Non-standard lexical and grammatical resources in Finland Twitter English

Steven Coats
University of Oulu
steven.coats@oulu.fi

Online social media such as Twitter have become important sites for informal communication. Some research suggests that in a global context, English is playing an increasingly important role as a language of informal online communication, including in communities where English has hitherto not been widely used as a language of local or informal communication. At the same time, the English typical of some genres of Computer Mediated Communication (CMC) exhibits high frequencies of non-standard lexical and grammatical features such as orthographical variants and symbols or emoticons. A characterization of some of the linguistic properties of Twitter English in a second-language environment such as Finland can shed light on the dynamics of global online English varieties as they continue to evolve and how sociolinguistic factors may contribute to the differentiation of Global Englishes.

In a first step, a corpus of geo-located English-language Twitter user messages from Finland was created by accessing the Twitter Streaming API and using an automated language detection tool to remove non-English user messages. Then, the frequencies of token types showing affective lengthening (multiple repetition of letter characters, e.g. *goooood*) or representing emoticon symbols were compared with those derived from a corpus of non-geo-located English-language Twitter messages subject to the same processing procedures. After disambiguating the messages by author/sex and gender (using automated methods), it can be demonstrated that Finland Twitter English is distinctive in some ways, and that sociolinguistic factors play a role in the emergence of the variety. The analysis may shed light on some properties of the lexis and grammar of global online English varieties as they continue to develop.