

SFU Opinion and Comments Corpus



Varada Kolhatkar and Maite Taboada
 Department of Linguistics, Simon Fraser University, Canada
 vkolhatk@sfu.ca, mtaboada@sfu.ca



1. Why comments?

- Online news comments are a rich resource for computational linguists
 - They are an excellent source of evaluative and argumentative language; dialogic structure
 - They contain examples of toxic language, sarcasm, and well-informed constructive language
 - They also contain information about people's opinion on important issues and policies
- Can be used to tackle research questions around
 - Journalism and public opinion
 - Online language
 - Human conversation

2. Comment thread example

How to preserve Canada's indigenous languages

tkwhyte
 Dec 30, 2016
 How could preserving and teaching 60 languages that no one speaks be a benefit to anyone?? Teaching young people to speak these languages to "overcome their alienation", give me a break!! What better way to alienate a young person than to teach them a language no one understands. More wisdom from the Globe, from people that have probably never met an aboriginal person in their life!

37 Reactions Flag React Reply

SCLEB
 Dec 30, 2016
 The better question to ask is why is it the obligation of the state to preserve these languages?

19 Reactions Flag React Reply

Rochdale68
 Dec 30, 2016
 In reply to:
 How could preserving and teaching 60 languages that no one speaks be a benefit to anyone?? Teaching young people to speak these languages to "overcome their alienation", give me a break!! What better...
 — tkwhyte

Just an excuse to soak more tax dollars, they need to join the rest of the country not hide in desolation.
 Yes Chrétien said it best "MOVE"

7 Reactions Flag React Reply

4. Constructiveness examples

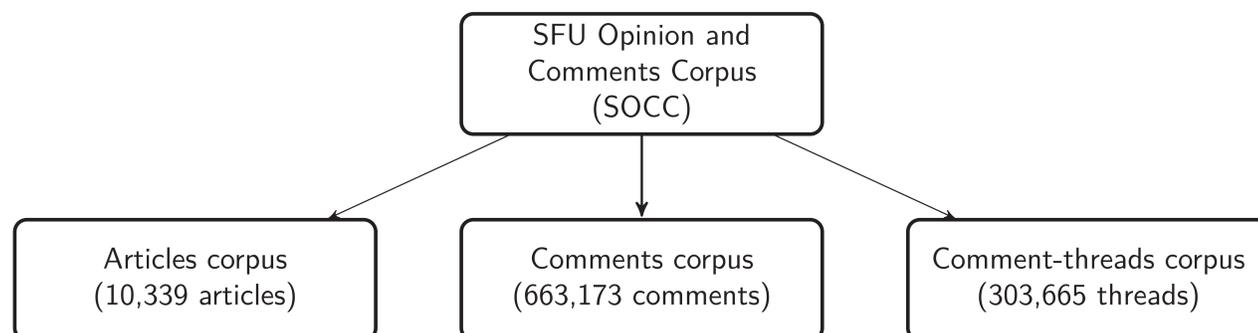
NDP thinks big with national daycare plan

Simpson is right: it's a political winner and a policy dud - just political smoke and mirrors. **Mulcair is power-hungry.** He wants Canada to adopt a national childcare model so he can hang on to seats in Quebec, that's all. Years ago I worked with a political strategist working to get a Liberal candidate elected in Conservative Calgary. He actually told his client to talk about national daycare - this was in the early 90's. The Liberal candidate said, 'Canada can't afford that!' to which the strategist responded 'Just say the words, you don't have to actually do it. It'll be good for votes.' I could barely believe the cynicism, but over the years I've come to realize that's what it is: vote getting and power politics. Same thing here.

Excellent!! Maggie, I think you are the only sane person in the world!

3. Raw corpus

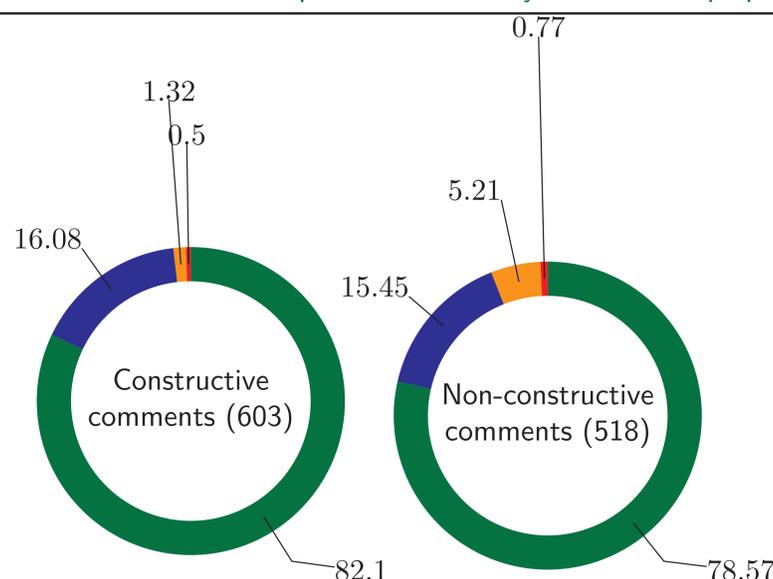
The raw corpus comprises **opinion articles** from the Canadian national newspaper *The Globe and Mail* for a 5-year period (January 2012–December 2016), plus all comments posted in response to the articles.



5. Constructiveness and toxicity annotations

Constructive comments target **specific points**; provide **appropriate evidence**; **offer a solution** to the issues discussed in the article; share a related **personal story or experience**; and **encourage** other readers to participate in the discussion.

- Very toxic** Please stop whining. Trump is a misogynist, racist buffoon and perhaps worse. Clinton is ethically challenged and craven in what she will tolerate in her lust for power. Neither of them is a stellar representative of their gender. Next time, put up a female candidate who outshines the male, not one who has sunk to his same level. Simple.
- Toxic** Why don't the NDP also promise 40 acres and a mule? They will never lead this country. Panderers to socialists and unionists.
- Mildly toxic** Is this a joke? Marie Henein as feminist crusader, advising us what to tell our daughters?? no thanks
- Not toxic** In my opinion, criticizing the new generation is not going to solve any problem. If you want to produce children, you should be prepared to pay for their care.



Number of instances	1,121
Inter-rater agreement	87.88%
Crowd-expert agreement	77.93%

6. Appraisal annotations

- What is the distribution of different types of evaluation?

The shallowness of this debate is depressing.	Affect, Negative
Thoughtful article! Thanks.	Appreciation, Positive
low-skilled immigrants...	Judgment, Negative

7. Negation annotations

- What kind of opinions are negated?
- What is most often in the focus of negation?
- How can we accurately identify the item being negated?

Clinton lost, **not** because <she's a woman> but because she was perceived to be an establishment candidate.

8. Further information

All corpora are available to download at: <https://github.com/sfu-discourse-lab/SOCC>

Ongoing work

- Carrying out large-scale annotations for constructiveness and toxicity
- Exploring the relation between toxicity and constructiveness
- Building computational models for constructiveness