# “There are no pronouns in the Constitution”

## A political and sentiment analysis of pronoun discourse in social media


**Michael Dow & Kelly Biers**

*Université de Montréal & University of North Carolina Asheville*


October 18, 2024

# Roadmap

# Introduction

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

  b. I'm agender and would like to only be referred to by my name, and no pronouns.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

  b. I'm agender and would like to only be referred to by my name, and no pronouns.

  c. I don't care what any liberal says, I'm not dating a girl with pronouns.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

  b. I'm agender and would like to only be referred to by my name, and no pronouns.

  c. I don't care what any liberal says, I'm not dating a girl with pronouns.

  d. Using pronouns is basic human decency.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

  b. I'm agender and would like to only be referred to by my name, and no pronouns.

  c. I don't care what any liberal says, I'm not dating a girl with pronouns.

  d. Using pronouns is basic human decency.

- Statements **(a, b)** show a disconnect between the use of "pronouns" and the word's technical sense, here disregarding first-person pronouns.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!

  b. I'm agender and would like to only be referred to by my name, and no pronouns.

  c. I don't care what any liberal says, I'm not dating a girl with pronouns.

  d. Using pronouns is basic human decency.

- Statements **(a, b)** show a disconnect between the use of "pronouns" and the word's technical sense, here disregarding first-person pronouns.
- Meanwhile, **(c, d)** attest to an understanding of "pronouns" extending from grammar to an acknowledgement (or refusal) of the possibility of gender identity beyond or outside of a binary determined by sex assigned at birth.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

# Introduction

- A **pronoun** has long been widely understood as, essentially, a word which can replace any noun.
- Recently, however—especially within the last decade—we notice changes in its meaning, giving rise to statements such as the following:

  a. There are no pronouns in the Constitution!
  b. I'm agender and would like to only be referred to by my name, and no pronouns.
  c. I don't care what any liberal says, I'm not dating a girl with pronouns.
  d. Using pronouns is basic human decency.

- Statements **(a, b)** show a disconnect between the use of "pronouns" and the word's technical sense, here disregarding first-person pronouns.
- Meanwhile, **(c, d)** attest to an understanding of "pronouns" extending from grammar to an acknowledgement (or refusal) of the possibility of gender identity beyond or outside of a binary determined by sex assigned at birth.
- All statements belie a politically charged atmosphere, with **(a, c)** being coded as conservative and **(b, d)** as liberal.

!! Please be aware that some examples from our corpus contain coarse, hateful, mocking or uncomfortable language regarding sexual orientation and gender identity.

# Framing

- The contemporary usage of the word "pronouns" can be interpreted along the lines of changes in its **quantity** and **quality**.
- In terms of its scope, it encompasses an arguably simultaneously *narrowed* (third-person pronouns) and *broadened* meaning (along the lines of non-cisgender, non-heteronormative identity, or allyship towards such folks).
- While these meanings overall have not supplanted the word's more technical meaning, there is anecdotal evidence for such a change:

  *My moms bf tore up my little sisters homework because it was about pronouns... he got so pissed off that the school was teaching this stuff that he ripped it up and... he started going on a rant about schools being too liberal and that pronouns didn't exist until my generation made them up.*

  *Source: Reddit, **r/mildlyinfuriating***

- The word also carries emotive connotations, either *positive* or *negative*, often aligning with one's political worldview (e.g., Gustafsson Sendén et al. 2021).

# Research questions

In light of these shifts in meaning, we wanted to look in a large, online media corpus to explore the following questions:

- How has discourse about pronouns changed over time, and what themes emerge?
- What is the relationship between politics and this discourse?
  - That is, how tdo people of either side of a political binary (liberal vs. conservative) talk about pronouns?
  - Do less politically overt people talk about pronouns in a similar or different way?
- How can we apply what we learn from this corpus towards protecting the rights of everyone and advocating for gender inclusivity?

# Some context

- Due to the impact of Hurricane Helene on Asheville, certain aspects of this talk remain to be developed more fully. For more information about local charities and recovery efforts:
  - **Beloved Asheville**
  - **Hearts with Hands**
  - An extensive list of ways to help can be found **here**
- Today, we'll be focusing on presenting the corpus we collected as well as *why* we collected it, plus the methodology and our preliminary results.
- This corpus has a lot to offer, so we're excited about where we can go from here and look forward to discussing with you!

# Background

# Our previous research

- Earlier results (Biers et al., 2023) using a 25% sample of the Twitter corpus and a simpler, binary political classifier showed:
  - An early spike in the relative number of tweets mentioning pronouns in 2017, and a major spike in 2020, and
  - A separation point in lower-likelihood conservative tweets, such that positive tweets began significantly outstripping negative tweets around 2018. (Higher-likelihood conservative tweets remain balanced throughout the corpus.)
- This study also included a qualitative study of other social media. We identified the following themes:
  - Possession of pronouns
  - Pronouns as an ideology (respect vs. control)
  - Pronouns as a threat to hegemonic masculinity
  - Moral panic
  - Absurdism (including reclamation through satire)
- Given the appearance of seasonality aligning with real-world political events, such as U.S. elections, we recommended:
  - Vigilance against being baited into amplifying opportunistic, hateful messaging, as well as
  - Efforts to include sharing joy as a means of resistance.

