

# Structural Ambiguity and Conceptual Relations

Philip Resnik  
Computer and Information Science  
University of Pennsylvania  
200 South 33rd Street  
Philadelphia, PA, 19104 USA  
resnik@linc.cis.upenn.edu

Marti A. Hearst  
Computer Science Division  
465 Evans Hall  
University of California, Berkeley  
Berkeley, CA 94720 USA  
marti@cs.berkeley.edu

## ABSTRACT

Lexical co-occurrence statistics are becoming widely used in the syntactic analysis of unconstrained text. However, analyses based solely on lexical relationships suffer from sparseness of data: it is sometimes necessary to use a less informed model in order to reliably estimate statistical parameters. For example, the "lexical association" strategy for resolving ambiguous prepositional phrase attachments [Hindle and Rooth, 1991] takes into account only the attachment site (a verb or its direct object) and the preposition, ignoring the object of the preposition.

We investigated an extension of the lexical association strategy to make use of noun class information, thus permitting a disambiguation strategy to take more information into account. Although in preliminary experiments the extended strategy did not yield improved performance over lexical association alone, a qualitative analysis of the results suggests that the problem lies not in the noun class information, but rather in the multiplicity of classes available for each noun in the absence of sense disambiguation. This suggests several possible revisions of our proposal.

## 1. Preference Strategies

Prepositional phrase attachment is a paradigmatic case of the structural ambiguity problems faced by natural language parsing systems. Most models of grammar will not constrain the analysis of such attachments in examples like (1): the grammar simply specifies that a prepositional phrase such as *on computer theft* can be attached in several ways, and leaves the problem of selecting the correct choice to some other process.

- (1) a. Eventually, Mr. Stoll was invited to both the CIA and NSA to brief high-ranking officers on computer theft.
- b. Eventually, Mr. Stoll was invited to both the CIA and NSA [to brief [high-ranking officers on computer theft]].
- c. Eventually, Mr. Stoll was invited to both the CIA and NSA [to brief [high-ranking officers] [on computer theft]].

As [Church and Patil, 1982] point out, the number of analyses given combinations of such "all ways ambiguous" constructions grows rapidly even for sentences of quite

reasonable length, so this other process has an important role to play.

Discussions of sentence processing have focused primarily on structurally-based preference strategies such as right association and minimal attachment [Kimball, 1973; Frazier, 1979; Ford *et al.*, 1982]; [Hobbs and Bear, 1990], while acknowledging the importance of semantics and pragmatics in attachment decisions, propose two syntactically-based attachment rules that are meant to be generalizations of those structural strategies.

Others, however, have argued that syntactic considerations alone are insufficient for determining prepositional phrase attachments, suggesting instead that preference relationships among lexical items are the crucial factor. For example:

- [Wilks *et al.*, 1985] argue that the right attachment rules posited by [Frazier, 1979] are incorrect for phrases in general, and supply counterexamples. They further argue that lexical preferences alone as suggested by [Ford *et al.*, 1982] are too simplistic, and suggest instead the use of preference semantics. In the preference semantics framework, attachment relations of phrases are determined by comparing the preferences emanating from all the entities involved in the attachment, until the best mutual fit is found. Their CASSEX system represents the various meanings of the preposition in terms of (a) the preferred semantic class of the noun or verb that proceeds the preposition (e.g., *move*, *be*, *strike*), (b) the case of the preposition (e.g., *instrument*, *time*, *loc-static*), and (c) the preferred semantic class of the head noun of the prepositional phrase (e.g., *physob*, *event*). The difficult part of this method is the identification of preference relationships and particularly determining the strengths of the preferences and how they should interact. (See also discussion in [Schubert, 1984].)
- [Dahlgren and McDowell, 1986] also suggests using preferences based on hand-built knowledge about the prepositions and their objects, specifying a simpler set of rules than those of [Wilks *et al.*, 1985].

She argues that the knowledge is needed for understanding the text as well as for parsing it. Like CASSEX, this system requires considerable effort to provide hand-encoded preference information.

