# Abstract

The major aim of our study is to improve phrase-based statistical machine translation (SMT) using syntactic information represented in constituent tree form. In recent years, there have been many studies about syntactic SMT. Most studies rely on formal grammars such as synchronous context-free grammars and tree transducers. The approaches can be different in a number of aspects such as input type for example string or tree, in rule form for example SCFG or xRs, in rule function including word reordering or word choice. Since these studies aim to improve both word reordering and word choice, their grammars have been fully lexicalized. We would like to make a distinction between word order and word choice when statistically modelling the translation process. We suppose that the input of a SMT system is a syntactic tree. Considering word order as a syntactic problem, we define syntactic transformation task which involves the word reordering, the deletion and the insertion of function words. We propose a syntactic transformation model based on the probabilistic context free grammar. By using this model, we studied a number of tree-to-string phrase-based SMT approaches which vary in the way syntactic information is used including preprocessing and decoding and the level of syntactic analysis including chunking and parsing. Our experimental results showed significant improvements in translation quality. Considering word choice as a semantic problem, we aim at incorporating WSD into phrase-based SMT. Our empirical study on this problem reveal various aspect of the integration. Our experiments showed a significant improvement in translation quality.

**Key words**: Computational Linguistics, Statistical Machine Translation, Syntactic Parsing, Word Reordering, Word Sense Disambiguation, Word Choice.