ABSTRACT

Language is socially situated: both what we say and what we mean depend on our identities, our interlocutors, and the communicative setting. The first generation of research in computational sociolinguistics focused on large-scale social categories, such as gender. However, many of the most socially salient distinctions are locally defined. Rather than attempt to annotate these social properties or extract them from metadata, we turn to social network analysis, which has been only lightly explored in traditional sociolinguistics. I will describe three projects at the intersection of language and social networks. First, I will show how unsupervised learning over social network labelings and text enables the induction of social meanings for address terms, such as “Ms” and “dude”. Next, I will describe recent research that uses social network embeddings to induce personalized natural language processing systems for individual authors, improving performance on sentiment analysis and entity linking even for authors for whom no labeled data is available. Finally, I will describe how the spread of linguistic innovations can serve as evidence for sociocultural influence, using a parametric Hawkes process to model the features that make dyads especially likely or unlikely to be conduits for language change.