

# Word Sense Disambiguation Based on Weight Distribution Model with Multiword Expression

Hee-Cheol Seo, Young-Sook Hwang, and Hae-Chang Rim

Dept. of Computer Science and Engineering, Korea University  
1, 5-ka, Anam-dong, Seongbuk-Gu, Seoul, 136-701, Korea  
{hcseo, yshwang, rim}@nlp.korea.ac.kr

**Abstract.** This paper proposes a two-phase word sense disambiguation method, which filters only the relevant senses by utilizing the multiword expression and then disambiguates the senses based on Weight Distribution Model. Multiword expression usually constrains the possible senses of a polysemous word in a context. Weight Distribution Model is based on the hypotheses that every word surrounding a polysemous word in a context contributes to disambiguating the senses according to its discrimination power. The experiments on English data in SENSEVAL-1 and SENSEVAL-2 show that multiword expression is useful to filter out irrelevant senses of a polysemous word in a given context, and Weight Distribution Model is more effective than Decision Lists.

## 1 Introduction

Word sense disambiguation(WSD) is the task of selecting the correct sense of a word in a context. Many applications of natural language processing(NLP), such as machine translation, information extraction, and question answering, require a semantic analysis where WSD plays a crucial role. With its importance, WSD has been known as a very important field of NLP and studied steadily since the advent of NLP in the 1950s.

One of the most successful current lines of the research is the corpus-based supervised learning approach. Most of the approaches([2][3][4][5]) try to disambiguate the word senses by utilizing the words which co-occur with a polysemous word in a sense tagged corpus. The words that frequently co-occur with a particular sense of the polysemous word are regarded as useful features.

Based on the same observation, multiword expression(MWE) like an idiom or a phrasal verb is particularly valuable at WSD. That is to say, the sense of a polysemous word can be constrained by the multiword expression with the word. For example, given a MWE *every day*, we immediately realize what the word *day* means. Hence, we use the multiword expression including a polysemous word to determine the sense of the polysemous word in a context.

MWE can be used as a sense tagger or a sense filter. As a sense tagger, MWE selects just one sense for a polysemous word in a context. However, if a MWE