

A qualitative analysis of sarcasm, irony and related #hashtags on Twitter

Big Data & Society
July–December: 1–15
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2053951720972735
journals.sagepub.com/home/bds
 SAGE

Martin Sykora , Suzanne Elayan and Thomas W Jackson

Abstract

As the use of automated social media analysis tools surges, concerns over accuracy of analytics have increased. Some tentative evidence suggests that sarcasm alone could account for as much as a 50% drop in accuracy when automatically detecting sentiment. This paper assesses and outlines the prevalence of sarcastic and ironic language within social media posts. Several past studies proposed models for automatic sarcasm and irony detection for sentiment analysis; however, these approaches result in models trained on training data of highly questionable quality, with little qualitative appreciation of the underlying data. To understand the issues and scale of the problem, we are the first to conduct and present results of a focused manual semantic annotation analysis of two datasets of Twitter messages (in total 4334 tweets), associated with; (i) hashtags commonly employed in automated sarcasm and irony detection approaches, and (ii) tweets relating to 25 distinct events, including, scandals, product releases, cultural events, accidents, terror incidents, etc. We also highlight the contextualised use of multi-word hashtags in the communication of humour, sarcasm and irony, pointing out that many sentiment analysis tools simply fail to recognise such hashtag-based expressions. Our findings also offer indicative evidence regarding the quality of training data used for automated machine learning models in sarcasm, irony and sentiment detection. Worryingly only 15% of tweets labelled as sarcastic were truly sarcastic. We highlight the need for future research studies to rethink their approach to data preparation and a more careful interpretation of sentiment analysis.

Keywords

Social media, sarcasm, irony, sentiment analysis, Twitter

Introduction

Sentiment analysis has seen significant and widespread applications in various domains ranging from national security and crisis response (Cheong and Lee, 2011; Kumar et al., 2011; Sykora et al., 2013), brand management (Pang and Lee, 2008), business intelligence (Fan and Gordon, 2014; Goonetilleke et al., 2014), public health (Gruebner et al., 2016; Kuehn, 2015) and more broadly across our daily lives (Ilieva and McPhearson, 2018). Indeed, sentiment analysis is a popular and recurring element in social media studies with automated analytics (Ravi and Ravi, 2015). In this context it refers to computational techniques that allow the detection of valence, activation, as well as other emotions in sparse-text commonly encountered on social media at scale. Given its popularity and these numerous applications, a substantial effort by the information retrieval (IR) and natural language processing (NLP) communities has been dedicated to developing techniques, algorithms and approaches to detect

affect more accurately in social media messages (Pang and Lee, 2008; Ravi and Ravi, 2015). Due to the prevalence of conducting sentiment analysis on large unstructured user generated content, accurate detection poses a challenge (Phan et al., 2020). Nevertheless, limitations on text length imposed by most social media and microblogging services such as Twitter arguably do nothing to discourage creative language use such as sarcasm and irony which allow strong sentiments to be expressed effectively (Ghosh et al., 2015).

Traditional NLP tools do not consider how words can be used playfully (ibid.). Hence, figurative language

Centre for Information Management, School of Business and Economics, Loughborough University, Loughborough, UK

Corresponding author:

Martin Sykora, Centre for Information Management, School of Business and Economics, Loughborough University, Loughborough, LE11 3TU, UK.
Email: M.D.Sykora@lboro.ac.uk



use on social media and especially the use of sarcasm and irony, in witty and inventive ways, pose significant challenges to these approaches. There have been efforts to automatically detect such language use in order to help improve overall sentiment analysis accuracy. Despite some relatively successful attempts in automated sarcasm and irony detection (Bamman and Smith, 2015; Joshi et al., 2016a), there is a distinct gap in the understanding of data quality used in training datasets that unfortunately can significantly impact supervised or semi-supervised¹ machine learning (ML) models, which are most often used in developing these kinds of approaches (e.g., Bamman and Smith, 2015; Khattri et al., 2015; Rajadesingan et al., 2015). Another issue is the qualitative appreciation of semantic characteristics of the individual social media messages which are little understood. Hence, the primary contribution of this paper is a manual semantic analysis of over 4000 Twitter messages (1820 of which relate to 25 different topics and events, ranging from cultural events, product releases, accidents, scandals to terror incidents), providing a much needed qualitative survey of sarcastic and ironic language use on Twitter. This is in addition to the presented literature, and while our review is not systematic, crucial links are drawn with existing research efforts in computational approaches for sarcasm and irony detection. For an exhaustive systematic review of specific machine learning models and feature sets, we direct readers to Sarsam et al. (2020). One of our resulting conclusions is that prior work in this area may have falsely attributed and relied on very low quality training, validation and testing datasets² with automated language models. It is hoped that this paper provides a much needed re-assessment and acts as a catalyst to stimulate substantive academic discussion in this area.

The remainder of this paper is organised as follows. The next section introduces background and relevant theory on sarcasm and irony as well as an overview of automated approaches for their detection, and a brief introduction to sentiment analysis. The subsequent section presents the manual semantic annotation analysis of two Twitter datasets, establishing the prevalence of sarcastic and ironic expressions on Twitter; not previously reported across such a varied dataset. The penultimate section provides a discussion and reflections on findings, issues and suggestions for potential future direction of research. The paper is finally concluded with conclusions.

Background, prior work and theory

Irony and sarcasm are subtle figures of speech that can be traced back to ancient Greece, even preceding the philosopher Socrates, who used irony to illustrate his

views (Lee and Katz, 1998). Smith (n.d.) draws upon definitions and explanations given by socio-psycholinguists such as Fowler (1965), Kreuz and Glucksberg (1989) and Gibbs (2000), to finally arrive at this definition; “Irony and sarcasm are used to portray meanings that differ from the literal meaning of an utterance; many times this can be an opposite or hyperbole”. Sarcasm unlike irony is an aggressive and often hostile type of humour (Norrick, 2003). Specifically, Norrick (ibid.) classified conversational humour into four types: (1) jokes, (2) anecdotes, (3) wordplay and (4) irony, while sarcasm and mockery do not fall under any of these four categories as they can be found across all four. The Oxford English Dictionary defines “irony” as “the funny or strange aspect of a situation that is very different from what you expect” and “sarcasm” as “a way of using words that are the opposite of what you mean in order to be unpleasant to somebody or to make fun of them” (OED, 2020). Although irony and sarcasm are very closely related, they are hence distinctly different, with sarcasm being a more aggressive form of humour.

Expressions of sarcasm and irony within social media posts pose major challenges to real-world applications of established sentiment analysis techniques. In their literature review on sentiment analysis, Ravi and Ravi (2015: 14) highlight that “ambiguity and vagueness have been considered as major issues [...]” especially with regards to challenges of “sarcasm (mock[ing] or convey[ing] irony), rhetoric [and] metaphor”. Ghosh et al. (2015: 1) state that such language “allow[s] strongly-felt sentiments to be expressed effectively, memorably and concisely [...]” although “NLP tools [...] do not take account of how words can be used playfully and in original ways”. This is a significant issue for automated techniques and the accurate assessment of creative language use, which is prevalent across social media platforms. Although there is a genuine lack of studies reviewing the prevalence of the above issue, Maynard and Greenwood (2014) have tentatively suggested that sarcasm alone may possibly account for as much as a 50% drop in accuracy when detecting sentiment automatically. Unfortunately, they (ibid.) did not clearly evidence this claim and only considered 134 tweets with a #sarcasm hashtag dataset. As we will illustrate in ‘Prevalence of sarcasm and irony: A manual semantic analysis’ section, a drop in accuracy for sentiment analysis is due to human communication involving expressions where the author may mean to convey the opposite of what he/she is communicating, and effectively transforming the polarity of the detected sentiment.

