #MindF #Irony.
- **Translation:** The saying ‘old is gold’ is old. Today’s generation thinks ‘gold is old’ but only the old one makes sense. #MindF #Irony.

Cignarella et al. (2020a) explored how mBERT performed in multiple languages’ irony detection tasks separately. Given it has been proved code-switching patterns are beneficial for NLP tasks like humor, sarcasm, and hate speech detection in RNNs settings (Bansal et al., 2020), A future direction is to merge the irony detection datasets from multiple languages (consider Karoui et al. (2017)) or even code-mixed texts, and explore how multilingual datasets could enhance irony understanding in mBERT.

5.2.3 Multitask Irony Processing

MTL is to make models learn several tasks simultaneously rather than independently once at a time. Recent work in figurative language processing proposed several MTL strategies to improve the performance interactively. Chauhan et al. (2020) proposed a MTL framework to do sentiment, sarcasm, and emotion analysis simultaneously and the framework yielded better performance with the help of MTL.

Generally, we will classify figurative language into several categories like metaphors, parodies, humors, ironies, and etc. However, noted that there are not clear differences between each other, a single task figurative language processing will only focus on one particular aspect and fail to capture the interactions. Recent work (Ao et al., 2022) also verified the combination of humor and sarcasm could improve political parody detection.

PLMs could understand figurative language better than random but apparently worse than human evaluation (Liu et al., 2022). We suggest that future work should consider domain adaptation towards figurative language as a whole via weak supervision. MTL strategy could utilize previous research in corpus linguistics, and design an appropriate proportion in summing the loss function. Besides, a unified framework to model figurative languages could be expected.

5.2.4 Multiagent Irony Processing

Human-like language generation is a central topic in multiagent interactive systems. Besides robots' ironical understanding, we are also curious about how robots could generate ironical expressions. Unlike transferring non-ironic sentences to ironic, multiagent irony measures the performance during the interactions in dialogues. Ritschel et al. (2019) improved the robots by introducing ironic expressions, which showed better user experiences in human evaluation.

Further explorations in multiagent irony could aim at better dialogue state tracking and understand when irony should be introduced.

5.3 New Tasks: Inspiration from Sarcasm

Compared to sarcasm, irony is rarely seen as a term in NLP conferences. Recently we have witnessed great improvements in sarcasm processing and in this part we will discuss how new tasks in sarcasm could motivate irony research.

Data Collection As discussed, datasets are highly dependent on hashtags as a signal to extract ironical expressions. Shmueli et al. (2020) proposed an algorithm to detect sarcastic tweets from a thread based on exterior cue tweets. A distant supervision based method for extracting ironies from platforms is crucial, given ironies in conversational contexts are central topic in the future.

Intended and Perceived Irony Oprea and Magdy (2019) explored how author profiling affected the perceived sarcasm (manual labelling)

versus the intended sarcasm (hashtags), and verified the difference between both. Further, Oprea and Magdy (2020) introduced iSarcasm dataset which divided intended sarcasms and perceived sarcasms. The state-of-the-art sarcasm detection models performed obviously worse than human evaluation on this dataset. Future work could focus on multimodal perceived and intended irony, especially across various cultures.

Target Identification Sarcasm target identification was firstly proposed in Joshi et al. (2018), in which sarcasm targets were classified as one target, several targets and outside. Patro et al. (2019) introduced sociolinguistic features and a deep learning framework, and improved target identification by a lot. For irony processing, most ironical expressions do not equip a specific target in itself as previously discussed. However, its ironical effects are likely in dialogue or visually grounded environment, which encourages us to enhance irony datasets in aforementioned ways.

Irony Explanation Irony, according to the definition, have opposite real meanings to literal meanings. However, this does not mean adding a single negation could interpret ironies well. Kumar et al. (2022) proposed a new task, sarcasm explanation in dialogue. Irony explanation might encounter more complex problems due to relatively low proportion of targets. Still, we should include irony explanation as a branch of multimodal irony processing like Desai et al. (2021).

5.4 Explainable Irony Processing

Explainable machine learning is of interest for most researchers to uncover the blackbox. In irony processing, we are also curious about why specific expressions are recognized as ironies. Buyukbas et al. (2021) explored explainability in irony detection using Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) methods. Results showed that punctuations and strong words play important roles in irony detection.

For future work, we suggest using explainable methods in multimodal settings and check how different modalities act various roles in making a class label.

