Computational Irony

Will Machines Have the Ability to Identify, Comprehend and Produce Irony?

Foreword

My paper was already being printed, when this still came to my mind:
I shortly remark in my paper, that, to answer the question wether machines will be able to identify, comprehend and produce irony, it should – within the bounds of the paper – suffice, if the machine shows human-like irony-abilities. When thinking about machines possessing real genuine irony-abilities, Chalmers and the easy and hard problem of consciousness come to my mind. Although I answer the question, that machines will display human-like irony-abilities, positively, I think of this case as “easy” and computationally solvable. But genuine irony-abilities still are – from my point of view – a “hard problem” and it is not yet possible to answer whether machines will have real, genuine irony-abilities.

Computational Irony

Will Machines Have the Ability to Identify, Comprehend and Produce Irony?

Anaïs Siebers
Cognitive Science, Ruhr-University, Bochum, Germany,
anais.siebers@rub.de

21st August 2022

Abstract

It is difficult to imagine a machine which makes ironic remarks and reacts appropriately to humans making ironic statements. But the amount of verbal human-machine interaction in everyday life increases constantly, and the experienced quality depends heavily on the naturalness of the conversation and relationship between user and machine. This interaction has been discovered to be positively influenced by machine's use of irony. Furthermore, increasing hate speech has started research into the sentiment and meaning of social media content, to be able to distinguish hateful from ironic messages and only restrict hateful messages. Apart from that, computational irony might also be a useful tool to shed new light on the debate about irony processing and give more insight into the cognitive processes underlying irony. Current approaches to computational irony focus on the detection of irony in text. Since irony has a very important purpose in communication, it is foreseeable that future research will focus more strongly on other aspects like irony comprehension and production – also in speech. This paper investigates whether machines will be able to identify, comprehend and produce irony by investigating the current level of research of irony and computational irony. It will be argued that it is likely that machines reach a human-like level of irony identification, comprehension and production. Further research concerning irony will be necessary especially in the context of background-information, social, cultural and general context as well as when it is appropriate to use irony.
Keywords: irony identification; irony comprehension; irony production; irony; machines; computational irony

Contents

1 Introduction

Ironic machines are yet difficult to imagine. Even machines which laugh correctly to ironic remarks are still science fiction. But irony is a very common phenomenon in everyday language and communication. It is often used in interpersonal communication and shapes the quality of conversations and the conversation participant's perception of another (Ritschel et al., 2019, 1). Irony plays a very important role because it is "situated in a social context and has the purpose to communicate the mental states of the speaker" (Fabry, 2021, 6454). The complex pragma-linguistic phenomenon is studied in philosophy and linguistics (Karoui et al., 2017, 262; Fabry, 2021, 6454), but also in psychology, neuroscience and, lately, computer science.

1.1 Relevance of Irony-Abilities for Machines

In fiction, machines and especially robots are usually portrayed as humourless beings even if they already are depicted with advanced skills at natural language or movement (Binsted et al., 2006, 22). But there are various reasons why the abilities to identify, comprehend and produce irony are relevant for machines and why computer scientists became interested in irony.

The first aspect is that humour[1] "affects attention and memory, facilitates social interaction, and ameliorates communication problems" (Binsted et al., 2006, 22). With current trends to integrate virtual assistants and social (ro-)bots into everyday life, the communication between humans and machines faces new challenges concerning the naturalness of speech and expressive body and face language. If computers should be integrated in human life, they have to be humorous, because it will make them more credible, natural, efficient and acceptable. Irony is essential because it is part of socially intelligent behaviour (Binsted et al., 2006, 22; Ritschel et al., 2019, 1). Furthermore, the creativity, displayed by the use of irony, is important in human friendships and to build bonds between a virtual assistant and his user. It is not sufficient to tell canned jokes (Winters, 2021, 4).

Another aspect of the research into *computational irony* is, that it can provide insights into the cognitive processes behind humour by testing and following a particular theory of irony (Winters, 2021, 4; Binsted et al., 2006, 22).

Lastly, irony detection has become increasingly important for sentiment analysis and the differentiation of potential threats from ironic comments on social media platforms (Zhang et al., 2018, 2). At the same time, the increase of available data for analysis (for example tweets with #irony), first enabled research in the area of irony detection. Because the influence of social media and respectively the amount of published content has increased strongly in the last view years, automatic filtering and classifying of sentiments and / or hate speech has become a research focus (Van Hee, 2017, 15; Ghanem et al., 2020, 1).

[1] At this point I am talking about humorous irony. The differences between humour and irony will be highlighted later on.

1.2 Identify, Comprehend and Produce Irony

Although irony is ubiquitous in human life and present from a very young age – irony comprehension can first be observed in 3- and 4-year-old children (Köder and Falkum, 2021, 2) – the comprehension of irony is not yet understood (Fabry, 2021, 6456). The paper at hand deals with three aspects of irony: *irony identification*, *irony comprehension* and *irony production*, and only touches upon the social aspect of *irony appreciation*. While irony identification and comprehension are concerned with the aspect of listening to someone and understanding that the person used irony (directed at the machine), irony production also takes the proactive generation of ironic content into account (directed from the machine). Most research is concerned with understanding and detecting irony, but there are voices stressing the importance of dynamical irony production (Ritschel et al., 2019, 1). The three aspects are further also referred to as *irony-abilities*. This paper distinguishes between identification and comprehension because irony can be identified without being understood: there are *multimodal markers* for ironic statements. These also have relevance for the production, as they can also be actively used to indicate that one is making an ironic remark.