Given the potential impact on sentiment analysis and related automated methods, work in this area deserves careful attention. In the field of sentiment

analysis, this has not received the level of scrutiny one may expect; for instance, in their comprehensive sentiment analysis review, Ravi and Ravi (2015) only found four papers on the topic out of over 160 reviewed articles. Similarly, Yadollahi et al. (2017) do not mention sarcasm and irony at all, despite proposing a useful nuanced taxonomy for sentiment analysis tasks, they do not acknowledge the potential role of humour (such as sarcasm or irony) in sentiment analysis. In fact, studying public mass media, Puschmann and Powell (2018) show how the framing and public expectations around analytical capabilities of sentiment analysis are often misinterpreted and overly optimistic, (p.9)

[...] public perceptions of sentiment analysis attribute qualities and capabilities to sentiment analysis that diverge from how this method has been developed and refined in practice. [...] This connects with a broader shift toward a perception of feeling as something to be computationally grasped.”

Sarcasm

The social context is crucial when interpreting sarcastic comments (Katz and Lee, 1993). Ducharme (1994) states that sarcasm promotes “group solidarity”, while Gibbs (2000) finds that sarcasm is used to “vent frustration”, a tool that Twitter is often used for. For instance, Kim et al. (2011) found that Twitter users bonded over expressing sympathy, worry and frustration. Consider the examples from trending British tweets about tax avoidance (see ‘Method’ section for dataset details) from early April 2016 in Table 1.

Most psycholinguists agree that a sarcastic utterance typically has a target. Davidov et al. (2010) found that because of Twitter’s context-poor and unstructured nature, it is not always possible to easily identify the targets sarcastic comments are aimed at. In the above

examples, the hashtags are an apparent indicator that leaves little doubt as to what/who the sarcastic tweet is aimed at. Text-based user generated content (UGC) is potentially quite limiting to sarcasm detection, at least compared to other modes of communication such as video or spoken language, where visual and acoustic cues significantly help distinguish such utterances, for instance speech features including speech rate, and other contextual cues including laughter (Cheang and Pell, 2009; Tepperman et al., 2006). With increasingly more multimedia-based social media content, there may arguably be some future potential in multi-modal ML approaches.

Riloff et al. (2013) were the first to attempt to leverage recognising contexts that contrast a positive sentiment with a negative activity or state to detect sarcasm in social media posts (e.g., “Great, how exciting that I have to see my dentist today...having my tooth taken out”). Their approach was based on the intuition around sarcasm that arises from the contrast between a positive sentiment referring to a negative situation, where the challenge is to recognise the stereotypically negative situations that are generally considered undesirable and unenjoyable. Of course, such situations are context sensitive and depend on a person’s demographics and social network. Riloff et al. (2013) did not pursue the latter contexts, but they did propose a rule-based classifier looking for positive verb phrase followed by a negative situation with a custom set of learned phrases achieving a 0.7 precision score, although recall was extremely low – 0.09 – with a very poor overall F measure of 0.15. The final proposed system was an ensemble of this rule-based approach and an SVM (support vector machine) model combined, which achieved a respectable F score of 0.51 (precision 0.62, recall 0.44), showing that such an approach could have some success.

Rajadesingan et al. (2015) exploits behavioural traits from a user’s past tweets, in addition to lexical and linguistic cues. In a similar approach, Khattri et al. (2015) also apply ideas from Riloff et al. (2013) and propose a classifier based on contrast-based identification and a historical tweet-based model that identifies if the sentiment expressed towards an entity in the target tweet agrees with sentiment expressed by the author towards that entity in the past. Comparing their approach against the same datasets that Riloff et al. (ibid.) used, Khattri et al. (2015) achieved an F score of 0.83 (precision 0.84, recall 0.81), which is a substantial improvement. Their work primarily shows that making use of text other than just the one believed to contain sarcasm, by leveraging historical text-based features within a supervised sarcasm detection framework, is a promising approach. Bamman and Smith (2015) also explore contextual features and point out

Table 1. Example tweets on tax avoidance (i.e. from early April 2016).

*I've got an ISA ... I'd better consider my resignation #Not
#TaxAvoidance #resigncameron*

*It's blatantly obvious that #offshoreaccounts are being reframed by an
ever-helpful media not as #taxavoidance but for business efficiency.*

*@David_Cameron and @George_Osborne are clearly perfect to
resolve #taxavoidance. Set a thief to catch a thief.*

*Everyone leave David Cameron alone. He's an honest man #pana-
mapapers #taxavoidance #tax #TaxHavens #CameronOut
#CameronResign*

*@George_Osborne Lol :D You should do stand up, you're a natural!
#taxavoidance*

that they may aid in accurately identifying sarcasm, providing a useful overview of some complex features, some of which were found to perform well in model training. However, it is questionable how efficient their approach would be in terms of running-time and memory footprint, as it likely would not perform well under real-time running requirements, or large volumes of data, hence making these features altogether impractical for larger³ social media datasets. The authors (Bamman and Smith, 2015) point out that users with historically negative sentiments tend to have higher likelihoods of sarcasm use, and some features in their approach were also based around the intuition that sarcasm is more likely to take place between people who are more familiar with each other, hence taking into account past @mention messages exchanged between users. Interestingly, in subsequent analysis, the authors (ibid.) found that the strongest audience feature predictors of sarcasm were actually the absence of mutual mentions. It is worth noting that they (ibid.) used sarcasm hashtags to generate their seed dataset, which is a method that, as we show in ‘Prevalence of sarcasm and irony: A manual semantic analysis’ section in this paper, is rather problematic.

Parde and Nielsen (2018) have shown that using sarcastic tweets (the #sarcasm hashtag) and enriching training datasets with additional annotated Amazon product review data can slightly improve the performance of automated sarcasm detection, where the F score improvements reported are 0.58 to 0.59 for tweets and 0.74 to 0.78 for Amazon reviews. Amazon-based sarcasm detection is more accurate than detection on Twitter, primarily due to longer messages.

Irony

Although irony and sarcasm are distinct from each other (see section 2), what we observed from the computational research literature is that sarcasm and irony are occasionally bundled together and used interchangeably. The reasons for this may be varied, but include some conceptual misunderstandings, similar characteristics (e.g., frequent use of hyperbole), and hence similar automated detection approaches being applicable. Also, more pragmatic considerations may play a role, such as the challenge of human annotators consistently differentiating between irony and sarcasm, hence lack of example cases of irony and availability of labelled datasets for the usually “data hungry” computational machine learning methods. In this section, we provide a brief overview of some prior work that specifically dealt with irony detection.

Reyes et al. (2009) were among the first to discuss the importance and principles behind detecting irony in terse-text⁴ in order to correctly assign fine-grained

polarity levels. Sarmento et al. (2009) developed a model, also based on pattern detection to locate positive and negative sentiment in Portuguese newspapers – finding that the only problems arising in their system were due to the automatic misinterpretation of irony. Carvalho et al. (2009) created an automatic system for detecting irony relying on emoticons and special punctuation in online newspaper comments. Reyes and Rosso (2011) developed a model that represents irony in Amazon customer reviews by integrating “different linguistic layers” (i.e. from simple n-grams to affective content). They (ibid.) point out the importance of correctly identifying irony because of its use in expressing opinions but acknowledge the difficulty in finding a solution to detect it, admitting that it is unrealistic to rely on a single technique or algorithm. Hence, Reyes et al. (2013) propose a multidimensional model of “textual elements”, identifying four dimensions that are found in ironic tweets; those are:

1. Signatures: focuses on typographical elements, capitalisations or punctuation marks that imply opposition within a text (such as having the word “love” and the word “hate” within the same sentence).
2. Unexpectedness: examines patterns such as “temporal imbalance” which mixes different tenses and contextual imbalance that locates inconsistencies in a text.
3. Style: monitors recurring textual sequences.
4. Emotional scenarios: the use of emoticons to display moods or emotions.

