

Abstract

Statistical-based machine translation is a natural language processing technique based on statistics by using learning model from a group of sentences (corpus) as training data. One of factor which has major influence in training process and determine the quality of translation result is the corpus's phrase translation process.

This final project conducted the implementation of phrase-level correction *rule* method in statistical-based machine translation to know and understand the significance of this method's influence in the quality of statistical-based machine translation's output if applied to translate sentences from English to Indonesian. The implementation was done by make the phrase-level correction rule module, integrate it to statistical-based machine translation, and then compare the output of this phrase-level correction rule's statistical machine translation with some other machine translation's output, and finally analyze the comparison result to make the conclusion of the research.

In this research, found that the phrase-level correction rules decrease the quality of statistical-based machine translation's output, because of it's context-independent and partial monotone alignment behavior, and also because of the better phrase correction method that has been applied in statistical-based machine translation.

Keywords: statistical-based machine translation, corpus, phrase, phrase-level correction rule