1.3 Epistemic and Argumentative Goal

The central question and epistemic goal of this paper is to investigate the current state of computational irony and discuss whether machines will be able to identify, comprehend and produce irony. To answer the question, first, irony will be examined with a focus on the purpose and kinds of irony, closing irony off from humour and sarcasm. Then, theories of irony processing and current computational approaches to irony are outlined, taking a look at the approaches and implementations. This, then, allows to subsequently discuss whether machines will have the abilities for irony identification, comprehension and production. The question will be answered positively, using humans' irony-abilities as a scale. Nevertheless, this is still a long way to go considering the current level of research. Especially, because irony still is hotly debated / researched and it is a not yet well understood phenomenon of language, which complicates the transfer to machines.

2 What is Irony?

After shortly introducing irony and its role in everyday life in the previous chapter, a more detailed look on irony is now taken. The definition of an acknowledged dictionary – Marriam Webster – roughly distinguishes three cases[2]:

1. a the use of words to express something other than and especially the opposite of the literal meaning

[2]Merriam-Webster. (n.d.). Irony. In Merriam-Webster.com dictionary. Retrieved August 20, 2022, from https://www.merriam-webster.com/dictionary/irony

b a usually humorous or sardonic literary style or form characterized by irony

c an ironic expression or utterance

2. a (1) incongruity between the actual result of a sequence of events and the normal or expected result

   (2) an event or result marked by such incongruity

   b incongruity between a situation developed in a drama and the accompanying words or actions that is understood by the audience but not by the characters in the play
   — called also *dramatic irony*

3. a pretense of ignorance and of willingness to learn from another assumed in order to make the other's false conceptions conspicuous by adroit questioning
   — called also *Socratic irony*

There are two aspects which reappear very often in irony research: the difference between *verbal* and *situational irony* (see 1. and 2.) and the role of *incongruity* and the opposite meaning of the literal meaning (see 1.a). In the literature, verbal, situational, *dramatic* and sometimes *Socratic irony* are distinguished (Van Hee, 2017, 10).

2.1 Different Kinds of Irony

The first kinds of irony explained are Socratic and dramatic irony. They can both be described as an incongruence between the observer's knowledge and the observed's pretended knowledge (considering performance aspects). Situational irony – also called irony of fate – rather describes the difference between two situations (often the expected and the actual situation), whereas verbal irony describes situations where a speaker intentionally states the opposite of his beliefs (Van Hee, 2017, 10). Most researchers, as for example Littman and Mey (1991, 131), only distinguish verbal and situational irony. There is a debate whether there is situational irony or whether by uttering an ironic statement about an ironic situation, it reduces and identifies the ironic situation only to the verbal: "Seule l'ironie du type 2 [ironie proprement verbale] va nous intéresser désormais" [Only irony of type 2 (verbal irony) is of interest for us.] (Kerbrat-Orecchioni 1975, 19 quoted in Littman and Mey, 1991, 133).

A distinction describes the forms in which situational irony is realised. It is made between *intentional*, *serendipitous* and *competence irony* (Littman and Mey, 1991, 137):

– **intentional irony** The irony if a reasonable action or thought to achieve a goal is taken, but the plan fails and has negative effect. For example, if a new check-out opens in a super-market and somebody rushes over to be the first and fastest of the queue, but then the cashier takes longer than the waiting time of the left queue would have been.

– **serendipitous irony** The irony of good-news/bad-news stories: the actor has no intention, but accidentally acts in a way which results in a possibility to fulfil his[3] goal, which nevertheless has a negative effect. An example would be a student who did not learn for a test and then gets sick.

– **competence irony** The irony of a competent actor who must fail, suffer in some way from the failure, and he has to be competent in the area of expertise required. A common example are "He should have known better"-stories. It is, for example, ironic if it burns at the fire brigade because a firefighter did not properly turn off the oven.

Concerning verbal irony, the intention behind an ironic utterance can be differentiated. Two basic types are discussed in the literature: *ironic praise* and *ironic criticism*. The difference is that ironic praise is marked by a negatively valenced utterance but a positive, appraising intention of the circumstances or listener (somebody who claims to be a beginner at cooking makes a great soup – "Of course, you cannot cook at all.") whereas ironic criticism is the opposite (the train is late again – "That's just great!") (Bruntsch and Ruch, 2017, 1f.). Ironic praise is less known and the "formerly neglected type of irony" (Bruntsch and Ruch, 2017, 13). But it is crucial when studying the role of humour, ability and personality in irony detection (Bruntsch and Ruch, 2017, 13). It often occurs during flirting, which leads to the next section: the purpose of irony and why it is used.

