New linguistic tools can predict your dialect characteristics

September 24 2014

A new linguistic study may make it possible to more accurately predict the dialect features people use based on their demographic characteristics and where they live. In a new article published in the September 2014 issue of *Language*, Martijn Wieling (University of Groningen) and colleagues used statistical modeling techniques to predict whether speakers in Tuscany use words from standard Italian or words unique to local dialects.

The article, "Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling", is available in a pre-print version at http://www.linguisticsociety.org/files/wieling.pdf.

In the article, Wieling *et al.* studied how over 2,000 speakers of Italian and Tuscan dialects referred to 170 different concepts. (The Italian word for 'cheese', for example, is formaggio; a Tuscan speaker may refer to this instead as cacio.) Using a technique known as generalized additive mixed modeling, the researchers examined how the location of a speaker, as well as demographic information such as their age, sex, and education level, are likely to affect whether a speaker will use the standard (Italian) or dialectal (Tuscan) form for a given concept. Though the effects of geography and social factors in shaping language use have previously been studied by many linguists, Wieling *et. al*'s study considers them together in a single and mathematically more sophisticated model.

Their findings reflected many previously-studied trends in dialect variation: for example, men, farmers, and speakers further from the city of Florence were more likely to use dialectal, Tuscan-specific words than women, while speakers with higher levels of education were more likely to use standard, Italian words. However, Wieling *et. al*'s model also provided new insight into dialect patterns. For example, old speakers were more likely than young speakers to use their local dialect's terms for frequently-used concepts, but both young and old speakers showed similar patterns of usage for less-frequently-used words. Additionally, there was great variability in the usage patterns across concepts. For some concepts, the standard Italian form was more likely to be used in smaller villages than larger villages—while for other concepts this pattern was reversed, with a greater likelihood of a dialect-specific form in smaller villages.

Though this study focused on a group of speakers in a single region of Italy, the modeling methods used in this study could be applied to predict how geography and demographics could affect the language used by speakers of other languages, such as American English.

Provided by Linguistic Society of America