# Twitter

- As of the time of data collection (Nov. 2022), Twitter (rebranded after data collection to X) was a microblogging social media platform, with text posts of maximally 140 characters before 2017 and 280 afterwards.
- The number of users around the time of data collection was estimated at 450 million accounts (Hirose, 2022), with around 500 million tweets sent per day (Sayce, 2022).
- User age is fairly evenly distributed, with 25-34 being the most represented age bracket of users. User sex is overwhelmingly male on average (Shepherd, 2023).
- Up until recently, the majority of Twitter content from the U.S. has come from Democrat-leaning accounts (Center, 2020).
- At the time of collection, Twitter offered an API access tier to academics which allowed collecting up to 10 million tweets + metadata per month from its entire historical archive.

# Reddit

- Reddit is a forum-based social media platform in the form of posts and associated comments (limit 10,000 characters). Individual forums are called subreddits, some of which offer users a choice of “flair.”
- These flairs are subreddit-specific and allow users to display aspects of their identity such as partisanship (e.g., to a political candidate or to a reality show contestant).
- A large corpus of all English-language comments of Reddit as of 2019 is available to researchers and contains information such as subreddit, timestamp, unique user ID and flair.
- At the time of the release of this database, users numbered more than 350 million (nearly half of whom are based in the U.S.) and, in the U.S., skew male and under 30. Users in the U.S. also lean to the left on average (Vogels et al., 2021).

# Methodology

# Data collection

- Using the `academictwitteR` package (Barrie and Chun-ting Ho, 2021) and academic API tier access, we collected all English-language tweets from April 2006 to November 2022 containing the word “pronoun(s)”, excluding retweets and duplicates.
- Metadata for these tweets were also gathered (e.g., timestamp, user description).
- We excluded users with empty or non-English user descriptions and ended up excluding tweets before January 1, 2012 due to a relative lack of tokens.
- This yielded a corpus of 5,512,999 tweets from 830,177 unique users.
- For fine-tuning our text classifiers, we used data from the large Reddit database release of 2019, as made available by the ConvoKit project (Chang et al., 2020).

# Machine learning

- Two separate models were fine-tuned for this study using the DistilRoBERTa base model (Sanh et al., 2019). The models were trained in Python using the Hugging Face `transformers` library (Wolf et al., 2020) for tokenization and sequence classification.
  - First, we separated users into overtly political and “neutral” (or *apolitical*) groups.
  - Second, we categorized the political users based on predicted binary affiliation (liberal vs. conservative).
- In both cases, Reddit data were used for training and validation, due to the ease of categorization based on community participation.
- Since we are interested in how people of different publicly professed political alignments talk about pronouns, our models were run on user descriptions, not the tweets themselves (*contra* our earlier work).

# Political overtness

- To determine which users are overtly political in their self-description, we trained a RoBERTa text classification model (see appendices for parameters and statistics).
- For the political half of our training data, we gathered the comments from 7 political and gender- and sexuality-related subreddits representing either side of the political spectrum.
- For the apolitical half, we collected the comments from 7 subreddits representing pets and vocation, as well as the expressly apolitical subreddit “CasualConversation”.
- We then randomly took 9000 English-language, non-moderator comments of 50 characters or more from each subreddit and assigned them labels of either political or apolitical.
- Our model performed well (with F1 and accuracy of both 0.96).

# Political affiliation

- We took as inspiration for our political affiliation model the text classifier of Alkiek et al. (2022).
- We took from each political subreddit another sample of 9000 English-language, non-moderator comments of 50 characters or more which were not present in the political overtness training data.
- Due to initial difficulties of the model to learn, each text was then expanded into sentences, and the proportion of texts per affiliation was balanced again.
- While this proved to be a more difficult task to generalize, the model performed relatively well, with F1 and accuracy both higher than 0.77 (see appendices for more information).
- In the end, we identified 2,079,854 liberal, 1,242,132 conservative and 2,191,013 neutral tweets (280,006, 176,474 and 373,697 unique users, respectively).

# VADER

- We used the VADER model (Hutto and Gilbert, 2014) in R via `tidyvader` to perform sentiment analysis on each tweet.
- VADER is specifically tailored for microblog-like social media and provides scores based on both emotional polarity and intensity, taking into account typographical features which are used to convey emotion, such as capitalization and punctuation.
- Each tweet is provided with **positive**, **negative** and **neutral** scores, as well as a **compound** score for the entire text (see appendices for more information).

# Putting it together

- After running our political overtness and affiliation models on the user descriptions of our database, we matched users with their tweets.
- We then averaged positive, negative and compound scores by affiliation (or lack thereof) within each month, starting in 2012 (due to low numbers in earlier years).
- These averages were passed to line graphs with a LOESS regression.

# Looking at themes

- Finally, we examined themes within our corpus according to affiliation and time by performing key term extraction and topic modelling in R.
- We are much indebted to the tutorials of Andreas Niekler and Gregor Wiedemann found **here**.
- For both methods, the following was done:
  - Tweets from 2012 or later were lemmatized, and mentions and URLs were removed.
  - Using the `quanteda` package (Benoit et al., 2018), a corpus object was created.
  - Punctuation, symbols numbers and stopwords were removed from terms, and text was passed to lowercase.
  - Collocations of a minimal frequency of 25 were added to the terms.
- In the interest of time, only topic modelling is discussed here, but see the appendices for more information about key term extraction.