They (ibid.) found that hashtags were one way tweeters use in order to avoid being taken literally. They annotated tweets with the hashtag #irony and compared them to tweets that were not ironic. The Monge Elkan edit distance score was used to allow gaps of unmatched characters to minimise noise from misspellings, and it is worth pointing out that such measures of text similarity⁵ are generally useful across sentiment analysis tasks. Barbieri and Saggion (2014) attempt to detect irony on Twitter using the same datasets that Reyes et al. (2013) prepared. They also measure word use frequency and apply supervised ML methods; their model included word frequency, style, adjective/adverb intensity and sentence structure. They also looked at sentiments, synonyms and ambiguities in order to “distinguish irony from as many different topics as possible” (ibid.: 66). The researchers only tested their model in controlled experiments using basic linguistic tools such as WordNet,⁶ and acknowledge the primary problem of disambiguation of meaning is an open problem. Barbieri et al. (2014a) adopted the same technique they had tried in their previous work, this time adding parts of speech (POS) tagging⁷

to increase accuracy, and most interestingly, they tackled the issue of ambiguity, concluding that the more meanings a word has, the more likely it is to be used ironically. Barbieri et al. (2014b) adapted their previous technique to Italian tweets using a Decision Tree classifier from machine learning to detect ironic tweets. They reported an F measure improvement of 0.11 over simple baselines. Barbieri et al. (2015) adapted their model to detect satire in Spanish tweets with comparable results. Buschmeier et al. (2014) used Davidov et al. (2010)'s method of monitoring the imbalance among the polarity of terms, with the addition of taking emoticons and interjections into account and applied their method on Amazon reviews after having five human annotators decide between ironic and non-ironic reviews. Their system achieved an F measure of 0.68. Hao and Veale (2010) proposed an algorithm for separating ironic from non-ironic similes, detecting common terms used in this ironic comparison. They found that web users often use figurative comparisons as a means to express ironic opinions. A set of example ironic tweets for the hashtag #GrandNational, which is an annual British horse racing event that has received some public criticism around animal welfare (see 'Method' section for dataset details), are shown in Table 2. In the next section, we present our focused manual semantic annotation analysis study of sarcasm and irony across two datasets.

Prevalence of sarcasm and irony: A manual semantic analysis

Using sarcasm and irony hashtags as “gold”⁸ labels for training sarcasm detection systems is a widely used method (e.g., Bamman and Smith, 2015; Khattri et al., 2015; Parde and Nielsen, 2018; Rajadesingan et al., 2015; Sarsam et al., 2020). Although distant supervision-based learning was found to work well for basic sentiment classification (Hogenboom et al., 2013; Pak and Paroubek, 2010), doubts about such approaches being used for sarcasm and irony detection

Table 2. Example tweets on trending hashtag #GrandNational (i.e. from early April 2016).

Have Channel 4 no sense of irony: if the horse in their #GrandNational advert fell in the race like it did in the ad, it would be shot.
 @BBCNews: #GrandNational-winning jockey Ryan Mania “may have been hit by another horse while he was on the ground” the irony gets better!
 There's an irony somewhere that Michael Owen is so in love with a sport where you are shot for being injured #grandnational
 The irony of a country outraged by the #horsemeat scandal but willing to watch them die for an hour of entertainment #grandnational
 Two horses die at #grandnational but trainer defends the race by telling people 'to get a life' O the irony. . .

have been raised some time ago by Davidov et al. (2010). Hashtag-based datasets especially for sarcasm and irony may result in a high proportion of noisy data. Besides, the issue has not received any attention in related literature, and our understanding of the prevalence of sarcasm and irony within various contexts on social media is lacking. To address this, we present the outcomes of a manual semantic analysis of 4334 Twitter messages associated with (i) hashtags commonly employed by the above-mentioned prior work and (ii) relating to a selection of 25 distinct events, from a dataset presented previously by Sykora et al. (2014).

Method

The manual semantic analysis was performed by a researcher with training in linguistics and discourse analysis. The task focused on the human expert deciding whether the intended meaning of a message was sarcastic or ironic, based on the surface text of the tweet, as well as identifying commonly used types of expressions within various contexts. This was a systematic process in line with closed coding content analysis approaches, as outlined in Thelwall (2013). A manual qualitative analysis of tweets allows for a deeper understanding of the shared material, particularly where the context is nuanced due to irony and general humour, as well as the use of out of vocabulary terms, slang, abbreviations, etc. The coding strictly focused on the surface level meaning of communicated message in the tweet. The reason for this is that we are interested in the use of humour and irony within the message content itself, as word/character-based models are primarily based on such features in automated detection approaches (e.g., Bamman and Smith, 2015; see 'Background, prior work and theory' section). To assure reliability, as is common practice in related literature (Mahoney et al., 2019), out of all tweets across the datasets i and ii described below, a uniform random sample of (480 tweets) was also coded by the second author, resulting in virtually perfect (99%) agreement with the linguistics/discourse expert.

The analysed data consisted of two separate datasets,⁹ containing Twitter messages associated with;

- (i) A set of predefined hashtags often employed in training ML models (e.g., Riloff et al., 2013, etc.), specifically, analysing a random sample of a maximum 300 messages for each hashtag on data collected via the Twitter Search API, over a one-week time period (i.e. March-April 2016), with 2514 tweets in total, namely:

- [1] - #sarcasm
- [2] - #sarcastic

- [3] - #not
- [4] - #notsarcasm
- [5] - #notsarcastic
- [6] - #irony
- [7] - #ironic
- [8] - #joke
- [9] - #humour
- [10] - #funny

- (ii) A random sample of 1820 tweets, sampled directly from a collection of 1,570,303 tweets retrieved via the Twitter Search API on 28 separate datasets relating to 25 unique events (65 random tweets from each of the 28 datasets were selected). An overview of this data and the dataset itself is discussed in more detail, including its emotional properties (derived using the EMOTIVE sentiment analysis system) within Sykora et al. (2014).

#Sarcasm, #irony and #joke hashtag use

The primary output from the manual analysis of the 10 hashtags (dataset i) from Twitter is presented in Table 3.

The annotation criteria for the dataset from Table 3 to be considered as ironic, sarcastic or humorous were strictly focused on the communicated message and the associated semantics, including any additional hashtags, other than the indicative hashtag used for the collection (e.g., #sarcasm, #ironic), which was purposefully ignored. #notsarcasm and #notsarcastic both have less than 300 tweets as the terms were not prevalent and not as many results were returned through the Twitter search API. From the table, it is evident that the proportion of actually sarcastic expressions in the sarcasm-related category was a low percentage of only 14.89% of the messages. To unpack the characteristics

of these tweets further, a brief qualitative discussion of each category of messages; sarcastic, ironic and humorous from Table 3, to illustrate their use (all taken from the dataset i), follows below.

