Language Effects in Surveys (LES) – A Research Project

Over the past decades, social science scholars have wrestled with the challenge of securing equivalence in survey measures of social and political phenomena across culturally and linguistically varied societies. The research project Language Effects in Surveys (LES) at the Political Science Department at University of Gothenburg aims at addressing this problem by developing/applying methods from natural language processing and computational semantics in order to empirically study translational effects. And to provide tools that can be used to correct for the impact of language discrepancies in cross-cultural comparative research. The approach will provide quantifiable measures of attitudinal expressions as well as of translation discrepancies – and methods for overcoming, or at least minimizing them.

The main investigative tool of the LES-project is a version of distributional semantics that automatically quantify language similarities between linguistic entities. The method collects distributional patterns for linguistic units from large collections of dynamic web-based data – mainly generated by social media (Big Data). The results give an understanding of the usage of words, and allow the finding of text-specific similarities and associations. Computing of such relations across languages will tell about similarities and differences in the meaning of words – and will highlight translational discrepancies.

The LES-project will apply the developed methods to a wide variety of languages – all principle Western languages, but as well languages like Russian, Arabic, Chinese, Hindi, and Japanese. Frequently used concepts in leading comparative surveys, like WVS, ISSP, CSES, ESS, Latinobarometer, Eurobarometer, and Afrobarometer – just to mention a few - have the highest priority to be analyzed. Examples are
concepts like Democracy, Corruption, Government, Happy, and Satisfied.

The LES-project is an interdisciplinary collaboration of researchers from political science and computational linguistics with a national background in Sweden, Norway, and Canada. The project is partly founded by a four year grant from the Swedish Science Foundation ending in 2019. Project manager is Stefan Dahlberg at the Political Science Department at Gothenburg University.