Civic Engagement with Local Government Videos: Comparing YouTube Transcripts with User Comments

Steven Coats
English, University of Oulu, Finland
steven.coats@oulu.fi

9th CMC-Corpora Conference, Santiago
September 29th, 2022

image source

Background

YouTube as a platform for mediated quasi-interaction with three communicative levels (Bou-Franch et al. 2012, Dynel 2014)

- First level: Face-to-face spoken interaction
- Second level: Corresponds to classifications of mass media
- *Third level*: Additional modalities/affordances of platform (e.g. commenting)

Mediated interaction can be important for societal stakeholders such as companies, organizations, and governments

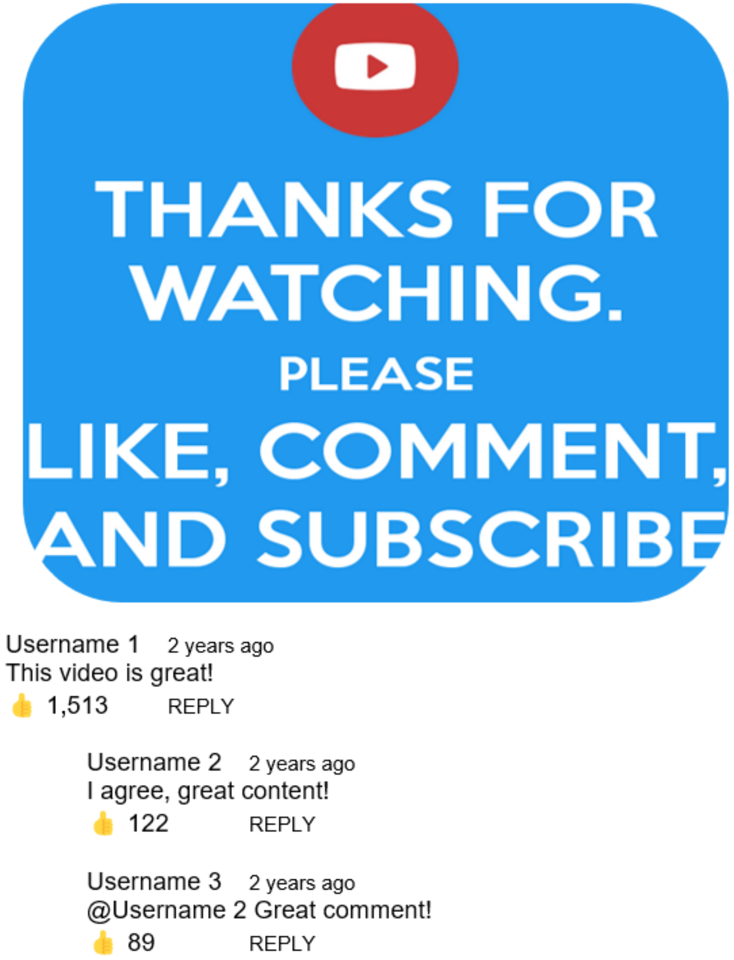
- How do people interact with content uploaded by local governments: what do they like and dislike?

Prior studies of YouTube comments

- Quality of comments (e.g. Goode et al. 2011)
- Typological classifications (e.g. Herring & Chae 2021, Häring et al. 2018)
- Sentiment of comments (e.g. Ksiazek 2018)
- Like ratio vs. text of comments (e.g. Schultes et al. 2012, Siersdorfer et al. 2014)

This study: **Transcripts of videos vs. comments**

- First large-scale comparison of discourse content of the videos and comments?
- Exploratory study to be developed

