Abstract

The thesis presents a method for diachronic comparison of synchronic corpora that reflect language of very close time periods. Its primary aim is the assessment of possibilities and limitations of language change detection based on the synchronic written SYN-series corpora. The approach is corpus-driven, based on a statistical evaluation of differences among normalized average reduced frequencies of lemmata and lexical combinations. There are several variants of the method applied on various subcorpora of corpus SYN and their results examined in detail.

Difficulty of the comparison lies in the influence of corpus composition and the interconnection of changes in language with changes in society. As it is not easy to distinguish the signs of diachronic shift from naturally existing synchronic variability, the statistically discovered significance of frequency differences is additionally verified by querying the base corpora. The interpretation of the results is also adjusted by the knowledge of their exact composition.

The conclusions are based mainly on the newspapers as a written text type that is most receptive to the changes. The changes can be characterized as a thematic diversion from the original political and economical orientation of the newspapers towards real-life and free-time topics associated with increasing informality of the language. The informality has an impact on shifts in part-of-speech frequencies, frequency increase of a number of core vocabulary lemmata, growing share of semantically weak verbs, substitution of some conventional expressions etc.

The thesis also contributes to the evaluation of composition of the SYN-series corpora, especially the representative corpora SYN2000, SYN2005 and SYN2010. As a result, a number of practical improvements of the concept of representativeness, text categorization and data composition are formulated. The suggestions constitute a valuable feedback for compilation of future SYN-series corpora.