RiCL Research in Corpus Linguistics

Commenting on local politics: An analysis of *YouTube* video comments for local government videos

Steven Coats
University of Oulu / Finland

Abstract – This study compares the content of transcripts of videos uploaded by local governments with the comments on those videos, utilizing three transformer-model-based techniques: summarization of the discourse content of video transcripts, topic modeling of summarized transcripts, and sentiment analysis of transcripts and of comments. The analysis shows that some types of video content, for example those dealing with music or education, are more likely to attract positive comments than content related to policing or government meetings. In addition to their potential relevance for local government outreach, the study may represent a viable exploratory method for comparison of online video content and written comments in the context of computational social science analyses of user interaction and commenting behavior.

Keywords – *YouTube*; comments; ASR transcripts; local government; transformer models; topic modeling; sentiment analysis

1. Introduction and background[1]

Comparison of the discourse content of video streams with comments on those streams represents an under-researched topic in studies of Computer-Mediated Communication (henceforth, CMC) and discourse. Over the last two decades, there has been a noticeable transition towards a greater reliance on CMC environments, a shift encompassing various forms of communicative interactions and interactive registers. Notably, civic engagement at the local community level is increasingly conducted online, a tendency facilitated by increased access to the communicative affordances of online platforms and accelerated in the early 2020s by the Covid-19 pandemic. While feedback via CMC provides citizens with a means to express their satisfaction and their concerns about the workings of local government and issues of local importance, online commenting differs from traditional forms of citizen engagement. Comments on video streams or recordings exhibit

[1] The author would like to extend thanks to Finland's *Centre for Scientific Computing* for providing computational resources, and to two anonymous reviewers for their helpful comments.

communicative features that reflect the interactive parameters of the medium as well as aspects of the online userbase in ways that make them difficult to compare with traditional feedback forms. Nevertheless, public comments are important sources of information for local governments and other organizations, and gauging public sentiment towards local government ordinances, initiatives, services, and news/information is an important aspect of responsible and successful governance.

YouTube comments have attracted substantial research attention and, in recent years, their linguistic and interactive properties have been the subject of qualitative, quantitative, and corpus-based analysis from a variety of theoretical and methodological perspectives. Studies of *YouTube* comments have investigated questions of commentator stance and addressee (e.g., Bou-Franch *et al.* 2012; Dynel 2014; Herring and Chae 2021), discourse pragmatic concerns such as impoliteness and flaming (e.g., Andersson 2021; Lehti *et al.* 2016), or the relationship between video content, popularity, and commenting behavior (e.g., Siersdorfer *et al.* 2014; Ksiazek *et al.* 2016), among other topics. However, despite the diversity of approaches, few studies have compared comments specifically with the language content of videos, and *YouTube* channels of governmental organizations have not been a primary focus.

For this study automatic speech recognition (henceforth, ASR) transcripts from the *Corpus of North American Spoken English* (CoNASE; Coats 2023), were assessed in terms of sentiment, and summarized using a transformer model. The summarized transcripts were then assigned to topics using BERTopic (Grootendorst 2022), a suite of topic modeling scripts that utilizes large, context-sensitive transformer models. All available comments for the corresponding videos were then retrieved and assessed in terms of sentiment using the same model as used for transcripts, namely, twitter-roberta-base-sentiment-latest (Camacho-Collados *et al.* 2022), a fine-tuned version of RoBERTa-large (Liu *et al.* 2019). The study represents an exploratory approach to the following research questions:

1) What are the main topical concerns of local government meetings in North America?
2) What is the relationship between topic and the discourse content of transcripts in terms of sentiment?
3) What is the relationship between topic and the discourse content of comments in terms of sentiment?

The study shows that certain video topics, as determined by the summarization-topic modeling procedure, are more likely to represent positive sentiment as well as to attract positive comments. The analysis demonstrates how transformer models can be applied to publicly accessible data in order to assess trends and attitudes in the public sphere, and as such represents a method that may be relevant not only for the study of local governance, but for investigations of many types of online interaction. In terms of civic engagement, the results may help policymakers direct their social media outreach efforts towards the creation of content that is more likely to elicit viewer responses such as commenting and liking. Because communities with engaged citizens are more likely to exhibit positive traits such as increased government accountability or societal inclusiveness (Gaventa and Barrett 2012), engagement represents a desideratum of local government policymakers.

In a broader perspective, the comparison of the discourse content of videos and streams with comment content is relevant for the empirical study of discourse in terms of multimodal communicative pragmatics. The study exemplifies the use of corpus data and transformer models for social research and demonstrates an analytical approach for understanding the relationship between comments and video content. As such, it also represents an example of linguistic data science research at the intersection between language studies, social science, quantitative data analysis, and corpus-based computational sociolinguistics (Schmid 2020; Grieve *et al.* 2023; Coats and Laippala, 2024).

The article is organized as follows: Section 2 discusses some previous research into commenting behavior and comments on *YouTube* videos. Section 3 describes the data used in the study and the methods used to gauge sentiment in the transcripts and comments and assign transcripts to topics. Section 4 presents the largest topics in the transcript material and compares sentiment scores in the transcript material with sentiment scores in comments. In Section 5, the results are discussed and interpreted and several caveats pertaining to the data, methods, and interpretations are noted.

2. Previous research

Commenting behavior in CMC has been extensively studied. Early research investigated communicative and linguistic aspects of comment threads on bulletin boards and as responses to edited texts, such as news articles. More recent studies, a few of which are discussed below, have focused on commenting behavior on image-, video-, or live stream-hosting platforms such as *Instagram*, *YouTube*, *Twitch*, *TikTok*, and others.

2.1. Interactivity, pragmatics, and modeling of comments

Classifications of comments in terms of addressivity patterns and pragmatic functions have been the focus of linguistic studies of commenting behavior, for example, utilizing the theoretical and methodological framework of Conversation Analysis (Bou-Franch *et al.* 2012). Analyses of comment structure and content can be complicated by the, at times, unclear addressivity patterns within comment threads: that is, individual comments can be directed towards page content in general, towards individuals identified in the content on a page, towards other commentors on the page in general, or towards specific users/commentors, among other configurations. A basic distinction can be drawn between comments which are directed to the main content of a page such as the video or news text it presents and comments directed towards other comments, a distinction for which Ksiazek *et al.* (2016) suggest the terms 'user–content interactivity' and 'user–user interactivity'. In addition to different addressivity configurations, comments for some online platforms often make use of emoticons, emoji, and animated graphicons whose semantic and pragmatic values are not always easy to analyze. Herring and Dainas (2017), for example, analyzed the use of graphicons such as emoji and reaction image gifs in a corpus of *Facebook* posts, classifying them into five pragmatic categories. 'Reaction' usages, in which a graphicon is used in a stand-alone manner without accompanying text, were most common, followed by 'tone' usages, in which the images could be interpreted to be modifying the text content of the post.