2.2 The Purpose of Irony

Irony is often associated with wit, intelligence and regarded as a "sophisticated, complex and prized mode of communication" (Attardo, 2000, 15). Roughly said, the purpose of irony is hidden in its social and rhetorical effects. Humans put extra effort into irony and encode a message in an ironical sentence. Moreover, they risk a misunderstanding, which is why uttering an ironical statement is also referred to as *risky bet* or *play* (Attardo, 2000, 15f.). Irony has been suggested as "language based social cognition task" (Kieckhäfer et al., 2019, 1). It is used because it allows to mutually estimate another and assort themselves socially because it shows the possession of certain knowledge required to decipher the implicit non-literal message in the ironic statement. This recognition is often accompanied by an indicator like a laugh (Gibbs et al., 2014, 589). In fMRI studies, the activation of the subcortical structures which are associated with the "reward processing of social events" has been observed (Obert et al., 2016, 1).

A lot of different abilities are required to identify and comprehend irony: the context, relationships and personalities of the interlocutors are just examples. Additionally, it challenges the perspective-taking capabilities of the communicative partners and the abilities of the listener to make inferences (Kieckhäfer et al., 2019, 1; Köder

[3]For reasons of readability, I will refrain from gendering.

and Falkum, 2021, 1). Often, irony is used to emphasise the plausibility and naturalness of our expectation and the absurdity of non-fulfilment (Valitutti and Veale, 2015, 153).

Littman and Mey (1991, 149) summarise two main purposes pursued by the use of irony: *social hedging goals* and *instructional goals*. The first was elaborated in detail in the paragraphs above: irony serves as a tool to mutually reveal the knowledge and values of people. The second is describes as a gentle approach to humorously comment and inform somebody (often parents their children) that there was a violation of a rule.

2.3 Irony, Sarcasm and Humour

Irony, sarcasm and humour are difficult to differentiate in everyday discourse. In academic research, there are various attempts to neatly define and distinguish between them (see for example Dynel2014, 619f.). Concerning irony and humour, they are often related to another in spoken and in written language (Gibbs et al., 2014, 575). There are certain similarities concerning ironic jokes like for example inappropriateness or a question-answer / rhetorical structure, understatement, jocularity (which was distinguished by Gibbs (2000, 2012)), hyperbole and so on (Gibbs et al., 2014, 576f.; Dynel, 2014, 620; Chłopicki, 2005, 962; Karoui et al., 2017, 262). But irony and humour are not the same. Humour is inter alia "that quality which appeals to a sense of the ludicrous or absurdly incongruous : a funny or amusing quality"[4]. Thus, irony can be humorous, but does not have to be. Furthermore, irony has usually two sides: a humorous one and a tragic one (Littman and Mey, 1991, 148).

It is important to define the factors which make irony humorous (Dynel 2018 quoted in Fabry, 2021, 6482). The difference between irony and humorous irony is that irony conveys an evaluative message from the speaker underneath the literal statement, which can but must not be humorous. Dynel (2014) proposes that there are two characteristics of irony: untruthfulness and negativity. There are various examples of non-ironic humour and phenomena such as (Dynel, 2014, 635): teasing, parody, absurdity, litotes and hyperbole, humorous lying, humorous metaphors, metonymy and sarcasm.

This leads to the controversial debate of the relation between sarcasm and irony. They are presumed to be quite identical. Some reasons speak for this relation: irony is one of the main components of sarcasm (Littman and Mey, 1991, 147f.). But one major difference is that there can be situational irony, but a situation cannot be sarcastic because sarcasm is always a speech act (Littman and Mey, 1991, 148). There are other differences as well. One is that sarcasm aims at hurting a listener – its object is an agent. Therefore, sarcasm is considered to be more aggressive and also shows some vocal clues as nasality (Van Hee, 2017, 15; Littman and Mey, 1991, 147; Karoui et al., 2017, 262).

[4]Merriam-Webster. (n.d.). Humor. In Merriam-Webster.com dictionary. Retrieved August 20, 2022, from https://www.merriam-webster.com/dictionary/humor

3 The Phenomenon of Irony

The previous chapter offered an overview of irony. Now, a stronger focus on the phenomenon of linguistic and spoken irony is taken. First, the most known and prominent theories of irony processing are introduced. Later on, the three abilities which are investigated in this paper – irony identification, comprehension and production – are examined with their respective multimodal markers, linguistic indicators and constituent features.

3.1 Theories of Irony Processing

The historically first theory of irony processing is called *Standard Pragmatic View* and stems from Grice's (1975) and Searle's (1979) philosophical work (Fabry, 2021, 6457) and is based on Grice's *Cooperative Principle* (CP) of a conversation. Grice puts forward four *maxims* (Grice, 1975, 45f.): quantity (be as informative but not more informative than is required), quality (do not say what you believe to be false or lack the adequate evidence for), relation (be relevant) and manner (be perspicuous), which have to be followed. If one of them is violated by the speaker, the listener has to infer why the utterances was flouted and the maxim violated and thus reassesses the pragmatic information to infer the nonliteral message (Grice, 1975, 45f.; Gibbs Jr., 2002, 457-459; Fabry, 2021, 6457).

Following research identified this view as inadequate, because it is neither necessary nor sufficient (Sperber and Wilson (1981) quoted in Van Hee, 2017, 12). Furthermore, there is experimental evidence speaking against it (reaction times (RTs) for irony detection would be longer than normal reaction times which is not always the case) (Fabry, 2021, 6462) and even neo-Griceans as Horn (1988) criticised it as "at best incomplete" (hua, 2017, 50).

