Disambiguating PP-attachment sites with graded semantic data, or how to rule out elephants in pajamas.

This work presents results from a novel approach to prepositional phrase (PP) attachment disambiguation in English. The 1930 Groucho Marx line, “One morning I shot an elephant in my pajamas. How he got into my pajamas I'll never know,” shows that while certain structures may be ambiguous, semantic information helps speakers calculate distinct likelihoods for different attachment sites. This system was designed to capture these graded intuitions rather than approach the problem as an absolute classification task. It uses data from an experiment in which subjects gave graded judgments on the acceptability of interpretations of sentences from the Penn Treebank arising from varied attachment sites. The learning algorithm determines attachment sites by training on these data and also resources from WordNet: the gloss corpus, teleological links, new teleological annotations, and the evocation database. Additionally, the system used semantic tags from dictionary entries of the stimulus words. The goals was to test the hypothesis that the closer the verb and object of the preposition are semantically, the more likely it is that the PP should attach to the verb phrase (VP) rather than intervening noun phrases. Though baseline accuracy, in which VP-attachment is always selected, is quite high (~80%), this system outperforms the baseline and many other previously implemented WordNet systems.