Predictive modeling has been used to interpret patterns of comments. Häring *et al.* (2018), for example, created two large corpora of German-language comments on news articles and used word embeddings to train a classifier to distinguish between 'non-meta' and 'meta' comments (i.e., comments which address the content of the news article and comments which are directed towards the article author, the publisher, readers on the

news platform, the moderator of the comment space, or others). Ksiazek (2018) analyzed 330,000 comments on almost 2,000 news articles about diverse topics from English-language news websites. After articles were categorized into 25 different topics, content word frequencies from *Linguistic Inquiry and Word Count* (LIWC),[2] a tool which assesses the linguistic and psychological dimensions of text based on aggregate content word counts (Tausczik and Pennebaker 2010), were used to measure the civility and hostility of comments. Using a hierarchical regression model, he found that some news topics, such as the Tea Party, healthcare, and government budgets, were more likely to generate larger numbers of comments overall, whereas others, such as gun control or foreign policy, were more likely to attract negative or hostile comments. Krohn and Weniger (2019) created a model to predict the size of comment threads based on data from *Reddit*, a platform in which much of the content consists of hierarchically arranged user comments. Their model, which included post title, author, and other properties of seed posts, predicted the size and temporal dynamics of comment threads; they report improved results compared to baseline models.

2.2. YouTube comments

Commenting on *YouTube*, which as a platform has been characterized as a kind of mediated quasi-interaction (Bou-Franch *et al.* 2012), can occur with a variety of addressivity configurations (Dynel 2014; Herring and Chae 2021). As of 2023, *YouTube* comment threads have a maximum depth of two. Top-level comments are shown in order of recency or popularity directly under a video; replies to comments are shown indented under top-level comments (see Figure 1).[3]

[2] https://www.liwc.app/

[3] Please note that Figure 1 does not represent a real video or real comments but was created for illustrative purposes.

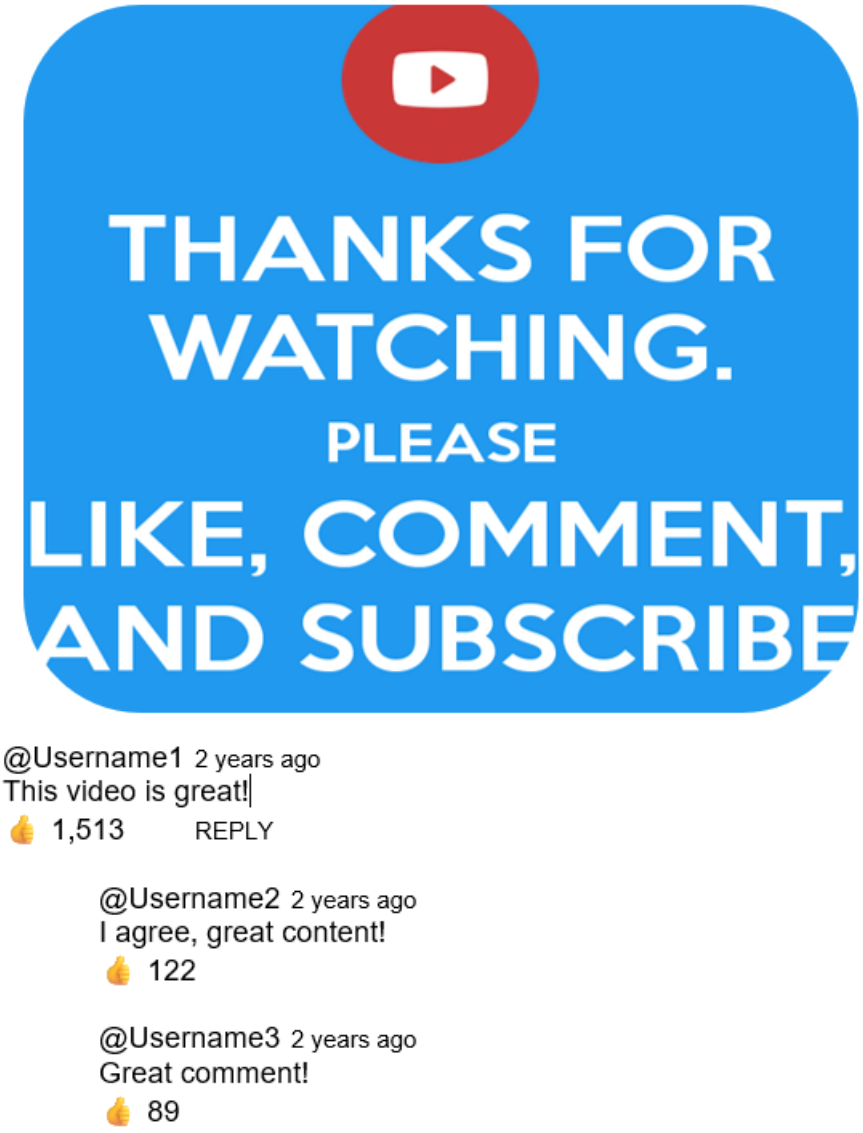

Figure 1: Schematic representation of a *YouTube* video and comment structure

As is the case with studies of other comment-based CMC, analyses of *YouTube* commenting behavior have been undertaken from qualitative and descriptive perspectives, as well as by building predictive models.

Qualitative studies include Goode *et al.* (2011), who analyzed 30 videos in the *YouTube* channels of eight mainstream news outlets. Investigating whether *YouTube* comments on news videos could represent an idealized Habermasian "public sphere" that enables positive civic participation and dialogue, they found that, on the contrary, *YouTube* comment sections tend to be an "unruly" place, characterized by expressions of anger, boredom, or vulgarity, with a "low signal-to-noise ratio" (Goode *et al.* 2011: 611). Bou-Franch *et al.* (2012) hand-coded 300 comments on two *YouTube* videos, comprising almost 12,000 words in total, for a variety of turn-maintenance devices described in

previous CMC research or derived from concepts developed in Conversation Analysis (essentially, whether a comment refers to the immediately preceding comment, to some other comment, or to the video on the page). They classified most comments as "adjacency turns" (Bou *et al*. 2012: 502) which referred to the immediately preceding comment. Lehti *et al.* (2016) described types of impoliteness in the comment thread of a well-known *YouTube* video from 2014. Herring and Chae (2021) discussed addressivity in comment threads on *YouTube*, noting that it is not always obvious to whom a comment is directed. Qualitatively analyzing 200 comments for each of three *YouTube* videos, they found that the largest proportion of comments are free-floating, without a specific addressee. Comments can be directed to speakers in the video, to other commenters on *YouTube*, to the *YouTube* platform, or to speakers in embedded videos, for video clips that include embedded content. Similarly, Cotgrove (2022) compiled a corpus of 3m *YouTube* comments from German-language youth-oriented videos to analyze lexical, grammatical, and discourse features of online youth language.

