Linguistic diversity and social media networks

Large-scale social media data have been widely used to study linguistic diversity in areal (Eisenstein 2017), spatiotemporal (Grieve et al. 2016, 2018), and socio-demographic variation (Gonzales 2024). Yet, the full potential of social media remains underutilized, since most studies overlook its role in community formation and network building. Research across disciplines highlights the fundamental importance of social networks for humans (Granovetter 1973; Dunbar 2020; Waldinger & Schultz 2023) and for understanding linguistic variation (Milroy 1987), making it timely to integrate network evidence into the study of linguistic diversity on social media

One possible reason for not including network information lies in methodological challenges: using complex social media data to construct networks is difficult (Eisenstein et al. 2014). This presentation addresses this challenge by introducing methods for building networks from interactional metadata in social media. We propose a framework for quantifying ego networks, applicable to any social media application where interactions form a directed graph (e.g. Bluesky, Twitter, etc.). Developed within a large digital infrastructure project, this methodology combines expertise from sociolinguistics and computer science.

We answer two research questions:

(1.) To what extent can we quantify the structure of ego networks?
(2.) What research potential emerges from combining social networks with large-scale social media data?

To illustrate the potential of networks, we examine lexical innovation in English. Prior work shows that new lexis originates in urban centers (Grieve et al. 2018), but little is known about the innovators or their social settings. Using three large datasets (3.42 billion words, 2020–2022) from American, Australian, and British English, we trace emerging lexis (e.g., *rizz*, *oomfies*, *girlies*) across networks. Our analysis demonstrates how social network information enriches accounts of areal variation in lexical emergence. The presentation offers methodological insights on how large-scale social media data can be accompanied by interaction-based social information.

References:

Dunbar, R. I. M. (2020). Structure and function in human and primate social networks: Implications for diffusion, network stability and health. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 476(2240), 20200446. https://doi.org/10.1098/rspa.2020.0446

Eisenstein, J. (2017). Identifying regional dialects in on-line social media. In C. Boberg, J. Nerbonne, D. Watt (Eds.). *The Handbook of Dialectology* (pp. 368–383). Wiley. https://doi.org/10.1002/9781118827628.ch21

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PLOS ONE*, 9(11), e113114. https://doi.org/10.1371/journal.pone.0113114

Gonzales, W. D. W. (2024). When to (not) split the infinitive: Factors governing patterns of syntactic variation in Twitter-style Philippine English. *English Language & Linguistics*, 28(2), 305–339. https://doi.org/10.1017/S1360674323000631

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6), 1360–1380.

Grieve, J., Nini, A., & Guo, D. (2018). Mapping Lexical Innovation on American Social Media. *Journal of English Linguistics*, 46(4), 293–319. https://doi.org/10.1177/0075424218793191

Grieve, J., Nini, A., & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language & Linguistics*, 21(1), 99–127. https://doi.org/10.1017/S1360674316000113

Milroy, L. (1987). *Language and Social Networks*. Wiley.

Waldinger, R., & Schulz, M. (2023). *The Good Life: Lessons from the World's Longest Scientific Study of Happiness*. Simon & Schuster.