Consider the example message from the dataset;

RT @__user__: Free #healthcare? This man Bernie Sanders is a communist, just like Canadians. #FeelTheBern #DemDebate __url__

The social media environmental context in this message's meaning is crucial, and associated with the hashtag #FeelTheBern, which given its supportive nature (FeelTheBern-org, n.d.) of Bernie Sanders (a US presidential candidate for 2016 US election), is the only hint that implies that this message was actually meant sarcastically, rather than literally. Similarly, context associated with @__user__ is potentially instrumental in being able to identify sarcasm in this message “@__user__ oohhh wow, u r such a genius...”. In the tweet “Tonight was wonderful. #suckedass” interpreting the polarity of the hashtag would be crucial. Unfortunately, virtually none of the current state-of-the-art sentiment analysis systems do this (i.e. analysing multi-word hashtags) and the issues around this have not been previously explored, hence we cover this in detail within the discussion, in sub-section ‘Multi-word hashtags in sarcasm and irony use’. On the other hand, the #not hashtag is likely an even noisier approach to collecting training datasets for ML approaches, as especially Twitter users frequently employ hashtags to form sentences using hashtags themselves, for instance;

#Trying to be a #GoodWomen in a #World where #BadB's get all the #credit ... #Not Made for this #World!! #Jesus

Table 3. Overview of hashtags-based tweets often employed in training ML models.

Dataset	Total # of tweets	# of unambiguously sarcastic/ironic /humorous tweets	Aggregate/total by category
#sarcasm	300	48 (16.00%)	134 (14.89%)
#sarcastic	300	49 (16.33%)	
#not	300	37 (12.33%)	34 (29.82%)
#notsarcasm	78	22 (28.20%)	
#notsarcastic	36	12 (33.33%)	
#irony	300	168 (56.00%) RT noise removed: 68 ironic tweets out of 200, 34%	140 (28.00%)
#ironic	300	72 (24.00%)	4 (00.44%)
#joke	300	1 (00.33%)	
#humour	300	1 (00.33%)	
#funny	300	2 (00.66%)	
Total	2514	412 (16.38%) RT noise removed: 312 (12.41%)	.

However, as illustrated with example messages in Table 4, this hashtag (#not) can sometimes work relatively well as a sarcasm marker.

Irony was considerably more prevalent with its associated hashtag, than sarcasm (i.e. 28% vs. 14.89% in Table 3). Several examples of typical ironic expressions are illustrated with tweets presented in Table 5. However, the high proportion of actual ironic tweets is partly due to a trending topic in the dataset with a high proportion of retweets, example tweet below;

RT @__user__: Flight 787 of Royal Brunei lands in Jeddah, piloted by an elite crew who are not allowed to drive there. #Irony __url__

Table 3 also shows statistics for #irony-related tweets once messages regarding the Flight 787 were ignored, which still contained a high prevalence of ironic tweets (34% for #irony). However, in the dataset of tweets based on the #ironic hashtag, this Flight 787 message did not take hold at all, and the sample contained a fairly well-balanced random collection of topics, which still showed a high proportion of 24% actual ironic tweets, compared to an average of 14.89% actual

sarcastic tweets for #sarcasm, #sarcastic and #not (i.e. right column in Table 3).

For comparison, we conducted the same study on humorous tweets collected via the #joke, #humour and #funny hashtags, it is apparent that a significant high proportion of tweets include links, without the surface text of the message itself being funny. Overall only 4 tweets were genuine attempts at jokes (i.e. Table 6) out of all 900 of tweets (i.e. the research associate was liberal in annotating this category as content that only remotely resembled an attempt at a humorous expression/joke would be counted as such).

The interesting characteristic of this particular category of the dataset (i.e. tweets collected from #joke, #humour and #funny) is that out of 900 tweets in total 720 contain url links (80%). This highlights the observation that in particular with humorous Twitter content, analysing the surface text of tweets is going to miss substantial context. Techniques that analyse embedded media, including url links with potentially media and rich context, are hence needed to deal with such expressions. In contrast, out of the 900 tweets associated with #sarcasm, #sarcastic and #not, 374 tweets (41.56%) contained url links. For the 600 tweets on #irony and #ironic, this was 297 (49.50%),

Table 4. Example tweets collated using the #not hashtag as a distant learning sarcasm indicator.

Car alarms going off at midnight . . . how I love to live in Br____ton!!!	True positives
Love having a cracked kneecap	(manual analysis
No better way to end the day than with a speeding ticket!	agrees with hashtag)
People that lie are just my favourite!!	
Totally love it when twitter tells me I have a notification when I don't	
We better not have Saturday class because I am #not going	False positives
I care about your feelings more than mine.	(manual analysis
Isn't this just lovely weather #rain #no #school	disagrees with hashtag)
Well, so funny __url__	
Very classy. __url__	

Table 5. Example tweets collated using the #ironic and #irony hashtags as indicators of irony.

I am panicking about not feeling any panic about Midterms ! #ThisMonday	True positives
RT @__user__: Health and Safety event CANCELLED due to #healthandsafety	(manual analysis
concerns __url__ __url__	agrees with hashtag)
Doctors make d worst patients	
Definition Of #Irony. Pro-gun mum shot by 4yo son __url__ via @ABCNews	
#GunViolence #YouCantChangeStupid #SecondAmendment	
#WhyIDidntGetIntoHeaven Cause you didn't listen to the crazy people in the	
world who are actually the only sane ones.	
@__user__ That was a scheduled tweet actually.	False positives
There is no God!	(manual analysis
It seems appropriate that Jurassic Park is on tonight given my earlier tweet	disagrees with hashtag)
#MotivationMonday #dinosaur #bestmovieever	
Bernie is sexist for asking Hillary not to interrupt. #ThingsHillarySupportersThink	
Is that what you thought when you did those things to me?	

Table 6. Example tweets collated using the #joke, #humour and #funny hashtags as indicators of humour.

When cooking Alphabet Soup, don't leave it unattended, it could spell > disaster. LOL	True positives (manual analysis agrees with hashtag)
Homework, forever bringing families together. #learning #school #education	
#teachers #students __url__	
@__user__ my daughter looks older then me :-), #screenshotsaturday #folloback	
__url__	False positives (manual analysis disagrees with hashtag)
@__user__ in this coffee house, carrot cake count as a vegetable. #quote #cap-	
puccino #carrotcake #boccamoka __url__	
We ended another busy week in waves of laughter #classassembly __url__	
Why would we expect anything different? __url__	
One more day till #Friday! #tdursday #countdown #lol	
Out in space two alien life forms are speaking with each other __url__ #lol #haha	

and out of the 114 tweets for #notsarcasm and #not-sarcastic only 16 (14.04%) contained a url link to an external resource, while on average these tweets were found to contain sarcastic content more often than the other hashtags (i.e. Table 3).

Event-related tweets

We conducted the analysis of dataset (ii) which consisted of 1820 Twitter messages on a broad range of heterogeneous, randomly chosen topics; some trending in the United Kingdom and others trending internationally. The aim was to provide an indicative overview of the prevalence of sarcastic, ironic and general humorous messages within fairly “standard” social media use, across wide range and types of topics and events in the English language. In previous work, Riloff et al. (2013) reported that out of 1600 random tweets obtained from the streaming Twitter API that did not contain a sarcasm hashtag, only 29 (1.8%) were judged to be sarcastic by their human expert annotator. In our random event-based approach, we found an even lower proportion of such expressions. Out of 1820 tweets, only 20 (1.09%) were either sarcastic or ironic, with sarcasm expressions in 14 and irony in only 6 distinct messages. There were only 28 tweets containing other types of humorous content.