Quantitative and predictive modeling approaches have also been employed for the study of *YouTube* comments. In Schultes *et al.* (2013), comments for a pseudo-random sample of 304 *YouTube* videos were assigned class labels ('discussion post', i.e., a post containing content directed at another comment/user; 'inferior comment', containing insults, offensive statements, or short, emotional replies; or 'substantial comment', non-offensive comments directed towards the video's content) on the basis of features such as comment length in number of tokens, presence of offensive or emotional words or of emoticons, lexical overlap with the title of the video, and other features. They found that labels generated in this manner could be used to train a classifier to achieve high internal consistency when predicting comment type. In addition, they considered the relationship between these labels and the like/dislike ratio of videos. Discussion post comments were found to be the strongest predictor of likes, while inferior comments were found to better predict dislikes. It should be remarked, however, that *YouTube* comments at the beginning of the 2010s were a wilder place than at the present time, with relatively unsophisticated automatic filters on the platform making it possible to post a wider variety of potentially objectionable content (see, e.g., Nycyk 2012, who provides examples of abusive comments that are no longer encountered on the platform). Siersdorfer *et al.* (2014) considered comments on *YouTube* videos and on *Yahoo News* articles in terms of their aggregate ratings (like/dislike ratios) and how these corresponded to comment textual

content. They found that comments with higher ratings tended to include positive terms such as *love*, *greatest*, or *perfect*, whereas those with low ratings included negative terms such as *retard* or *idiot*.

Khan (2017) used a survey-based method to investigate *YouTube* behaviors. Participants responded to questions about their uploading, liking, disliking, commenting, and sharing activity on *YouTube* on a Likert scale; questions were designed to address a variety of motives such as seeking information, social interaction, or relaxing and seeking entertainment. A regression of survey results showed that the social interaction had the largest coefficients for commenting; information seeking and giving information were also positively correlated with commenting on videos. Andersson (2021) considered impoliteness in comment threads for ten *YouTube* videos with negative words (e.g., *terrible* or *hysterical*) in their titles featuring climate activist Greta Thunberg. Using word2vec on the ~33,000 comments and ~500,000 words, she examined which words were closest to Greta in semantic space, finding that most of these words had negative evaluative content. The results were interpreted as an indication that impoliteness serves to consolidate similar views.

Overall, although several studies have considered the addressivity and interaction patterns of comments or used quantitative and predictive methods to explore aspects such as the relationship between metadata fields, few studies have compared the spoken discourse of videos and the discourse of the comments thereupon. In the next section, the methods used to evaluate the content and sentiment of videos as well as the comments on those videos are described.

3. DATA AND METHODS

The starting point for the analysis was transcripts of videos indexed in CoNASE (see Section 1), a 1.3-billion-word corpus of ASR transcripts of videos uploaded to the *YouTube* channels of municipalities and other local government entities in the US and Canada.[4] Much of the content of CoNASE consists of transcripts of public meetings of local councils in which local government and community issues are discussed, but other

[4] https://cc.oulu.fi/~scoats/CoNASE.html

content types, including interviews, sporting events, performances, and news reports are included.

In this study, only those videos which had comments in CoNASE were considered. Assessment of the content of the transcripts and the comments on the corresponding videos, summarization of the transcripts, and topic modeling of transcript content were undertaken in four principal steps, schematically illustrated in Figure 2.

First, after retrieval of all available comments, a sentiment score was calculated for each transcript and for each comment using the twitter-roberta-base-sentiment-latest transformer model (Camacho-Collados *et al.* 2022; Loureiro *et al.* 2022). Next, the transcripts were summarized into short texts, ranging in length from one to ten short paragraphs, using the distilbart-cnn-12-6-samsun model (Schmid 2023). This step was undertaken to create more coherent topics (see below). Topic modeling was then undertaken on the summarized content, using the BERTopic library (Grootendorst 2022). Finally, the sentiment scores, as values along a cline negativity-neutrality-positivity, were analyzed for the eight largest topics in terms of transcript and comment sentiment.

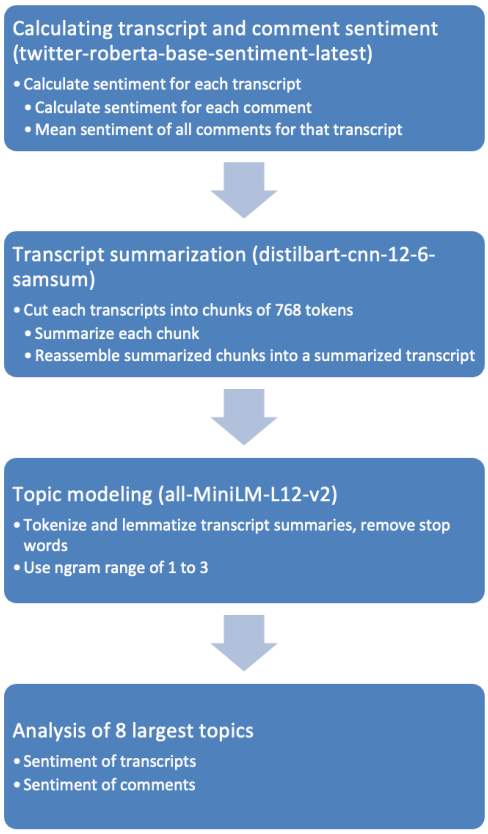

Figure 2: Schematic illustration of the processing and analysis steps. Transformer models are shown in parentheses

3.1. Transcript and comment retrieval and processing

The open-source *YouTube*-comment-downloader (Bouman 2022) was used to retrieve comments, via the innertube API, from videos whose transcripts are available in CoNASE. The vast majority of the videos in CoNASE have no user comments (and very few views), a fact which is unsurprising, considering the predictable nature of local government meetings and other municipal channel content. In total, of the 301,846 videos indexed in CoNASE, comments could be retrieved for 20,965. In addition, a small number of videos had been removed or made private (i.e., the comments were not available) in the time between the collection of the CoNASE data (2017–2021) and the time the comments were downloaded (mid-2022). The 190,097 downloaded comments ranged in length from 1 to 2,010 word tokens, with a mean value of slightly over 28 tokens.

3.2. Sentiment analysis

Sentiment analysis assigns negative, neutral, or positive sentiment to texts. Older, bag-of-words models, in which texts are assigned a value based on aggregate scores for individual lexical items, can perform poorly due to word order and contextual factors. A negative evaluation such as *he said it was great, wonderful, and fantastic, but it is really terrible* may be assigned a positive value based on the presences of three items with positive values and one item with a negative value. Similarly, language transformer models are typically better able to disambiguate the meanings of homonyms, determine the scope of negators, and correctly represent pronominal deixis due to their sensitivity to word-order and contextual considerations. A number of transformer-based sentiment analysis packages exist for text classification, but as *YouTube* comments tend to be rich in emoji, an analysis pipeline sensitive to emoji was selected, namely, the twitter-roberta-base-sentiment-latest transformer model (Camacho-Collados *et al.* 2022),[5] a fine-tuned sentiment model trained on 124m tweets (Loureiro *et al.* 2022), ultimately based on the RoBERTa pretraining approach (Liu *et al.* 2019).

