6. PREPOSITIONAL PHRASE DISAMBIGUATION

Syntactic ambiguity is a notoriously difficult problem. In this chapter we show that lexical level commonsense knowledge can be used, in conjunction with syntactic and preposition-specific knowledge, to handle 99% of the post-verbal prepositional phrase ambiguity in a corpus of geography text. The method discards spurious syntactic ambiguities introduced by post-verbal prepositional phrase attachment during parsing. A completely naive parser will generate three parses for sentences of the form NP-V-NP-PP. The prepositions alone are insufficiently precise in meaning to guide selection among competing parses. The method employs commonsense knowledge of concepts in preference strategies which appeal to the meaning of the preposition combined with information about the verbs and nouns associated with it drawn from the text and from the generic and ontological databases. These determine which syntactic structures generated by a semantically naive parser are commonsensically plausible. The method was successful in 93% of cases tested.

6.1. Semantically Implausible Syntactic Ambiguities

A problem for text understanding systems is that syntactic rules alone produce numerous ambiguities, many of which are not semantically possible (or likely) interpretations. Consider sentence (66), for which any standard parser would produce three distinct syntactic structures. Figure 13 on page 125 is a syntactic tree showing the parse for (66) in which the key belongs to the lock. The *with*-phrase is a constituent of the noun phrase headed by *lock* (NP constituency). Figure 14 on page 126 is a syntactic tree showing the *with*-phrase as a constituent of the verb phrase (VP constituency). Figure 15 on page 127 shows the parse in which the *with*-phrase modifies the sentence (S-
modification), so that the event of buying the lock takes place with the key. Only one of these syntactic possibilities is semantically possible for (66), namely the one in which the prepositional phrase is a complement of the NP whose head is lock. Similarly, only VP constituency is semantically possible for (67), and only S-modification for (68).

(66) John bought the lock with the key.

(67) John bought the lock with five dollars.

(68) John bought the lock in the afternoon.

(69) John took the key to the lock.

Clearly, the semantically implausible syntactic ambiguities generated for (66) - (68) are spurious. On the other hand, some syntactic ambiguities correspond to possible semantic ambiguities. In sentence (69), both the VP constituency and NP constituency parses are semantically possible. It is easy to imagine a situation in which John physically carries the key over to the lock. However, in this case the preferred reading maps to NP constituency because the head of the to-phrase is typically 'a part of' or 'used for' the head of the direct object NP. A text understanding system that can guess NP constituency in this case is not only practical and workable, it is also superior to one which chooses randomly. The commonsense disambiguation method to be described in this chapter assigns constituency for prepositional phrases according to commonsense preference, and the only ambiguities which remain after the preference strategy has been invoked are the semantically and commonsensically possible ambiguities, such as those below. In (70) either the seeing event or the man might have been in the park, and in (71) either the building event or the houses might have been by the sea.

(70) John saw the man in the park.

(71) John built the houses by the sea.