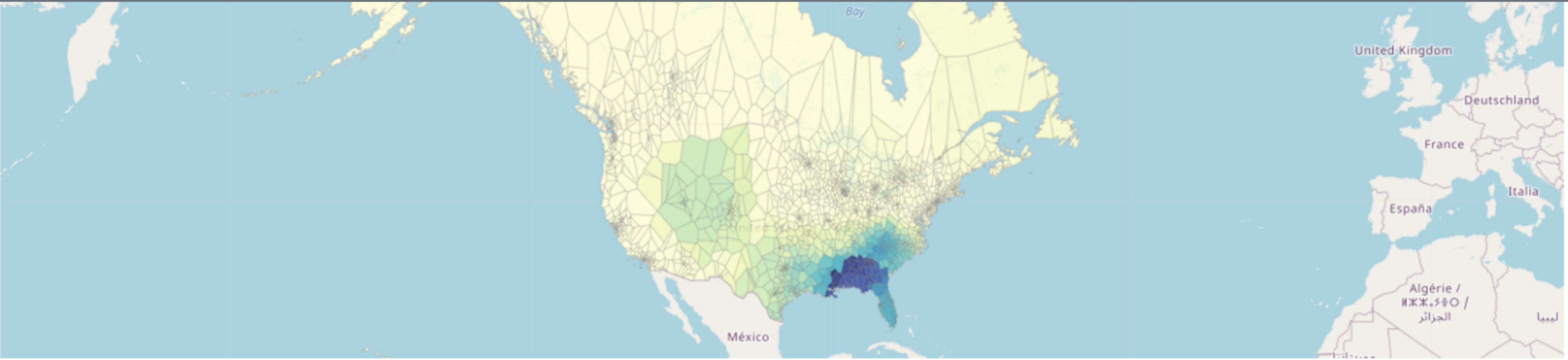


Data: transcripts

Corpus of North American Spoken English (CoNASE): 1.25b-word corpus of 301,846 word-timed, part-of-speech-tagged Automatic Speech Recognition (ASR) transcripts (Coats forthcoming a)

- Mostly transcripts of meetings and other local government content
- Freely available for research use; download from the Harvard Dataverse

Corpus of North American Spoken English v. 1.1 | Home | Description | Format | Locations | Sample script | Access | Contact



The Corpus of North American Spoken English (CoNASE) is a 1.29-billion-word corpus of geolocated automatic speech recognition (ASR) YouTube transcripts from the United States and Canada. It was created for the study of lexical, grammatical, and discourse-pragmatic phenomena of spoken language, including their geographical distribution, in North American English. The size of the corpus allows rare phenomena to be considered, and because the annotation includes the video IDs of transcripts, search hits can be manually inspected and

Focus on regional and local government channels

Many recordings of meetings of elected councillors: advantages in terms of representativeness and comparability

- Speaker place of residence (cf. videos collected based on place-name search alone)
- Topical contents and communicative contexts comparable
- In most jurisdictions government content is in the public domain

Data: comments

- For all videos in CoNASE, retrieve all available comments with youtube-comment-downloader
- = 190,079 total comments, for 20,965 videos (6.95% of CoNASE videos), 116,009 unique users, 5,334,096 word tokens
- Most of these videos have few views/likes/comments
- Local government does not engage people as much as music videos, video game streaming, makeup tutorials and other popular YouTube content

Sentiment

- Transformer models outperform "bag-of-words"-based sentiment analysis
- YT comments are rich in emoji, so sentiment models need to include emoji

```
string = "incredible 😭👎🤮"
doc = nlp(string)
print(doc._.blob.polarity,"\t", sentiment_task(string))
```

```
0.9	[{'label': 'Negative', 'score': 0.9354274272918701}]
```

Model: Twitter-roBERTa-base, trained on ~124m tweets from January 2018 to December 2021 (Loureiro et al. 2022, Barbieri et al. 2020)

- BERT-derived transformer model processing pipelines can typically only handle texts up to 512 tokens long
- Code to chunk transcripts, assign values to chunks, take mean values
- Assign sentiment values in the range 0 (negative) to 2 (positive) to all video transcripts and all comments

Topic Modeling

BERTopic (Grootendorst 2022)

- Groups lexical items and documents together in "topics"
- A form of dimensionality reduction that can give insight into discourse/content of text data
- BERTopic uses embeddings from sentence transformer models (take word context into account)