6 Conclusion

In this paper, we reviewed the development in automatic irony processing from underexplored theo-

retical and cognitive science to computational perspectives, and offered a comprehensive analysis in future directions. We hope that our work and thinking will encourage further interdisciplinary research between linguistics and human language technology, motivate the research interests in irony and even, figurative languages.

Acknowledgement

This review is based on the first author's previous research proposal. We would like to express our thanks to Professor Chu-Ren Huang and Dr. Yat Mei Lee for their suggestions.

References

- Meyer Howard Abrams and Geoffrey Harpham. 2014. *A glossary of literary terms*. Cengage Learning.
- Xiao Ao, Danae Sánchez Villegas, Daniel Preotjiuc-Pietro, and Nikolaos Aletras. 2022. [Combining humor and sarcasm for improving political parody detection](#).
- Srijan Bansal, Vishal Garimella, Ayush Suhane, Jasabanta Patro, and Animesh Mukherjee. 2020. [Code-switching patterns can be an effective route to improve performance of downstream NLP applications: A case study of humour, sarcasm and hate speech detection](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1018–1023, Online. Association for Computational Linguistics.
- Katharina Barbe. 1995. *Irony in context*, volume 34. John Benjamins Publishing.
- Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, and Viviana Patti. 2016. Overview of the evalita 2016 sentiment polarity classification task. In *Proceedings of third Italian conference on computational linguistics (CLiC-it 2016) & fifth evaluation campaign of natural language processing and speech tools for Italian. Final Workshop (EVALITA 2016)*.
- Francesco Barbieri and Horacio Saggion. 2014. [Modelling irony in Twitter](#). In *Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 56–64, Gothenburg, Sweden. Association for Computational Linguistics.
- Wayne C Booth. 1974. *A rhetoric of irony*. University of Chicago Press.
- Andrea Bowes and Albert Katz. 2011. When sarcasm stings. *Discourse Processes*, 48(4):215–236.
- Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*, volume 4. Cambridge university press.
- Ege Berk Buyukbas, Adnan Harun Dogan, Asli Umay Ozturk, and Pinar Karagoz. 2021. [Explainability in irony detection](#). In *International Conference on Big Data Analytics and Knowledge Discovery*, pages 152–157. Springer.
- Santiago Castro, Devamanyu Hazarika, Verónica Pérez-Rosas, Roger Zimmermann, Rada Mihalcea, and Soujanya Poria. 2019. [Towards multimodal sarcasm detection \(an _Obviously_ perfect paper\)](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4619–4629, Florence, Italy. Association for Computational Linguistics.
- Tuhin Chakrabarty, Debanjan Ghosh, Smaranda Muresan, and Nanyun Peng. 2020. [R³: Reverse, retrieve, and rank for sarcasm generation with commonsense knowledge](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7976–7986, Online. Association for Computational Linguistics.
- Basilis Charalampakis, Dimitris Spathis, Elias Kouslis, and Katia Kermanidis. 2015. Detecting irony on greek political tweets: A text mining approach. In *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, pages 1–5.
- Dushyant Singh Chauhan, Dhanush S R, Asif Ekbal, and Pushpak Bhattacharyya. 2020. [Sentiment and emotion help sarcasm? a multi-task learning framework for multi-modal sarcasm, sentiment and emotion analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4351–4360, Online. Association for Computational Linguistics.
- Zheng Lin Chia, Michal Ptaszynski, Fumito Masui, Gniewosz Leliwa, and Michal Wroczynski. 2021. [Machine learning and feature engineering-based study into sarcasm and irony classification with application to cyberbullying detection](#). *Information Processing Management*, 58(4):102600.
- Alessandra Teresa Cignarella, Valerio Basile, Manuela Sanguinetti, Cristina Bosco, Paolo Rosso, and Farah Benamara. 2020a. [Multilingual irony detection with dependency syntax and neural models](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1346–1358, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Alessandra Teresa Cignarella, Simona Frenda, Valerio Basile, Cristina Bosco, Viviana Patti, Paolo Rosso, et al. 2018. Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In *Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018)*, volume 2263, pages 1–6. CEUR-WS.
- Alessandra Teresa Cignarella, Manuela Sanguinetti, Cristina Bosco, and Paolo Rosso. 2020b. [Marking](#)

