

Umashanthi Pavalanathan, Jim Fitzpatrick, Scott Kiesling, and Jacob Eisenstein

Reddit.com -- “the front page of the Internet” -- is a very large Internet community that is a conglomeration of subreddits. Originally designed for posting items of interest from the Internet, Reddit offers the ability to submit comments, a design which leads to the creation of small conversation-like threads. In addition, each subreddit tends to have loyal followers who create a community of practice. Reddit is thus ideal for studying conversational and linguistic norms across communities. We use a massive Reddit corpora to explore the ways that stance and English stance markers are used in various Reddit communities and types of threads on Reddit. We find that subreddits that have distinct conversational structure tend to use different kinds of stance markers, and that some kinds of stances are correlated with different interactional norms across subreddits.

Our corpus is an archive of 530 million comments posted on Reddit in 2014. Using a list of stance markers based on Biber and Finegan (1989), we constructed a term-subreddit matrix and performed singular value decomposition (SVD) to automatically extract the principal dimensions that explain the most variation. Each dimension consists of a set of stance markers that are grouped together based on their linguistic functions.

The preliminary analysis of stance markers suggest several *interactional style* norms for specific types of subreddits. The first is an ‘argumentative’ style that focuses on an issue or topic. These subreddits evince high levels of epistemic markers, high levels of certainty, and interaction among commenters rather than comments on the single original post. These tend to be argumentative type subreddits such as */r/philosophy*, */r/atheism*, and */r/politics* with complex conversation thread structure. A second interactional style is characterized by a correlation with more interpersonally-focused stance markers (such as *nah*, *please*, *oh*, *um*, and *kidding*); this style is characteristic of subreddits which have higher levels of user engagement: more posts per user and a more uniform contribution by all commenters rather than the typical pattern of a minority of active commenters. We suggest that these posts lead to a ‘personal’ style that has more expressions of emotion and a focuses on the commenters themselves rather than discussions about knowledge.

Increased user engagement is a property of successful online communities (Kraut et al, 2012); therefore interpersonal style and higher level of certainty can be attributed to the success of the community. We also find that subreddit threads with uncertain discussions tend to be longer, and they are more sincere when the post initiator has higher engagement with commenters.

This study has implications for understanding how communities form linguistic interactional norms, and how genre or speech event characteristics shape those stances as well. Most importantly, we show that stance can be used in quantitative analyses of interaction to a yield significant understanding of the patterning of linguistic variation.

References:

Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text-interdisciplinary Journal for the study of discourse*, 9(1), 93-124.

Kraut, R. E., Resnick, P., Kiesler, S., Burke, M., Chen, Y., Kittur, N., ... & Riedl, J. (2012). *Building successful online communities: Evidence-based social design*. MIT Press.