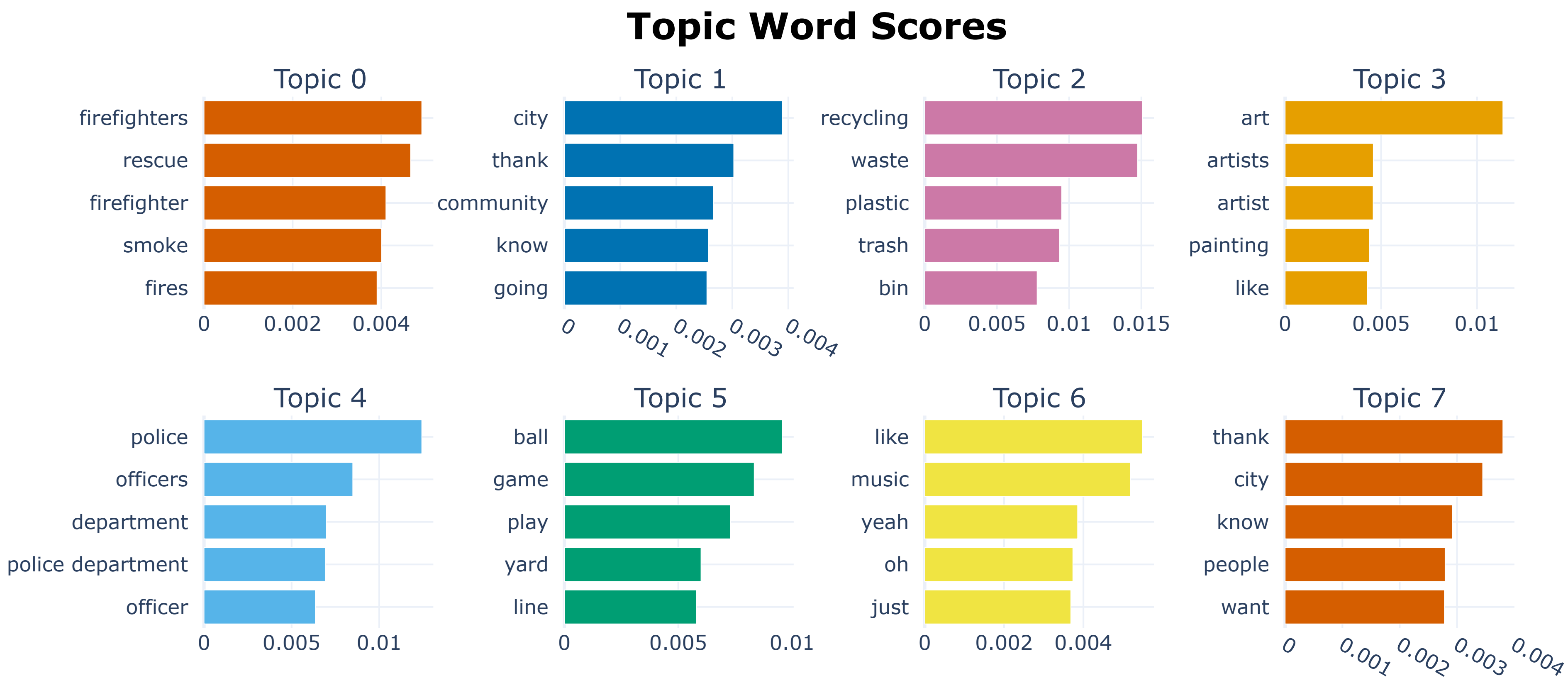
```
from sklearn.feature_extraction.text import CountVectorizer

vectorizer_model = CountVectorizer(ngram_range=(1, 3), stop_words="english")

topic_model = BERTopic(language="english", embedding_model="all-MiniLM-L12-v2",
                       vectorizer_model=vectorizer_model, nr_topics="auto")
topics, probs = topic_model.fit_transform(list(conase_subset.text))
```

- Used all-MiniLM-L12-v2, a model trained on 1.7 billion words of web texts from various genres

Research questions

- What kind of discourse content is represented in the videos?
- What does topic modeling tell us about the content of the transcripts?
- Which content attracts positive/negative comments?