- [Jensen and Binot, 1987] resolve prepositional phrase attachments by using preferences obtained by applying a set of heuristic rules to dictionary definitions. The rules match against lexico-syntactic patterns in the definitions — for example, if confronted with the sentence *She ate a fish with a fork*, the system evaluates separately the plausibility of the two proposed constructs *eat with fork* and *fish with fork* based on how well the dictionary supports each. By making use of on-line dictionaries, the authors hope to create a system that will scale up; however, they report no overall evaluation.

An empirical study by [Whittemore *et al.*, 1990] supports the main premise of these proposals: they observe that in naturally-occurring data, lexical preferences (e.g., *arrive at*, *flight to*) provide more reliable attachment predictions than structural strategies. Unfortunately, it seems clear that, outside of restricted domains, hand-encoding of preference rules will not suffice for unconstrained text. Information gleaned from dictionaries may provide a solution, but the problem of how to weight and combine preferences remains.

## 2. Lexical Association

[Hindle and Rooth, 1991] offer an alternative method for discovering and using lexical attachment preferences, based on corpus-based lexical co-occurrence statistics. In this section, we briefly summarize their proposal and experimental results.

An “instance” of ambiguous prepositional phrase attachment is taken to consist of a verb, its direct object, a preposition, and the object of the preposition. Furthermore, only the heads of the respective phrases are considered; so, for example, the ambiguous attachment in (1) would be construed as the 4-tuple (*brief, officer, on, theft*).<sup>1</sup> We will refer to its elements as *v*, *n1*, *p*, and *n2*, respectively.

The attachment strategy is based on an assessment of how likely the preposition is, given each potential attachment site; that is, a comparison of the values  $\Pr(p|n1)$  and  $\Pr(p|v)$ . For (1), one would expect  $\Pr(on|brief)$  to be greater than  $\Pr(on|officer)$ , reflecting the intuition that *brief X on Y* is more plausible as a verb phrase than *officer on Z* is as a noun phrase.

Hindle and Rooth extracted their training data from a corpus of Associated Press news stories. A robust parser

<sup>1</sup> Verbs and nouns are reduced to their root forms, hence *officer* rather than *officers*.

[Hindle, 1983] was used to construct a table in which each row contains the head noun of each noun phrase, the preceding verb (if the noun phrase was the verb’s direct object), and the following preposition, if any occurred. Attachment decisions for the training data in the table were then made using a heuristic procedure — for example, given *spare it from*, the procedure will count this row as an instance of *spare from* rather than *it from*, since a prepositional phrase cannot be attached to a pronoun. Not all the data can be assigned with such certainty: ambiguous cases in the training data were handled either by using statistics collected from the unambiguous cases, by splitting the attachment between the noun and the verb, or by defaulting to attachment to the noun.

Given an instance of ambiguous prepositional phrase attachment from the test set, the lexical association procedure for guessing attachments used the t-score [Church *et al.*, 1991] to assess the direction and significance of the difference between  $\Pr(p|n1)$  and  $\Pr(p|v)$  — *t* will be positive, zero, or negative according to whether  $\Pr(p|n1)$  is greater, equal to, or less than  $\Pr(p|v)$ , respectively, and its magnitude indicates the level of confidence in the significance of this difference.

On a set of test sentences held out from the training data, the lexical association procedure made the correct attachment 78.3% of the time. For choices with a high level of confidence (magnitude of *t* greater than 2.1, about 70% of the time), correct attachments were made 84.5% of the time.

## 3. Prepositional Objects

The lexical association strategy performs quite well, despite the fact that the object of the preposition is ignored. However, Hindle and Rooth note that neglecting this information can hurt in some cases. For instance, the lexical association strategy is presented with exactly the same information in (2a) and (2b), and is therefore unable to distinguish them.

- (2) a. Britain reopened its embassy in December.
- b. Britain reopened its embassy in Teheran.

In addition, [Hearst and Church, in preparation] have conducted a pilot study in which human subjects are asked to guess prepositional phrase attachments despite the omission of the direct object, the object of the preposition, or both. The results of this study, though preliminary, suggest that the object of the preposition contributes an amount of information comparable to that contributed by the direct object; more important, for some prepositions, the object of the preposition appears to be *more* informative.

