

# A Multi-View Sentiment Corpus

Debora Nozza   Elisabetta Fersini   Enza Messina

University of Milano-Bicocca,

Viale Sarca 336, 20126, Milan, Italy

{debora.nozza, fersini, messina}@disco.unimib.it

## Abstract

Sentiment Analysis is a broad task that involves the analysis of various aspect of the natural language text. However, most of the approaches in the state of the art usually investigate independently each aspect, i.e. Subjectivity Classification, Sentiment Polarity Classification, Emotion Recognition, Irony Detection. In this paper we present a Multi-View Sentiment Corpus (MVSC), which comprises 3000 English microblog posts related the *movie domain*. Three independent annotators manually labelled MVSC, following a broad annotation schema about different aspects that can be grasped from natural language text coming from social networks. The contribution is therefore a corpus that comprises five different views for each message, i.e. *subjective/objective*, *sentiment polarity*, *implicit/explicit*, *irony*, *emotion*. In order to allow a more detailed investigation on the human labelling behaviour, we provide the annotations of each human annotator involved.

## 1 Introduction

The exploitation of user-generated content on the Web, and in particular on the social media platforms, has brought to a huge interest on Opinion Mining and Sentiment Analysis. Both Natural Language Processing (NLP) communities and corporations are continuously investigating on more accurate automatic approaches that can manage large quantity of noisy natural language texts, in order to extract opinions and emotions towards a topic. The data are usually collected from Twitter, the most popular microblogging platform. In this particular environment, the posts, called tweets,

are constrained to a maximum number of characters. This constraint, in addition to the social media context, leads to a specific language rich of synthetic expressions that allow the users to express their ideas or what happens to them in a short but intense way.

However, the application of automatic sentiment classification approaches, in particular when dealing with noisy texts, is subjected to the presence of sufficiently manually annotated dataset to perform the training. The majority of the corpora available in the literature are focused on only one (or at most two) aspects related to Sentiment Analysis, i.e. Subjectivity, Polarity, Emotion, Irony.

In this paper we propose a Multi-View Sentiment Corpus<sup>1</sup>, manually labelled by three independent annotators, that makes possible to study Sentiment Analysis by considering several aspects of the natural language text: *subjective/objective*, *sentiment polarity*, *implicit/explicit*, *irony* and *emotion*.

## 2 State of the Art

The work of Go et al. (2009) was the first attempt to address the creation of a sentiment corpus in a microblog environment. Their approach, introduced in (Read, 2005), consisted to filter all the posts containing emoticons and subsequently label each post with the polarity class provided by them. For example, :) in a tweet indicates that the tweet contains positive sentiment and :( indicates that the tweet contains negative sentiment. The same procedure was also applied in (Pak and Paroubek, 2010), differently from the aforementioned works they introduced the class of objective posts, retrieved from Twitter accounts of popular newspapers and magazines. Davidov et

<sup>1</sup>The proposed corpus is available at [www.mind.disco.unimib.it](http://www.mind.disco.unimib.it)

al. (2010) maintained the idea of distant supervision by combining 15 common emoticons and 50 sentiment-driven hashtags for automatic labelling. However, an intervention of human experts was needed to annotate the sentiment of frequent tags. Kouloumpis et al. (2011) extended their work in order to perform a 3-way polarity classification (positive, negative and neutral) on the Edinburgh Twitter corpus (Petrović et al., 2010).

Mohammad (2012) and Wang et al. (2012) applied the same distant supervision approach for the construction of a large corpus for emotion classification. They collect the data retrieving tweets by considering as keywords a predefined list of emotion hashtags. In (Mohammad, 2012), the authors used the Ekman's six basic emotions (#anger, #disgust, #fear, #joy, #sadness, and #surprise), while in (Wang et al., 2012) the authors expanded this list by including both basic and secondary emotions and their lexical variants, for a total of 131 keywords.

Hashtags have also been used to create datasets for irony detection purposes. The work of Reyes et al. (2013) proposed a corpus of 40000 tweets, 10000 ironic and 30000 non ironic tweets respectively retrieved with the hashtags #irony for the former and #education, #humor, #politics for the latter.