The next approach to irony, presented here, is the *Direct Access View*, strongly influenced by Gibbs. It claims that the listener does not need to analyse the literal meaning before identifying a violation of a CP and then identifying the speaker's "real" message (Gibbs Jr., 2002, 460; Fabry, 2021, 6457). Nevertheless, the RTs to process irony would sometimes still be longer following the Direct Access View because some expressions in the context might be novel and have to be integrated. Compared to the Standard Pragmatic View, context influences the linguistic processing (Gibbs Jr., 2002, 460, 462). But there also is experimental evidence speaking against the Direct Access View (Fabry, 2021, 6462) and some researches criticise that the literal interpretation is neglected (Gibbs Jr., 2002, 458).

The *Graded Slience Hypothesis* assumes that there are so-called "mental lexica" where the most salient meaning of a word or expression is represented, which would explain that there are some ironical remarks where the RT is shorter than for the literal meaning (Giora and Fein, 1999, 202f.; Fabry, 2021, 6458). If there is a word with multiple meanings, the more popular, prototypical or frequently used meaning is salient and activated. The degree of salience depends on familiarity, frequency of use

and conventionality. There are psycholinguists which stress the influence of context on this selective access (for example Sperber and Wilson, 1995; Gibbs, 1994 quoted in Giora, 1991, 921). The results of empirical research at this moment are "the most compatible with the Graded Salience Hypothesis", but still find gaps (Filik et al., 2014, 825).

Three other theories, which are less prominent but nevertheless widely cited, are worth mentioning. First, there is the *Echoic Mention Theory* in which "a speaker echoes a remark in such a way as to suggest that he finds it untrue, inappropriate, or irrelevant" (Sperber and Wilson, 1981, 307 quoted in Van Hee, 2017, 12). Second is the *Pretence Theory*, which suggests that – rather than echoic mention – pretence is involved in verbal irony (Gibbs et al., 1991, 527). The act of pretence is staged and non-serious and the speaker plays an imaginary speaker and listener (Gibbs et al., 2014, 578). Lastly, there is the *Indirect Negation Theory* by Giora (1995) which goes back to the Gricean approach. Instead of only processing the literal interpretation first, the ironic and literal interpretation are accessed simultaneously, and the underlying discrepancy is present when hearing an ironic utterance (Van Hee, 2017, 13).

3.2 Irony Identification

To begin with, it has to be considered that there are markers of irony which involve multimodal bodily clues and not only the linguistic statement (Attardo et al., 2003, 243). There are indicators through gestures, facial expressions, vocal and visual cues in verbal irony. These markers should not be confused with the actual phenomenon of irony, since they are optional and only ease the recognition of irony if present (Attardo, 2000, 15). Attardo et al. (2003, 246) thus proposes a hierarchy of cues: behavioural cues $\rightarrow$ intonational clues $\rightarrow$ semantic clues.

The bodily markers mainly concentrate on the face. Among others, there is gaze aversion – simple horizontal saccades – and a decreased eye contact (Williams et al., 2009, 4). A facial expression called "blank face" is also a known marker of irony. Furthermore, the eyebrows are raised or lowered, the eyes are wide open or squinting or rolling, there is winking, nodding or smiling (Attardo et al., 2003, 245f.).

Apart from the bodily markers, there are a lot of vocal cues like the tone of voice, pitch (Attardo et al., 2003, 243). There is no "ironical intonation per se", but the tone of voice can differ during ironic utterances and become parodic or pretending towards children (Köder and Falkum, 2021, 1). Furthermore, the intonation is reported as flat contour question intonation which neither rises nor falls. Frequently, a nasalisation is described as a marker of irony, as well as a slower speech rate or syllable lengthening (Attardo et al., 2003, 245). There are also laughter syllables in the utterance, indicating the irony (Gibbs et al., 2014, 588; Attardo et al., 2003, 245). Valitutti and Veale (2015, 158) discovered a form of semantic slanting which is brought upon typically positive words.

It can be said that there are two forms of communication in spoken irony which

consists of a *metacommunicative* layer which is constituted by the markers such as vocal or bodily cues, and the ironic statement itself. The markers are metacommunicative in that they facilitate understanding the intention of the speaker, but they can also appear in a *paracommunicative* manner, accompanying the ironical statement, but not communicating about the ironical statement (Attardo et al., 2003, 257).

Coming back to the linguistic content of the ironic statement, various cues indicate that the sentence is not meant literally. Most prominent is probably the previously discussed (see section 3.1) incongruence between the literal and meant (often opposite) meaning of a statement.

3.3 Irony Comprehension

The difference between irony comprehension and identification is less important for humans, since identification and comprehension of irony go hand in hand for humans. But machines need a stricter differentiation because they sometimes can identify or *classify* statements as ironical without understanding what exactly the irony is (see sections 4 and 5). In this paper, linguistic cues and markers are considered for the identification of irony. But aspects which require more background knowledge are relevant and subsumed under irony comprehension.

