# Unlexicalized Dependency Parser for Variable Word Order Languages based on Local Contextual Pattern

Hoojung Chung and Hae-Chang Rim

Department of Computer Science
Korea University, Seoul 136-701 Korea
{hjchung, rim}@nlp.korea.ac.kr

**Abstract.** We investigate the effect of unlexicalization in a dependency parser for variable word order languages and propose an unlexicalized parser which can utilize some contextual information in order to achieve performance comparable to that of lexicalized parsers. Unlexicalization of an early dependency parser makes performance decrease by 3.6%. However, when we modify the unlexicalized parser into the one which can consider additional contextual information, the parser performs better than some lexicalized dependency parsers, while it requires simpler smoothing processes, less time and space for parsing.

## 1 Introduction

Lexical information has been widely used to achieve a high degree of parsing accuracy, and parsers with lexicalized language models [1–3] have shown the state-of-the-art performances in analyzing English. Most of parsers developed recently use lexical features for syntactic disambiguation, whether they use a phrase structure grammar or a dependency grammar, regardless of languages they deal with.

However, some researchers recently insisted that the lexicalization did not play a big role in parsing with probabilistic context-free grammars (PCFG). [4] showed that the lexical bigram information does not contribute to the performance improvement of a parser. [5] concluded that the fundamental sparseness of the lexical dependency information from parsed training corpora is not helpful to the lexicalized parser, and proposed an accurate unlexicalized parsing model.

This is the story of analyzing fixed word order languages, e.g. English, with a phrase structure grammar. What about parsing other type of languages with other type of grammars, without lexical dependency information? For instance, can an unlexicalized dependency parser for languages with variable word order achieve high accuracy as the unlexicalized PCFG parser for English does?

This paper investigates the effect of the unlexicalization in a dependency parser for variable word order languages and suggests a new unlexicalized parser which can solve the problems of the unlexicalized dependency parser.