

Dialect variation in online social media

JACOB EISENSTEIN

Georgia Institute of Technology, USA

While social media language is sometimes described as a dialect in itself, in fact it displays a remarkable amount of internal variation, aligning with both geography and ethnicity. Such variation can be revealed by computational statistical methods that search for patterns of association between language and geography from large corpora of unannotated text. I will also discuss the relationship between written language in social media and traditional dialect variation. It is well known that online writing contains phonetically-inspired spellings, but perhaps more surprising is that these spellings reproduce some of the systematic context-sensitivity of the spoken language variables that they transcribe. Finally, I will present new research on the social properties of online dialect variation, with evidence that authors modulate their use of social media variables depending on both the context and their audience.

[This is joint work with David Bamman, Brendan O'Connor, Umashanthi Pavalanathan, Tyler Schnoebelen, Noah Smith, and Eric P. Xing.]