Thus, there appears to be good reason to incorporate the object of the preposition in lexical association calculations. The difficulty, of course, is that the data are far too

sparse to permit the most obvious extension. Attempts to simply compare  $\Pr(p, n2|n1)$  against  $\Pr(p, n2|v)$  using the t-score fail dismally.<sup>2</sup>

#### 4. Word Classes

We are faced with a well-known tradeoff: increasing the number of words attended to by a statistical language model will in general tend to increase its accuracy, but doing so increases the number of probabilities to be estimated, leading to the need for larger (and often impractically larger) sets of training data in order to obtain accurate estimates. One option is simply to pay attention to fewer words, as do Hindle and Rooth. Another possibility, however, is to reduce the number of parameters by grouping words into equivalence classes, as discussed, for example, by [Brown *et al.*, 1990].

[Resnik, 1992] discusses the use of word classes in discovering lexical relationships, demonstrating that WordNet [Beckwith *et al.*, 1991; Miller, 1990], a broad-coverage, hand-constructed lexical database, provides a reasonable foundation upon which to build class-based statistical algorithms. Here we briefly describe WordNet, and in the following section describe its use in resolving prepositional phrase attachment ambiguity.

WordNet is a large lexical database organized as a set of word taxonomies, one for each of four parts of speech (noun, verb, adjective, and adverb). In the noun taxonomy, the only one used here, each word is mapped to a set of word classes, corresponding roughly to word senses, which Miller *et al.* term *synonym sets*. For example, the word *paper* is a member of synonym sets [*newspaper, paper*] and [*composition, paper, report, theme*], among others. For notational convenience, we will refer to each synonym set by its first word, sometimes together with a unique identifier — for example (*paper, 2241323*) and (*newspaper, 2202048*).<sup>3</sup>

The classes in the taxonomy form the nodes of a semantic network, with links to superordinates, subordinates, antonyms, members, parts, etc. In this work only the superordinate/subordinate (i.e., IS-A) links are used — for example, (*newspaper, 2202048*) is a subclass of (*press, 2200204*), which is a subclass of (*print\_media, 2200360*), and so forth.

Denoting the set of words subsumed by a class  $c$  (that is, the set of all words that are a member of  $c$  or any subordinate class) as  $\text{words}(c)$ , the frequency of a class can be estimated as follows:

$$f(c) = \sum_{n \in \text{words}(c)} f(n) \quad (1)$$

<sup>2</sup>This experiment was attempted using expected likelihood estimates, as in [Hindle and Rooth, 1991], with data extracted from the Penn Treebank as described below.

<sup>3</sup>These identifiers differ depending upon the version of WordNet used; the work described in this paper was done using version 1.2.

Owing to multiple inheritance and word sense ambiguity, equation (1) represents only a coarse estimate — for example, each occurrence of a word contributes equally to the count of all classes of which it is a member. However, [Resnik, 1992] estimated class frequencies in a similar fashion with acceptable results.

#### 5. Conceptual Association

In what follows, we propose to extend Hindle and Rooth's lexical association method to take advantage of knowledge about word-class memberships, following a strategy one might call *conceptual association*. From a practical point of view, the use of word classes reduces the sparseness of the training data, permitting us to make use of the object of the preposition, and also decreases the sensitivity of the attachment strategy to the specifics of the training corpus. From a more philosophical point of view, using a strategy based on conceptual rather than purely lexical relationships accords with our intuition that, at least in many cases, much of the work done by lexical statistics is a result of the semantic relationships they indirectly encode.

Our proposal for conceptual association is to calculate a measure of association using the classes to which the direct object and object of the preposition belong, and to select the attachment site for which the evidence of association is strongest. The use of classes introduces two sources of ambiguity. The first, shared by lexical association, is word sense ambiguity: just as lexically-based methods conflate multiple senses of a word into the count of a single token, here each word may be mapped to many different classes in the WordNet taxonomy. Second, even for a single sense, a word may be classified at many levels of abstraction — for example, even interpreted solely as a physical object (rather than a monetary unit), *penny* may be categorized as a (*coin, 3566679*), (*cash, 3566144*), (*money, 3565439*), and so forth on up to (*possession, 11572*).