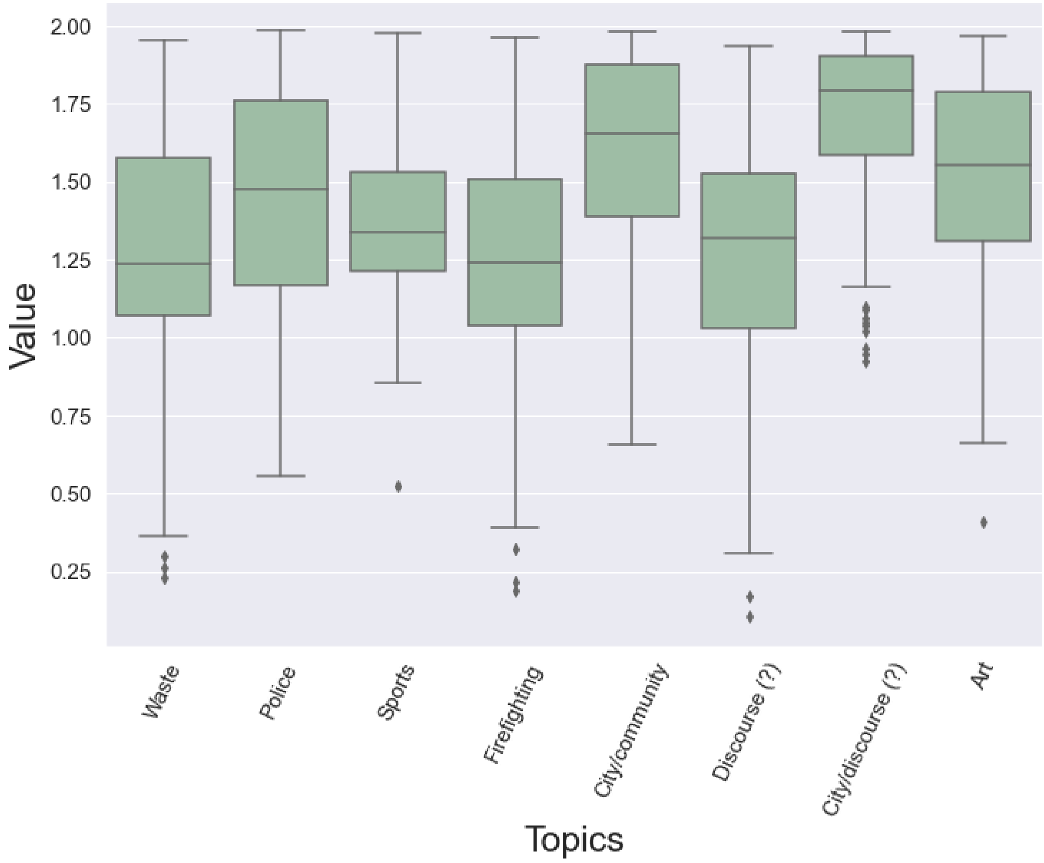
8 largerst topics (transcripts)