- irony activators in a Universal Dependencies treebank: The case of an Italian Twitter corpus. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5098–5105, Marseille, France. European Language Resources Association.
- Andrea Cimino, Lorenzo De Mattei, and Felice Dell’Orletta. 2018. Multi-task learning in deep neural networks at evalita 2018. In *EVALITA@CLiC-it*.
- Herbert L Colston. 2017. Irony and sarcasm. In *The Routledge handbook of language and humor*, pages 234–249. Routledge.
- Ulisses Brisolará Corrêa, Leonardo Coelho, Leonardo Santos, and Larissa A. de Freitas. 2021. Overview of the IDPT task on irony detection in portuguese at iberlef 2021. *Proces. del Leng. Natural*, 67:269–276.
- María del Pilar Salas-Zárate, Giner Alor-Hernández, José Luis Sánchez-Cervantes, Mario Andrés Paredes-Valverde, Jorge Luis García-Alcaraz, and Rafael Valencia-García. 2020. Review of english literature on figurative language applied to social networks. *Knowledge and Information Systems*, 62(6):2105–2137.
- Poorav Desai, Tanmoy Chakraborty, and Md Shad Akhtar. 2021. Nice perfume. how long did you marinate in it? multimodal sarcasm explanation.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Shelly Dews and Ellen Winner. 1995. Muting the meaning a social function of irony. *Metaphor and Symbol*, 10(1):3–19.
- 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’12)*, pages 392–398, Istanbul, Turkey. European Language Resources Association (ELRA).
- Ruth Filik, Alexandra Țurcan, Christina Ralph-Nearman, and Alain Pitiot. 2019. What is the difference between irony and sarcasm? an fmri study. *cortex*, 115:112–122.
- Simona Frenda. 2016. Computational rule-based model for irony detection in italian tweets. In *EVALITA 2016*, volume 1749, pages 1–6.
- Bilal Ghanem, Jihen Karoui, Farah Benamara, Véronique Moriceau, and Paolo Rosso. 2019. Idat at fire2019: Overview of the track on irony detection in arabic tweets. In *Proceedings of the 11th Forum for Information Retrieval Evaluation*, pages 10–13.
- Debanjan Ghosh and Smaranda Muresan. 2018. "with 1 follower i must be awesome: P." exploring the role of irony markers in irony recognition. In *Twelfth International AAAI Conference on Web and Social Media*.
- Debanjan Ghosh, Elena Musi, Kartikeya Upasani, and Smaranda Muresan. 2020. Interpreting verbal irony: Linguistic strategies and the connection to the type of semantic incongruity.
- Raymond W Gibbs. 2000. Irony in talk among friends. *Metaphor and symbol*, 15(1-2):5–27.
- Rachel Giora. 1995. On irony and negation. *Discourse processes*, 19(2):239–264.
- Melanie Glenwright and Penny M Pexman. 2010. Development of children’s ability to distinguish sarcasm and verbal irony. *Journal of Child Language*, 37(2):429–451.
- José-Ángel González, Lluís F. Hurtado, and Ferran Plà. 2019. Elirf-upv at irosva: Transformer encoders for spanish irony detection. In *IberLEF@SEPLN*.
- Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.
- John Haiman et al. 1998. *Talk is cheap: Sarcasm, alienation, and the evolution of language*. Oxford University Press on Demand.
- Mark Heitmann, Christian Siebert, Jochen Hartmann, and Christina Schamp. 2020. More than a feeling: Benchmarks for sentiment analysis accuracy. Available at SSRN 3489963.
- Suzana Ilić, Edison Marrese-Taylor, Jorge Balazs, and Yutaka Matsuo. 2018. Deep contextualized word representations for detecting sarcasm and irony. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 2–7, Brussels, Belgium. Association for Computational Linguistics.
- Aditya Joshi, Pranav Goel, Pushpak Bhattacharyya, and Mark Carman. 2018. Sarcasm target identification: Dataset and an introductory approach. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Aditya Joshi, Vinita Sharma, and Pushpak Bhattacharyya. 2015. Harnessing context incongruity for sarcasm detection. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 757–762, Beijing, China. Association for Computational Linguistics.
- Aditya Joshi, Vaibhav Tripathi, Pushpak Bhattacharyya, Mark Carman, Meghna Singh, Jaya Saraswati, and