In our experiments, we adopted the simplest possible approach: we consider *each* classification of the nouns as a source of evidence about association, and combine these sources of evidence to reach a single attachment decision.

**Algorithm 1.** Given  $(v, n1, p, n2)$ ,

1. Let  $C1 = \{c \mid n1 \in \text{words}(c)\}$   
Let  $C2 = \{c \mid n2 \in \text{words}(c)\} = \{c_{2,1}, \dots, c_{2,N}\}$
2. For  $i$  from 1 to  $N$ ,
 
$$c_{1,i} = \operatorname{argmax}_{c \in C1} I(c; p, c_{2,i})$$

$$I_i^n = I(c_{1,i}; p, c_{2,i})$$

$$I_i^v = I(v, p, c_{2,i})$$

3. For  $i$  from 1 to  $N$ ,

$$S_i^n = \text{freq}(c_{1,i}, p, c_{2,i}) I_i^n$$

$$S_i^v = \text{freq}(v, p, c_{2,i}) I_i^v$$

4. Compute a paired samples t-test for a difference of the means of  $S^n$  and  $S^v$ . Let "confidence" be the significance of the test with  $N - 1$  degrees of freedom.
5. Select attachment to  $n1$  or  $v$  according to whether  $t$  is positive or negative, respectively.

Step 1 of the algorithm establishes the range of possible classifications for  $n1$  and  $n2$ . For example, if the algorithm is trying to disambiguate

- (3) But they foresee little substantial progress in exports...

the word *export* can be classified alternatively as  $(\text{export}, 248913)$ ,  $(\text{commerce}, 244370)$ ,  $(\text{group\_action}, 241055)$ , and  $(\text{act}, 10812)$ .

In step 2, each candidate classification for  $n2$  is held fixed, and a classification for  $n1$  is chosen that maximizes the association (as measured by mutual information) between the noun-attachment site and the prepositional phrase. In effect, this answers the question, "If we were to categorize  $n2$  in this way, what would be the best class to use for  $n1$ ?" This is done for each classification of  $n2$ , yielding  $N$  different class-based interpretations for  $(n1, p, n2)$ .  $I_i^n$  is the noun-attachment association score for the  $i$ th interpretation. Correspondingly, there are  $N$  interpretations  $I_i^v$  for  $(v, p, n2)$ .

At this point, each of the  $N$  classifications for  $n1$  (*progress*) and  $n2$  (*export*) provides one possible interpretation of  $(\text{foresee}, \text{progress}, \text{in}, \text{export})$ , and each of these interpretations provides associational evidence in favor of one attachment choice or the other. How are these sources of evidence to be combined?

As a first effort, we have proceeded as follows. Each of the values for  $I_i^n$  and  $I_i^v$  are not equally reliable: values calculated using classes low in the taxonomy involve lower frequencies than those using higher-level classes. In an attempt to assign more credit to scores calculated using higher counts, we weight each of the mutual information scores by the corresponding trigram frequency — thus in step 3 the association score for noun-attachment is calculated as the product of  $f(c_{1,i}, p, c_{2,i}) I_i^n$ . The corresponding verb-attachment score is  $f(v, p, c_{2,i}) I_i^v$ . This leaves us with a table like the following:

foresee...	$S^n$	$S^v$
(situation) in (export)	67.4	39.8
(rise) in (commerce)	178.3	23.8
(advance) in (group_action)	104.9	19.9
(advance) in (act)	149.5	40.6

In step 4 the  $N$  different sources of evidence are combined: a t-test for the difference of the means is performed, treating  $S^n$  and  $S^v$  as paired samples (see, e.g., [Woods *et al.*, 1986]). In step 5 the resulting value of  $t$  determines the choice of attachment site, as well as an estimate of how significant the difference is between the two alternatives. (For this example,  $t(3) = 3.57$ ,  $p < 0.05$ , yielding the correct choice of attachment.)

## 6. Combining Strategies

In addition to evaluating the performance of the conceptual association strategy in isolation, it is natural to combine the predictions of the lexical and conceptual association strategies to make a single prediction. Although well-founded strategies for combining the predictions of multiple models do exist in the speech recognition literature [Jelinek and Mercer, 1980; Katz, 1987], we have chosen a simpler "backing off" style procedure:

**Algorithm 2.** Given  $(v, n1, p, n2)$ ,

1. Calculate an attachment decision using Algorithm 1.
2. If significance  $< 0.1$ , use this decision.
3. Otherwise, use lexical association.

