The truth is out there, but you might not find it: multidimensional signals and analytic flexibility

Michael Proctor

Different approaches to linguistic data analysis can lead to substantially different conclusions based on the same data, raising concerns that researcher degrees of freedom might lead to claims that do not stand the test of time. The multidimensional nature of speech makes it especially prone to analytic variability, from decisions about what and how to quantify, to which models to use. I report findings from a multi-analyst study of German prosody designed to examine these issues in more detail, and reflect on the implications for the Open Science movement.

46 teams of researchers attempted to answer the same research question, using the same dataset: *Do speakers acoustically modify utterances to signal atypical word combinations?* Acoustic recordings of 900 utterances produced by 30 German speakers were provided to each team to analyze as they saw fit. Analyses were peer-review and compiled. Substantial variability was found in reported effect sizes and their interpretation. Bayesian meta-analysis finds little to no evidence that the observed variability can be explained by analysts' prior beliefs, expertise, or perceived quality of their analyses. We recommend that researchers more transparently share details of their analysis, strengthen the link between theoretical construct and quantitative system, and calibrate their (un)certainty in their conclusions.

More about the project: https://many-speech-analyses.github.io

Coretta, S., et al. (2022). *Multidimensional signals and analytic flexibility: Estimating degrees of freedom in human speech analyses*. https://doi.org/10.31234/osf.io/q8t2k