

Tato bakalářská práce se zabývá automatickou identifikací českých ekvivalentů německých substantivních kompozit. Skládání slov je doloženo v obou jazycích, ačkoliv v němčině je více produktivní než v češtině, kde převládá odvozování.

V první části práce jsme se věnovali identifikaci českých protějšků německých kompozit za pomoci paralelních korpusů a nástrojů na statistický strojový překlad založený na frázích. Poté jsme rozdělili české protějšky na jednoslovné, dvouslovné a víceslovné. České ekvivalenty byly analyzovány podle jejich slovnědruhového zařazení. Více než polovina německých substantivních kompozit odpovídá sekvenci dvou nebo více slov v češtině, většina sekvencí je tvořena přídavným a podstatným jménem. U jednoslovných ekvivalentů byla zkoumána jejich morfologická struktura a byla mezi nimi rozlišena kompozita a odvozená slova, kde druhá část německého kompozita odpovídá příponě v českém protějšku.