However, each of these resources have been created either fully automatically or in a semi-supervised way based on the assumption that single words and symbols are representative of the whole document. Moreover, the use of hashtags and emoticons for exploiting distant-supervision approaches can definitely create a bias towards posts that do not use these forms of expression to communicate opinions and emotions. Adopting a manual annotation approach is crucial for dealing with these issues and obtaining high quality labelling. In this direction the SemEval corpora (Nakov et al., 2013; Rosenthal et al., 2014; Nakov et al., 2016) have provided a fundamental contribution. These datasets have been labelled by taking advantage of crowdsourcing platforms, such as Amazon Mechanical Turk and CrowdFlower. Although the size of these corpora is very high (around 15-20K posts), Mozetič et al. (2016) overly exceeded these dimensions proposing a set of over 1.6 million sentiment labelled tweets. This corpus, that is the largest manually-labelled dataset reported in the literature, was an-

notated in 13 European languages.

Regarding emotion classification, Roberts et al. (2012) introduced a corpus of tweets manually labelled with the Ekman's six basic emotions and love. In (Liew et al., 2016), the authors extended their work by considering a fine-grained set of emotion categories to better capture the richness of expressed emotions.

The only manually-annotated corpus on irony detection was proposed by (Gianti et al., 2012). They studied the use of this particular device on Italian tweets, focusing on the political domain.

In this paper, we present a Multi-View Sentiment Corpus (MSVC) on English microblog posts that differs from the state of the art corpora for several reasons:

- The proposed corpus is the first benchmark that collects implicit or explicit opinions. This contribution will allow researchers to develop sentiment analysis approaches able to model opinions not directly expressed.
- The corpus provides different annotations simultaneously: subjectivity/objectivity, polarity, implicitness/explicitness, emotion, irony. This characteristic allows researchers to perform wide-ranging studies on the users' opinions, instead of considering each of this view as independent from the others.
- The corpus will show the label provided by each annotator, instead of producing a final label obtained by a majority voting rule. Given the different expertise of the annotators involved, a detailed investigation on single behaviours can be performed to improve the knowledge about the annotation procedures.
- This is the first corpus that explicitly labels emojis. We aim to prove that the role of the emojis is strictly related to the context where they appear: their contribution in terms of conveyed sentiment (or conveyed topic) strictly depends on the domain where they are used.

### 3 Annotation Procedure

The corpus has been annotated by considering different views related to the same post: subjectivity/objectivity, polarity, implicitness/explicitness, presence of irony and emotion. In this section, we provide a definition and examples for each of these

views. Moreover, we present the characteristics of the annotators in order to have more insights on their behaviour.

### 3.1 Annotation of Subjectivity/Objectivity

Given a post  $p$  about a given topic  $t$ , its subjectivity or objectivity can be defined as follows (Liu, 2012):

**Definition 1.** An *objective* post  $p_o$  presents some factual information about the world, while a *subjective* post  $p_s$  expresses some personal feelings, views, or beliefs.

In microblogs contexts the recognition of objective posts can be easily misled by the presence of hashtags and other linguistic artefacts that aim to show the post as more appealing. The reported examples are very similar, despite they belong to different classes:

[Objective] “Tonight @CinemaX #SuicideSquad!! Come to see #HarleyQuinn :)”

[Subjective] “-1 to #Deadpool...that’s tomorrow!!!! I can’t wait!”

### 3.2 Annotation of polarity

Given a subjective post  $p_s$  that expresses an opinion about a topic  $t$ , we want to determine its polarity between positive, negative and neutral classes. While the definition of positive and negative classes is commonly clear, the neutral label is differently treated in the state of the art. As in Pang and Lee (2008), we use neutral only in the sense of a sentiment that lies between positive and negative.

Posts that express a sentiment about specific aspects of a given topic  $t$ , such as actors, scenes, commercials for a film are considered part of the topic. Moreover, it is important to understand what is the target of the opinion, because it can lead to completely different interpretations.

[Positive] “Best Joker EVER!! #suicidesquad”

[Negative] “Deadpool is so childish! I slept during the movie”

[Neutral] “Good movie, @VancityReynolds worst actor ever #deadpool” (neutral - mixed sentiment)

[Neutral] “I love my boyfriend! We are watching deadpool tonight” (positive about the boyfriend - neutral about the film)

### 3.3 Annotation of explicit/implicit opinion

Given a subjective post  $p_s$  that expresses an opinion about a topic  $t$ , we can define its implicitness or explicitness as follows (Liu, 2012):

**Definition 2.** An *explicit opinion* is a subjective statement that gives an opinion.

**Definition 3.** An *implicit (or implied) opinion* is an objective statement that implies an opinion. Such an objective statement usually expresses a desirable or undesirable fact.

