Abstract

Besides the social survey data, texts have been an important source of sociological data since the beginning of the development of sociological methodology. Text analysis methods contain two main branches of development: Bernard Berelson's content analysis and Hans-Georg Gadamer's hermeneutic analysis. Both these methodological branches have been influenced by the development of information technologies in the last twenty years. The thesis presented here deals with one of the methods of computer text analysis (CATA), which stands on the border between these two methodological streams, a method of analyzing words' collocations in texts. The thesis presents the method in the context of other methods of text analysis, and mentions sources of inspiration for further development of these methods - corpus linguistics and text mining. The second part discusses the different steps of words' collocation analysis: building a text corpus, dictionary compilation, calculation of data matrix and visualisation of words' distances using multidimensional scaling (MDS). The method is also applied to a specific data, two text corpora compiled from transcripts of biographical interviews with actors of Czechoslovak normalization - with dissidents and Communist functionaries. Quality of the models is assessed, depending on the choice of parameters (distance coefficient, size of the context unit). Subsequently, these models are interpreted. To assess the validity, the interpretations are also confronted with the results of qualitative hermeneutic analysis performed on the same data.