While this model was appropriate for most of the comments in the data, which tend to be shorter in length, video transcripts are often much longer than the maximum input length for BERT models (often 512 tokens). An iterative procedure was therefore developed for texts longer than 512 tokens. They were converted to chunks of 512 tokens,

[5] https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest

then each chunk fed to the transformer model. The mean of the output vector values was then taken to be the sentiment for the entire text. The twitter-roberta-base-sentiment-latest model generates a vector of the likelihood of a given input being categorized as negative, neutral, or positive, outputting the argmax as a discrete value 0 (negative), 1 (neutral), or 2 (positive). This vector was converted directly to a continuous value in the range of 0 to 2 by calculating the dot product of the weighted values. For example, a model output of [0, .45, .55] indicates that according to the model, the text has zero percent probability of being negative, a 45 percent probability of being neutral, and a 55 percent probability of being positive. This corresponds to a score of 1.55, or mostly positive. The procedure was used to assign sentiment scores to all transcripts and comments in the study. An example is the video with the YouTube ID a1WSkvlw7zQ, entitled 'Meridian's new "Storey Bark Park"', uploaded by the YouTube channel of the City of Meridian, Idaho. The short video (1m 35s in length) consists of footage from the opening of a new community dog park, with a voice-over providing information on the event and about the park, and a speaker in the video making remarks at the opening ceremony. The calculated sentiment value for the transcript, which records a celebratory event, is 1.98, based on the overwhelmingly positive evaluative terms in the transcript (e.g., *huge success*, *enjoying the park*, *celebrate*). The single comment for the video is also positive and reads *Hands down the best dog park in the Treasure valley*, which was also assigned a value of 1.98.

3.3. Transcript summarization

Initial experimentation with topic modeling using the raw ASR transcripts produced inconsistent results. The BERTopic pipeline, which converts textual content to vector representations, is designed to process text with standard sentencization (i.e., periods or other sentence-ending punctuation) and is optimized for short texts such as sentences or paragraphs. The CoNASE transcripts with viewer comments which are analyzed in this study, however, have no sentence-ending punctuation, and vary greatly in length, from 100 to almost 80,000 tokens. To improve the quality of topics, a summarization step was undertaken for the video transcripts, using a recent transformer pipeline trained on transcripts of conversational speech (distilbart-cnn-12-6-samsum).[6] The model, based on

[6] https://huggingface.co/philschmid/distilbart-cnn-12-6-samsum

the BART architecture (Lewis *et al.* 2019), captures the essential discourse content of longer text passages and recapitulates it as short paragraphs.

First, transcripts were tokenized using *spaCy* (Honnibal *et al.* 2020) and split into 768-token chunks for summarization. The output for each transcript, consisting of 40-token summaries of the 768-token chunks, was then aggregated to generate the full summary for each transcript. The procedure reduced the variability in the length of the transcripts and introduced standard punctuation conventions. The resulting short texts, which retained the essential content of the longer transcripts, produced consistent topics, which upon manual inspection were found to correspond to most of the video content in the underlying video. For example, the video FUXTWgIqSfQ, entitled '"Sounds of Christmas" Christmas Band Concert', is a 47-minute recording of a school band performance. The transcript of the video, which is 1,649 tokens long, mainly consists of comments made by the band conductor. It comprises words of welcome and introduction to the audience, expressions of thanks to colleagues, parents, pupils, and band musicians, and introductions to each piece being performed. The summarized content of the video, which is 120 tokens long, foregoes expressions of welcome and thanks, beginning *The Bruton middle school intermediate band is playing the Nutcracker at the Bruton Christmas concert tonight.*

3.4. Topic modeling

Topic modeling (Blei *et al.* 2003) is an approach for the automatic identification of co-occurring word patterns, or topics, in sets of texts, which themselves can be defined in terms of the extent to which they participate in each topic. The technique, which can be considered a dimensionality reduction procedure, can be useful for the classification and interpretation of large sets of documents by distilling them into semantically interpretable topics. The default topic modeling approach utilizes relative word frequencies or term frequency-inverse document frequency values as input parameters for the algorithm. Traditionally, topic modeling is undertaken using 'bag-of-words' approaches based on word frequencies. While these can generate good results, they fail to account for sentence context. Transformer models such as BERT (Devlin *et al.* 2019), in which individual lexical items as well as immediate collocational contexts are represented by embeddings, or distributed vectors of numerical values, have been shown to be useful for a wide range of language processing tasks, including more robust topic modeling. This study utilized

BERTopic (Grootendorst 2022) for topic modeling. CoNASE transcripts were first summarized, as described above. Then, topic modeling was undertaken with all-MiniLM-L12-v2,[7] a model derived from miniLM (Wang *et al.* 2020), trained on 1.7 billion words of web texts from various genres and designed to map sentences and paragraphs to a multidimensional vector space for tasks like clustering or semantic search.

4. Results and analysis

4.1. Topics

The main input argument to the BERTopic algorithm is an array of textual content, in this case, a list of the 20,965 transcript summaries generated according to the procedure described above. In addition, the user can specify the underlying transformer model, the text tokenization procedure, the dimensionality reduction method, the words to be ignored (stopwords), and many other settings and parameters. For this analysis, tokenization was undertaken with the default CountVectorizer from scikit-learn (Pedregosa *et al.* 2011) and the English stopwords from NLTK (Bird *et al.* 2009). The output of the algorithm is the model, which can be inspected and visualized in many ways. One way to interpret the resulting topic model is to inspect the words which are most strongly associated with the topics in the model.

The eight largest topics, shown in Figure 3, represent the kinds of discourse that is typical for CoNASE transcripts, a large proportion of which are records of public meetings. The largest topic, Topic 0, is related to fire and rescue, services that are typically organized and funded by municipal governments in the United States. The provision of these crucial services often accounts for a considerable proportion of local government budgets, and discussion of, for example, hiring firefighters or the purchase of new equipment such as vehicles is a common discourse element in local government meetings. The words with the highest representation values in the topic include *firefighter*, *rescue*, *station*, and *department*, the latter two of which are likely collocates of *fire*, and *metro*, a term often appearing in the official names of municipal fire departments.

[7] https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2

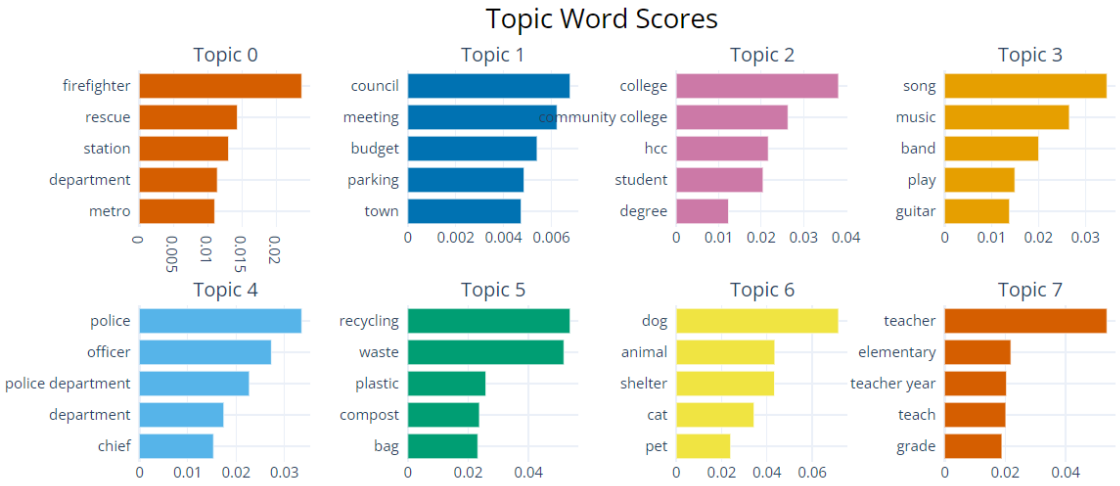

Figure 3: Words most strongly associated with the eight largest topics

Topic 1 includes words strongly associated with the content of municipal meetings: *council*, *meeting*, and *town* denote the activity of the municipal body itself, while *budget* and *parking* represent issues that are typical concerns of municipal governments. The items with the highest values in Topic 2 are from discourse pertaining to tertiary education: *college* and *community college* are places where the *student* can receive a *degree*. *Hcc*, in this context, is an initialism used to refer to several community colleges referenced in the discourse of CoNASE transcripts, including Houston Community College, Texas. In the United States, community colleges, which typically offer 2-year degrees, are often subsidized by municipalities. The videos from which the transcripts in this topic were taken include promotional content and interviews with community college presidents and staff members.

