

*Language Comparison and Corpus Data:
From Monolingual through Parallel to Comparable Corpora*
Marc Kupietz and Beata Trawiński

Since the empirical turn in linguistics, corpus data has become increasingly important both in language-particular and cross-linguistic research. While the choice of a corpus as a data source for a language-specific study is naturally limited to monolingual corpora, several options regarding types of corpora are available for multilingual studies: Multilingual studies can be conducted using multiple monolingual corpora, parallel / translational corpora or comparable corpora. In this talk, we will address all these possibilities and point out their advantages and shortcomings for language comparison. In addition, an initiative will be presented on creating a European Reference Corpus (EuReCo), a language resource using existing national corpora as a base for comparable corpora. We will argue that EuReCo provides a unique linguistic resource offering new perspectives for fine grained comparative research on authentic cross-linguistic data.