

This bachelor thesis deals with automatic identification of Czech equivalents of German nominal compounds and their linguistic analysis. Compounding is a word formation process which is exploited in both languages, however, in German it is much more productive than in Czech, where the derivation word formation process predominates.

The first part of the thesis copes with identification of Czech counterparts of German compounds with the help of parallel corpora and tools for phrase-based statistical machine translation. After the identification, one-word, two-word and multi-word Czech equivalents were distinguished. The Czech equivalents were analysed according to their part-of-speech tags. Over a half of the German nominal compounds correspond to a sequence of two or more words in Czech, most of the sequences are made up of an adjective and a noun. Morphological structure of one-word equivalents was studied and these equivalents were distinguished into compounds and derivatives, in which the second part of the German compound corresponds to a suffix in the Czech counterpart.