

Linguistic divergence in American English along socio-political polarities

Andres Karjus | ERA Chair for Cultural Data Analytics, Tallinn University, Estonia | @andreskarjus | andreskarjus.github.io
Christine Cuskley | Centre for Behaviour and Evolution, Newcastle University, UK | @nerdpro | ccuskley.github.io

Questions:

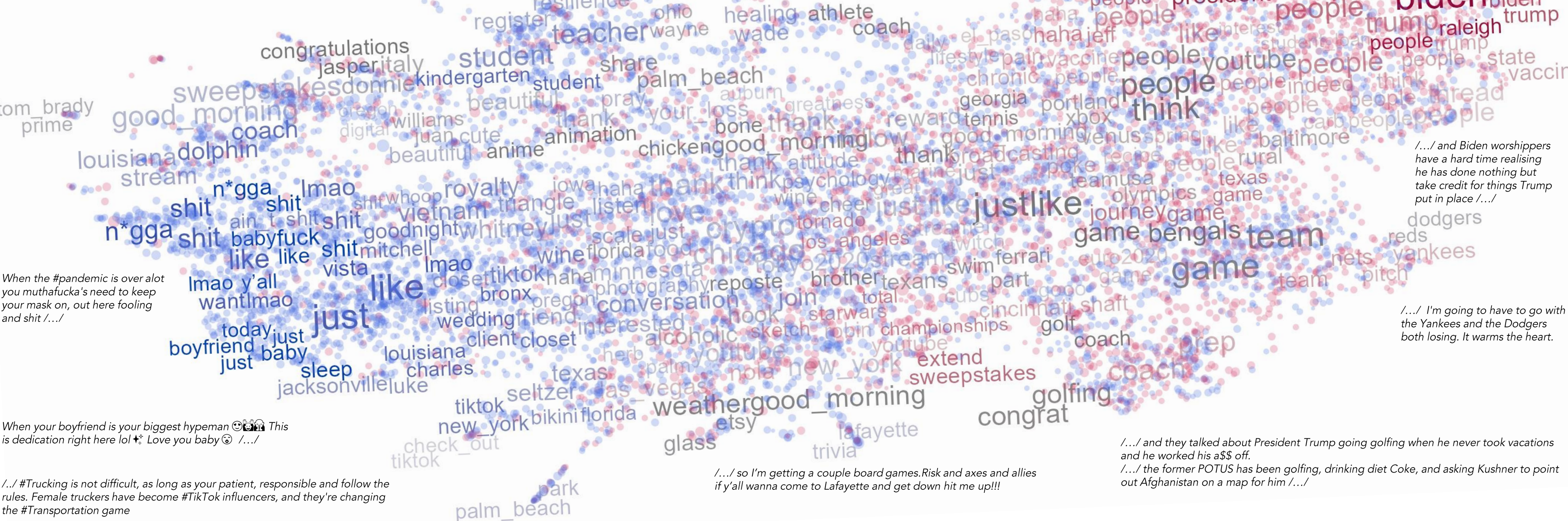
- Explorative: Is soc-pol polarization reflected in language use? What are the biggest differences between left/right?
- Methodological: can we reliably model semantic divergence using comp. models e.g. word embeddings?

Data:

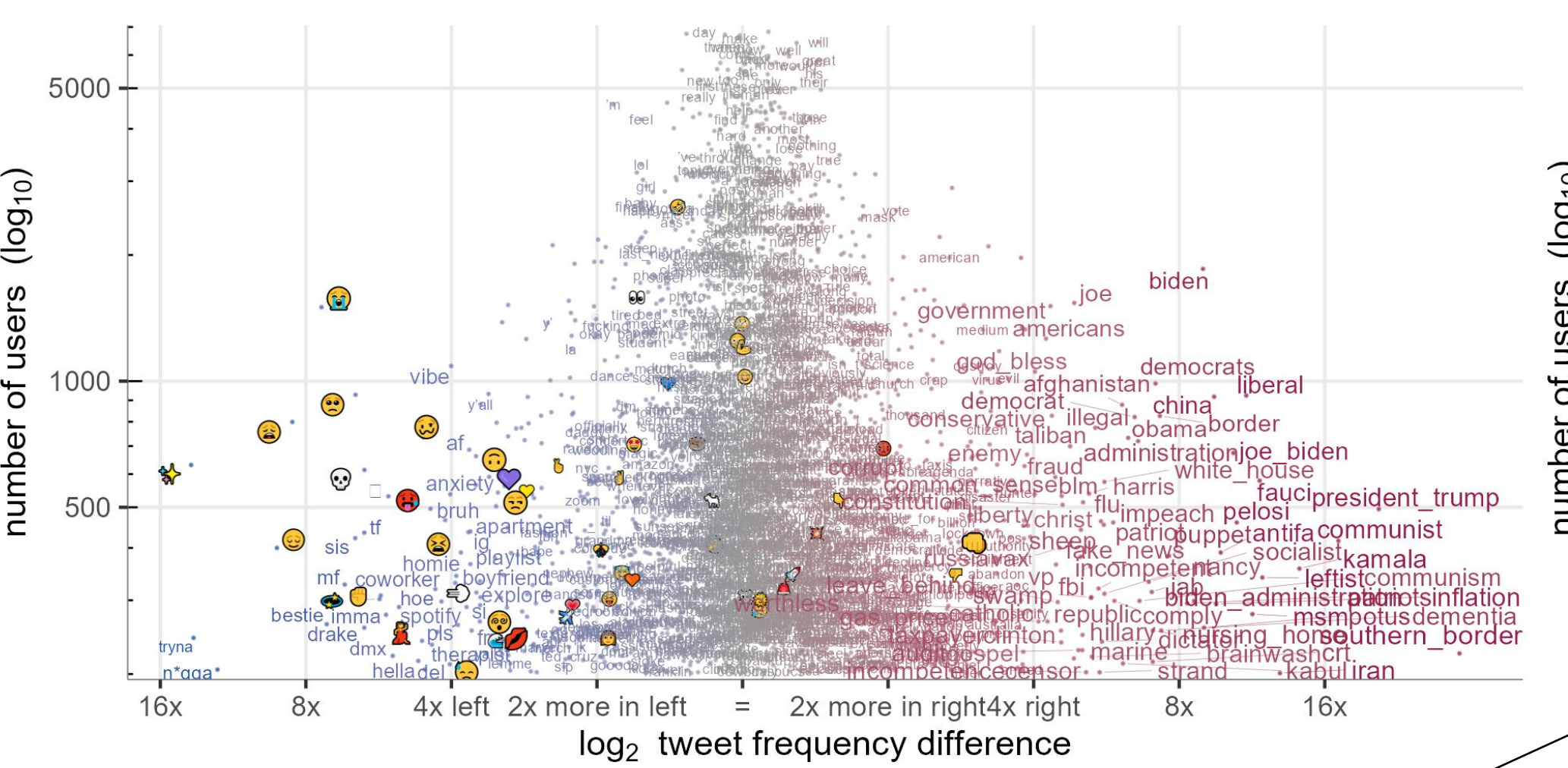
- 10,986 US Twitter users; from 422m follower listings
- "Left" (n=6202) and "right" (n=4784): only follow media on "their side" (based on Allsides 2021)
- Corpus: 1,483,570 tweets; Feb-Sept 2021

Users by topic similarity:

- LDA+UMAP+TFIDF

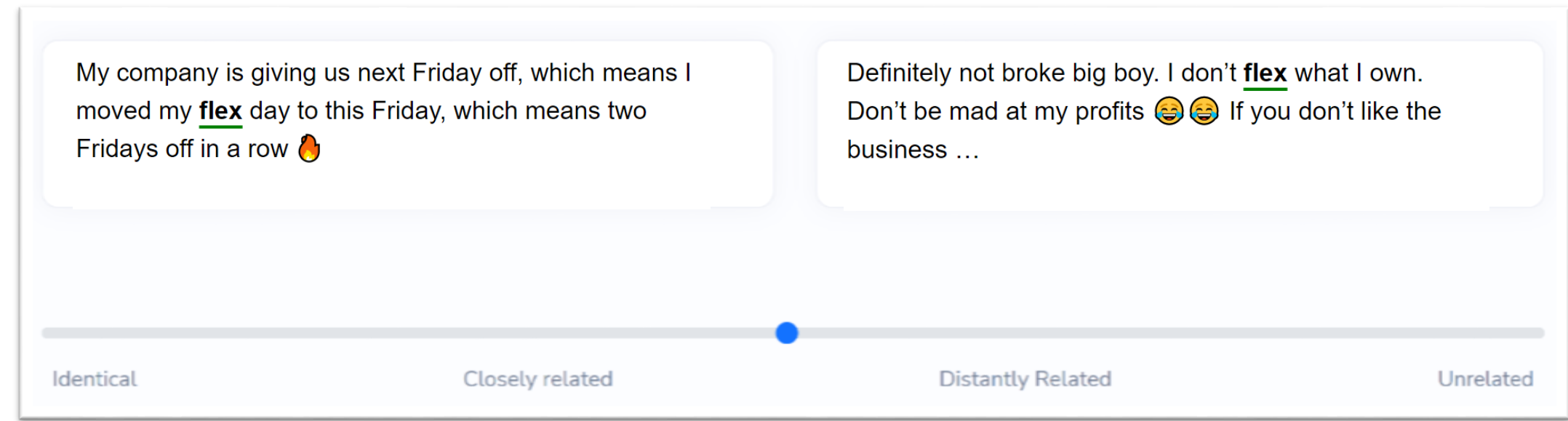


Divergence in word usage by frequency:

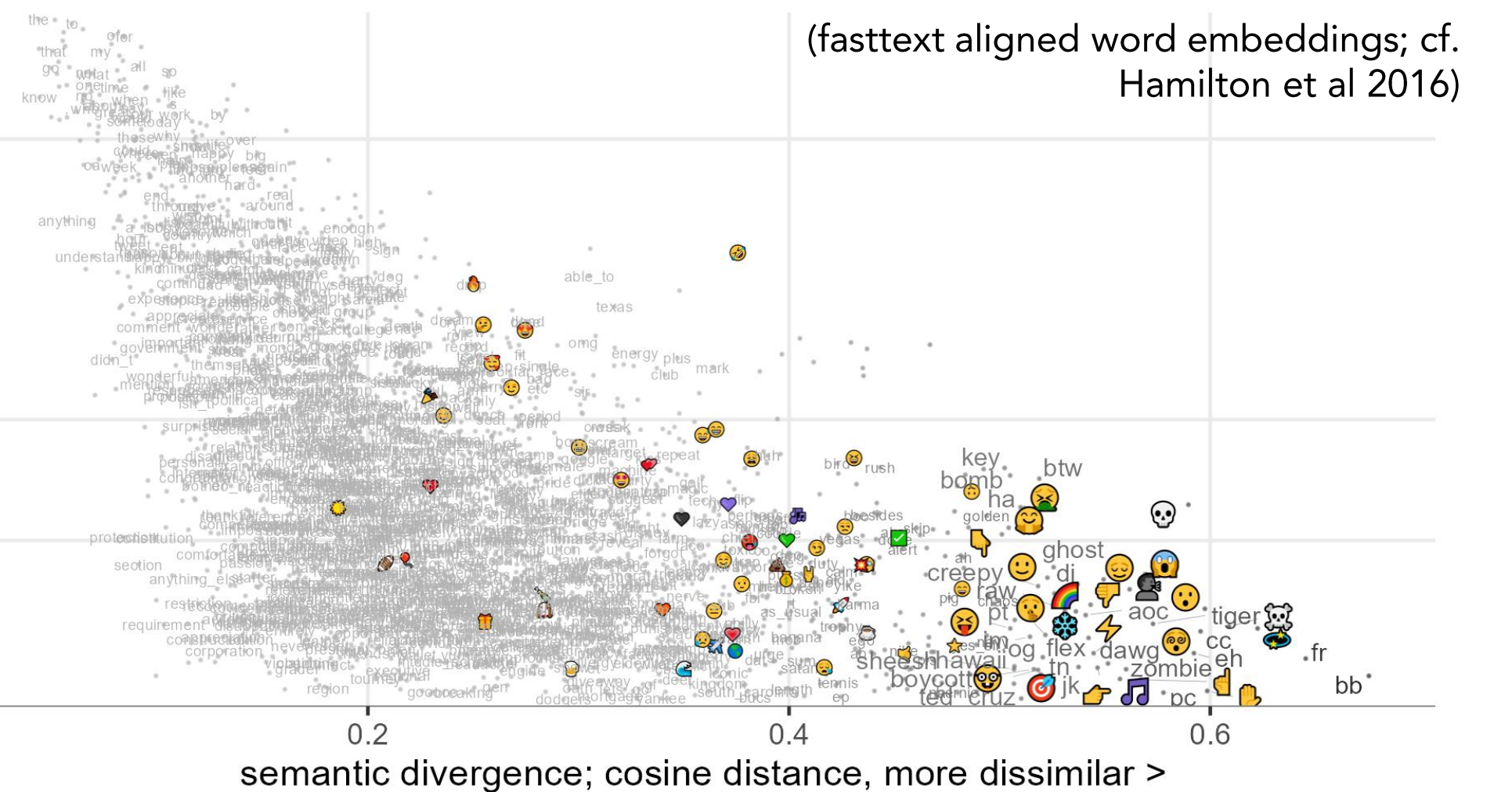


Work in progress:

- Could we test how well these computational models capture meaning and semantics shifts?
- The DUREL annotation task (Schlechtweg et al 2018) for measuring semantic change + Mechanical Turk



Divergence in word meanings:



The short version

- We mined large corpus of tweets by users grouped by political polarities, according to what news media outlets they follow
- Preliminary results: differences in topics of conversation, certain words and phrases used in different senses
- Ongoing work: validate the corpus findings using an annotation experiment
- Future work: tease apart sources of divergence

References: Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: Divided they blog. • Albertson, B. L. (2015). Dog-whistle politics: Multivocal communication and religious appeals. • Allsides Media Bias Ratings, version 4. Available from: <https://www.allsides.com> • Bhat, P., & Klein, O. (2020). Covert Hate Speech: White Nationalists and Dog Whistle Communication on Twitter. • Brown, J. R., & Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. • Chen, T. H. Y., Salloum, A., Gronow, A., Via-Antilla, T., & Kivela, M. (2021). Polarization of climate politics results from partisan sorting: Evidence from Finnish Twitter. • Densky, D., Gang, N., Voigt, R., Zou, J., Gentzkow, M., Shapiro, J., & Jurafsky, D. (2019). Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. • Gentzkow, M., Shapiro, J. M., & Taddy, M. (2019). Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech. • Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. • Halpern, D. J., & Rodriguez, P. L. (2018). Partisan Differences in Semantic Representations and their Role in Attitude Judgments. • Jurkowitz, M. A., Mitchell, E., Shearer, S., & Walker (2020). US media polarization and the 2020 election: A nation divided. Online report (www.pewresearch.org). • Khudabukhsh, A. R., Sarkar, R., Karim, M. S., & Mitchell, T. M. (2020). We don't speak the same language: Interpreting polarization through machine translation. • Schlechtweg, Dominik, Sabine Schulte im Walde, and Stefanie Eckmann. 2018. "Diachronic Usage Relatedness (DUREL): A Framework for the Annotation of Lexical Semantic Change."