Topic 3 pertains to music. In the CoNASE corpus, it corresponds mostly to video transcripts of news announcements of upcoming musical performances, and occasionally of the performances themselves, for example those held during holiday or commemorative events, as well as performances organized by schools, universities, and other local organizations. The words most strongly linked to this topic denote music, instruments, and those who create music. Topic 4 represents another vital service of local governments. As is the case with fire and rescue, in the United States, most municipalities maintain a local police force and use local tax revenues for hiring and staffing the force as well as for procuring equipment such as uniforms and vehicles. The words in this topic denote police officers and the head of the police force, the *chief*.

Topic 5 deals with waste management, another service organized mostly by local governments. In addition to the words *trash* and *waste*, the words most strongly associated with this topic include *recycling*, *plastic*, and *compost*, indicating a concern for the environmental consequences of municipal waste and a desire to implement greener waste management policies. Topic 6 pertains to animals, as indicated by the words *animal*, *dog*, *cat*, and *pet*. The discourse for this topic relates to another service typically provided by municipalities in the United States and funded by local taxes, namely, animal control services, or the provision of facilities (in the form of a *shelter*) for stray and abandoned pets. Videos in the corpus with this topic include many in which animals at a shelter are introduced and offered for adoption. Topic 7 includes words used to discuss primary education, such as *teacher*, *teach*, and *grade*; *teacher year*, a word used in budgeting to describe the working hours of schoolteachers, and *elementary*, likely as a collocate of *school*. Primary education, in the US, is organized by municipalities and is therefore a frequent subject of discussion in municipal government meetings. In addition, the transcripts in this topic include content produced by school districts and schools themselves.

Overall, seven of the eight largest topics represent discourse that clearly pertains to local government decision-making: *firefighting*, *meetings*, *community colleges*, *police*, *waste disposal*, *animal control*, and *primary education*. These topics correspond to services that are provided at the local level by most municipalities in the US and Canada and whose concrete forms and budget allocations are the subject of much discussion by government representatives. The topic modeling procedure therefore accurately captures the fact that videos uploaded to municipal government channels are mostly about the immediate concerns of local governments, as captured in the discourse content of government meetings. In the next subsection, the sentiment expressed in those meetings, as well as in comments on the *YouTube* pages hosting those videos, are examined.

4.2. Sentiment

Both the transcripts and the comments in the data are more positive than negative, corresponding to the expected pattern for the sentiment of public discourse: communicators, in general, tend to accentuate positive sentiment and avoid expression of negative sentiment (Dodds *et al.* 2015). For this data, the mean sentiment value for transcripts was 1.20; for comments 1.29.

The distribution of sentiment values for transcripts and comments in Figure 4 shows peaks for comment sentiment near 0, 1, and 2. These peaks correspond to very short (mostly single-word or single-emoji) comments that are assigned a discrete value by the algorithm with a high probability score. Thus, a comment such as *great!!!* is assigned a value of 2 (positive), with 98 percent likelihood, whereas a comment such as *terrible!!!* or 🤮 would be assigned a value of 0 (negative) with high probability. As shown in Figure 4, single-token comments expressing positive sentiment are more common than neutral or negative single-token comments.

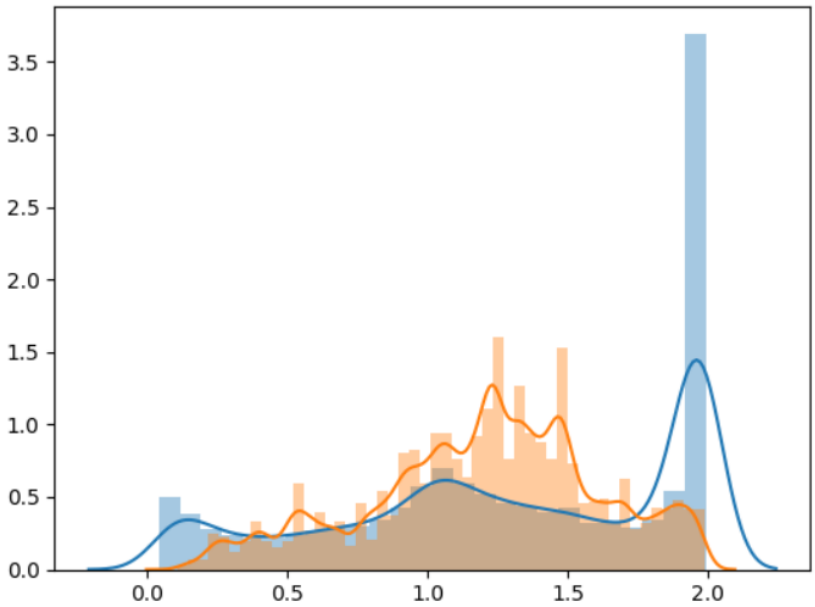

Figure 4: Distribution of sentiment values for transcripts (orange) and comments (blue)

The sentiment expressed in the video transcripts varies between the topics. Figure 5 depicts sentiment values for the eight largest topics in the transcript data. The median sentiment values for the topics, calculated on the basis of all the videos in the data assigned to that topic, range from 1.14, for the topic *meetings*, to 1.87, for the topic *school*. The sentiments expressed in the videos assigned to the topics *waste*, *firefighting*, and *police* have slightly lower median sentiment values of 1.23, 1.28, and 1.52. The topics *animal control*, *music*, and *community college* have higher median values: 1.54, 1.63, and 1.67.

