

# Abstract

Statistical machine translation for the English-German language pair is prone to errors due to the structural differences between the two languages. Frequent translation errors that result from these differences are: missing words, in particular verbs, incorrect word order and wrong lexical choices. This work investigates two approaches to these problems that target different components of a statistical machine translation system.

The first approach is pre-reordering of the source sentence, such that its word order becomes similar to the word order of the target language. In a recent work, Tromble and Eisner (2009) proposed a framework for modeling the reordering of words as the Linear Ordering Problem. Following this framework, the present work investigates two state-of-the-art linear ordering algorithms as applied to the task of word reordering. Results of automatic evaluation indicate an improvement in translation quality, while error analysis shows that this approach reduces the number of missing verbs, incorrect words and improved word order.

The second approach is to improve the language model component by modeling the dependencies between words in adjacent sentences. Momtazi et al. (2010) proposed the Within-and-Across Sentence Boundary language model (ASB), which models this relation by estimating unigram probabilities based on the word distribution in previous sentences. This thesis presents results of integrating the ASB language model in the phrase-based statistical machine translation system Moses. Automatic evaluation indicates a small contribution from the language model, but further experiments and error analysis show that this model can improve word disambiguation, lexical choices and reduce the number of missing words.