

# A Corpus for the Analysis of Online News Comments

**Varada Kolhatkar**

Discourse Processing Lab  
Simon Fraser University  
Burnaby, Canada  
vkolhatk@sfu.ca

**Maita Taboada**

Discourse Processing Lab  
Simon Fraser University  
Burnaby, Canada  
mtaboada@sfu.ca

## Abstract

There is a growing interest among journalists, social scientists and computational linguists in studying online interactions and mediating them to improve their quality. We present the SFU Opinion and Comments Corpus which can be used to study online interactions in a systematic way. The corpus comprises 10,339 opinion articles from the Canadian national newspaper *The Globe and Mail* and 663,173 comments posted in response to those articles. We also present a description of four layers of annotations carried out on a subset of the corpus: constructiveness, toxicity, negation, and Appraisal.

## 1 Introduction

Online commenting allows for direct communication among people and organizations from diverse socioeconomic classes and backgrounds on important issues. Popular news articles receive thousands of comments. These comments create a rich resource for computational linguists, as they are an excellent source of evaluative, abusive and argumentative language; sarcasm; dialogic structure; and occasionally well-informed constructive language. They contain information about people's opinion or stance on important issues, policies, popular topics, and public figures. Such a resource can be useful to answer a number of interesting research questions about journalism, online language, and human conversation. Potential questions include:

1. How can we best organize comments to encourage constructive and civil conversations online?
2. Do the comments express varying views on issues and policies? What is the most popular view? What is the most informative and constructive view supported by evidence?
3. How can we create a succinct summary of different views on an article or issues or policies in the article?
4. Do people engage more in emotionally-driven conversation or in discussion based on facts and evidence?
5. How often do commenters target authors and other commenters, rather than the arguments or issues in the article?

In order to study these questions systematically, we need a large, well-curated corpus of reader comments. Currently, only a few such corpora are available. Two prominent ones are: the Yahoo News Annotated Comments Corpus (YNACC) (Napoles et al., 2017)<sup>1</sup>, and the SENSEI Social Media Annotated Corpus<sup>2</sup> (Barker and Gaizauskas, 2016). The former contains 522,000 comments from 140,000 threads posted in response to Yahoo News articles. Among these, 9,200 comments and 2,400 threads have been annotated for phenomena such as sentiment, persuasiveness, tone of each comment, and the quality of the overall thread. The SENSEI Social Media Annotated Corpus was created with the goal of generating summaries of reader comments. Accordingly, they have created an annotated corpus of 1,845 comments posted on 18 articles from the British newspaper *The Guardian*. These corpora contain rich sources of information. We are interested, however, in the link between articles and comments, and in particular in how evaluative language varies between articles that contain opinion and their comments.

<sup>1</sup><https://webscope.sandbox.yahoo.com/catalog.php?datatype=1&did=83>

<sup>2</sup><http://sensei.group.shef.ac.uk/sensei/corpus.html>

Constructive comments	Non-constructive comments
target specific points and provide appropriate evidence	present opinions without support
offer a solution to the issues discussed in the article	dismiss the terms of debate
share a related personal story or experience	are provocative or excessively flattering
encourage other readers to participate in the discussion	irrelevant or unsubstantial

Table 1: Some prototypical characteristics of constructive comments

## 2 SFU Opinion and Comments Corpus

We present the SFU Opinion and Comments Corpus (SOCC), a collection of opinion articles and the comments posted in response to the articles. The articles include all the opinion pieces published in the Canadian newspaper *The Globe and Mail* in the five-year period between 2012 and 2016, a total of 10,339 articles and 663,173 comments from 303,665 comment threads. The corpus contributes by providing a pairing of articles and comments, and introducing the largest dataset of this kind to date. Furthermore, the articles are all opinion articles, not hard news articles. This is important, because it allows for comparisons of evaluative language in both text types, opinion articles and reader comments. Opinion articles are generally subjective and evaluative, but their language tends to be more formal and argumentative. The comments are also subjective; they, however, tend to be more informal and personal in nature. The corpus has been collected with attention to preserving reply structures and other metadata. We have made this corpus publicly available for non-commercial use through GitHub.<sup>3</sup>

## 3 Annotations

We have been carrying out annotations and analyses on a subset of SOCC (Kolhatkar and Taboada, 2017a,b). In this section, we summarize the results of four annotation experiments.

**Constructiveness.** There is a growing interest in automatically organizing reader comments in a sensible way (Napoles et al., 2017; Llewellyn et al., 2014). One useful way to organize comments is based on their *constructiveness*, i.e., by identifying which comments provide insight and encourage a healthy discussion. For instance, *The New York Times* manually selects and highlights comments representing a range of diverse views, referred to as *NYT Picks*.

<sup>3</sup><https://github.com/sfu-discourse-lab/socc>

To understand constructiveness in online comments, we annotated a subset of SOCC containing 1,121 comments for constructiveness using crowdsourcing.<sup>4</sup> We define constructiveness in terms of prototypical characteristics which we obtained from a crowdsourced survey, shown in Table 1.

