

Czech and Vietnamese are the national languages of the Czech Republic and Vietnam, respectively. The distinctive features and the shortage of resources renders Czech-Vietnamese machine translation into a difficult task, leading to the fact that no effort has been put into developing a translation tool specifically for the language pair. In this thesis, we develop phrase-based statistical machine translation systems for the language pair and investigate the potential to improve the translation quality with pivoting. Pivoting refers to a set of machine translation approaches through which a natural language, called pivoting language, is introduced to solve the problem of data scarcity between source and target languages, one of the most challenging problems of statistical machine translation. Selecting English as the sole pivoting language for Czech-Vietnamese translation, we prepare training and testing corpora for the three language pairs. All possible corpus sources are explored regarding each specific language pair. The next step is to improve quality of the training corpora through normalizing and filtering. Various experiments with pivoting methods are carried out to analyse the performance of pivoting methods in a realistic working condition.