

Abstract

The meaning of a sentence can be conveyed in different forms. Pronunciation, word choices, and constructions can differ and when the variation is linked to social factors like age or gender, then it is called a social variation. In this work we investigate if syntactic variation occurs between different socio-economic groups. We work on user-generated text and we use an automatic method to collect the data. We use a machine learning approach to test our hypothesis. We report positive significant results. We also test whether the results hold across languages and we confirm this hypothesis too. To the best of our knowledge, this work is the first computational study to investigate syntactic variation along the socio-economic axis.