# Topic modelling

The following procedure was done separately for each political affiliation.

- Due to the difficulty of topic modelling to run over short documents (Murshed et al., 2023), tweets were concatenated by week. (This unfortunately makes it harder to match individual tweets to topics in a principled way.)
- A document-term matrix was created from the corpus object, and terms were limited to those appearing in at least 1% of tweets in that subcorpus and in no more than 90% of those tweets.
- A Latent Dirichlet Allocation model was then run on the matrix using the `topicmodels` package (Grün and Hornik, 2011) with a thematic resolution of 20 themes and 500 iterations of Gibbs sampling.
- We extracted the 10 most probable terms per topic and passed them to a network graph. The topic terms are reported in the appendix; we summarize them in the results.

# Results

# Positive & negative sentiment

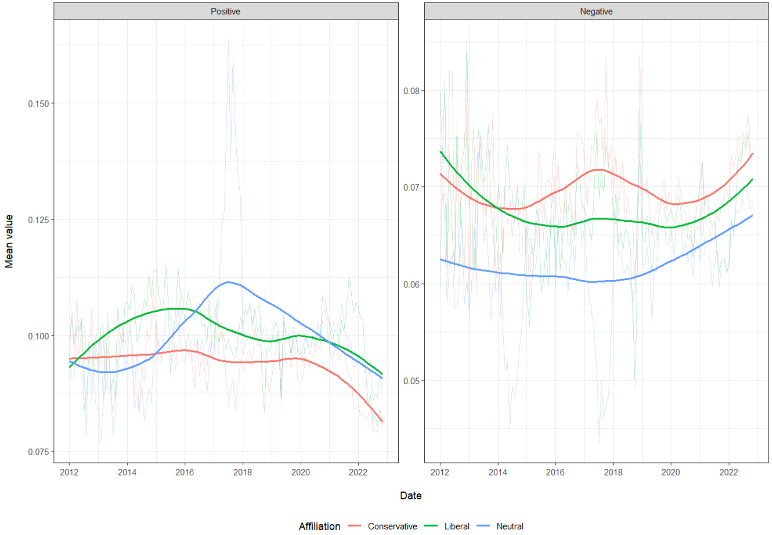



Figure 1: Positive and negative sentiment words over time, by political affiliation (averaged by month). (Note different axis ranges)

# Positive & negative sentiment

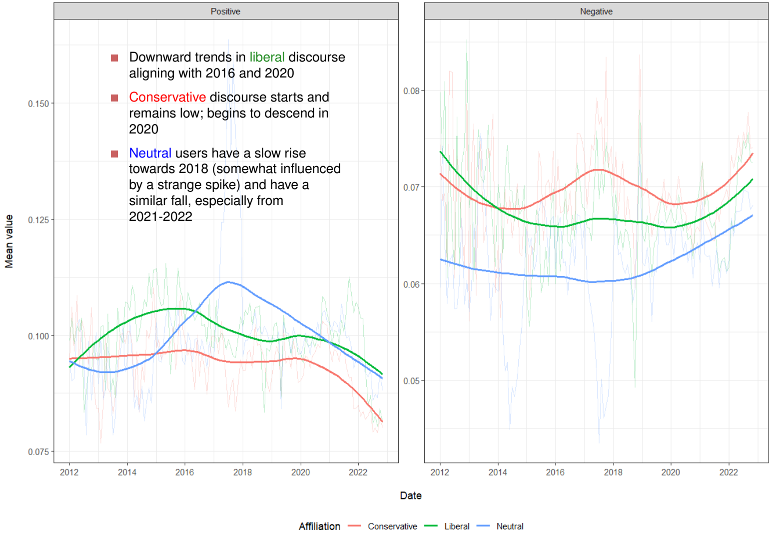



Figure 1: Positive and negative sentiment words over time, by political affiliation (averaged by month). (Note different axis ranges)

# Positive & negative sentiment

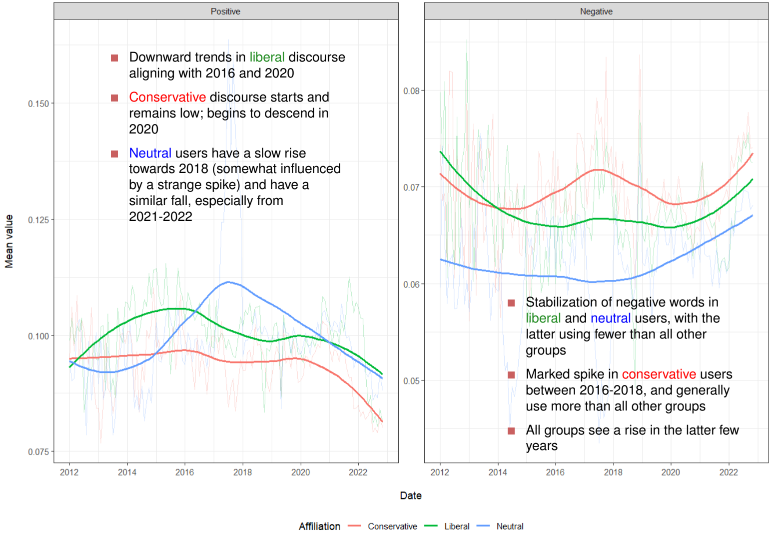



Figure 1: Positive and negative sentiment words over time, by political affiliation (averaged by month). (Note different axis ranges)