## 7. Experimental Results

### 7.1. Experiment 1

An experiment was conducted to evaluate the performance of the lexical association, conceptual association, and combined strategies. The corpus used was a collection of parses from articles in the 1988–89 *Wall Street Journal*, found as part of the Penn Treebank. This corpus is an order of magnitude smaller than the one used by Hindle and Rooth in their experiments, but it provides considerably less noisy data, since attachment decisions have been performed automatically by the Fidditch parser [Hindle, 1983] and then corrected by hand.

A test set of 201 ambiguous prepositional phrase attachment instances was set aside. After acquiring attachment choices on these instances from a separate judge (who used the full sentence context in each case), the test set was reduced by eliminating sentences for which the separate judge disagreed with the Treebank, leaving a test set of 174 instances.<sup>4</sup>

<sup>4</sup>Of the 348 nouns appearing as part of the test set, 12 were not covered by WordNet; these were classified by default as members of the WordNet class  $(\text{entity}, 2383)$ .

Lexical counts for relevant prepositional phrase attachments (v,p,n2 and n1,p,n2) were extracted from the parse trees in the corpus; in addition, by analogy with Hindle and Rooth's training procedure, instances of verbs and nouns that did not have a prepositional phrase attached were counted as occurring with the "null prepositional phrase." A set of clean-up steps included reducing verbs and nouns to their root forms, mapping to lowercase, substituting the word *someone* for nouns not in WordNet that were part-of-speech-tagged as proper names, substituting the word *amount* for the token % (this appeared as a head noun in phrases such as *rose 10 %*), and expanding month abbreviations such as *Jan.* to the full month name.

The results of the experiment are as follows:

	LA	CA	COMBINED
% Correct	81.6	77.6	82.2

When the individual strategies were constrained to answer only when confident ( $|t| > 2.1$  for lexical association,  $p < .1$  for conceptual association), they performed as follows:

STRATEGY	ANSWERED (%)	ACCURACY (%)
LA	44.3	92.8
CA	67.2	84.6

The performance of lexical association in this experiment is striking: despite the reduced size of the training corpus in comparison to [Hindle and Rooth, 1991], performance exceeds previous results — and although fewer test cases produce confident predictions (as might be expected given generally lower counts), when the algorithm is confident it performs very well indeed.

The performance of the conceptual association strategy seems reasonable, though it is clearly overshadowed by the performance of the lexical association strategy. The tiny improvement on lexical association by the combined strategy suggests that including the conceptual association strategy may improve performance overall, but further investigation is needed to determine whether such a conclusion is warranted; the experiments described in the following two sections bear on this issue.

## 7.2. Experiment 2

Although the particular class-based strategy implemented here might not provide great leaps in performance — at least as judged on the basis of Experiment 1 — one might expect that a strategy based upon a domain-independent semantic taxonomy would provide a greater degree of robustness, reducing dependence of the attachment strategy on the training corpus.

We set out to test this supposition by considering the performance of the various associational attachment

strategies when tested on data from a corpus *other than* the one on which they were trained. First, we tested performance on a test set drawn from the same genre. Of the test cases drawn by Hindle and Rooth from the Associated Press corpus, we took the first 200; eliminating those sentences for which Hindle and Rooth's two human judges could not agree on an attachment reduced the set to 173. Several minor clean-up steps were taken to make this test set consistent with our training data: if the object of the preposition was a complementizer or other word introducing a sentence (e.g. *begin debate on whether*), it was replaced with the word *something*; proper names (e.g. *Bush*) were replaced with *someone*; some numbers were replaced with *year* or *amount*, (e.g. 1911 and 0, respectively); and "compound" prepositions were replaced by a "normal" preposition consistent with what appeared in the full sentence (e.g. *by/about* was replaced with *by* for the phrase *outnumbered losers by about 6 to 5*).