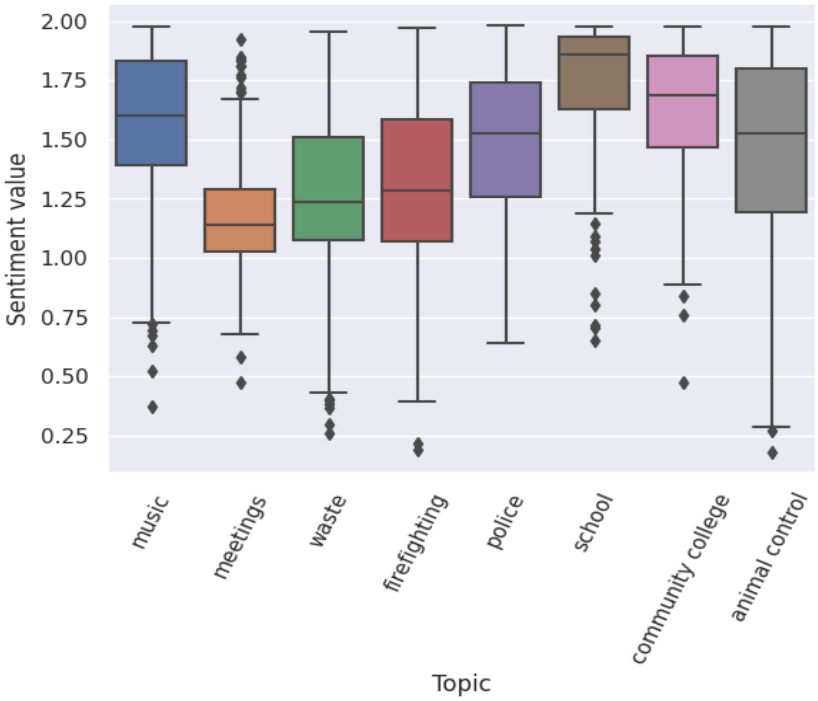

Figure 5: Transcript/summary sentiment for the eight largest topics

Comment sentiment, calculated as the mean for individual videos, tends to recapitulate the sentiment of the transcript summaries (Figure 6). For comments, median values per topic range from 0.93, for *meetings*, to 1.76, for *music*. The topics *police*, *waste*, and *firefighting* have median values of 1.08, 1.22, and 1.30, and the topics *school*, *animal control*, and *community college* median values of 1.50, 1.64, and 1.76.

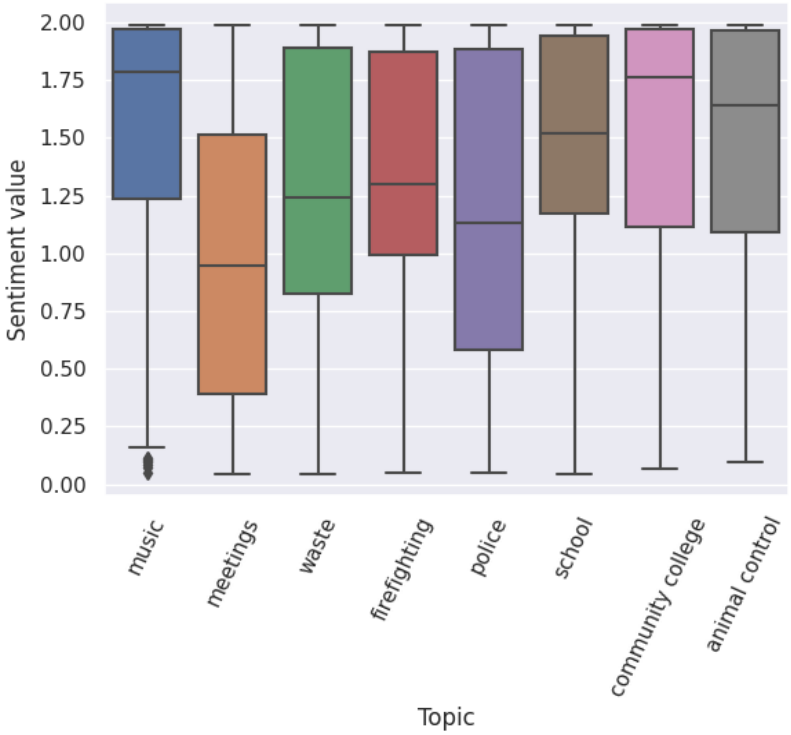

Figure 6: Comment sentiment for the eight largest topics

The difference between median sentiment values for the topics in terms of transcript content and aggregate comments may provide insight into the general contours of public perception of local government activities in the US and Canada.

Figure 7 shows the differences, per topic, between median comment sentiment value and median transcript sentiment values. Here, the topics that resonate positively with local communities become apparent: *music*, which as a topic exhibits positive sentiment on the basis of the transcript content, tends to attract comments that are even more positive. Likewise, transcripts with the topics *community college* and *animal control* attract comments that are more positive than the (already positive) sentiment contained in the transcript discourse.

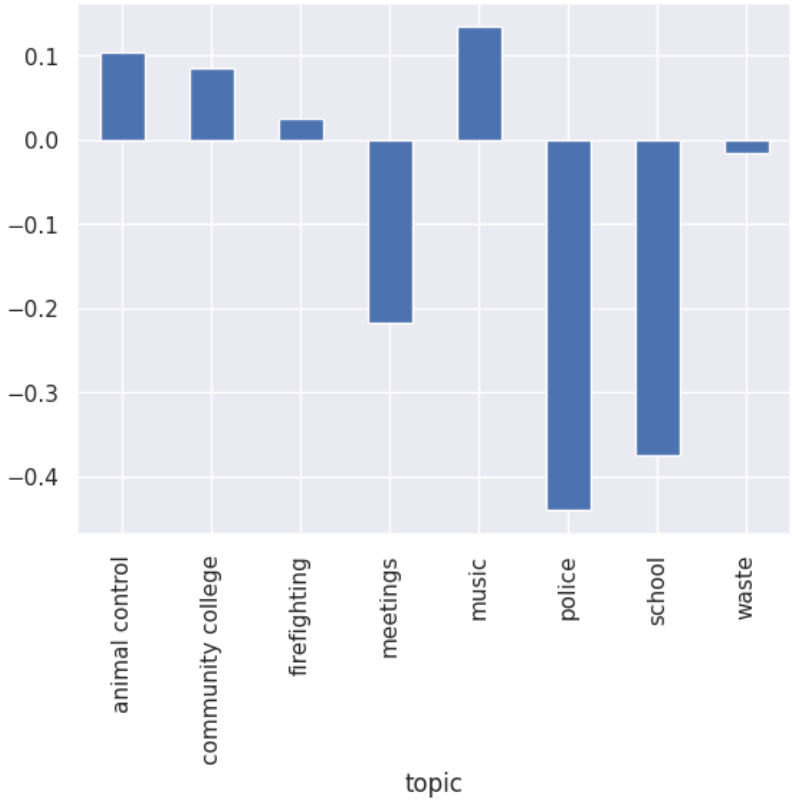

Figure 7: Difference in comment and transcript sentiment, per topic

The topics *firefighting* and *waste* attract comments that are approximately equivalent, in terms of median values, with the transcript content for those topics. Comments on *firefighting* are slightly more positive than the corresponding transcripts, while comments on videos with the topic *waste* are slightly more negative.

A different picture emerges for the topics *meetings*, *school*, and *police*. Here, the median sentiment of comments is significantly lower than the median sentiment for the videos. It is likely that the transcripts in the topic meetings are videos of local government

council meetings in which political decisions are discussed and debated, whereas transcripts in topics such as *music* or *animal control* include recordings of performances and informational videos about local organizations such as orchestras and animal adoption centers. The former type of video represents an interactive situation, both in the council chamber and in the comments section on the video's page, where critical and negative sentiments are more likely to receive expression. Disagreement is an important part of the political process, and council discussions are more likely to attract viewers who are critical of local government policies than are musical performances. Transcripts dealing with music and animals are more often informational, rather than discussion oriented. Videos showing musical performances or animals up for adoption are less likely to be criticized or discussed in a negative manner, not only because of their content, but because they are possibly less appropriate venues for the airing of disagreement.