# Compound sentiment

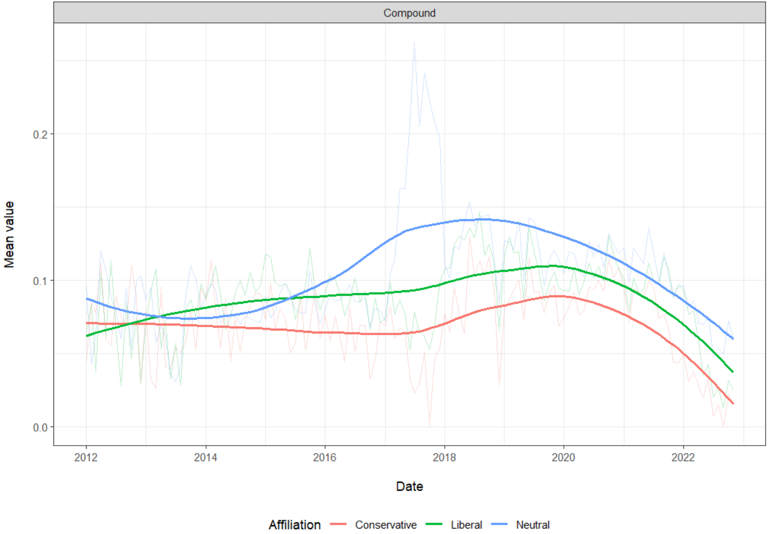



Figure 2: Compound sentiment over time, by political affiliation (averaged by month)

# Compound sentiment

- ...all of which translates to a slight increase in compound sentiment between 2018-2020 and a decrease right after in all groups
- But the reason for these trends in conservative speakers is slightly different. Recall the stability of positive words and the relative decrease of negative words after the 2017 peak

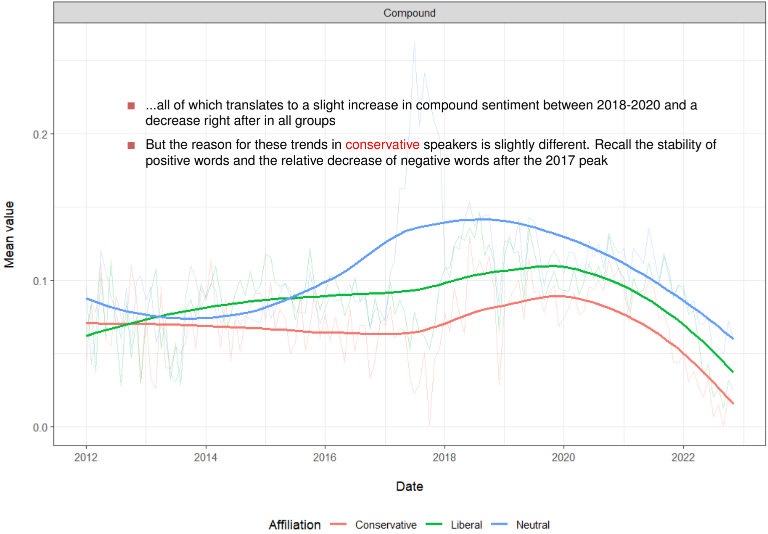



Figure 2: Compound sentiment over time, by political affiliation (averaged by month)

# Top sentiment contributors

Note that these results are presented just by raw frequency and valence in the dictionary; it does not take into account context, which the algorithm will do.

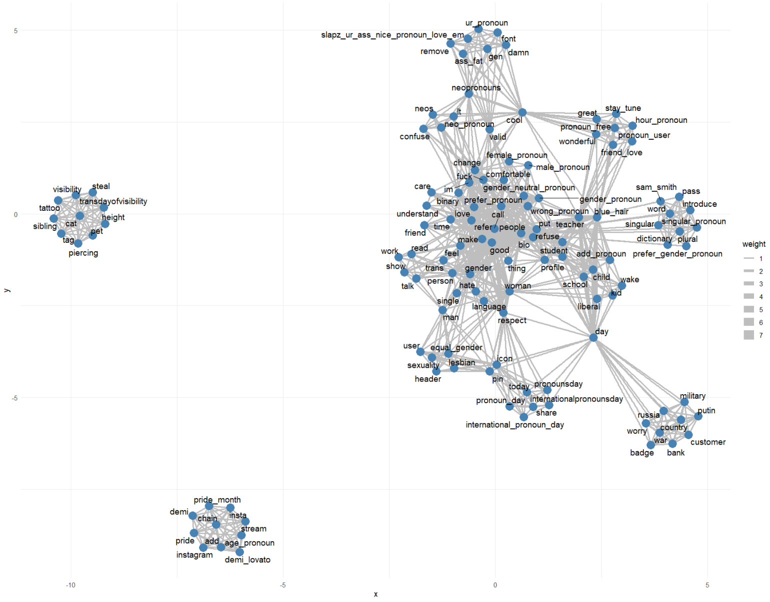
| Conservative | | Liberal | | Neutral | |
|---|---|---|---|---|---|
| **Negative** | **Positive** | **Negative** | **Positive** | **Negative** | **Positive** |
| fuck | good | fuck | good | fuck | love |
| hard | respect | hard | respect | hard | respect |
| shit | love | shit | love | shit | care |
| wrong | care | wrong | care | wrong | friend |
| bad | friend | hate | friend | confuse | cool |
| stop | matter | confuse | cool | hate | **happy** |
| hate | lol | bad | **comfortable** | bad | fine |
| confuse | **support** | idk | lol | idk | matter |
| **leave** | **god** | stop | matter | stop | **comfortable** |
| **problem** | fine | forget | fine | forget | lol |