Each comment was annotated by at least three annotators. As we were interested in the verdict of native speakers of English, we limited the allowed demographic region to English-speaking countries. The percentage agreement on a random sample of 100 annotations was 87.88%, suggesting that constructiveness can be reliably annotated. Among 1,121 comments, 603 comments (53.79%) were classified as constructive, 517 (46.12%) as non-constructive, and the annotators were not sure in only one case. To examine the quality of the crowd annotations we asked a moderator to evaluate the acceptability of the crowds answers. For that, we randomly selected 222 instances from the crowd-annotated data and asked the expert whether they agree with the crowds answer or not. Overall the expert agreed with the crowd 77.93% of the time. We carefully curated the crowd annotated corpus, removing duplicates and instances with very low agreement. Our curated corpus contains 1,043 instances and is available for non-commercial use via GitHub.<sup>5</sup>

**Toxicity.** We are interested in exploring the relationship between *constructiveness* and *toxicity* in reader comments. For that purpose, we added a layer of toxicity annotations on the same subset of 1,121 comments above. We defined four levels of toxicity: very toxic, toxic, mildly toxic, and not toxic. In our annotation guidelines we provide prototypical characteristics of toxicity, as shown in Table 2. Again we used crowdsourcing for annotation and each comment was annotated by at least three annotators.

Among all comments, 203 (19.46%) comments

<sup>4</sup><https://www.figure-eight.com>

<sup>5</sup><https://github.com/sfu-discourse-lab/socc#constructiveness>

Toxicity level	Prototypical characteristics
Level 4	contain harsh, abusive or offensive language are inflammatory contain a personal attack, insult or condemnation
Level 3	ridicule, tease, or poke others cause embarrassment or disrespect disagreeing aggressively or joking inappropriately
Level 2	express frustration and anger likely to be perceived as toxic by some people in some contexts
Level 1	unlikely to be perceived as toxic

Table 2: Some prototypical characteristics of toxic comments

had some toxicity in them. In our data, we did not find any significant differences in toxicity levels between constructive and non-constructive comments, i.e., constructive comments were as likely to be toxic as non-constructive comments. Current comment filtering approaches filter comments primarily based on toxicity. Our results suggest that we need to consider constructiveness along with toxicity while filtering or organizing reader comments, as aggressive debate might be good as long as it is constructive.

**Negation and Appraisal.** We have enriched our constructiveness corpus with two more layers of annotations: negation and Appraisal (Martin and White, 2005) annotations. For both phenomena, we have developed detailed guidelines. For negation, we have annotated keyword, scope, and focus (e.g., *I (cannot)<sub>keyword</sub> ((believe)<sub>focus</sub> (that the Globe publishes such articles)<sub>scope</sub>.*); and for Appraisal, we have annotated Attitude (Affect, Appreciation, or Judgement), polarity (positive, negative or neutral) and Graduation (Force or Focus). These annotations are at the word, phrase or sentence level, i.e., only words and expressions that convey evaluation were annotated, unlike the constructiveness and toxicity annotations, which were at the comment level. The negation annotations include 1,397 instances of a negative keyword (*not, never, nowhere*), with scope and focus also identified when one was present. The Appraisal annotations contain 6,623 instances of evaluative language, further labelled as Affect (expression of emotion, 3.4% of the total), Judgement (evaluation of a person, 42.6%), or Appreciation (evaluation of objects, 54%). We found that comments contained overwhelmingly negative Appraisal, 73.6% of the time. This set of annotations is useful to investigate different types of

evaluative language, and the interplay of evaluation and negation.

## 4 Discussion and conclusion

We present the SFU Opinion and Comments Corpus, a corpus of opinion articles and the comments posted in response to these articles. We summarize four annotation layers carried out on a subset of this corpus: constructiveness, toxicity, negation, and Appraisal. While carrying out annotations for constructiveness and toxicity, we learned that these phenomena are an interplay between a variety of other phenomena of interest in computational linguistics, such as argumentation, relevance of the comment to the article, the tone of the comment, and sarcasm. We believe that we may obtain better quality annotations if we ask specific questions leading to constructiveness and toxicity (e.g., whether the claims made in the article are supported by evidence), instead of asking binary questions, and in our current work, we are pursuing this research direction.

We believe that our raw and annotated corpora make an invaluable resource for those interested in the language of evaluation, a host of linguistic phenomena, and how public opinion is expressed through comments. The annotations can also be used in supervised machine learning approaches. We are making all these corpora freely available for non-commercial use via GitHub.<sup>6</sup>

## References

Emma Barker and Robert Gaizauskas. 2016. Summarizing multi-party argumentative conversations in reader comment on news. In *Proceedings of ACL 2016*, pages 12–20, Berlin.

<sup>6</sup><https://github.com/sfu-discourse-lab/SOCC>

- Varada Kolhatkar and Maite Taboada. 2017a. Constructive language in news comments. In *Proceedings of the First Workshop on Abusive Language Online*, pages 11–17, Vancouver.
- Varada Kolhatkar and Maite Taboada. 2017b. Using New York Times Picks to identify constructive comments. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 100–105, Copenhagen.
- Clare Llewellyn, Claire Grover, and Jon Oberlander. 2014. Summarizing Newspaper Comments. In *Proceedings of ICWSM*, Ann Arbor, MI.
- James R. Martin and Peter R. R. White. 2005. *The Language of Evaluation*. Palgrave, New York.
- Courtney Napoles, Joel Tetreault, Aasish Pappu, Enrica Rosato, and Brian Provenzale. 2017. Finding good conversations online: The Yahoo News Annotated Comments Corpus. In *Proceedings of the 11th Linguistic Annotation Workshop, EACL*, pages 13–23, Valencia.