



Language
Technologies
Institute

Thesis Defense

Zoom| Thursday, July 22, 2021| 2:00 pm



COMMITTEE:

Carolyn P. Rosé, (chair)

Yulia Tsvetkov

Geoff Kauffman

**David Jurgens
(Univ. of Michigan)**

Computational Models of Identity Presentation in Language

Michael Miller Yoder

Abstract

Researchers in computer science and natural language processing have built models of how language use reflects the latent, stable identities of language users. However, such a conception of identity is ill-equipped to investigate dynamic, contextual expressions of identity in online communities. This thesis draws on theories of identity in language from sociolinguistics, linguistic anthropology, and the social sciences that view identity not as fixed and predetermined, but constructed in language and interaction. We pair these theories with techniques from machine learning, statistics, and natural language processing in a new framework for computational investigations of identity presentation in online communities. This framework relates identity presentation to social interaction and outcomes, such as the sharing of content within social media or correlations with social movements offline. We demonstrate this framework on datasets of linguistic and social interaction from two online contexts: Tumblr, a social media site known for identity talk, and fanfiction, narratives that transform and expand on original media.

https://raw.githubusercontent.com/michaelmilleryoder/michaelmilleryoder.github.io/master/files/yoder_thesis.pdf