Table 1: Top 10 sentiment contributors (negative and positive), per political affiliation

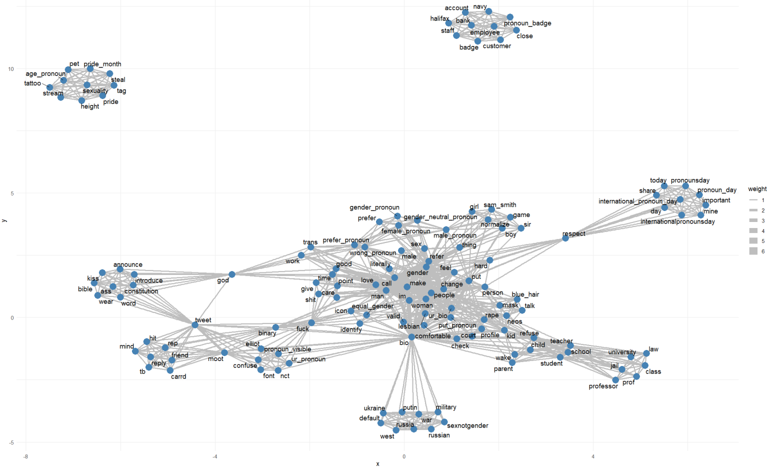
# Emergent topics

- In general, the classifier seemed to have more difficulty identifying conservatives than liberals. While some conservative-coded discourse is found in the liberal subcorpus, there appears to be more liberal discourse falsely present in the conservative subcorpus.
- Topics remain to be matched to sentiment, but certain themes emerge in the liberal and conservative subcorpora with distinctly different points of view, especially with respect to:
  - **Sharing pronouns**: Liberal discourse largely celebrates its normalization, while conservative discourse mocks it and expresses feelings of coercion.
  - **Pronoun awareness in children** or educational settings: Liberal discourse seems somewhat divided but also expresses disbelief when authority figures refuse to use one's self-identified pronouns. (This is also true outside of this topic.) Conservative discourse focuses on narratives of repression (e.g., jail time for refusing to use one's pronouns) and coercion.
- Discourse in both subcorpora explores the intersections among any of the following, without immediately clear "party lines": sexual orientation, transness, gender binarity (or lack thereof) and self-expression.
- The same also applies to grammatical discussions about singular *they*; that is, no clear consensus emerges at the present state.
- Neutral users appear to engage in several of the same topics, but impressionistically in a more positive way. Their subcorpus also includes generic introductions, apolitical grammatical discussion and publishing advice.

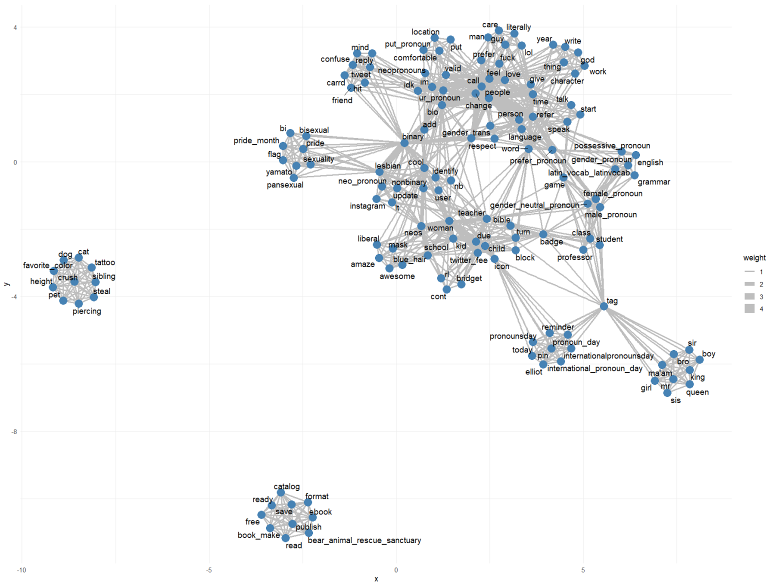
# Liberal topic networks



# Conservative topic networks



# Neutral topic networks



# Discussion & conclusion

# Discussion

- While sentiment remains stratified, clearly “the vibes are deteriorating.” Negative sentiment is on the rise in all groups and positive sentiment declining, even in the more positive neutral users.
- It is still too early to attribute any causes with a high level of certainty, but certain topics suggest different experiences behind these emotional trends, namely the narratives of coercion on the conservative side and the expression of frustration on the liberal side about a lack of respect or understanding from others.
- We find further evidence for seasonality in negative discourse from conservative users. As tempting as it is to link this with political events, this warrants a closer look.