The detection of an implicit opinion can be complex because it does not rely on specific words (e.g. amazing, awful), as in the following examples:

[Explicit - Positive] “Suicide Squad is a great movie and an awesome cast”

[Implicit - Positive] “I’ve already watched Deadpool three times this month”

[Implicit - Negative] “I went out the cinema after 15 minutes #suicidesquad”

### 3.4 Annotation of Irony

Given a subjective post  $p_s$  that expresses an opinion about a topic  $t$ , the presence of irony can be detected focusing on the definition given by Wilson and Sperber (2007):

**Definition 4.** *Irony* is basically a communicative act that expresses the opposite of what is literally said.

Irony is one of the most difficult figurative language to comprehend, and a person can perceive it differently depending on several factors (e.g. culture, language).

[Ironic] “Hey @20thcenturyfox remember when you didn’t want anything to do with #Deadpool and now it’s your biggest opening weekend ever?”

### 3.5 Annotation of Emotion

A post  $p$  about the topic  $t$  can be associated to an emotion  $e$  corresponding to the eight Plutchik primary emotions (shown in Figure 1): anger, anticipation, joy, trust, fear, surprise, sadness and disgust. We provide an example for each emotion.

[Anger] “#Deadpool I wasted time and money grrrrrrrr”

[Anticipation] “Can’t wait to see Deadpool”

[Joy] “Deadpool was A-M- A-Z- I-N- G”

[Trust] “Best movie ever #Deadpool! Trust me!”

[Fear] “Saw #Deadpool last night. I was frightened during some crude scenes!”

[Surprise] “Much to my surprise, I actually liked *Deadpool*.”

[Sadness] “i finally got to watch *deadpool* and im so sad this is so boring”

[Disgust] “*Deadpool* is everything I hate about our century combined in the trashiest movie possible.”

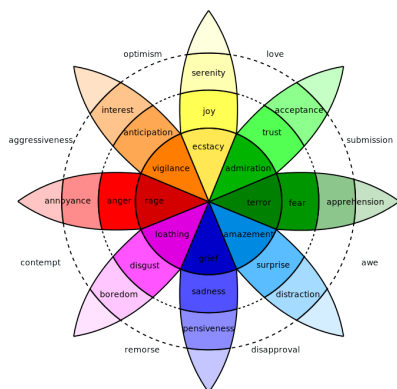


Figure 1: Plutchik’s wheel of emotions.

### 3.6 Annotation of Emojis

Given a post  $p$  related to a specific topic  $t$ , each emoji (if present) has been labelled as *positive*, *negative*, *neutral* or *topic-related* according to the context where it has been used. We provide an example for each label.

[Positive] “Love Jared Leto 🤩❤️”

[Negative] “*Suicide Squad* was awful 🤮”

[Neutral] “#*Suicide Squad*...Still thinking 🤔”

[Topic] “*Deadpool* today !! 🍷👤🎥”

### 3.7 Annotators

The complete set of posts has been labelled by three different annotators. Each annotator is a very proficient English speaker and he/she has a different level of NLP background and topic knowledge from the others. We distinguish these two types of knowledge because they are equally important and necessary for annotating a dataset, especially in a movie domain. A topic expert can be very confident on understanding the meaning of the text, but without any NLP knowledge he/she would not be able to perform a confident annotation, especially when dealing with the implicitness/explicitness and subjectivity/objectivity labels. On the other hand, being only a NLP expert

is not sufficient when in the text subtle and sophisticated references to the topic are present, resulting in an incorrect annotation because of an improper understanding.

The first annotator  $A_1$  is a NLP expert while he/she is not very confident on the topic selected, the second annotator  $A_2$  has a good expertise in NLP and a good knowledge about the topic, the third annotator  $A_3$  is a beginner in the field of NLP but he/she is competent on the topic.

## 4 Dataset

The data has been retrieved by monitoring different keywords on the Twitter microblogging platform related to two popular movies: *Deadpool* and *Suicide Squad*. This choice was motivated by the intention to increase the number of opinionated posts and therefore to have a variety of aspects to be analysed. Also, both the movies were massive blockbuster successes with popular actors and this led to a very wide and diverse audience.

This case study is experimentally convenient for our purposes because it represents a domain where people are more willing to express opinions, so that the final corpus will have a variety of opinionated tweets expressed in diverse ways. The collection of the data has been performed in the narrow days of the release date, *Deadpool* 18<sup>th</sup> February 2016 and *Suicide Squad* 1<sup>th</sup> August 2016.