One of the background factors which is assumed to influence irony understanding are social and cultural information about the context and speaker. Moreover, the relationship of the interlocutors is of importance (Kieckhäfer et al., 2019, 3). It has been discovered that the *Theory of Mind* (ToM) has a vital role in filling the void between the literal and the speaker's meaning (Spotorno et al., 2012, 25). Current results bring into focus that the personality of the recipient is – apart from the mentalisation abilities and the context – very relevant (Kieckhäfer et al., 2019, 12). The same goes for social knowledge as stereotypes which appear to be activated during irony understanding (Champagne-Lavau and Charest, 2015, 1). Nevertheless, researchers still cannot agree on the temporal role of context in irony comprehension (Giora and Fein, 1999, 201). To summarise, different social abilities and knowledge seem to be required to be able to comprehend irony (Kieckhäfer et al., 2019, 2): social context, world knowledge / common sense, meta-representation, cultural knowledge and familiarity.

There are interesting findings by Gibbs et al. (1991, 523), who concluded that humans can comprehend what was meant by the ironic statement without identifying that it was ironic, ironic statements are found to be especially ironic if they echo social norms or expectations and that concerning situational irony, that ironic statements can be understood as ironic because of the situation without the speaker's intention to be ironic.

3.4 Irony Production

Irony production heavily relies on both, irony identification and comprehension, because to generate irony, the same rules and mechanism have to be used. Apart from a suitable language, non-verbal behaviour like prosody, facial expressions and gestures have to be tuned to fit the dialogue (Ritschel et al., 2019, 1). Signalling the listener that the interaction is different from the previous conversation and marking play by vocal signals also highlights that speakers, producing irony, use more contrasts over more dimensions at the same time (Gibbs et al., 2014, 584f.). It has been outlined before, that laughter accompanies ironic statements. However, it is conspicuous that the majority of laughter in a conversation is created by the speaker and not the listener (Gibbs et al., 2014, 588). Making an ironic statement also requires extensive world, contextual and social knowledge because the statement has to be created violating expectations of the listener by using the opposite of the expected or using common ironic expressions at the right moment.

4 Computational Approaches to Irony

Compared to the very elaborate research on irony, less can be yet said concerning the computational approaches to irony. One of the first attempts to create a computational theory was undertaken by Littman and Mey (1991). They proposed three tasks a computational theory of irony would have to fulfil (Littman and Mey, 1991, 131): 1) distinguish irony from non-irony, 2) describe why a situation is (not) ironic and 3) generate description of ironic situations. Their tasks are similar to the aspects dealt with in this paper: number 1 with irony identification, number 2 and irony comprehension and number 3 with irony production. But until now, this computational irony in the context of human-machine and human-robot interaction is not realised (Ritschel et al., 2019, 1).

4.1 Main Approaches

There are two main approaches to tackle the identification of irony, as for other natural language problems in computer science. One approach are rule-based methods and the other are machine learning-based methods (either supervised or unsupervised) (Van Hee, 2017, 16f.). The difference between rule-based and machine-learning methods is that rule-based approaches are based on lexical information and the knowledge is engineered whereas machine-learning enables to exploit different types of features like bag of words, syntax, sentiment and semantic information without hand-engineering. Nowadays, deep leaning techniques become very popular for this task because they allow to integrate word embeddings and semantic relatedness (Van Hee, 2017, 17).

As outlined before, mainly social media content is used because there is a lot of data available, which makes training with machine-learning algorithms possible.

Another data resource are Amazon reviews. Furthermore, self-describing hashtags like #irony and #sarcasm can be used and this reduced manual annotation efforts. Nevertheless, this has an impact on the quality of the ironic and non-ironic message data sets.

Most approaches to computational irony are centred around irony identification / detection. There are few which deal with the generation of irony and not many concerning the comprehension of irony, in other words, the reproduction or explanation of the irony of an ironic utterance or situation. All approaches concentrate on verbal, written irony yet.

4.2 Concrete Implementations

There are various forms of how irony detection was implemented, always trying to increase accuracy and performance of the algorithm. The first algorithm mentioned here was a semi-supervised algorithm which made use of punctuation and syntactic patterns. A similar approach was to extract part-of-speech tags and to make use of different feature types like syntactic, semantic and lexical information. Among them were punctuation, emoticons, polarity n-grams, character n-grams, verb tenses, semantic similarity, emotional context (Van Hee, 2017, 17f.). Onomatopoeic expressions for laughter, positive interjections and specific morphosyntactic constructions have also been used as irony features (Frenda, 2016, 2). Another algorithm used the hashtags and trained a classifier which brought the insight that not all tweets are correctly hashtagged and that sarcasm is different on Twitter (Van Hee, 2017, 18). Some researchers developed pipelines to extract different types of features for the detection of irony (Van Hee, 2017, 18). Other approaches made use of Word2Vec or convolutional neural networks (CNN). They tried to create semantic clusters from large background corpora and also include features as sentiment, emotion and personality (Van Hee, 2017, 19).

Further recent work includes even more contextual information such as "author profiles, conversational threads, or querying external sources of information" (Karoui et al., 2017, 268). Exploratory research investigated whether many languages share similar irony features, based on studies revealing that irony is a universal phenomenon. Their multilingual algorithmic approach resulted in a performance comparable to the monolingual algorithms(Ghanem et al., 2020, 2f.).