# Future directions

- Some amount of noise and error is inevitable in a corpus this large and “widely cast” (i.e., sole criterion being the presence of “pronoun(s)”), but additional steps may clear up the picture, such as:
  - Fine-tuning the affiliation model with manually labelled descriptions from our Twitter corpus,
  - Removing swear words when performing sentiment analysis, and
  - Attempting topic modelling with fewer themes.
- We look forward to nuancing our sentiment analysis by examining the interaction of affiliation and sentiment with individual topics, as well as interacting with the literature on gender, semantics, social media and discourse analysis, among other areas.
- The API access used for data collection was abruptly removed in 2023. As such, collecting follow-up data is largely unfeasible.
- We did, however, collect a number of additional tweets from a small number of crucial actors in the corpus, as well as some information (retweeting and quoting interactions) which may be useful for a future network analysis.

# Thanks


- Thank you to my colleagues François Lareau, Patrick Drouin and Ayla Rigouts Terryn for their help with machine learning and/or term extraction.
- Thanks to Yutaka Suzuki and the authors of Alkiek et al. (2022) for their thoughts on the structure and training of the political classifier.
- This research is funded by the Social Sciences and Humanities Research Council of Canada (*Harnessing Twitter for morphophonological variation*)

# References I


Alkiek, Kenan, Bohan Zhang, and David Jurgens. 2022. Classification without (proper) representation: Political heterogeneity in social media and its implications for classification and behavioral analysis. In **Findings of the association for computational linguistics: Acl 2022**. 504–522.

Barrie, Christopher and Justin Chun-ting Ho. 2021. academictwitteR: an R package to access the Twitter Academic Research Product Track v2 API endpoint. **Journal of Open Source Software** 6(62): 3272. doi:10.21105/joss.03272. URL https://github.com/cjbarrie/academictwitteR.

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller, and Akitaka Matsuo. 2018. quanteda: An r package for the quantitative analysis of textual data. **Journal of Open Source Software** 3(30): 774. doi:10.21105/joss.00774. URL https://quanteda.io.

Biers, Kelly, Michael Dow, Eric Beuerlein, Kimaya Guthrie, Sam McIntosh, and Heather Roberts-VanSickle. 2023. Of pronoun havers and havers-not: The semantic shift of 'pronoun' in online political discourse. In **Unca queer studies conference**. Asheville, North Carolina, USA.

Center, Pew Research. 2020. Differences in how Democrats and Republicans behave on Twitter. URL https://www.pewresearch.org/politics/2020/10/15/differences-in-how-democrats-and-republicans-behave-on-twitter/, accessed on March 1, 2023.

Chang, Jonathan P., Caleb Chiam, Liye Fu, Andrew Wang, Justine Zhang, and Cristian Danescu-Niculescu-Mizil. 2020. ConvoKit: A Toolkit for the Analysis of Conversations. In **Proceedings of sigdial**. URL https://convokit.cornell.edu.

Grün, Bettina and Kurt Hornik. 2011. topicmodels: An R package for fitting topic models. **Journal of Statistical Software** 40(13): 1–30. doi:10.18637/jss.v040.i13.

Gustafsson Sendén, Marie, Emma Renström, and Anna Lindqvist. 2021. Pronouns beyond the binary: The change of attitudes and use over time. **Gender & Society** 35(4): 588–615.

# References II


Hirose, Alyssa. 2022. 24 Twitter demographics that matter to marketers in 2023. URL https://blog.hootsuite.com/twitter-demographics/, accessed on March 1, 2023.

Hutto, Clayton and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In **Proceedings of the international aaai conference on web and social media**, vol. 8. 216–225.

Murshed, Belal Abdullah Hezam, Suresha Mallappa, Jemal Abawajy, Mufeed Ahmed Naji Saif, Hasib Daowd Esmail Al-Ariki, and Hudhaifa Mohammed Abdulwahab. 2023. Short text topic modelling approaches in the context of big data: taxonomy, survey, and analysis. **Artificial Intelligence Review** 56(6): 5133–5260.

Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. **ArXiv** abs/1910.01108.

Sayce, David. 2022. The number of tweets per day in 2022. URL https://www.dsayce.com/social-media/tweets-day/, accessed on March 1, 2023.

Shepherd, Jack. 2023. 23 essential twitter statistics you need to know in 2023. URL https://thesocialshepherd.com/blog/twitter-statistics, accessed on June 1, 2023.

Vogels, Emily A., Brooke Auxier, and Monica Anderson. 2021. Partisan differences in social media use show up for some platforms, but not facebook. URL https://www.pewresearch.org/short-reads/2021/04/07/partisan-differences-in-social-media-use-show-up-for-some-platforms-but-accessed: September 27, 2024.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In **Proceedings of the 2020 conference on empirical methods in natural language processing: System demonstrations**. Online: Association for Computational Linguistics, 38–45. URL https://www.aclweb.org/anthology/2020.emnlp-demos.6.

