

# 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

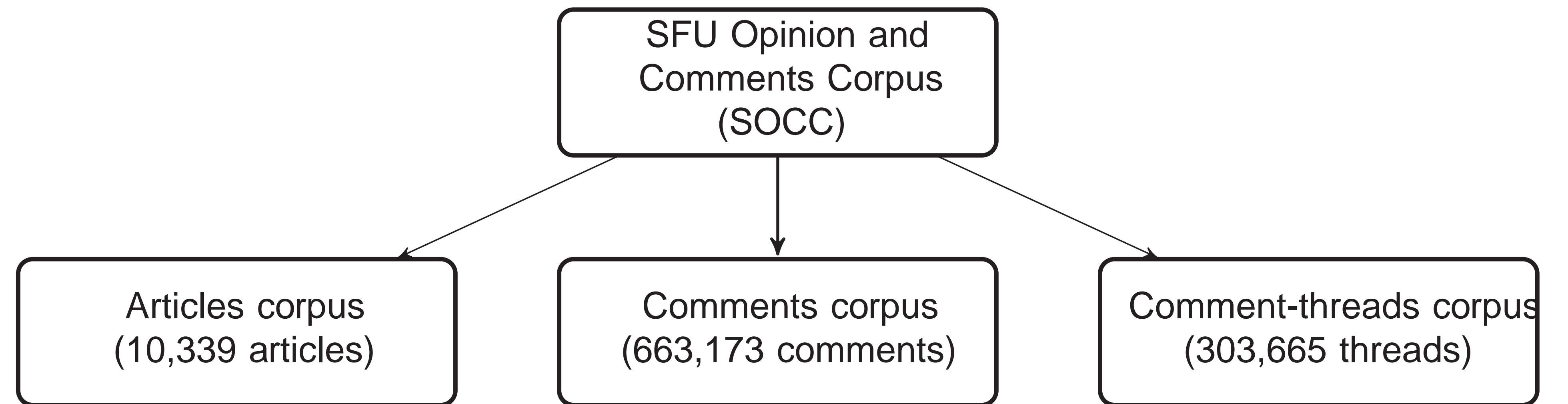
Journalism and public opinion

Online language

Human conversation

## 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