Given the significantly smaller proportion of such language-use, this is in stark contrast to dataset (i) with messages where a sarcasm or irony hashtag was included (i.e. Table 3), pointing to some relative usefulness of the #hashtag heuristic with this type of content being more likely. Table 7 presents an overview of all the 28 different datasets with the count of distinct tweets with sarcastic, ironic and humorous language highlighted in the three right-most columns. The emotional variables associated with tweets from this events dataset (i.e. Sykora et al., 2014) allow us to tentatively explore potential associations with humour, sarcasm and irony. Spearman's rho correlation over all the 28 datasets highlights a potential yet weak relationship

between (absolute) emotionality of tweets and the likelihood of humorous content ($\rho = 0.131$). This would be consistent with Rajadesingan et al. (2015) who found that there was a higher probability of users using sarcasm when they were in an emotionally charged “mood”. Sarcasm was potentially positively correlated with irony ($\rho = 0.278$) and humour ($\rho = 0.284$), while humour was, as expected, negatively correlated to irony ($\rho = -0.234$). However, none of these correlations were actually significant at p (two-tailed) $< .05$; keeping in mind recent concerns around drawing premature conclusions from non-statistically significant p -values (Amrhein et al., 2019), these results are at most indicative of potential associations and need to be interpreted cautiously.

Several important observations can be drawn from results in Table 7. For instance that #horsemeat and #tubestrike both contained a high proportion of humorous language, 10 (15.38%) and 6 (9.23%), respectively. These are by far more prevalent than even tweets tagged with a hashtag implying humorous language from dataset (i), which indicates that certain events evoke humorous content much more frequently. As was highlighted in ‘Prevalence of sarcasm and irony: A manual semantic analysis’ section, a widely accepted theory of humour proposes that it is a tool used for coping, and both these events were particularly significant in that they (potentially) affected a large population in a (likely) negative way. Both these events also had the highest proportion of sarcastic tweets in their respective samples of tweets. The results also indicate that events centred around deaths of well-known individuals and the tragic murder of Daniel Pelka contained no sarcasm nor irony, except for one ironic message associated with #JamesGandolfini, “Why is it that nice things are never spoken of when the person is alive, but only after their death? #JamesGandolfini”. Hence, it is possible that depending on the nature of events, there may be scope for identifying likely types and styles of irony, sarcasm and/or humour. However, further investigation over

Table 7. Overview of the analysed datasets, with frequency information on sarcasm, irony and humour over a random sample of 65 tweets for each event, and the related event description.

Dataset	Total (N)	Emotional tweets (%)	Event	Event type	Sarcasm	Irony	Humour
helicopter crash	25,387	13.99	Helicopter crashes into crane in central London (16th Jan)	Accident	2	2	0
#september11	88,739	9.62	September 11th 2013 anniversary	Anniversary	0	0	0
#twintowers	28,168	16.32	September 11th 2013 anniversary	Anniversary	0	0	1
#ChineseNewYear	22,466	36.13	Chinese New Year, 31st Jan 2014	Cultural event	0	0	1
#bankholiday	7862	11.71	Bankholiday - public holiday in the UK	Daily life	0	0	1
#sleep	36,139	3.65	An eight day long period	Daily life	0	0	1
#tired	79,253	4.49	An eight day long period	Daily life	0	0	1
#JamesGandolfini	11,975	18.92	Death of actor James Gandolfini	Death	0	1	0
Ariel Sharon	90,603	8.18	Death of the ex-prime minister of Israel	Death	0	0	0
Nelson Mandela	108,794	12.51	Death of Nelson Mandela	Death	0	0	1
"Daniel Pelka"	11,708	21.54	Sentencing of the killers in the brutal murder of school boy Daniel Pelka	Death/murder	0	0	0
#RoyalMail	4309	6.75	Privatisation of the British Royal Mail, 12th Sep announcement	Economic/controversial	0	0	0
#tubestrike	41,176	8.47	London February tube strike by RMT and TSSA unions	Economic/controversial	3	0	6
#LFW	43,509	4.27	London Fashion Week	Fashion event	0	0	0
Anjem Choudary	1047	5.44	Controversial comments from a radical cleric on BBC	Hate speech incident	0	1	0
#2DayFM	10,898	20.43	Royal prank by Australian 2DayFM – suicide of Nurse Jacintha Saldanha	Incident/death	1	0	1
#jacinthasaldanha	1216	37.91	Royal prank by Australian 2DayFM – suicide of Nurse Jacintha Saldanha	Incident/death	1	0	1
#royalprank	10,459	23.17	Royal prank by Australian 2DayFM – suicide of Nurse Jacintha Saldanha	Incident/death	0	0	0
g8 summit	32,676	4.24	39th G8 Summit in UK on 17th–18th June	Political/controversial	0	0	0
#iPhone5C	8824	3.90	Announcement of new iPhone on 10th Sep	Product release	1	0	0
#iPhone5S	14,638	5.70	Announcement of new iPhone on 10th Sep	Product release	0	0	0
gta5	130,748	4.22	Release of computer game GTA 5 on 17th Sep	Product release	1	0	0
#NSA	381,402	5.08	National Security Agency PRISM surveillance program (initially leaked early Jun)	Scandal	0	0	0
#prism	106,432	4.96	National Security Agency PRISM surveillance program (initially leaked early Jun)	Scandal	1	1	1
Horsemeat	56,970	7.47	Horsemeat missold as beef (issue came to light on 15th Jan)	Scandal	3	0	10
#ClosingCeremony	87,943	11.55	London 2012 Olympics – Closing ceremony	Sport event	0	0	3

(continued)

Table 7. Continued

Dataset	Total (N)	Emotional tweets (%)	Event	Event type	Sarcasm	Irony	Humour
#paralympics	27,993	13.97	London 2012 Olympics – Paralympic games (29th Aug – 9th Sep)	Sport event	0	0	0
#woolwich	98,969	12.63	Attack and murder of Drummer Lee Rigby in Woolwich, by extremists	Terror incident/ murder	1	1	0

larger corpora of events and topics on social media would indeed be needed to validate such a hypothesis.

Discussion

Based on the outcomes of our presented analysis we need to highlight two primary observations, some related issues, as well as limitations associated with our study's method, and suggestions for future work. An intriguing qualitative observation around the role of multi-word hashtags, which has not been covered in prior research, is also discussed in detail within section 'Multi-word hashtags in sarcasm and irony use'.

First from analysis of the hashtag marker-based dataset (i) we can conclude that the proportion of genuinely humorous, sarcastic and ironic surface-level content in such tweet messages is generally very low. Although semi-supervised learning has been applied with some success on large datasets across range of problems where labelled data were not readily available (Zhu and Goldberg, 2009); given our findings across commonly employed hashtag markers, this observation has implications for such automatic approaches. The assumption that linguistic features that are indicative of humorous, sarcastic and ironic messages can be leveraged when in reality this is hardly the case, is hence worrying. Social media messages, especially from strictly character restricted platforms such as Twitter, are of primary concern, given their often short sparse nature and lack of context. This is to some extent empirically illustrated in the work by Parde and Nielsen (2018), where sarcasm detection over Amazon reviews, compared to Twitter, is substantially more accurate due to the longer length of content and hence richer context available to the sarcasm detection model. Another implication stems from our observation of a very high proportion of url links (80%) seen among humorous tweets, as well as sarcastic (41.56%) and ironic tweets (49.50%). However, our study was strictly limited to the surface-level semantics of a tweet and we did not explore the linked url content. Therefore, little can be said with certainty regarding the url linked content, but it is indeed possible that at times all or part of the

humorous, sarcastic or ironic message was embedded within such linked sources. Especially given the low frequency of surface-level humour in tweets and their relative high proportion of url links is intriguing. Despite this, the majority of current approaches in sentiment analysis and automated sarcasm/irony/humour detection still do not support non-textual modalities of data analysis, nor their linked textual content (Ravi and Ravi, 2015; Sarsam et al., 2020; Yadollahi et al., 2017). Studying such content in tandem, and techniques that analyse embedded media or potentially rich multi-modal and textual content from url links, referred to in tweets, would seem to be a worthwhile area of further future work.