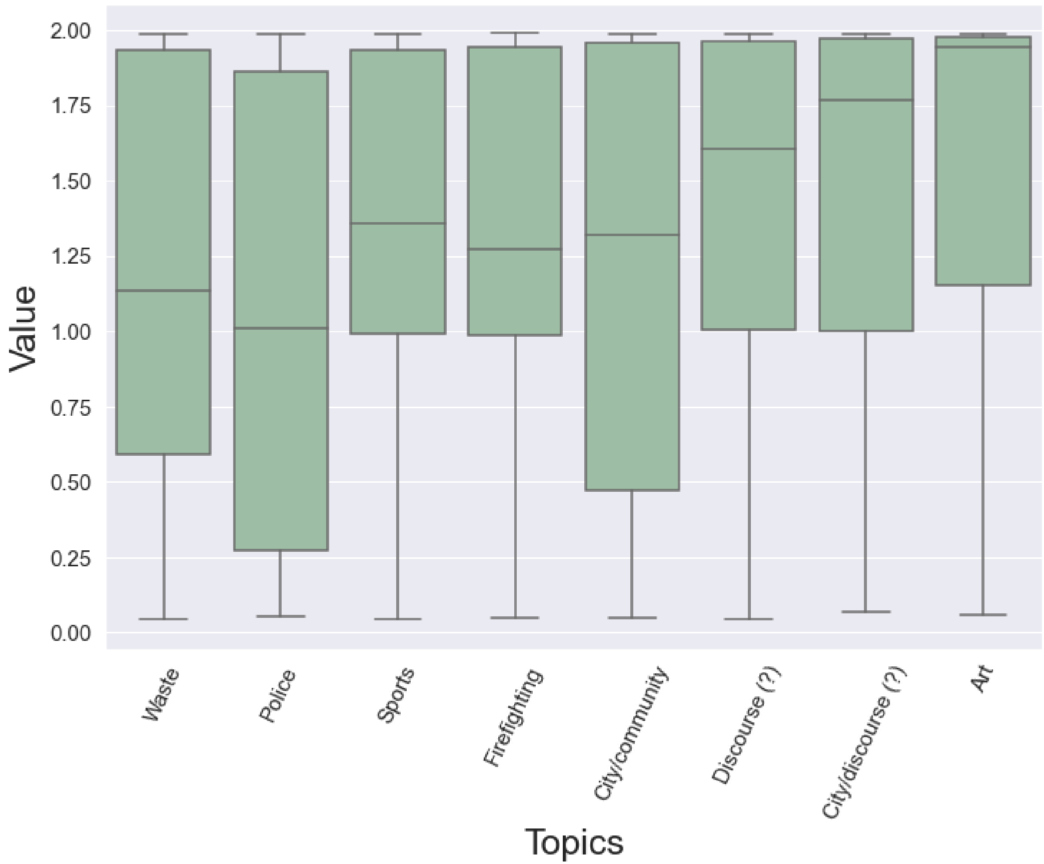
Topics interpretation

- Firefighting
- City/community
- Waste disposal and management
- Art
- Police
- Sports
- Discourse (?)
- City/discourse (?)

Sentiment of transcripts by topic



Sentiment of comments by topic



Mutual love score

For a given channel with n videos, each of which has m comments, the mutual $lovescore$ is the mean of the transcript sentiment times the mean of the comment sentiment:

$$love\ score = \frac{1}{n}\sum_{i}^{n} st_i \cdot \frac{1}{m}\sum_{j}^{m} sc_j$$

- Ranges from 0 (negative) to 4 (postive videos and positive comments)

Lovefest ratio

	Channel	Score
1	TownofRaymondVideos	3.56
2	Summerland Chamber of Commerce	3.29
3	Houston Community College	3.22
4	PlattevilleWISC	3.16
5	Village of Schaumburg	3.13
6	City of Cordova Alaska	3.01
7	Tourism Squamish	2.98
8	Guelph Arboretum	2.94
9	Official Westchester Gov Videos	2.9
10	Glenview Television	2.77

Showing 1 to 100 of 344 entries

Previous 1 2 3 4 Next

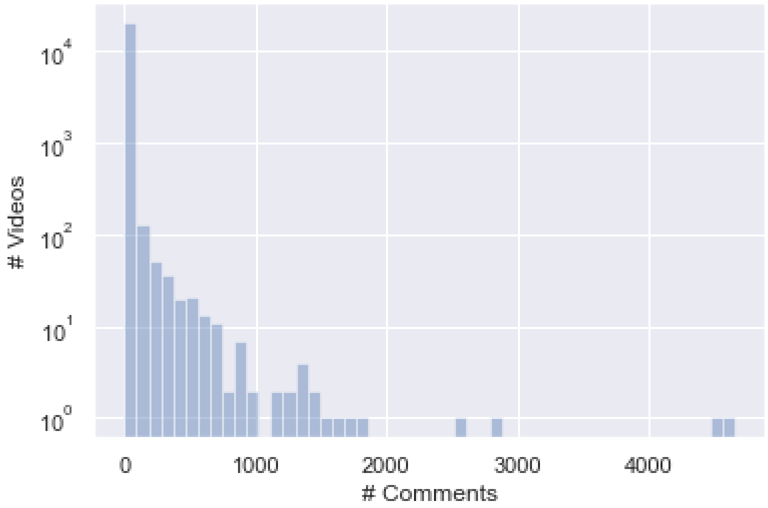
Comment sentiment map

Implications for local government

- Citizen engagement leads to better communities (Gaventa & Barrett 2012)
- More engagement in the form of art/food/outreach videos, fewer police videos?

