

Dissertation:
“Multilingual Automated Content Analysis for
Comparative Communication Research”
By Fabienne Lind

Abstract

Communication scientists have made fast progress in the automated content analysis of large quantities of media data. This cumulative dissertation contributes to this field with a comprehensive methodological work on automated content analysis strategies specially for the analysis of multilingual text corpora. It does so with three overview articles, each introduces, compares, and discusses multiple strategies for the analysis of a multilingual text corpus: Article 1 for rule-based dictionary methods, article 2 for supervised machine learning, and article 3 for topic modeling methods. As the methodological handling of several languages is particularly important for comparative communication research, the dissertation further provides recommendations for this discipline. These developed guidelines concern the valid and resource-saving design of methodological strategies for scenarios in which the cross-language comparability of automated instruments and measurements is important. To highlight the language sensitivity of automated instruments, article 4 considers the implications of more complex language-specific structures and their translatability. A key conclusion derived from this dissertation is that automated strategies are appropriate methodological strategies to capture the essence of multilingual text corpora. Protocols for validating instruments that are adapted to the multiple languages and to the needs of comparative communication research are, however, essential and cannot do without the adequate expenditure of resources. In concluding, this dissertation provides solid ground for the application of the methodological insights to various urgent substantial research questions of comparative and multilingual dimensions. It joins in the call to strengthen international research collaborations, research infrastructures, and open science initiatives to stimulate further work in multilingual automated content analysis.