Prepositional Phrase Attachment and Interlingua

Rajat Kumar Mohanty, Ashish Francis Almeida, and Pushpak Bhattacharyya

Department of Computer Science and Engineering
Indian Institute of Technology Bombay
Mumbai - 400076, India
Emails: {rkm, ashishfa, pb}@cse.iitb.ac.in

Abstract. In this paper, we present our work on the classical problem of *prepositional phrase attachment.* This forms part of an interlingua based machine translation system, in which the semantics of the source language sentences is captured in the form of *Universal Networking Language (UNL)* expressions. We begin with a thorough linguistic analysis of six common prepositions in English, namely, *for, from, in, on, to* and *with.* The insights obtained are used to enrich a *lexicon* and a *rule base*, which guide the search for the correct attachment site for the prepositional phrase and the subsequent generation of accurate semantic relations. The system has been tested on British National Corpus, and the accuracy of the results establishes the effectiveness of our approach.

1 Introduction

No natural language processing system can do a meaningful job of analyzing the text, without resolving the prepositional phrase (PP) attachment. There are two fundamental questions related to this problem:

(1) *Given a sentence containing the frame*
[V-NP$_1$-P- NP$_2$]
does NP$_2$ attach to V or to NP$_1$?
(2) *What should be the semantic relation that*
links the PP with the rest of the concept graph of the sentence?

Our work is motivated by seeking answers to these questions. We focus our attention on six most common prepositions of English, *viz., for, from, in, on, to* and *with* (for the motivation, please see Table 5 in section 5).

In order to resolve these issues, we have taken linguistic insights from the following works [1–4]. Other related and motivating works specific to the PP-attachment problem are [5–9].

The roadmap of the paper is as follows: Section 2 provides a linguistic analysis of the six prepositions in question. The UNL system is introduced in Section 3. Section 4 discusses the design and implementation of the system. Evaluation results are given in Section 5. Section 6 concludes the paper and is followed by the references.

Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 241–253.