

The Role of PP Attachment in Preposition Generation

John Lee¹ and Ola Knutsson²

¹ Spoken Language Systems

MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, USA
jsylee@csail.mit.edu

² School of Computer Science and Communication
Royal Institute of Technology, Stockholm, Sweden
knutsson@csc.kth.se

Abstract. This paper is concerned with the task of preposition generation in the context of a grammar checker. Relevant features for this task can range from lexical features, such as words and their part-of-speech tags in the vicinity of the preposition, to syntactic features that take into account the attachment site of the prepositional phrase (PP), as well as its argument/adjunct distinction. We compare the performance of these different kinds of features in a memory-based learning framework. Experiments show that using PP attachment information can improve preposition generation accuracy on Wall Street Journal texts.

1 Introduction

Preposition usage is among the more frequent types of errors made by non-native speakers of English. In an analysis of texts [1], written by students in English-as-a-Second-Language classes, errors involving prepositions form the largest category, at about 29%³. A system that can automatically detect and correct preposition usage would be of much practical and educational value. Research efforts towards building such a grammar checking system have been described in [2], [3], and [4].

When dealing with preposition errors, the system typically makes two decisions. First, a *preposition generation* model needs to determine the best preposition to use, given its context in the input sentence. It should, for example, predict the preposition “*in*” to be the most likely choice for the input sentence:

Input: *He participated at? the competition.*

Corrected: *He participated in the competition.*

If the predicted preposition differs from the original one, a *confidence* model would then need to decide whether to suggest the correction to the user. In this case, confidence in the predicted preposition “*in*” should be much higher than the original “*at*”, and correction would be warranted.

³ As cited in [2].