Comment sentiment for the topic *school* is more negative than transcript sentiment, likely due to negative comments by pupils and parents. Comments on videos from this topic include remarks such as *School lunches suck* or *It's my opinion that [teacher name] gives way too much homework*, among others. While the topics *community college* and *school* both deal with education, school is sometimes perceived by pupils to be a burden imposed on them, against which *YouTube* comments may provide an opportunity for protest. Students at community colleges, on the other hand, choose to enroll in the college and usually must pay tuition fees, factors which may make them less likely to post negative comments.

The discourse pertaining to the topic *police* is also substantially more negative than the corresponding transcript material. Comments on the topic include general expressions of negative sentiment towards policing (*fuck the police*), as well as concern over overzealous and unprofessional policing practices, the awareness of which has increased in the last decade in the United States (comments for the topic include *unfortunately, lack of common sense and other far more disturbing behaviors with police officers seem to be commonplace* and *I beg you stop racial profiling it's evil and wrong racial profiling almost killed me*, among others). Commenting practices on videos pertaining to this topic appear to portray awareness of the fact that at community level, the practices and policies of many police forces in the US show room for improvement.

5. Discussion, outlook, and conclusion

It is perhaps unsurprising that popular sentiment is more positive towards topics such as *music*, *higher education*, or *pets*, compared to topics such as *waste management*, *meetings*, or *policing*. Higher education and music are universally acknowledged to be worthwhile and noble expressions of culture, and pets are objects of our love and affection. Furthermore, these topics represent areas where people can exercise agency: we choose to attend or view performances of music, to pursue higher education, and to own pets. Waste management, and policing, in contrast, are mostly perceived as external forces over which we have little influence, and which, in some cases, can be associated with unpleasant sensations, in the case of waste, or potentially dangerous situations, in the case of encounters with unprofessional police.

Although the automated methods of transformer-based sentiment analysis utilized in this study for gauging public attitudes may be new, they essentially recapitulate observations of previous generations towards music, perhaps most succinctly expressed by the English philosopher Herbert Spencer in 1854: “music must take rank as the highest of the fine arts—as the one which, more than any other, ministers to human welfare” (Spencer 2015 [1854]: 33).

The role of music as an uplifting and inspiring aspect of human existence, evident even in comments on *YouTube* channels of local governments, may have practical implications for the community outreach and engagement activities of municipalities. Local governments may be able to increase positive engagement with administrations by including content in their social media channels that reflects future-oriented aspects of communal life, such as education, music, and animals. In a broader perspective, the study represents an example of how transformer-based pipelines for text processing, including summarization, sentiment analysis, and topic modeling, can be used, in concert with ASR, to automatically gauge and assess aspects of communication and discourse.

Several caveats, however, should be noted, pertaining to the underlying data as well as the methods of analysis. The transcripts in the corpus contain ASR errors, with a mean Word Error Rate (WER) of approximately 15 percent (Coats 2024). Quality of ASR transcripts is influenced by both acoustic and dialect features, as highlighted in studies by Tatman (2017), Meyer *et al.* (2020), and Markl and Lai (2021). For this study, the sentiment analysis and summarization steps undertaken for the ASR transcripts ultimately rely on aggregate frequencies of word and n-gram types, as well as contexts. Although

inaccurate input data containing ASR errors may affect the precision of the results of these steps, it is unlikely to misrepresent overall trends in the data as long as the majority of automatically transcribed lexical items correspond to the correct types (see e.g., Agarwal *et al.* 2007).

The summarization step, in which long, unpunctuated ASR transcripts were converted to short paragraphs with standard punctuation, has not been validated for this kind of content (error-containing ASR transcripts). A validation of the accuracy of the summarization output for ASR transcripts would make the findings of the study more robust.

The topics generated by the BERTopic model are subject to a large number of variable input parameters, including the tokenization and lemmatization procedures for the text input, the underlying transformer (or other) architecture used to represent the input as numerical values, the algorithms for dimensionality reduction, as well as other parameters. Experimentations with various configurations of parameters showed that most input parameter settings resulted in the same largest topics. Nevertheless, the extent to which parameter variability can affect the model output, and hence the ensuing analysis, has not been assessed in this study.

The commenting behavior in the sample is not consistent. Some videos exhibit a very large number of comments, but most videos have just a few or one comment. A few comments are longer, in terms of number of tokens, but most comments are very short. This variability undoubtedly has an effect on the sentiment scores for the topics, and the significance of the calculated sentiment values has not been estimated. The method of comparing ASR transcript discourse with comment discourse, demonstrated in this study, may be better validated by selecting channels or videos with large numbers of comments. In addition, random sampling techniques for both videos and comments could help to demonstrate the relationship between transcript and comment content more robustly. In this respect —and considering the fact that municipal channel videos (such as those in CoNASE) typically have few comments, future studies, which do not necessarily need to consider engagement with local government— could target highly popular channels with extensive comments.

From a technical perspective, a few caveats should be remarked pertaining to the language models themselves. The twitter-roberta-base-sentiment-latest model, used to calculate comparable sentiment scores for transcripts and comments, was trained on

tweets, rather than on long, unpunctuated ASR transcripts. The accuracy of the model in predicting sentiment for longer texts remains unvalidated.

Despite these caveats, the study has demonstrated that large transformer models can be used in the context of computational social science for discovering the topical content of streamed or recorded meetings and for investigating the sentiment expressed therein, as well as for gauging the sentiment of comments on recordings of those meetings. While this finding has implications for media outreach for municipal governments or other kinds of organizations, the methods used in the study are not limited to analyzing organizational discourse. The potential utility of transformer models for research into communication and online interaction practices in general is great, and the comparison of speech content with commenting practices represents just the tip of the iceberg.

References

Agarwal, Sumeet, Shantanu Godbole, Diwakar Punjani and Shourya Roy. 2007. How much noise is too much: A study in automatic text classification. In Naren Ramakrishnan, Osmar R. Zaïane, Yong Shi, Christopher W. Clifton and Xindong Wu eds. In Naren Ramakrishnan, Osmar R. Zaïane, Yong Shi, Christopher W. Clifton and Xindong Wu eds. *Proceedings of the Seventh IEEE International Conference on Data Mining.* Los Alamitos; IEEE Computer Society. https://doi.org/10.1109/ICDM.2007.21

Andersson, Marta. 2021. The climate of climate change: Impoliteness as a hallmark of homophily in YouTube comment threads on Greta Thunberg's environmental activism. *Journal of Pragmatics* 178: 93–107.

Bird, Steven, Edward Loper and Ewan Klein. 2009. *Natural Language Processing with Python*. Beijing: O'Reilly Media.

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3: 993–1022.

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

Bouman, Egbert. 2022. *YouTube-Comment-Downloader*. https://github.com/egbertbouman/*YouTube*-comment-downloader

Camacho-Collados, Jose, Kiamehr Rezaee, Talayeh Riahi, Asahi Ushio, Daniel Loureiro, Dimosthenis Antypas, Joanne Boisson, Luis Espinosa Anke, Fangyu Liu and Eugenio Martínez Cámara. 2022. TweetNLP: Cutting-edge natural language processing for social media. In Wanxiang Che and Ekaterina Shutova eds. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Abu Dhabi: Association for Computational Linguistics, 38–49.