After the streaming collection phase, we filter out the non-English tweets, duplicates and retweets resulting in a dataset of millions of posts. Then, we randomly sampled 3000 tweets equally distributed between the topics, maintaining the original daily distribution. This sample has been manually annotated, obtaining a final corpus composed of 1500 posts about *Deadpool* and 1500 posts about *Suicide Squad*.

On average, a tweet is composed of about 14 words of which one is a hashtag. Although this number can lead to conclude that hashtags are an important language expression and therefore they can be used for automatically collecting opinions and emotions, we found that most of them are strongly related only to the topic, e.g. #ShowTimeAtGC, #Joker, #HarleyQuinn, #DC. A preliminary analysis of the user mentions has shown that users are inclined to directly mention the actors or the entertainment companies for complaining or complimenting, and this, together with hashtags, can be particularly helpful when per-

forming aspect-based sentiment analysis.

Regarding emojis, the most frequent ones are (as expected) sentiment-driven, i.e. joy, heart eyes, sob.

## 5 Annotation Evaluations

The annotation of emotions, sentiment and other emotional aspects in a microblog message is not an easy task, and strongly depends on subjective judgement of human annotators. Annotators can disagree between themselves, and sometimes an individual cannot be consistent with her/himself. The disagreement depends on the complexity of the annotation task, the use of complex language (e.g., slang), or simply on the poor annotator work.

Table 1: Label distribution per annotator

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>
<b>Subjective</b>	0.671	0.748	0.657
<b>Objective</b>	0.330	0.252	0.343
<b>None</b>	0.330	0.252	0.343
<b>Positive</b>	0.509	0.476	0.426
<b>Neutral</b>	0.038	0.146	0.131
<b>Negative</b>	0.123	0.126	0.100
<b>None</b>	0.330	0.252	0.343
<b>Explicit</b>	0.254	0.512	0.416
<b>Implicit</b>	0.416	0.236	0.242
<b>Not Ironic</b>	0.980	0.988	0.971
<b>Ironic</b>	0.020	0.012	0.029
<b>None</b>	0.374	0.355	0.507
<b>Joy</b>	0.317	0.328	0.243
<b>Anticipation</b>	0.144	0.108	0.078
<b>Disgust</b>	0.070	0.071	0.047
<b>Surprise</b>	0.038	0.024	0.038
<b>Sadness</b>	0.035	0.044	0.036
<b>Anger</b>	0.014	0.034	0.028
<b>Trust</b>	0.008	0.034	0.022
<b>Fear</b>	0.001	0.001	0.002

In Table 1, we report some statistics that summarize behaviours of the involved annotators. By analysing the distributions, we can observe different attitudes: A<sub>1</sub> is inclined to label more posts as positive against the neutral ones; A<sub>2</sub> shows a predisposition to identify a high number of explicit expressions; A<sub>3</sub> has a low sensibility to capture the emotions behind the text. Moreover, we can highlight a balanced distribution for implicit/explicit opinions.

For those tweets encoding one of the eight emotions, there is a predominance of the joy label. Concerning the remaining classes the distributions are skewed towards a specific label, i.e. Subjective, Positive and Not Ironic.

An analogous consideration can be drawn for the emoji distribution (see Table 2). It turns out that most of the emojis are positive, especially the most popular ones and their presence provide an insight of the human emotional perceptions.

Table 2: Emojii distribution

	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>
<b>Topic</b>	0.141	0.129	0.095
<b>Positive</b>	0.559	0.601	0.593
<b>Negative</b>	0.141	0.187	0.173
<b>Neutral</b>	0.160	0.083	0.139

By a detailed analysis of the emoji annotations, it emerges that the role of the emojis is closely related to the context where they appear: their contribution in terms of conveyed sentiment (or conveyed topic) strictly depends on the domain where they are used. In Table 3, we report a comparison between the label distribution of two emojis in our corpus and the corresponding distribution in a state of the art emoji sentiment lexicon (Novak et al., 2015).

In the proposed corpus, the *fire* emoji has been mainly labelled as positive because it represents the word “hot”, whose meaning is intended as something beautiful and trendy. However, in the emoji sentiment lexicon the same emoji primarily corresponds to a neutral sentiment. Similar considerations can be drawn for the *pistol* emoji: in our corpus it represents the topic underlying the two movies, while in the state of the art lexicon it is frequently used to denote a negative sentiment orientation. As conclusion, any emoji should be not considered as independent on the context and therefore evaluated according to its semantic.