The second set of observations based on our analysis of dataset (ii) confirms an exceedingly low proportion of sarcastic content, just over 1%, as previously reported by Riloff et al. (2013) across typical, large collections of topics on Twitter data. Although Riloff et al.'s (2013) dataset consisted of randomly collected tweets, and our own consisted of a random selection of mostly UK-based events, we confirm this result, and we additionally find that the prevalence of sarcasm, as well as irony and humour, can vary significantly across different types of events and topics. We report on events ranging from scandals, product releases, cultural events, to accidents and terror incidents. While most of the 28 datasets reported here, do not show any more than 1.5–2% of such language use, in line with Riloff et al. (ibid.); for some events, the prevalence of humour and sarcasm was found to be much higher, in range of 6–20%, for the London helicopter crash, tube strike and the Horsemeat scandal. What this likely means in practice is that most of the time, across large cross-topic datasets, sarcasm, irony and the use of humour may be much less of a concern as is generally thought (Puschmann and Powell, 2018); yet, the performance of sentiment analysis and related tools will strongly depend on the specific topic or event covered and discussed on Twitter. This is similar to findings from work by Ribeiro et al. (2016), who reported how performance of twenty-four state-of-the-art sentiment analysis tools is highly context sensitive, and

varies significantly across different datasets, even when these are from the same social media platforms. A natural limitation of our study is the small size of the analysed dataset, while still larger than prior work; in future work, we plan to substantially extend the variety and number of events/datasets. The extent of the annotation effort could also be increased. This will allow to undertake a more substantial analysis across the types of events and topics to evaluate their nuanced characteristics that are likely to evoke more humour, sarcasm or irony on Twitter. Such bottom-up empirical research is important across social media platforms in order to further our understanding around the nature of discourse and properties of events and themes that are likely to elicit conversations with higher levels of humour, sarcasm or irony, and its implications for sentiment analysis being taken into account.

There are also cross-cultural considerations. For instance, Joshi et al. (2016b) present early empirical work illustrating some differences between what various cultures consider to be sarcastic on social media. Sarcasm and irony detection in different cultural contexts and especially non-English languages have been quite rare, with only a few exceptions, such as work in Dutch by Liebrecht et al. (2013), Italian by Barbieri et al. (2014b) and Spanish, again by Barbieri et al. (2015). Due to the sensitive nuanced nature of language and culture specific differences in figurative language use, multi-lingual issues pose an interesting future research area, also echoed in Abulaish et al. (2020), especially when it comes to the use of humour, sarcasm and irony; and how these empirically differ across populations.

Multi-word hashtags in sarcasm and irony use

Multi-word hashtags, such as #suckedass (see section ‘#Sarcasm, #irony and #joke hashtag use’) pose a problem for sarcasm and irony detection. Zappavigna (2015), who has conducted one of the most detailed linguistic-based analyses of hashtag use on Twitter to date, suggests that the issue is that hashtags are not only used to help organise social streams by their topic, but have a more experiential and social function, such as “#soawesome”, “#Ihatemornings” “#ff” (i.e. follow Friday – used to mention accounts worth following). This poses issues in practice as the semantic context is not straightforward to interpret for most sentiment analysis and automated analytics tools. Generally, hashtags can be considered to be a form of metadata (i.e. data about data) embedded within the social media message and have meta-semantic value, effectively a sort of meta-commentary. They may also have a limited lifetime, such as for instance #ObamaCare, which relates to a health-care system

reform by the ex-US president and remains short-lived only while public debate around this political issue remains. Interestingly new hashtags can sporadically evolve and emerge (Maity et al., 2015), such as #ObamacareWorks. Zappavigna (2015: 277) points out that

[.] #ObamacareWorks while opportunistic is an example of meta-evaluation used to emphasize a particular political perspective and perhaps also to rhetorically imply that there is an ambient audience of microbloggers who agree with the point being made (and who might potentially search for or use this otherwise idiosyncratic tag).

Hence, where such semantically rich hashtags are used in a sarcastic, or ironic context, the primary challenge is to accurately interpret the meaning of such content. According to Zappavigna (ibid.), these kinds of meta-evaluations are quite common. Pointing out (ibid.: 285) the meta-commentary may often provide additional information that is not contained in the message itself, for instance;

Hard to believe summer is almost done. And school is right around the corner. #sad #toofast

In the example above, the negative affect is only expressed in the hashtags. Zappavigna (ibid.) further highlights uses of hashtags in communicating judgements, such as in the example below (coincidentally also sarcastic);

Yah no worries, just dump your load of mulch in the middle of the road #Idiot <http://t.co/CKPD9HMI1A1>

In these examples, most sentiment analysis approaches would have failed in detecting the sentiment as hashtags are ignored from analysis in the vast majority of approaches and algorithms (e.g., Hutto and Gilbert, 2014). Often, if these are not ignored, then they are only used in semantically poor ways, such as placeholder tokens in language models, for instance as in Montejo-Ráez et al. (2014) where any words prefixed by “#” were effectively normalised with the token “_HASHTAG_” in their model, ignoring any explicitly stated semantics within such tokens. This model and skip gram deep learning-related models may still learn to classify certain #hashtag polarities for frequent n-grams/expressions accurately but would tend to fail with less common multi-word hashtags. To put this to the test, we selected a sample of 20 exemplary hashtags containing affective and judgmental content from Zappavigna (ibid.: 285–286, incl. Table 8) and we ran 20 state-of-the-art sentiment classifiers on the messages

using the i-Feel 2.0 sentiment benchmarking platform (Araújo et al., 2016; iFeel, n.d.). On average 3.6 classifiers were correct out of 20 classifiers over 20 different hashtags. The best performance was found for “#happy” (10 out of 20 classifiers correctly detect positive sentiment), and “#sad” (9 out of 20 classifiers). However as soon as the hashtag contained multiple words, for example “#sohappy” and “#veryhappy”, only one classifier identified the correct sentiment polarity. Less common expression such as with “#totalystoked” or “#BenefitsOfALongDistanceRelationship,” not a single classifier correctly identified the sentiment polarity. This small scale evaluation, and the linguistic interpretation of hashtags as outlined in (Zappavigna, 2015) are indicative of the importance of processing the natural language within hashtags.

Future work

We believe that future research of automated tools may benefit from carefully considering such concerns, and a promising future direction may well be the development of semantic models in the form of knowledge bases. Hee et al. (2018) have specifically shown how introducing knowledge bases of common-sense, also known as general world-knowledge semantics, can significantly improve performance in detecting irony (in their case achieving a respectable F-measure of just over 0.7). Indeed, an error analysis of misclassifications for automated sarcasm detection by Parde and Nielsen (2018) highlights three particular issues and areas for future work; (i) introducing models of world or common-sense knowledge to help detect sarcasm/irony, (ii) better approaches to normalising text, such as splitting up compound hashtags into individual words, to ultimately improving performance of sentiment detection approaches by appropriately reversing the polarity of detected sentiment, and (iii) issues around evolving language and slang on social media pose a continued challenge. These three future research challenge areas directly relate to also addressing the multi-word hashtag issue we have pointed out above. The fourth (iv) challenge area will be how to effectively resolve the temporal sense associated with complex hashtags and expression over certain specific time periods and individuals/community of users, where the meaning behind certain language use may be very specific to certain social network of users, or individuals.

