Evaluation of Feature Combination
for Effective Structural Disambiguation

So-Young Park, Yong-Jae Kwak, Joon-Ho Lim, and Hae-Chang Rim

NLP Lab., Dept. of CSE, Korea University,
5-ka 1, Anam-dong, Seongbuk-ku, SEOUL, 136-701, KOREA
{ssoya,yjkwak,jhlim,rim}@nlp.korea.ac.kr

Abstract. In this paper, we present the useful features of a syntactic constituent for a probabilistic parsing model and analyze the combination of the features in order to disambiguate parse trees effectively. Unlike most of previous works focusing on the features of a single head, the features of a functional head, the features of a content head, and the features of size are utilized in this study. Experimental results show that the combination of different features such as the functional head feature and the size feature is prefered to the combination of similar features such as the functional head feature and the content head feature. Besides, it is remarkable that the function feature is more useful than the combination of the content feature and the size feature.

1 Introduction

Natural language parsing is regarded as a task of finding the parse tree for a given sentence. A probabilistic approach such as PCFG selects the best parse tree with the highest probability, which is generated by the production rules. However, it cannot select the best parse tree between the parse trees in Figure 1 because of the same CFG rules.

In order to improve the syntactic disambiguation, most of recent parsing models have been lexicalized[1–4] so that they can discriminate between $P(NP/mother \rightarrow NP/mother\ PP/in)$ and $P(NP/portrait \rightarrow NP/portrait\ PP/in)$. Besides, some of them also utilize the inner contexts[2, 5], the outer contexts[3] or the derivational history[4, 6]. Still, the parse tree type selected in Figure 1 is the same as the parse tree type selected for a noun phrase "the portrait of my mother in oil" since the previous models don't consider the relationship between "portrait" and
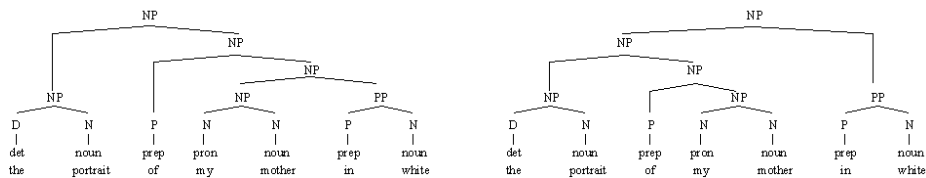

Fig. 1. Syntactic Ambiguities Of A Noun Phrase By The Syntactic Tag