- Rajita Shukla. 2016. [How challenging is sarcasm versus irony classification?: A study with a dataset from English literature](#). In *Proceedings of the Australasian Language Technology Association Workshop 2016*, pages 123–127, Melbourne, Australia.
- Jihen Karoui, Farah Benamara, Véronique Moriceau, Viviana Patti, Cristina Bosco, and Nathalie Aussenac-Gilles. 2017. [Exploring the impact of pragmatic phenomena on irony detection in tweets: A multilingual corpus study](#). In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 262–272, Valencia, Spain. Association for Computational Linguistics.
- Shivani Kumar, Atharva Kulkarni, Md Shad Akhtar, and Tanmoy Chakraborty. 2022. [When did you become so smart, oh wise one?! sarcasm explanation in multi-modal multi-party dialogues](#).
- 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.
- Anna Xenia Laszlo. 2017. *A corpus-based textual analysis of irony and sarcasm in scripted discourse*. Ph.D. thesis, Hong Kong Polytechnic University.
- Christopher J Lee and Albert N Katz. 1998. The differential role of ridicule in sarcasm and irony. *Metaphor and symbol*, 13(1):1–15.
- John S Leggitt and Raymond W Gibbs. 2000. Emotional reactions to verbal irony. *Discourse processes*, 29(1):1–24.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- An-Ran Li and Chu-Ren Huang. 2020. [A method of modern chinese irony detection](#). In *From Minimal Contrast to Meaning Construct*, pages 273–288. Springer.
- Anran Li. 2021. *A corpus-based research of Chinese verbal irony*. Ph.D. thesis, Hong Kong Polytechnic University.
- 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.
- Emmy Liu, Chen Cui, Kenneth Zheng, and Graham Neubig. 2022. [Testing the ability of language models to interpret figurative language](#).
- Yaochen Liu, Yazhou Zhang, Qiuchi Li, Benyou Wang, and Dawei Song. 2021. [What does your smile mean? jointly detecting multi-modal sarcasm and sentiment using quantum probability](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 871–880, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. [Roberta: A robustly optimized bert pretraining approach](#).
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. [Distributed representations of words and phrases and their compositionality](#). In *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.
- Abhijit Mishra, Tarun Tater, and Karthik Sankaranarayanan. 2019. [A modular architecture for unsupervised sarcasm generation](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6144–6154, Hong Kong, China. Association for Computational Linguistics.
- Debora Nozza, Elisabetta Fersini, and Enza Messina. 2016. Unsupervised irony detection: A probabilistic model with word embeddings. In *KDIR*.
- Silviu Oprea and Walid Magdy. 2019. [Exploring author context for detecting intended vs perceived sarcasm](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2854–2859, Florence, Italy. Association for Computational Linguistics.
- Silviu Oprea and Walid Magdy. 2020. [iSarcasm: A dataset of intended sarcasm](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1279–1289, Online. Association for Computational Linguistics.
- Reynier Ortega-Bueno, Francisco Rangel, D Hernández Farias, Paolo Rosso, Manuel Montes-y Gómez, and José E Medina Pagola. 2019. Overview of the task on irony detection in spanish variants. In *Proceedings of the Iberian languages evaluation forum (IberLEF 2019), co-located with 34th conference of the Spanish Society for natural language processing (SEPLN 2019)*. CEUR-WS. org, volume 2421, pages 229–256.
- Hongliang Pan, Zheng Lin, Peng Fu, Yatao Qi, and Weiping Wang. 2020. [Modeling intra and inter-modality incongruity for multi-modal sarcasm detection](#). In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 1383–1392, Online. Association for Computational Linguistics.
- Jasabanta Patro, Srijan Bansal, and Animesh Mukherjee. 2019. [A deep-learning framework to detect sarcasm](#)