### 5.1 Agreement Measures

The kappa coefficient (Cohen, 1960) is the most used statistic for measuring the degree of reliability between annotators. The need for consistency among annotators immediately arises due to the

Table 3: Emoji label distribution



			
Multi-View Sentiment Corpus	Topic	0.127	<b>0.476</b>
	Negative	0.079	0.143
	Neutral	0.111	0.190
	Positive	<b>0.683</b>	0.190
Emoji Sentiment Ranking (Novak et al., 2015)	Negative	0.124	<b>0.493</b>
	Neutral	<b>0.613</b>	0.209
	Positive	0.263	0.298

Table 4: Inter-agreement (PABAK-OS)

	Subjective/Objective	Sentiment Polarity	Implicit/Explicit	Irony	Emotion
$\mathbf{A}_1$ vs $\mathbf{A}_2$	0.606	0.598	0.354	0.949	0.590
$\mathbf{A}_2$ vs $\mathbf{A}_3$	0.670	0.596	0.476	0.923	0.601
$\mathbf{A}_1$ vs $\mathbf{A}_3$	0.592	0.585	0.416	0.912	0.551

Table 5: Self-agreement (PABAK-OS)

	Subjective/Objective	Sentiment Polarity	Implicit/Explicit	Irony	Emotion
$\mathbf{A}_1$	0.920	0.920	0.640	1.000	0.820
$\mathbf{A}_2$	0.878	0.867	0.670	0.960	0.865
$\mathbf{A}_3$	1.000	0.920	0.850	0.878	0.820

variability among human perceptions. This **inter-agreement** measure can be summarized as:

$$k = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}} \quad (1)$$

However, considering only this statistic is not appropriate when the prevalence of a given response is very high or very low in a specific class. In this case, the value of kappa may indicate a low level of reliability even with a high observed proportion of agreement. In order to address these imbalances caused by differences in prevalence and bias, Byrt et al. (1993) introduced a different version of the kappa coefficient called prevalence-adjusted bias-adjusted kappa (PABAK). The estimation of PABAK depends solely on the observed proportion of agreement between annotators:

$$PABAK = 2 \cdot \text{observed agreement} - 1 \quad (2)$$

A more reliable measure for estimating the agreement among annotators is PABAK-OS (Parker et al., 2011), which controls for chance agreement. PABAK-OS aims to avoid the peculiar, unintuitive results sometimes obtained from Cohen’s Kappa, especially related to skewed annotations (prevalence of a given label).

We report in Table 4, the inter-agreement between couples of annotators distinguished for each label. We can easily note that the highest agreement is related to the irony/not-irony labelling. This is due to the predominance of non-ironic messages identified by all the annotators. Thus, we perform a detailed analysis on the disagreement between each couple of annotators regarding only the ironic messages. From the results, reported in Table 6, we can confirm that  $\mathbf{A}_1$  and  $\mathbf{A}_2$  annotators are more willing to interpret irony similarly (as already stated in Table 4).

Concerning the implicit/explicit labels, the inter-agreement measure highlights the difficulties

Table 6: Count of disagreement on the irony label

	Disagreement ( <i>irony</i> )
$\mathbf{A}_1$ vs $\mathbf{A}_2$	76
$\mathbf{A}_2$ vs $\mathbf{A}_3$	114
$\mathbf{A}_1$ vs $\mathbf{A}_3$	132

encountered by the annotators to distinguish “objective statements” (see Definition 1) from “objective statements that imply an opinion” (see Definition 3). Regarding the remaining labels, we can assert that there is a moderate agreement between the labellers. An analogous conclusion can be derived for the consensus about the emoji annotation, where the inter-agreement is 0.731 for  $\mathbf{A}_1$  vs  $\mathbf{A}_2$ , 0.771 for  $\mathbf{A}_2$  vs  $\mathbf{A}_3$ , and 0.647 for  $\mathbf{A}_1$  vs  $\mathbf{A}_3$ .

When dealing with complex annotations, the perception of the same annotator on the same post can change over time, resulting in inconsistent labelling. In order to estimate the uncertainty of the annotation of each labeller, we sampled a portion of tweets to be annotated twice by the same annotator. We report in Table 5 the **self-agreement** measure, that is a valid index to quantify the quality of the labelling procedure. The resulting statistics show that there is a high self-agreement for almost all the labels. The annotators can be considered moderately reliable for implicit/explicit annotations and very accurate for the remaining labels.