5 Will Machines Identify, Comprehend and Produce Irony?

After summarising the current level of research concerning the basics of what irony is and the computational approaches towards irony, whether machines will be able to "have irony" can be discussed. A machine should show these abilities because the speech act of irony is a complex strategy which is manifested at different levels of

pragma-linguistic analysis with the primary perlocutionary effect to break patterns of expectations (Gibbs et al., 1991, 523) and a common device for social hedging, friendship and estimating intelligence. Moreover, a lot of people appreciate a good sense of humour (including ironic remarks).

Summarising the sections about irony, it can be said that there are two main forms of irony: verbal and situational irony. Verbal irony can be further distinguished into spoken and written irony, which is relevant because it requires a very different handling in computational irony. Furthermore, the optional yet facilitating markers usable to identify the ironic statements greatly differ between spoken and written irony. Whereas verbal irony requires identificatory, comprehensive and productive abilities, situational irony does not necessarily require productive abilities but is – at the same time – ultimately necessary to be able to produce irony in the sense that it offers the possibility for an ironic statement. Except for the special case of dramatic irony where a writer has to create fictional ironic situations, humans do not create ironic situations. Thus, in this context, the following cases are distinguished (the actual phenomenon is the same for spoken and written irony and partly even for situational irony because it touches upon the underlying character of irony):

1. **verbal irony** (identification & comprehension & production)
    a *spoken irony*: markers (→body language & vocal cues), actual phenomenon
    b *written irony*: markers (→punctuations & emoticons), actual phenomenon
2. **situational irony** (identification & comprehension & *production*)

To be able to answer whether machines will be able to have irony, it first has to be considered that they need not understand situational irony because it is less frequent as situational irony. In section 2.1, it is outlined that many researchers neglect the notion of situational irony because they claim it is contained in verbal irony. If irony is the negation of the literal meaning, there is a metaphysical problem that the ironic statement regresses ad infinitum by ironising itself by another ironic statement. The solution opted for by some linguists and philosophers is to apply the ironic statement to the speaker's attitude and not the world as such (Littman and Mey, 1991, 133f.). Which accentuates that irony is an affair of using language which is only done in a concrete situation. Thus, there is always a situation underlying verbal irony. Hence, in the following, the focus will only be on verbal irony, keeping in mind that there are ironic situations functioning as material for verbal irony.

5.1 Scope and Limitations of Computational Irony

Foremost, it has to be considered that people with, for example, autism or schizophrenia and people with specific personality traits do not all show the same competence for irony identification, comprehension and production. A lot of people are also called "humourless". Thus, as with every criterion, it is difficult to define "the norm"

and what level of proficiency in each ability indicates irony abilities – for humans as well as for machines. A good indicator is a test like the Turing test evaluating the naturalness of the interaction: It "is sometimes used more generally to refer to some kinds of behavioural tests for the presence of mind, or thought, or intelligence in putatively minded entities"[5]. Nevertheless, all possible means to evaluate abilities like irony identification, comprehension and production are not yet sufficiently developed to claim: "Now the machine has irony-abilities". When talking about computational irony, it has to be considered that computers, respectively algorithms, are utilised around the world. As there is cultural specific irony, algorithms will either be language specific or utilise monolingual approaches, which makes comparison more difficult. Ghanem et al. (2020) discovered, for example, that Arabic is more difficult to deal with than other languages.

Be that as it may, it can be argued that there already are good approaches showing indicators that machines will be able to have or simulate human-like irony[6]. Current approaches to identify written irony in tweets are already very advanced and can be considered as at least average compared to humans. But there still is a lack of real comprehension of what the irony is about and identifying irony in a greater context. Written irony is mostly researched by using markers as well as linguistic features to identify ironic statements. However, spoken irony is not yet used in the computational approaches towards irony identification, comprehension nor production. This might be due to the availability of data as well as the level of research and progress comparing written and spoken text. Since there are a lot of recent advances and huge progress in the area of speech analysis, a change of focus in the research of computational irony is to be expected. Furthermore, it is likely that the development of computational irony identification in spoken irony will start off with the linguistic content and then make a similar development as the development of the computational irony identification of written irony, starting to use irony markers and contextual information (the current focus of research (Zhang et al., 2018, 3)) as well.

Concerning contextual knowledge, a lot of research still has to be done for written irony identification – but even more so for irony comprehension and artificial intelligence in general (Littman and Mey, 1991, 144). Approaches like the knowledge graph Atomic (Hwang et al., 2021) are first attempts to build machine knowledge necessary and required to be able to build a Theory of Mind or situational expectations. Another currently strongly researched aspect important to irony comprehension is meta-representational reasoning (Gibbs et al., 2014, 579). Although the algorithms presented in this paper already show astonishing performance in identifying irony, irony comprehension, as many other reasoning tasks, is difficult for machines and not yet realised. Comprehension would require to understand and be able to explain or

[5]Oppy, Graham and David Dowe, "The Turing Test", The Stanford Encyclopedia of Philosophy (Winter 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/win2021/entries/turing-test/.

[6]I will not distinguish between really possessing and imitating human intellectual abilities as irony. The superficial presence of non-distinguishable natural behaviour will be sufficient in this context.

reproduce the irony of the statement – also considering the situation. Using tone of voice and the salience of the expression "That's just great", it is easy to identify the statement as ironic. But to say, why it is ironic, might require the background information that the person ran to the platform, fearing to miss his train and arrives at the platform, just to be informed that the train was cancelled.