- targets. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6336–6342, Hong Kong, China. Association for Computational Linguistics.
- Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. [Deep contextualized word representations](#). In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.
- Rolandos Alexandros Potamias, Georgios Siolas, and Andreas Georgios Stafylopatis. 2020. [A transformer-based approach to irony and sarcasm detection](#). *Neural Computing and Applications*, 32(23):17309–17320.
- 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 Davide Buscaldi. 2012. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12.
- Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language resources and evaluation*, 47(1):239–268.
- Hannes Ritschel, Ilhan Aslan, David Sedlbauer, and Elisabeth André. 2019. Irony man: Augmenting a social robot with the ability to use irony in multimodal communication with humans. In *AAMAS*.
- Omid Rohanian, Shiva Taslimipour, Richard Evans, and Ruslan Mitkov. 2018. [WLV at SemEval-2018 task 3: Dissecting tweets in search of irony](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 553–559, New Orleans, Louisiana. Association for Computational Linguistics.
- Andrea Santilli, Danilo Croce, and Roberto Basili. 2018. A kernel-based approach for irony and sarcasm detection in italian. In *EVALITA@CLiC-it*.
- Rossano Schifanella, Paloma de Juan, Joel Tetreault, and Liangliang Cao. 2016. Detecting sarcasm in multimodal social platforms. In *Proceedings of the 24th ACM international conference on Multimedia*, pages 1136–1145.
- Ken-ichi Seto. 1998. On non-echoic irony. *Relevance theory: Applications and implications*, pages 239–255.
- Boaz Shmueli, Lun-Wei Ku, and Soumya Ray. 2020. [Reactive Supervision: A New Method for Collecting Sarcasm Data](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2553–2559, Online. Association for Computational Linguistics.
- Pavel Skrelin, Uliana Kochetkova, Vera Evdokimova, and Daria Novoselova. 2020. Can we detect irony in speech using phonetic characteristics only?—looking for a methodology of analysis. In *International Conference on Speech and Computer*, pages 544–553. Springer.
- Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and cognition*, volume 142. Cite-seer.
- Emilio Sulis, Delia Irazú Hernández Farías, Paolo Rosso, Viviana Patti, and Giancarlo Ruffo. 2016. [Figurative messages and affect in twitter: Differences between irony, sarcasm and not](#). *Knowledge-Based Systems*, 108:132–143. New Avenues in Knowledge Bases for Natural Language Processing.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *NIPS*.
- Martin Sykora, Suzanne Elayan, and Thomas W Jackson. 2020. [A qualitative analysis of sarcasm, irony and related #hashtags on twitter](#). *Big Data & Society*, 7(2):2053951720972735.
- Yi-jie Tang and Hsin-Hsi Chen. 2014. Chinese irony corpus construction and ironic structure analysis. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1269–1278.
- Deborah Tannen et al. 2005. *Conversational style: Analyzing talk among friends*. Oxford University Press.
- Maggie Toplak and Albert N Katz. 2000. On the uses of sarcastic irony. *Journal of pragmatics*, 32(10):1467–1488.
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. 2016. [Exploring the realization of irony in Twitter data](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pages 1794–1799, Portorož, Slovenia. European Language Resources Association (ELRA).
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. 2018a. Semeval-2018 task 3: Irony detection in english tweets. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 39–50.
- Cynthia Van Hee, Els Lefever, and Véronique Hoste. 2018b. [We usually don’t like going to the dentist: Using common sense to detect irony on Twitter](#). *Computational Linguistics*, 44(4):793–832.

- Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar, and Manish Shrivastava. 2018. A dataset for detecting irony in hindi-english code-mixed social media text. *EMSASW@ ESWC*, 2111:38–46.
- Rob Voigt and Dan Jurafsky. 2012. [Towards a literary machine translation: The role of referential cohesion](#). In *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pages 18–25, Montréal, Canada. Association for Computational Linguistics.
- Byron C Wallace. 2015. Computational irony: A survey and new perspectives. *Artificial intelligence review*, 43(4):467–483.
- Byron C. Wallace, Do Kook Choe, Laura Kertz, and Eugene Charniak. 2014. [Humans require context to infer ironic intent \(so computers probably do, too\)](#). In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 512–516, Baltimore, Maryland. Association for Computational Linguistics.
- Deirdre Wilson. 2006. The pragmatics of verbal irony: Echo or pretence? *Lingua*, 116(10):1722–1743.
- Deirdre Wilson and Dan Sperber. 2012. Explaining irony. *Meaning and relevance*, pages 123–145.
- Chuhan Wu, Fangzhao Wu, Sixing Wu, Junxin Liu, Zhigang Yuan, and Yongfeng Huang. 2018. [THU_NGN at SemEval-2018 task 3: Tweet irony detection with densely connected LSTM and multi-task learning](#). In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 51–56, New Orleans, Louisiana. Association for Computational Linguistics.
- Rong Xiang, Xuefeng Gao, Yunfei Long, Anran Li, Emmanuele Chersoni, Qin Lu, and Chu-Ren Huang. 2020. [Ciron: a new benchmark dataset for Chinese irony detection](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 5714–5720, Marseille, France. European Language Resources Association.
- Shiwei Zhang, Xiuzhen Zhang, Jeffrey Chan, and Paolo Rosso. 2019. [Irony detection via sentiment-based transfer learning](#). *Information Processing Management*, 56(5):1633–1644.
- Mengdi Zhu, Zhiwei Yu, and Xiaojun Wan. 2019. [A neural approach to irony generation](#).