The results of the experiment are as follows:

	LA	CA	COMBINED
% Correct	69.9	72.3	72.8

When the individual strategies were constrained to answer only when confident:

STRATEGY	ANSWERED (%)	ACCURACY (%)
LA	31.8	80.0
CA	49.7	77.9

The conceptual association strategy, not being as dependent on the specific lexical items in the training and test sets, sustains a somewhat higher level of overall performance, although once again the lexical association strategy performs well when restricted to the relatively small set of predictions that it can make with confidence.

## 7.3. Experiment 3

Wishing to pursue the paradigm of cross-corpus testing further, we conducted a third experiment in which the training set was extracted from the Penn Treebank's parsed version of the Brown corpus [Francis and Kucera, 1982], testing on the *Wall Street Journal* test set of Experiment 1.

The results of the experiment are as follows:

	LA	CA	COMBINED
% Correct	77.6	73.6	79.3

When the individual strategies were constrained to answer only when confident:

STRATEGY	ANSWERED (%)	ACCURACY (%)
LA	35.6	85.5
CA	59.2	81.6

In this experiment it is surprising that all the strategies perform as well as they do. However, the pattern of results leads us to conjecture that the conceptual association strategy, taken in combination with the lexical association strategy, may permit us to make more effective use of general, corpus-independent semantic relationships than does the lexical association strategy alone.

## 8. Qualitative Evaluation

The overall performance of the conceptual association strategy tends to be worse than that of lexical association, and the combined strategy yields at best a marginal improvement. However, several comments are in order.

First, the results in the previous section demonstrate that conceptual association is doing some work: when the strategies are constrained to answer only when confident, conceptual association achieves a 50-60% increase in coverage over lexical association, at the cost of a 3-9% decrease in accuracy.

Second, it is clear that class information is providing some measure of resistance to sparseness of data. As mentioned earlier, adding the object of the preposition *without* using noun classes leads to hopelessly sparse data — yet the performance of the conceptual association strategy is far from hopeless. In addition, examination of what the conceptual association strategy actually did on specific examples shows that in many cases it is successfully compensating for sparse data.

- (4) To keep his schedule on track, he flies two personal secretaries in from Little Rock to augment his staff in Dallas.

For example, verb *augment* and preposition *in* never co-occur in the WSJ training corpus, and neither do noun *staff* and preposition *in*; as a result, the lexical association strategy makes an incorrect choice for the ambiguous verb phrase in (4). However, the conceptual association strategy makes the correct decision on the basis of the following classifications:

augment..	$S^n$	$S^v$
{gathering} in {dallas}	38.18	45.54
{people} in {urban.area}	1200.21	28.46
{personnel} in {region}	314.62	23.38
{personnel} in {geo..area}	106.05	26.80
{people} in {city}	1161.22	28.61
{personnel} in {location}	320.85	22.83

Third, mutual information appears to be a successful way to select appropriate classifications for the direct object, given a classification of the object of the preposition (see step 2 in Algorithm 1). For example, despite the fact that *staff* belongs to 25 classes in WordNet

— including {musical\_notation, 2332528} and {rod, 1613297}, for instance — the classes to which it assigned in the above table seem appropriate given the context of (4).

Finally, it is clear that our method for combining sources of evidence — the paired t-test in step 4 of Algorithm 1 — is hurting performance in many instances because (a) it gives equal weight to likely and unlikely classifications of the object of the preposition, and (b) the significance of the test is overestimated when the object of the preposition belongs to many different classes.

- (5) Goodrich's vinyl-products segment reported operating profit for the quarter of \$30.1 million.

For example, given the ambiguous attachment highlighted in (5), the contribution of the time-related classifications of *quarter* ({time\_period, 4014263}, {time, 9819}, etc.) is swamped by numerous other classifications in which *quarter* is interpreted as a physical object (coin, animal part), a number (fraction, rational number), a unit of weight (for measuring grain), and so forth. As a result, the conceptual association strategy comes up with the wrong attachment and identifies its decision as a confident one.

## 9. Conclusions

The conceptual association strategy described here leaves room for a number of improvements. The use of mutual information as an association measure, and the weighting of the mutual information score in order to bias the computation in favor of large counts, warrant further consideration — mutual information has been criticized for, among other things, its poor behavior given low frequencies, and an alternative measures of association may prove better.