# Apolitical model

**Training Options**

| Option | Value |
|---|---|
| num_train_epochs | 1 |
| train_batch_size | 8 |
| eval_batch_size | 8 |
| gradient_accumulation_steps | 8 |
| warmup_steps | 500 |
| weight_decay | 0.01 |

**Model Statistics**

| Metric | Value |
|---|---|
| eval_loss | 0.0981 |
| eval_accuracy | 0.9634 |
| eval_f1 | 0.9633 |
| eval_precision | 0.9653 |
| eval_recall | 0.9614 |

# Affiliation model

**Training Options**

| Option | Value |
|---|---|
| num_train_epochs | 3 |
| train_batch_size | 32 |
| eval_batch_size | 32 |
| eval_steps | 200 |
| warmup_steps | 500 |
| weight_decay | 0.01 |

**Model Statistics**

| Metric | Value |
|---|---|
| eval_loss | 0.50795 |
| eval_accuracy | 0.7741 |
| eval_f1 | 0.7722 |
| eval_precision | 0.7816 |
| eval_recall | 0.7631 |

# Calculating sentiment analysis

- For multiple-word strings, VADER provides a positive score (sum of sentiment scores of positive words over sum of sentiment scores for all words), neutral and negative scores (idem, but for sum of sentiment scores neutral and negative words, respectively) and a composite score (sum of sentiment scores over square root of sum of sentiment scores squared plus a constant).
- The composite score ranges from -1 (very negative) to 1 (very positive). For instance, “I feel happy and sad”:

$$\frac{2.7 - 2.1}{\sqrt{(2.7 - 2.1)^2 + 15}} = 0.153$$

- Again, certain typographical features like exclamation points will also impact the calculation.

# Liberal topics

| Topic | Terms |
|---|---|
| 1 | bio, put, people, make, call, refer, im, comfortable, valid, love |
| 2 | singular, people, singular_pronoun, introduce, plural, word, sam_smith, prefer_gender_pronoun, dictionary, pass |
| 3 | respect, single, hate, people, call, language, person, good, woman, make |
| 4 | kid, blue_hair, teacher, school, woman, day, child, student, liberal, wake |
| 5 | neopronouns, ass_fat, remove, ur_pronoun, font, cool, damn, slapz_ur_ass_nice_pronoun_love_em, gen, valid |
| 6 | people, gender, call, make, trans, respect, woman, feel, binary, good |
| 7 | pet, tattoo, steal, transdayofvisibility, height, tag, piercing, cat, sibling, visibility |
| 8 | pronounsday, internationalpronounsday, international_pronoun_day, pronoun_day, respect, today, pin, share, day, icon |
| 9 | feel, love, make, im, people, good, time, friend, call, gender |
| 10 | blue_hair, pronoun_user, cool, hour_pronoun, stay_tune, teacher, great, pronoun_free, wonderful, friend_love |

# Liberal topics (cont.)

| Topic | Terms |
| --- | --- |
| 11 | people, make, call, change, care, good, refer, prefer_pronoun, time, understand |
| 12 | military, badge, war, worry, country, russia, day, customer, putin, bank |
| 13 | bio, people, put, profile, call, make, fuck, thing, good, gender |
| 14 | instagram, pride, pride_month, stream, add, age_pronoun, chain, demi_lovato, insta, demi |
| 15 | prefer_pronoun, call, gender_neutral_pronoun, make, student, wrong_pronoun, good, gender_pronoun, teacher, refuse |
| 16 | lesbian, woman, icon, gender, user, header, equal_gender, man, pin, sexuality |
| 17 | im, binary, fuck, lt, neos, neo_pronoun, neopronouns, cool, confuse, change |
| 18 | refer, prefer_pronoun, gender_neutral_pronoun, male_pronoun, wrong_pronoun, call, people, female_pronoun, gender_pronoun, change |
| 19 | school, teacher, refuse, bio, put, blue_hair, child, profile, day, add_pronoun |
| 20 | person, make, talk, refer, work, gender, people, man, show, read |

# Conservative topics

| Topic | Terms |
|---|---|
| 1 | pet, age_pronoun, pride_month, pride, tag, height, tattoo, steal, stream, sexuality |
| 2 | word, bible, constitution, tweet, god, introduce, wear, ass, announce, kiss |
| 3 | bio, im, make, call, feel, gender, people, refer, love, literally |
| 4 | game, gender, girl, boy, male_pronoun, put, thing, normalize, sir, sam_smith |
| 5 | carrd, reply, mind, hit, tweet, friend, moot, tb, binary, rep |
| 6 | bio, make, refuse, people, put_pronoun, court, check, profile, rape, profile |
| 7 | lesbian, bio, woman, gender, man, identify, binary, icon, call, equal_gender |
| 8 | bio, school, teacher, child, kid, parent, change, refuse, wake, student |
| 9 | gender_neutral_pronoun, prefer_pronoun, gender_pronoun, male_pronoun, female_pronoun, prefer, refer, gender, change, wrong_pronoun |
| 10 | bio, pronoun_visible, confuse, tweet, fuck, nct, ur_pronoun, font, elliot, moot |

