WHERE’RE Y’ALL TWEETING?

INTRODUCTION

Data mining social media sites such as Facebook and Twitter has recently allowed the creation of massive linguistic corpora in a matter of minutes instead of having to tediously transcribe thousands of hours of speech. This has led to several coarse-grained studies of large-scale dialects in the US which have verified—and challenged—previously established trends.

The goal of this study is to look at how the second person plural ‘y’all’ is used by Twitter users across the US and try to use its spatial distribution to locate regional dialects of American English in the contiguous US. This region will then be compared to previous linguistic findings from Labov, Ash and Boberg (2006) as well as traditional cultural boundaries from Carver (1987). Additionally this study will investigate the relationship between race, specifically looking at Black and Hispanic users, and ‘y’all’ usage by using a regression model with 2010 Census data.

METHODOLOGY

The geocoded tweets used in this study are from Cheng, Caverlee and Lee (2010) who collected 5,516,047 tweets from 5,136 Twitter users using the Twitter Streaming API. Coordinates are unique to users giving this study 5,136 data points. Twitter users that tweeted ‘y’all’ were filtered from the 5.5 million tweets in addition to those who tweeted ‘you guys’ or ‘you all’. Users considered in this study either used ‘y’all’ or ‘you guys’/‘you all’. Users who used all three were not considered. In total 798 users used ‘y’all’ and 598 you all and/or you guys. ‘Y’all’ and you guys/you all users were mapped and spatially joined to counties. Counties with only one ‘y’all’ user were deleted. ‘Y’all’ users per county was normalized by dividing it by you guys/you all + ‘y’all’ users which was then divided by the county’s population using 2010 Census data. Centroids of these counties were made using the feature to point tool which were then interpolated using Gaussian kriging with the normalized ‘y’all’ usage adjusted for county population as a weight.

DISCUSSION & CONCLUSION

The model used in this study shows similar trends established by the existing literature, supporting previous sociolinguistic trends and a major cultural boundary in the US between the North and South. Therefore, ‘y’all’ is likely a feature most closely associated with the South and “Southernmost” (Lipski 1993). The results of this study do not show an extension of ‘y’all’s’ range in the US as was previously hypothesized.

Looking at the statistical analysis done using GeoDa, ‘y’all’ on Twitter seems correlated with Black users while, you guys/you all associated with Hispanic users. However considering the overall weak correlations, there cannot be much taken from this study about how Twitter user’s choice in second personal plural pronouns is influenced by race. This could be likely due to only having 103 significant points and only 5,136 users on a national level.

sources


Huge help from Julia Fowler for their coding prowess. Data projection: WGS_1984_UTM