The language of online commentary: Extracting information from news comments

Discourse Processing Lab
Luca Cavasso • Jennifer Fest • Emilie Francis • Varada Kolhatkar • Maite Taboada • Hanhan Wu

News comments
- Readers’ reactions posted after a news article
- Often informative, part of a dialogue – constructive
- Sometimes offensive, abusive, derogatory – toxic

Media and technology
- Toronto Star decides to close comments as of December 2015 [1]
- CBC News does not allow comments for stories on indigenous people [2]
- New Yorker: psychological toll on human monitors [3]
- Google: Jigsaw, devoted to helping monitor comments [4]

Online news comments
- The Globe and Mail
  - 10,339 articles
  - 77,238,179 words (comments)
  - 5 years: January 2012
- All opinion articles + all comments

SOCC: The SFU Opinion and Comments Corpus

The Globe and Mail
- All opinion articles + all comments
- 5 years: January 2012 – December 2016
- 10,339 articles – 1,280,454 comments
- 6,895,696 words (articles) + 77,238,179 words (comments)

Annotation

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The shallowness of this debate on government finances is depressing</td>
<td>Affect</td>
<td>Appreciation, Positive</td>
<td>Positive</td>
<td>Judgement, Negative</td>
</tr>
<tr>
<td>Great piece! Thanks</td>
<td>Affect</td>
<td>Appreciation, Positive</td>
<td>Positive</td>
<td>Judgement, Negative</td>
</tr>
<tr>
<td>Why does the Globe publish this kind of trash?</td>
<td>Affect</td>
<td>Appreciation, Positive</td>
<td>Positive</td>
<td>Judgement, Negative</td>
</tr>
</tbody>
</table>

1. Clinton lost, because she’s a woman but because she was perceived to be an establishment candidate.
2. This article was a big disappointment. Thank you Ms Henein. Now women know that wasting their time reading your emotion-based opinion piece is not an option.

Current work: Constructiveness & toxicity

Which comment is constructive?

<table>
<thead>
<tr>
<th>Comments</th>
<th>Constructive vs. Non-constructive</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARI HENEIN</td>
<td>Thank you, Hillary. Now women know retreat is not an option</td>
</tr>
</tbody>
</table>

Operationalizing constructiveness

- Argumentation [9]: Constructive comments more likely to
  - Contain modal verbs and argumentative connectives
- Rhetorical relations [10]: Constructive comments more likely to contain persuasive and presentational relations
  - Cause, Comparison, Condition, Contrast, Explanation

Operationalizing toxicity

- Toxic comments are likely to offend or cause distress
- Personal attacks, harassment, profanity, abuse, hate speech [11, 12]
- Sentiment analysis [13]: toxic comments contain (extreme) negative sentiment

Crowd annotation

- CrowdFlower. 1,200 comments from 10 articles. Labels:
  - Constructive vs. Non-constructive
  - Not toxic – Mildly toxic – Toxic – Very toxic
  - Results: no difference in toxicity between constructive and non-constructive comments

Future work

- Register analysis [14]:
  - Are comments more like conversations or monologues?
- Topic modelling in articles and comments [15]:
  - What is more frequently discussed in articles?
  - Which topics are discussed in the comments?
  - Which article topics are discussed in the comments?
  - What is more like conversations or monologues?
  - Link to sentiment

Applications

- Comment organization: sort by topic and sentiment
- Filtering: delete toxic comments
- Public opinion: what topics are discussed the most?

References
[8] Coherence relations.

Acknowledgements
Thanks to Emma Wilner, Rada Tomicic, Silvia Vázquez and Kazuki Yabe for their help in collecting and testing crowd questions. Thanks to Yue Wang for the poster template.

Funding: NSERC Insight Grant (“Opinion, subjectivity and point of view in language”); NSERC Discovery Grant (“A computational treatment of negation and speculation in natural language”).