# Conservative topics (cont.)

| Topic | Terms |
|---|---|
| 11 | woman, man, people, call, refer, gender, thing, sex, male, make |
| 12 | people, good, time, fuck, give, make, care, point, shit, change |
| 13 | badge, customer, staff, bank, pronoun_badge, close, account, halifax, employee, navy |
| 14 | people, make, refer, call, person, gender, respect, feel, hard, change |
| 15 | love, make, good, prefer_pronoun, time, wrong_pronoun, trans, call, work, god |
| 16 | refuse, student, professor, university, teacher, jail, law, school, class, prof |
| 17 | bio, putin, war, ukraine, russia, default, russian, military, sexnotgender, west |
| 18 | bio, put, people, make, love, valid, comfortable, put_pronoun, im, ur_bio |
| 19 | bio, blue_hair, person, talk, kid, people, mask, woman, lesbian, neos |
| 20 | pronounsday, internationalpronounsday, international_pronoun_day, pronoun_day, today, share, day, important, respect, mine |

# Neutral topics

| Topic | Terms |
|---|---|
| 1 | neos, instagram, add, binary, cool, neo_pronoun, lt, teacher, update, lesbian |
| 2 | pet, tattoo, height, piercing, steal, dog, cat, sibling, crush, favorite_color |
| 3 | blue_hair, woman, awesome, amaze, cool, mask, school, kid, neos, liberal |
| 4 | talk, person, people, game, time, speak, give, start, language, word |
| 5 | lesbian, bio, woman, gender, icon, identify, binary, user, nonbinary, nb |
| 6 | love, write, thing, work, god, time, feel, word, character, year |
| 7 | queen, sir, king, girl, mr, bro, sis, boy, tag, ma'am |
| 8 | call, people, gender, change, refer, prefer, feel, time, im, fuck |
| 9 | student, teacher, gender_neutral_pronoun, male_pronoun, female_pronoun, class, game, professor, badge, prof |
| 10 | due, woman, kid, word, tag, twitter_fee, bible, child, badge, block |

# Neutral topics (cont.)

| Topic | Terms |
|---|---|
| 11 | carrd, confuse, bio, hit, reply, mind, tweet, binary, friend |
| 12 | bio, people, call, care, literally, man, guy, refer, lol, give |
| 13 | bio, put, people, comfortable, valid, feel, call, put_pronoun, add, location |
| 14 | people, call, respect, refer, person, gender, prefer_pronoun, binary, trans, language |
| 15 | rl, due, school, teacher, cont, bridget, twitter_fee, child, woman, neos |
| 16 | gender_neutral_pronoun, male_pronoun, gender_pronoun, female_pronoun, prefer_pronoun, refer, latin_vocab_latinvocab, grammar, english, possessive_pronoun |
| 17 | pride, sexuality, pride_month, lesbian, yamato, bisexual, bi, binary, pansexual, flag |
| 18 | publish, bear_animal_rescue_sanctuary, format, ebook, save, read, free, ready, book_make, catalog |
| 19 | neopronouns, people, im, love, valid, cool, feel, ur_pronoun, idk, bio |
| 20 | pronounsday, pronoun_day, internationalpronounsday, international_pronoun_day, elliot, tag, pin, today, icon, reminder |

# Pairwise key term differences

For top 100 features in each of the three subcorpora.

| |
|---|
| **Liberal terms not in Conservative** |
| cool, character, nice, boy, forget, comfortable, singular, ur |
| **Conservative terms not in Liberal** |
| speak, student, article, contraction, address, great, fact, noun |
| **Liberal terms not in Neutral** |
| nice, boy, forget, refuse, life, important, comfortable, trans_people, ur |
| **Neutral terms not in Liberal** |
| noun, grammar, sentence, possessive_pronoun, add, verb, easy, case, book |
| **Neutral terms not in Conservative** |
| character, grammar, sentence, cool, possessive_pronoun, singular, add, verb, easy, case, book |
| **Conservative terms not in Neutral** |
| refuse, speak, student, article, contraction, address, great, fact, life, trans_people, important |

# Key term differences

For top 500 features in each of the three subcorpora.

| **Liberal terms** |
| --- |
| pronoun_kumoshibot, pronoun_make, phe_pronoun, dadbot, masculine_pronouns, presentation_tip, indirect, yall, ugh, location, angry, respond, sister, cuz, slip, pronoun_preference, coworker, pet, feel_comfortable, news, acknowledge, pronoun_wrong, accidentally, specifically, partner, transphobia |
| **Conservative terms** |
| possesive_pronoun, sweden, chelsea_man, win, attack, actual, offensive, concern, fall, capitalize, gender_specific_pronoun, train, suggest, prof, remove, feminist, misuse, hit, swedish, high, early, view, normal, statement, son, side, english_language, dick, doctor, inclusive, heart, dear, recently, ridiculous |
| **Neutral terms** (sample) |
| latin_vocab_latinvocab, publish, relative_pronoun, format, lesson, save, ready, ebook, reader, japanese, object_pronoun, express, pay, embarrass, blog, ebook_distribution, author, fill, subject_pronoun, improve, book_make, true_story, prepare, depend, enjoy, catalog, exercise, text, edit, pronounce, internet, short, numb, animal, antecedent, brush, usage, sit_write, apostrophe, specific |