Evolving innovations: A real time study of changing slang in Twitter
Rebecca Maybaum, University of Haifa

Session 12: Computer-Mediated Communication

As an incubator for innovation, the Internet is rife with linguistic changes in progress, but has thus far remained under-exploited as a resource for sociolinguistic study. As Herring (2004) and Squires (2010) note, the technologies that define computer-mediated communication (CMC) are far less novel than researchers like to think, and CMC data can be used to contribute to broader discussions of linguistic theory. In particular, the relative ease of collecting large amounts of text and time data provide a unique opportunity to study the large-scale variation of linguistic innovations as they diffuse over time.


As evidenced by the above example, published in October 2010 on the popular online microblogging service Twitter, there is considerable variation and little consensus as to the appropriate term to refer to fellow Twitter users. The natural vacuum of what to call “Twitter people” created the conditions for the spontaneous and simultaneous innovation of a variety of synonymous forms. In this paper I describe and analyze the ongoing negotiation of form, function, and meaning of these new words as they are adopted by Twitter users and spread throughout the community over time.

The 15-million-word corpus includes ten subcorpora, each of which contains the entire published history of a single Twitter-specific slang term, from the inauguration of the service in March 2006 through January 1, 2012: tweeps, tweeple, tweeple, tweethearts, tweeties, tweetheads, tweeps, twerps, twitterbugs, twittertwatters. Some of the terms were much more successfully diffused throughout the Twitter population than others: tweeps appeared in approximately 800,000 distinct posts, while twittertwatters was only used 16 times.

Each subcorpus was divided into five sections by time, based on Rogers’ adopter categories (1995), which correspond to the five common stages of innovation diffusion: innovators, early adopters, early majority, late majority, and laggards. A corpus analysis revealed variation among the “Twitter people” posts over time. For example, earlier posts were more likely to contain RT (a Twitter convention indicating that the post is a “retweet,” or re-post, of another user’s original post) than were later posts, for all variants. The greeting and departing cluster containing goodnight, morning, and night increased over time in the tweebs data, but decreased over time in the tweetheads data. This seems to indicate a divergence over time, as users perhaps began to associate each of the variants with a different function.

According to Eckert (2005)’s third wave of variation studies, wherever variation exists, there is also the potential for the construction of social meaning. This study supports the view that language users have agency in the construction of social meaning, and provides evidence of this phenomenon at work as innovative variants are adopted and changed by the language users themselves.