A few caveats

- Videos of local government not representative of speech in general
- ASR errors (mean WER after filtering ~14%), quality of transcript related to quality of audio as well as dialect features (Tatman 2017; Meyer et al. 2020; Markl & Lai 2021)
  - Low-frequency phenomena: manually inspect corpus hits
  - High-frequency phenomena: signal of correct transcriptions will be stronger (Agarwal et al. 2009) → classifiers
- High variability in discourse in different videos, high variability in number of comments (most few contents)



A few caveats

Sample is small and probably not statistically reliable. Better approach:

- Get channels with many, many videos
- Randomly sample large number of videos
- randomly sample large number of comments

- need bigger datasets (coming...)

- Transformer models (like all-MiniLM-L12-v2) are trained on segmented text with clear boundaries, but transcript text mostly has no punctuation such as periods or commas.

Thank you!

References

Barbieri, Francesco, Jose Camacho-Collados, Leonardo Neves & Luis Espinosa-Anke. 2020. TweetEval: Unified benchmark and comparative evaluation for tweet classification. *arxiv*:2010.12421 [cs.CL].

Bou-Franch, Patricia, Nuria Lorenzo-Dus & Pilar Garcés-Conejos Blitvich. (2012). Social interaction in YouTube text-based polylogues: A study of coherence. *Journal of Computer-mediated Communication* 17, 501–521.

Coats, Steven. (Forthcoming). Dialect corpora from YouTube. In: B. Busse & N. Dumrukcic (eds.). *Proceedings of ICAME.*

Dynel, Marta. (2014). Participation framework underlying YouTube interaction. *Journal of Pragmatics* 73, 37–52.

Gaventa, John & Gregory Barrett. (2012). Mapping the outcomes of citizen engagement. *World Development* 40 (12), 2399–2410.

Goode, Luke, Alexis McCullough & Gelise O'Hare. (2011). Unruly publics and the fourth estate on YouTube. *Participations: Journal of Audience and Reception Studies* 8(2), 594-615.

Grootendorst, Maarten. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*:2203.05794 [cs.CL].

Häring, Mario, Wiebke Loosen & Walid Maalej. (2018). Who is addressed in this comment? automatically classifying meta-comments in news comments. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1-20.

Herring, Susan & Seung Woo Chae. (2021). Prompt-rich CMC on YouTube: To what or to whom do comments respond?. In *Proceedings of the 54th Hawaii International Conference on System Sciences*, 2906-2915.

Ksiazek, Thomas B. (2018). Commenting on the News. *Journalism Studies* 19(5), 650-673.

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer & Veselin Stoyanov (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv*:1907.11692 [cs.CL].

References 2

Loureiro, Daniel, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke & Jose Camacho-Collados. (2022). TimeLMs: Diachronic Language Models from Twitter. arXiv:2202.03829v2 [cs.CL].

Markl, Nina & Catherine Lai. (2021). Context-sensitive evaluation of automatic speech recognition: considering user experience & language variation. In *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing*, 34–40.

Meyer, Josh, Lindy Rauchenstein, Joshua D. Eisenberg & Nicholas Howell. (2020). Artie bias corpus: An open dataset for detecting demographic bias in speech applications. In *Proceedings of the 12th Language Resources and Evaluation Conference*, 6462–6468.

Schultes, Peter, Verena Dorner & Franz Lehner. (2013). Leave a comment! An in-depth analysis of user comments on YouTube. In *11th International Conference on Wirtschaftsinformatik, 27th February – 1st March 2013, Leipzig, Germany*, 659-673.

Siersdorfer, Stefan, Sergiu Chelaru, Jose San Pedro, Ismail Sengor Altingovde, & Wolfgang Nejdl. (2014). Analyzing and mining comments and comment ratings on the social web. *ACM Transactions on the Web* 8 (3), Article 17.

Tatman, Rachel. (2017). Gender and dialect bias in YouTube's automatic captions. In *Proceedings of the First Workshop on Ethics in Natural Language Processing, April 4th, 2017, Valencia, Spain*, 53–59.