With current approaches, the generation of irony is far from natural, but already indicates that machine irony production is perceived as positive: increasing the acceptance and willingness to interact with the machine. But the results are only valid for written irony, since there are no attempts to produce spoken irony yet. To create an ironic statement, first a situation as to be identified and comprehended as ironic. Therefore, again, a lot of contextual information and reasoning capabilities are required. Without comprehension of the irony, it will be impossible to create irony. Furthermore, another aspect, which is less discussed in this paper until now, would be required: creativity. To generate ironic statements, one has to be creative and use a statement like "That's just great", which – at first glance – seems like a relatively easy task for machines. The main problem with this is the appropriateness and suiting of an ironic remark at a given situation. Irony at the wrong time is harmful and even inappropriate (Ritschel et al., 2019, 1). Laughing at the wrong time or identifying something as irony which is not, does not increase the relationship of speaker and listener, but has an opposite effect and decreases acceptance. Additionally, ironic praise and ironic criticism have been identified as two very different forms of irony. It is not clear when humans apply, which form and ironic praise has not yet been considered for computational irony production. Generally, the in section 2.1 introduced ironies: intentional, serendipitous or competence irony, can be used as rules of thumb when to make an ironic statement. Nevertheless, the description is far from a realisable algorithmic procedure to identify the appropriateness of irony and moreover does not cover all aspects.

To put it in a nutshell, as with the research of irony, the research and development of computational irony is still at the beginning. Nonetheless, it shows promising approaches and successes. Major problems arise at currently intensively researched areas of artificial intelligence as context knowledge, world knowledge, Theory of Mind and reasoning. The naturalness of artificially generated speech and also facial and bodily gestures is also investigated and improved in the field of robotics. Since computational irony is part of human interaction and will thus also be part of the – to be expected increasingly natural – human-machine / human-robot interaction, the research of computational irony will not cease to an end in the foreseeable future. These are good preconditions that machines will indeed in future exhibit human-like irony-abilities.

6 Conclusion, Implications and Outlook

Irony is of great importance because it can be used as a generator of emotions, serving as a creative and motivational tool which would enhance usability, productiv-

ity and the pleasantness of human-machine interaction. Moreover, humorous irony affects psychological states such as attention, memorisation, decision-making and is an instrument which can be used in motivation and persuasion. As such, it also bears dangers like bots influencing people's opinions over social media. But, at the same time, irony emphasises the inconsistency in clichés and stereotypes, enabling people to become more open and creative (Binsted et al., 2006, 28).

Computational irony has a large potential not only for human-machine interaction and perceived social intelligence (Ritschel et al., 2019, 1), but also sentiment analysis of texts (Van Hee, 2017, 19) and helps to understand the cognitive processes behind irony. Furthermore, if it is possible to write an algorithm which identifies, comprehends and produces irony, it might also be of help for people who struggle with using irony. Nonetheless, irony is a highly subjective device and there is nothing as "one standard default" rule or algorithm to irony. An interesting aspect to mention is that autistic people, although non-autistic and autistic people have difficulties understanding another, understand other autistic people well as do non-autistic and non-autistic people (Crompton et al., 2021, 4): this is very important in regard of the current trend to personalise machines and robots. Human understanding and, thus, the use of irony are highly individual.

Although strongly debated amongst philosophers, psychologists and linguists, the underlying character of the phenomenon of irony is still unclear. There is a common consensus that irony has to do with the incongruence between the meant and literal meaning of a spoken sentence. Most research concerning computational irony until now has focused on irony identification, neglecting the aspect of incongruence with the context, but are starting to include the context more strongly. This might in future shed a new light on the debate around irony processing.

To summarise, the current development in the area of computational irony indicates that machines will in future show human-like irony. The identified markers for irony should be used in irony identification and are likely to be imitated without great effort in the production of irony with advances in the fields of artificial intelligence and robotics. But since they are not necessary for irony, they should be objective for advanced computational irony research. As for the identification of irony in written text and also the production of an ironic sentence given a specific situation, the theories of irony processing agree on a strong difference between the literal and meant, which can be and is exploited in the identification and production of irony. The comprehension of irony and the appropriateness of making an ironic statement are the most difficult problems which will need to be tackled for the development of computational irony. Further research is needed to explain in how far irony in text and speech are different and when we decide to make what kind of ironic comment.

References

(2017). Neo-Gricean Pragmatics. In Huang, Y., editor, *The Oxford handbook of pragmatics*, Oxford handbooks, pages 47–78. Oxford University Press, Oxford ; New York, NY, first edition edition. OCLC: ocn972539357.

Attardo, S. (2000). Irony markers and functions: Towards a goal-oriented theory of irony and its processing. *Rask*, 12(1):3–20.

Attardo, S., Eisterhold, J., Hay, J., and Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor - International Journal of Humor Research*, 16(2).

Binsted, K., Bergen, B., Coulson, S., Nijholt, A., Stock, O., Strapparava, C., Ritchie, G., Manurung, R., Pain, H., Waller, A., and O'Mara, D. (2006). Computational Humor. *IEEE Intelligent Systems*, 21(2):59–69.