Conclusion

This paper reviews core approaches to automated detection, the issues and actual prevalence of humorous, sarcastic and ironic language use within social media posts, in order to further understand and help

researchers interpret and potentially mitigate the effects on sentiment analysis techniques. The main contribution of this paper is observations from a substantial semantic analysis of over 4000 social media posts. Within our analysis, we also explore 25 events and topics, from a random sample of 1820 tweets, taken from a large 1.5 million tweet dataset collection, across a range of events including, scandals, product releases, cultural events, accidents and terror incidents. Our findings highlight the lack of genuine use of sarcasm, humour and irony with certain hashtag markers. For instance, hashtags like “#sarcasm” are not necessarily good indicators of sarcasm. Machine learning models for automated sarcasm and irony detection, as well as related performance of sentiment analysis tools are likely to be impacted by this observation, as these are often trained on training data of hence highly questionable quality and little qualitative appreciation of the underlying data. We further highlight the contextualised use of multi-word hashtags in the communication of general humour, while focusing specifically on sarcasm and irony. After reviewing the state of the art, we found that most sentiment analysis systems simply fail to recognise such hashtag-based expressions. A study that qualitatively investigates surface meaning of humour, sarcasm and irony use over a sizeable set of tweets was until now missing. We believe our findings are of importance to the research community for academics that use and apply, as well as develop, such tools.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Martin Sykora  <https://orcid.org/0000-0002-5363-5857>

Notes

1. Machine learning models can be broadly categorised into (1) supervised and (2) unsupervised models, where labels/dependent variables are respectively (1) known/available or (2) unknown, while in (3) semi-supervised learning, weak labels are employed (Manning and Schütze, 1999), where usually some heuristics around some property in the dataset are employed to act as labels/dependent variable (e.g., happy/sad emoticons being used as positive/negative sentiment labels).

2. When developing ML-based models, datasets are usually split into training, validation (for model parameter tuning) and testing datasets, which can be used under various schemes, such as k-fold cross validation or leave-one-out validation.
3. The content produced every day by social media platform users is vast and carries the typical characteristics associated with big-data, such as high volume, velocity and variety in data (Kitchin, 2014; Laney, 2001).
4. Terse-text refers to brief textual messages as opposed to more traditional longer texts with richer context. Terse-text messages are often encountered on character restricted social media platforms, e.g. Twitter (140 and now 280 characters in most regions). Due to the lack of context/word-use, this presents challenges to traditional NLP/IR language approaches.
5. An overview of, and easy to use Python library of over 30 text similarity measures; <https://github.com/orsinium/textdistance>.
6. Wordnet is widely used among the computational NLP and IR research communities, and is a rich taxonomy/lexical database of English words and expression; available at <https://wordnet.princeton.edu/>.
7. POS information about words within a sentence, that is whether a word is an adjective, noun, pronoun, etc., are frequently used as features in machine learning models, as they often improve model performance.
8. Golden data, a.k.a. “ground truth data” (both terms are used commonly in NLP/IR) refer to the desired labels assigned to a data-point (i.e. piece of text) that a machine learning model is meant to learn from in the training and validation phases.
9. The tweet IDs are available at <https://doi.org/10.17028/rd.lboro.c.5173559> as well as by reasonable request to the authors.

References

- Abulaish M, Kamal A and Zaki MJ (2020) A survey of figurative language and its computational detection in online social networks. *ACM Transactions on the Web* 14(1): 1–52
- Amrhein V, Greenland S and McShane B (2019) Scientists rise up against statistical significance. *Nature* 567: 305–307
- Araújo M, Diniz JP, Bastos L, et al. (2016) *iFeel 2.0: A multilingual benchmarking system for sentence-level sentiment analysis*. In: *9th international conference on social media and weblogs ICWSM*, Cologne, Germany, 18–21 August.
- Bamman D and Smith NA (2015) Contextualized sarcasm detection on Twitter. In: *Proceedings of the 9th international conference on web and social media*, Oxford, UK, 26–29 May.
- Barbieri F, Ronzano F and Saggion H (2014b) Italian irony detection in Twitter: A first approach. In: *Proceedings of the 1st Italian conference on computational linguistics CLiC-IT & the 4th international workshop EVALITA*, Pisa, Italy, 9 December.
- Barbieri F, Ronzano H and Saggion F (2015) Is this tweet satirical? A computational approach for satire detection in Spanish. *Procesamiento del Lenguaje Natural* 55(1): 135–142.
- Barbieri F and Saggion H (2014) Modelling irony in Twitter. In: *Proceedings of the 14th conference of the European chapter of the association for computational linguistics*, Gothenburg, Sweden, 26–30 April.
- Barbieri F, Saggion H and Ronzano F (2014a) Modelling sarcasm in Twitter: A novel approach. In: *Proceedings of the 52nd annual meeting of the association for computational linguistics*, Baltimore, USA, 27 June.
- Buschmeier K, Cimiano P and Klinger R (2014) An impact analysis of features in a classification approach to irony detection in product reviews. In: *Proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis*, Baltimore, USA, 27 June.
- Carvalho P, Sarmiento L, Silva M, et al. (2009) Clues for detecting irony in user-generated contents: Oh. !! It's “so easy”. In: *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion*, Hong Kong, China, 6 November.
- Cheang HS and Pell MD (2009) Acoustic markers of sarcasm in Cantonese and English. *Journal of the Acoustical Society of America* 126(3): 1394–1405
- Cheong M and Lee VCS (2011) A microblogging-based approach to terrorism informatics. *Journal of Information Systems Frontiers* 13(1): 45–59.
- Davidov D, Tsur O and Rappoport A (2010) Semi-supervised recognition of sarcastic sentences in Twitter and Amazon. In: *Proceedings of the 14th conference on computational natural language learning CoNLL*, Uppsala, Sweden, 15–16 July.
- Ducharme LJ (1994) Sarcasm and interactional politics. *Symbolic Interaction* 17(1): 51–62
- Fan W and Gordon MD (2014) The power of social media analytics. *Communications of the ACM* 57(6): 74–81.
- FeelTheBern-org (n.d.) Endorse Bernie Sanders for President. Available at: <https://feelthebern.org/> (accessed 7 April 2017).
- Fowler H (1965) *A Dictionary of Modern English Usage*. 2nd ed. Oxford, UK: Oxford University Press.
- Ghosh A, Li G, Veale T, et al. (2015) Semeval-2015 task 11: Sentiment analysis of figurative language in Twitter. In: *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, Denver, Colorado, 4–5 June.
- Gibbs RW (2000) Irony in talk among friends. *Metaphor and Symbol* 15(1–2): 5–27
- Goonetilleke O, Sellis T, Zhang X, et al. (2014) Twitter analytics: A big data management perspective. *ACM SIGKDD Explorations Newsletter* 16(1): 11–20.
- Gruebner O, Sykora M, Lowe S, et al. (2016). Mental health surveillance after the terrorist attacks in Paris. *The Lancet* 387(10034): 2195–2196.
- Hao Y and Veale T (2010) An ironic fist in a velvet glove: Creative mis-representation in the construction of ironic similes. *Minds and Machines* 20(4): 635–650
- Hee v-C, Lefever E and Hoste V (2018) We usually don't like going to the dentist. Using common sense to detect irony on Twitter. *Computational Linguistics* 44(4): 1–63.

