

186: Social and Distributional Predictors of the Success of Lexical Innovations in Online Writing
Ian Stewart, Jacob Eisenstein
Georgia Institute of Technology

Lexical innovations such as the intensifier *af* (e.g., “that’s cool af”) emerge constantly in online communities (Grieve, Nini & Guo 2016). The success of an innovation often relies on the diversity of social and linguistic situations in which it is used (Metcalf 2004). The diversity of lexical context, or “context dissemination,” remains understudied yet could be important in explaining how linguistically useful an innovation will be: if the intensifier *af* remains bound to a few contexts, it may be abandoned in favor of a more useful intensifier.

We study lexical innovations that emerged on the online community Reddit from June 2013 to June 2016, focusing on two questions related to the influence of contextual factors:

- RQ1. Do innovations with higher context dissemination succeed more often and survive longer?
- RQ2. When controlling for context dissemination, do words with higher social dissemination succeed more often and survive longer?

We identify as successful innovations a set of 1,467 nonstandard words whose frequency grew consistently from June 2013 to June 2016, which include respellings, compounds and acronyms. Context dissemination (D^C) is defined as the count of unique trigram contexts in which a word appears, minus the expected trigram count, as computed by a linear regression on log-frequency. For RQ2, social dissemination is defined as the dissemination across users, subreddits, and discussion threads (Altmann, Pierrehumbert & Motter 2011). We match the successful innovations with a set of 601 slang words that grew and later declined over the same time period, but which had similar frequency during the pre-decline phase. We use logistic regression to compare the relative importance of context and social dissemination when controlling for initial frequency.

Addressing RQ1, higher trigram context dissemination indicates a higher probability of success ($\beta=3.824$, $p<0.01$). Among frequency-matched pairs, context dissemination alone distinguishes successful from failed innovations with 61.7% accuracy. For RQ2, higher subreddit dissemination ($\beta=1.94$, $p<0.01$) and a higher user dissemination ($\beta=3.287$, $p<0.01$) indicate a higher probability of innovation success. The fact that thread dissemination does not contribute to word success contrasts with prior findings (Altmann, Pierrehumbert, and Motter 2011) and suggests that adoption across users and subreddits is more important than adoption across threads. Lastly, we find that a predictive model incorporating both context and social dissemination outperforms context-only and social-only models (69.4% vs. 61.7% and 62.8%), showing that the linguistic and social contribute differently to word success.

References:

- Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2011). Niche as a determinant of word fate in online groups. *PLoS ONE*, 6(5), 1–12.
- Grieve, J., Nini, A., & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, 20(2), 1–29.
- Metcalf, A. (2004). *Predicting new words: The secrets of their success*. Boston: Houghton Mifflin.