Bruntsch, R. and Ruch, W. (2017). Studying Irony Detection Beyond Ironic Criticism: Let's Include Ironic Praise. *Frontiers in Psychology*, 8:606.

Champagne-Lavau, M. and Charest, A. (2015). Theory of Mind and Context Processing in Schizophrenia: The Role of Social Knowledge. *Frontiers in Psychiatry*, 6.

Chłopicki, W. (2005). The Linguistic Analysis of Jokes. *Journal of Pragmatics*, 37(6):961–965.

Crompton, C. J., DeBrabander, K., Heasman, B., Milton, D., and Sasson, N. J. (2021). Double Empathy: Why Autistic People Are Often Misunderstood. *Frontiers for Young Minds*, 9:554875.

Dynel, M. (2014). Isn't it ironic? Defining the scope of humorous irony. *HUMOR*, 27(4).

Fabry, R. E. (2021). Getting it: A predictive processing approach to irony comprehension. *Synthese*, 198(7):6455–6489.

Filik, R., Leuthold, H., Wallington, K., and Page, J. (2014). Testing theories of irony processing using eye-tracking and ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(3):811–828.

Frenda, S. (2016). Computational rule-based model for Irony Detection in Italian Tweets. Napoli.

Ghanem, B., Karoui, J., Benamara, F., Rosso, P., and Moriceau, V. (2020). Irony Detection in a Multilingual Context. arXiv:2002.02427 [cs].

Gibbs, R. W., Bryant, G. A., and Colston, H. L. (2014). Where is the humor in verbal irony? *HUMOR*, 27(4).

Gibbs, R. W., O'Brian, J., and O'Brian, J. (1991). Psychological aspects of irony understanding. *Journal of Pragmatics*, 16:523–530.

Gibbs Jr., R. W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics*, 34:457–486.

Giora, R. (1991). On the priority of salient meanings: Studies of literal and figurative language. *Journal of Pragmatics*, 31:919–929.

Giora, R. and Fein, O. (1999). Irony: context and salience. *Metaphor and Symbol*, 14:241–257.

Grice, H. P. (1975). Logic and Conversation. In *Speech arts*, number 3 in Syntax and semantics, pages 41–58. Elsevier.

Hwang, J. D., Bhagavatula, C., Bras, R. L., Da, J., Sakaguchi, K., Bosselut, A., and Choi, Y. (2021). COMET-ATOMIC 2020: On Symbolic and Neural Commonsense Knowledge Graphs. arXiv:2010.05953 [cs].

Karoui, J., Benamara, F., Moriceau, V., Patti, V., Bosco, C., and Aussenac-Gilles, N. (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, pages 262–272, Valencia, Spain. Association for Computational Linguistics.

Kieckhäfer, C., Felsenheimer, A. K., and Rapp, A. M. (2019). A New Test for Irony Detection: The Influence of Schizotypal, Borderline, and Autistic Personality Traits. *Frontiers in Psychiatry*, 10:28.

Köder, F. and Falkum, I. L. (2021). Irony and Perspective-Taking in Children: The Roles of Norm Violations and Tone of Voice. *Frontiers in Psychology*, 12:624604.

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

Obert, A., Gierski, F., Calmus, A., Flucher, A., Portefaix, C., Pierot, L., Kaladjian, A., and Caillies, S. (2016). Neural Correlates of Contrast and Humor: Processing Common Features of Verbal Irony. *PLOS ONE*, 11(11):e0166704.

Ritschel, H., Aslan, I., Sedlbauer, D., and Andre, E. (2019). Irony Man: Augmenting a Social Robot with the Ability to Use Irony in Multimodal Communication with Humans. Montreal, Canada.

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J.-B., and Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: The case of irony. *NeuroImage*, 63(1):25–39.

Valitutti, A. and Veale, T. (2015). Inducing an ironic effect in automated tweets. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 153–159, Xi'an, China. IEEE.

Van Hee, C. (2017). *Can machines sense irony? – Exploring automatic irony detection on social media*. PhD thesis, Universiteit Gent, Fakulteit Letteren en Wijsbegeerte, Gent.

Williams, J. A., Burns, E. L., and Harmon, E. A. (2009). Insincere Utterances and Gaze: Eye Contact during Sarcastic Statements. *Perceptual and Motor Skills*, 108(2):565–572.

Winters, T. (2021). Computers Learning Humor Is No Joke. *Harvard Data Science Review*.

Zhang, S., Zhang, X., Chan, J., and Rosso, P. (2018). Irony Detection via Sentiment-based Transfer Learning.

Competing Interests

This paper was written in the context of the course "Humour and Irony: Perspectives from Philosophy and Cognitive Science" at the Ruhr-University Bochum. The paper is rewarded with credit points and evaluated with a grade.

Statement of Authorship

I hereby certify under oath that the paper I am submitting is entirely my own original work except where otherwise indicated. I have not used any auxiliary means other than those listed in the bibliography or identified in the text and any use of the works of any other author, in any form, is properly acknowledged at their pint of use with indication of the source.

Anaïs Siebers (firstname, surname)

21.08.2022, Siegburg (date and place)

(signature)