- Hogenboom A, Bal D, Frasinicar F, et al. (2013) Exploiting emoticons in sentiment analysis. In: *Proceedings of the 28th annual ACM symposium on applied computing*, Coimbra, Portugal, 18 March.
- Hutto CJ and Gilbert E (2014) VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: *Proceedings of the international conference on web and social media – ICWSM-2014*, Ann Arbor, USA, 2–4 June.
- iFeel (n.d.) Welcome to iFeel 2. Available at: <http://blackbird.dcc.ufmg.br:1210/> (accessed 5 April 2017).
- Ilieva RT and McPhearson T (2018) Social-media data for urban sustainability. *Nature Sustainability* 1(10): 553–565.
- Joshi A, Bhattacharyya P and Carman MJ (2016a) Automatic sarcasm detection: A survey. *ACM Computing Surveys* 50(5): 1–22.
- Joshi A, Bhattacharyya P, Carman M, et al. (2016b) How do cultural differences impact the quality of sarcasm annotation? A case study of Indian Annotators and American Text. In: *The 10th SIGHUM workshop on language technology for cultural heritage, social sciences, and humanities (LaTeCH)*, Berlin, Germany, 11 August.
- Joshi A, Sharma V and Bhattacharyya P (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*, Beijing, China, 26–31 July.
- Katz A and Lee C (1993) The role of authorial intent in determining verbal irony and metaphor. *Metaphor and Symbolic Activity* 8(1): 257–279.
- Khattari A, Joshi A, Bhattacharyya P, et al. (2015) Your sentiment precedes you: Using an author's historical tweets to predict sarcasm. In: *Proceedings of 6th workshop on computational approaches to subjectivity, sentiment & social media analysis*, Lisbon, Portugal, 17 September.
- Kim Y, Sohn D and Choi SM (2011) cultural difference in motivations for using social network sites: A comparative study of American and Korean college students. *Computers in Human Behavior* 27(1): 365–372.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. London: Sage.
- Kreuz RJ (1996) The use of verbal irony: Cues and constraints. In: Mio JS and Katz AN (eds) *Metaphor: Implications and Applications*. Mahwah, NJ, USA, pp.23–38.
- Kreuz RJ and Glucksberg S (1989) How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology: General* 118(4): 374–385.
- Kuehn BM (2015) Twitter streams fuel big data approaches to health forecasting. *JAMA* 314(1): 2010–2012.
- Kumar S, Barbier G, Abbasi MA, et al. (2011) TweetTracker: an analysis tool for humanitarian and disaster relief. In: *Proceedings of the fifth international conference on weblogs and social*, Barcelona, Spain, 17–21 July.
- Laney D (2001) 3D data management: Controlling data volume, velocity and variety. Available at: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf> (accessed 19 December 2018).
- Lee CJ and Katz AN (1998) The differential role of ridicule in sarcasm and irony. *Metaphor and Symbol* 13(1): 1–15.
- Liebrecht C, Kunneman F and van den Bosch A (2013) The perfect solution for detecting sarcasm in tweets #not. In: *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis – WASSA*, Atlanta, USA, 14 June.
- Mahoney J, Le Moignan E, Long K, et al. (2019) Feeling alone among 317 million others: Disclosures of loneliness on Twitter. *Computers in Human Behavior* 98(1): 20–30.
- Maity SK, Saraf R and Mukherjee A (2015) # Bieber+# Blast = # BieberBlast: early prediction of popular hashtag compounds. arXiv, arXiv–1510.
- Manning CD and Schütze H (1999) *Foundations of Statistical Natural Language Processing*. New York, US: MIT Press.
- Maynard D and Greenwood MA (2014) Who cares about sarcastic tweets? In: *Investigating the impact of sarcasm on sentiment analysis, language resources and evaluation conference (LREC)*, Reykjavik, Iceland, 6 March.
- Montejo-Ráez A, Díaz-Galiano MC, Martínez-Santiago F, et al. (2014) Crowd explicit sentiment analysis. *Knowledge-Based Systems* 69(1): 134–139.
- Norrick NR (2003) Issues in conversational joking. *Journal of Pragmatics* 35(9): 1333–1359.
- OED (2020) Oxford University Press, December 2019. Web. Available at: <https://www.oed.com> (accessed 14 February 2020).
- Pak A and Paroubek P (2010) Twitter as a corpus for sentiment analysis and opinion mining. In: *Proceedings of LREC conference*, Malta, 17 May.
- Pang B and Lee L (2008) Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1–2): 1–135.
- Parde N and Nielsen R (2018) Detecting sarcasm is extremely easy. In: *Proceedings of the at workshop on computational semantics beyond events and roles*, New Orleans, USA, 5 June.
- Phan HT, Tran VC, Nguyen NT, et al. (2020) Improving the performance of sentiment analysis of tweets containing fuzzy sentiment using the feature ensemble model. *IEEE Access* 8(1): 14630–14641.
- Puschmann C and Powell A (2018) Turning words into consumer preferences: How sentiment analysis is framed in research and the news media. *Social Media + Society* 4(3): 1–12.
- Rajadesingan A, Zafarani R and Liu H (2015) Sarcasm detection on Twitter: A behavioral modelling approach. In: *Proceedings of the 8th ACM international conference on web search and data mining – WSDM*, Shanghai, China, 2–6 February.
- Ravi K and Ravi V (2015) A survey on opinion mining and sentiment analysis: Tasks. *Approaches and Applications, Knowledge-Based Systems* 89(1): 14–46.

- Reyes A, Buscaldi D and Rosso P (2009) The impact of semantic and morphosyntactic ambiguity on automatic humour recognition. In: *Proceedings of the 14th international conference on applications of natural language to information systems NLDB 2009*, Manchester, UK, 24 June.
- Reyes A and Rosso P (2011) Mining subjective knowledge from customer reviews: A specific case of irony detection. In: *Proceedings of the 2nd workshop on computational approaches to subjectivity and sentiment analysis*, Portland, USA, 24 June.
- Reyes A, Rosso P and Veale T (2013) A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation* 47(1): 239–268.
- Ribeiro FN, Araújo M, Gonçalves P, et al. (2016) SentiBench – a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science* 5(23): 1–29.
- Riloff E, Qadir A, Surve P, et al. (2013) Sarcasm as contrast between a positive sentiment and negative situation. In: *Proceedings of empirical methods in natural language processing EMNLP conference*, Seattle, Washington, 18–21 October.
- Sarmiento L, Carvalho P, Silva MJ, et al. (2009) Automatic creation of a reference corpus for political opinion mining in user-generated content. In: *Proceedings of the 1st international CIKM workshop on topic-sentiment analysis for mass opinion*, Hong Kong, China, 6 November.
- Sarsam SM, Al-Samarraie H, Alzahrani AI, et al. (2020) Sarcasm detection using machine learning algorithms in Twitter: A systematic review. *International Journal of Market Research* 62(5): 578–598.
- Smith (n.d.) The Perception and Interpretation of Irony and Sarcasm: How Varying Levels of Common Ground Affects Perception and Interpretation, Available at: <http://www.smithmichael.com/docs/ironyandsarcasm.doc> (accessed 7 April 2017).
- Sykora M, Jackson TW, O’Brien A, et al. (2014) Twitter based analysis of public, fine grained emotional reactions to significant events. In: *European conference on social media – ECSM 2014*, Brighton, UK, 10–11 July.
- Sykora M, Jackson TW, O’Brien A, et al. (2013) National security and social media monitoring: A presentation of the EMOTIVE and related systems. In: *IEEE European intelligence and security informatics conference*, Uppsala, Sweden, 12 August.
- Tepperman J, Traum D and Narayanan S (2006) “Yeah right”: Sarcasm recognition for spoken dialogue systems. In: *Proceedings of the INTERSPEECH – ICSLP 9th international conference on spoken language processing*, Pittsburgh, USA, 17–21 September.
- Thelwall M (2013) *Big Data and Social Web Research Methods*. Wolverhampton, UK: University of Wolverhampton.
- Yadollahi A, Shahraki AG and Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. *ACM Computing Surveys Journal* 50(2): 1–33.
- Zappavigna M (2015) Searchable talk: The linguistic functions of hashtags. *Social Semiotics* 25(3): 274–291.
- Zhu X and Goldberg AB (2009) *Introduction to Semi-Supervised Learning*. UK: Morgan & Claypool Publishers.