## 6 Conclusion

In this paper we presented a Multi-View Sentiment Corpus (MVSC), which simultaneously considers different aspects related to sentiment analysis, i.e. *subjectivity*, *polarity*, *implicitness*, *irony*, *emotion*. We described the construction of the corpus, together with annotation schema, statistics and some interesting remarks. The proposed corpus is aimed at providing a benchmark to develop sentiment analysis approaches able to model opinions not directly expressed. Researchers can also take advan-

tage of the complete label set given by the annotators to investigate their behaviours and the underlying annotation procedures. We finally provided some interesting conclusions related to the use of emojis, highlighting that their role is strictly related to the context where they appear. As future work, we aim at defining novel machine learning models able to simultaneously take advantage of the multiple views available. Moreover, an annotation scheme at a fine-grained level will be investigated.

## References

- Ted Byrt, Janet Bishop, and John B. Carlin. 1993. Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, 46(5):423–429.
- Jacob Cohen. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1):37–46.
- Dmitry Davidov, Oren Tsur, and Ari Rappoport. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 241–249. Association for Computational Linguistics.
- Andrea Gianti, Cristina Bosco, Viviana Patti, Andrea Bolioli, and Luigi Di Caro. 2012. Annotating irony in a novel italian corpus for sentiment analysis. In *Proceedings of the 4th Workshop on Corpora for Research on Emotion Sentiment and Social Signals*, pages 1–7.
- Alec Go, Lei Huang, and Richa Bhayani. 2009. Twitter sentiment analysis. *Final Projects from CS224N for Spring 2008/2009 at The Stanford Natural Language Processing Group*.
- Efthymios Kouloumpis, Theresa Wilson, and Johanna D. Moore. 2011. Twitter sentiment analysis: The good the bad and the omg!. *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*, 11:538–541.
- Jasy Suet Yan Liew, Howard R. Turtle, and Elizabeth D. Liddy. 2016. Emotweet-28: A fine-grained emotion corpus for sentiment analysis. In *Proceedings of the 10th International Conference on Language Resources and Evaluation*. European Language Resources Association (ELRA).
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.
- Saif M. Mohammad. 2012. #emotional tweets. In *Proceedings of the 1st Joint Conference on Lexical and Computational Semantics - Volume 1: Proceedings of the Main Conference and the Shared Task, and Volume 2: Proceedings of the 6th International Workshop on Semantic Evaluation*, pages 246–255. Association for Computational Linguistics.
- Igor Mozeti, Miha Grar, and Jasmina Smailovi. 2016. Multilingual twitter sentiment classification: The role of human annotators. *PLOS ONE*, 11(5):1–26.
- Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Alan Ritter, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. *Proceedings of the 7th International Workshop on Semantic Evaluation*, pages 312–320.
- Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov. 2016. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, pages 1–18. Association for Computational Linguistics.
- Petra Kralj Novak, Jasmina Smailović, Borut Sluban, and Igor Mozetič. 2015. Sentiment of emojis. *PLOS ONE*, 10(12):1–22.
- Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*. European Language Resources Association.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.
- Richard I. Parker, Kimberly J. Vannest, and John L. Davis. 2011. Effect size in single-case research: A review of nine nonoverlap techniques. *Behavior Modification*, 35(4):303–322.
- Saša Petrović, Miles Osborne, and Victor Lavrenko. 2010. The edinburgh twitter corpus. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics in a World of Social Media*, pages 25–26. Association for Computational Linguistics.
- Jonathon Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, pages 43–48. Association for Computational Linguistics.
- Antonio Reyes, Paolo Rosso, and Tony Veale. 2013. A multidimensional approach for detecting irony in twitter. *Language Resources and Evaluation*, 47(1):239–268.
- Kirk Roberts, Michael A. Roach, Joseph Johnson, Josh Guthrie, and Sanda M. Harabagiu. 2012. Empatweet: Annotating and detecting emotions on twitter. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*. European Language Resources Association.

Sara Rosenthal, Alan Ritter, Preslav Nakov, and Veselin Stoyanov. 2014. Semeval-2014 task 9: Sentiment analysis in twitter. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, pages 73–80. Association for Computational Linguistics and Dublin City University.

Wenbo Wang, Lu Chen, Krishnaprasad Thirunarayan, and Amit P. Sheth. 2012. Harnessing twitter “big data” for automatic emotion identification. In *Proceedings of the 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*, pages 587–592. IEEE Computer Society.

Deirdre Wilson and Dan Sperber. 2007. On verbal irony. *Irony in language and thought*, pages 35–56.