Coats, Steven. 2023. Dialect corpora from YouTube. In Beatrix Busse, Nina Dumrukcic and Ingo Kleiber eds. *Language and Linguistics in a Complex World*. Berlin: De Gruyter, 79–102.

Coats, Steven. 2024. Noisy data: Using automatic speech recognition transcripts for linguistic research. In Steven Coats and Veronika Laippala eds. *Linguistics Across Disciplinary Borders: The March of Data*. London: Bloomsbury Academic, 17–39.

Coats, Steven and Veronika Laippala eds. 2024. *Linguistics across Disciplinary Borders: The March of Data.* London: Bloomsbury Academic.

Cotgrove, Louis A. 2022. *#GlockeAktiv: A Corpus Linguistic Study of German Youth Language on YouTube*. Nottingham: University of Nottingham dissertation.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran and Thamar Solorio eds. *Proceedings of 2019 Conference of the North American Association for Computational Linguistics: Human Language Technologies.* Minneapolis: Association for Computational Linguistics, 4171–4186.

Dodds, Peter Sheridan, Eric M. Clark, Suma Desu and Christopher M. Danforth. 2015. Human language reveals a universal positivity bias. *PNAS* 112/8: 2389–2394.

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

Gaventa, John and Gregory Barrett. 2012. Mapping the outcomes of citizen engagement. *World Development* 40: 2399–2410.

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

Grieve, Jack, Dirk Hovy, David Jurgens, Tyler S. Kendall, Dong Nguyen, James N. Stanford and Meghan Sumner eds. 2023. *Computational Sociolinguistics*. Lausanne: Frontiers Media. https://doi.org/10.3389/978-2-8325-1760-4

Grootendorst, Maarten. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv*: 2203.05794 [cs.CL]. https://doi.org/10.48550/arXiv.2203.05794

Häring, Mario, Wiebke Loosen and Walid Maalej. 2018. Who is addressed in this comment? Automatically classifying meta-comments in news comments. In Karrie Karahalios, Andrés Monroy-Hernández, Airi Lampinen and Geraldine Firzpatrick eds. *Proceedings of the ACM on Human-Computer Interaction*. New York: Association for Computing Machinery, 1–20.

Herring, Susan and Ashley R. Dainas. 2017. "Nice picture comment!" Graphicons in Facebook comment threads. In Tung X. Bui and Ralph Jr. Sprague eds. *Proceedings of the 50th Hawaii International Conference on System Sciences*. Hawai: University of Hawaii at Manoa, 2185–2194.

Herring, Susan and Seung Woo Chae. 2021. Prompt-rich CMC on YouTube: To what or to whom do comments respond? In Dan Suthers and Ravi Vatrapu eds. *Proceedings of the 54th Hawaii International Conference on System Sciences*. Hawai: University of Hawaii at Manoa, 2906–2915.

Honnibal, Matthew, Ines Montani, Sofie Van Landeghem and Adriane Boyd. 2020. *spaCy: Industrial-strength Natural Language Processing in Python*. https://doi.org/10.5281/zenodo.1212303

Khan, M. Laeeq. 2017. Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior* 66: 236–247.

Krohn, Rachel and Tim Weninger. 2019. Modeling online comment threads from their start. *arXiv*: 1910.08575v1 [cs.SI]. https://doi.org/10.48550/arXiv.1910.08575

Ksiazek, Thomas B. 2018. Commenting on the news. *Journalism Studies* 19/5: 650–673.

Ksiazek, Thomas B., Limor Peer and Kevin Lessard. 2016. User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media & Society* 18/3: 502–520.

Lehti, Lotta, Johanna Isosävi, Veronika Laippala and Matti Luotolahti. 2016. Linguistic analysis of online conflicts: A case study of flaming in the Smokahontas comment thread on YouTube. *Wider Screen* 19. http://widerscreen.fi/numerot/2016-1-2/linguistic-anaead-on-*YouTube*/

Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov and Luke Zettlemoyer. 2019. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv*: 1910.13461 [cs.CL]. https://doi.org/10.48550/arXiv.1910.13461

Liu, Yinhan, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized BERT pretraining approach. *arXiv*: 1907.11692 [cs.CL]. https://doi.org/10.48550/arXiv.1907.11692

Loureiro, Daniel, Francesco Barbieri, Leonardo Neves, Luis Espinosa Anke and Jose Camacho-Collados. 2022. TimeLMs: Diachronic language models from Twitter. *arXiv*: 2202.03829v2 [cs.CL]. https://doi.org/10.48550/arXiv.2202.03829

Markl, Nina and Catherine Lai. 2021. Context-sensitive evaluation of automatic speech recognition: considering user experience & language variation. In Su Lin Blodgett, Michael Madaio, Brendan O'Connor, Hanna Wallach and Qian Yang eds. *Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing* . Association for Computational Linguistics, 34–40. https://aclanthology.org/2021.hcinlp-1.6

Meyer, Josh, Lindy Rauchenstein, Joshua D. Eisenberg and Nicholas Howell. 2020. Artie bias corpus: An open dataset for detecting demographic bias in speech applications. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk and Stelios Piperidis eds. *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille: European Language Resources Association, 6462–6468.

Nycyk, Michael. 2012. *Tensions in Enforcing YouTube Community Guidelines: The Challenge of Regulating Users' Flaming Comments*. Perth, Australia: Curtin University of Technology dissertation.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and Édouard Duchesnay. 2011. Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* 12: 2825–2830.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.

Schmid, Phillip. 2023. *Distilbart-cnn-12-6-samsum*. https://huggingface.co/philschmid/distilbart-cnn-12-6-samsum

Schultes, Peter, Verena Dorner and Franz Lehner. 2013. Leave a comment! An in-depth analysis of user comments on YouTube. In Rainer Alt and Bogdan Franczyk eds. *Proceedings of the 11th International Conference on Wirtschaftsinformatik*. Leipzig: University of Leipzig, 659–673.

Siersdorfer, Stefan, Sergiu Chelaru, Jose San Pedro, Ismail Sengor Altingovde and Wolfgang Nejdl. 2014. Analyzing and mining comments and comment ratings on the social web. *ACM Transactions on the Web* 8/3: 1–39

Spencer, Herbert. 2015 [1854]. The origin and function of music. In John Shepherd and Kyle Devine eds. *The Routledge Reader on the Sociology of Music*. London: Routledge, 27–34.

Tatman, Rachel. 2017. Gender and dialect bias in YouTube's automatic captions. In Dirk Hovy, Shannon Spruit, Margaret Mitchell, Emily M. Bender, Michael Strube and Hanna Wallach eds. *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*. Valencia: Association for Computational Linguistics, 53–59.

Tausczik, Yla R. and James W. Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29/1: 24–54.

Wang, Wenhui, Furu Wei, Li Dong, Hangbo Bao, Nan Yang and Ming Zhou. 2020. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. *Advances in Neural Information Processing Systems* 33: 5776–5788.

Corresponding author
Steven Coats
University of Oulu
Faculty of Humanities
Pentti Kaiteran katu 1
Linnanmaa
P.O. Box 8000.
90014 Oulu
Finland
E-mail: steven.coats@oulu.fi

received: November 2023
accepted: February 2024