In addition, as noted in the previous section, combining evidence using the paired t-test is problematic, essentially because of word-sense ambiguity. One alternative might be to perform sense disambiguation in advance — the results of [Yarowsky, 1993] demonstrate that a significant reduction in the number of possible noun classifications is possible using only very limited syntactic context, rather than global word co-occurrence statistics. Another related alternative would be to select a single *best* classification — for example, using the measure of selectional association proposed in [Resnik, 1993] — rather than considering all possible classifications.

Another possibility to investigate is the incorporation of structurally-based attachment strategies along with lexical and conceptual association. Such a fusion of structural and lexical preference strategies is suggested in [Whittemore *et al.*, 1990], and [Weischedel *et al.*, 1989]

have found that a structural strategy ("closest attachment") performs well in combination with a class-based strategy, although they use a relatively small, domain-specific taxonomy of classes and assume each word has a pointer to a unique class.

Still another direction for future work involves the application of similar techniques to other problems like prepositional phrase attachment for which the resolution of ambiguities would seem to require some form of semantic knowledge. The problems discussed in [Church and Patil, 1982] — including ambiguous prepositional phrase attachment, noun-noun modification, and coordination — would seem to form a natural class of problems to investigate in this manner. Although there will always be ambiguities that can be resolved only by appeal to complex inferences or highly domain-dependent facts, we believe the combination of domain-independent, knowledge-based resources such as WordNet with corpus-based statistics may provide the semantic power necessary for solving many instances of such problems, *without* the need for general reasoning about world knowledge.

### References

- [Beckwith *et al.*, 1991] Richard Beckwith, Christiane Fellbaum, Derek Gross, and George Miller. WordNet: A lexical database organized on psycholinguistic principles. In Uri Zernik, editor, *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pages 211–232. Erlbaum, 1991.
- [Brown *et al.*, 1990] Peter F. Brown, Vincent J. Della Pietra, Peter V. deSouza, and Robert L. Mercer. Class-based n-gram models of natural language. In *Proceedings of the IBM Natural Language ILL*, pages 283–298. Paris, France, March 1990.
- [Church and Patil, 1982] Kenneth W. Church and Ramesh Patil. Coping with syntactic ambiguity or how to put the block in the box on the table. *American Journal of Computational Linguistics*, 8(3-4):139–149, 1982.
- [Church *et al.*, 1991] Kenneth Church, William Gale, Patrick Hanks, and Donald Hindle. Using statistics in lexical analysis. In Uri Zernik, editor, *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pages 116–164. Erlbaum, 1991.
- [Dahlgren and McDowell, 1986] K. Dahlgren and J. McDowell. Using commonsense knowledge to disambiguate prepositional phrase modifiers. In *AAAI-86*, pages 589–593, 1986.
- [Ford *et al.*, 1982] Marilyn Ford, Joan Bresnan, and Ronald Kaplan. A competence-based theory of syntactic closure. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, 1982.
- [Francis and Kucera, 1982] W. Francis and H. Kucera. *Frequency Analysis of English Usage*. Houghton Mifflin Co.: New York, 1982.
- [Frazier, 1979] L. Frazier. *On comprehending Sentences: Syntactic Parsing Strategies*. PhD thesis, University of Massachusetts, 1979.
- [Hearst and Church, in preparation] Marti A. Hearst and Kenneth W. Church. An investigation of the use of lexical associations for prepositional phrase attachment, (in preparation).
- [Hindle and Rooth, 1991] D. Hindle and M. Rooth. Structural ambiguity and lexical relations. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, June 1991. Berkeley, California.
- [Hindle, 1983] Donald Hindle. User manual for Fidditch, a deterministic parser. Technical memorandum 7590-142. Naval Research Laboratory, 1983.
- [Hobbs and Bear, 1990] Jerry R. Hobbs and John Bear. Two principles of parse preference. In *Proceedings of 13th COLING*, pages 162–167, Helsinki, 1990.
- [Jelinek and Mercer, 1980] Frederick Jelinek and Robert L. Mercer. Interpolated estimation of Markov source parameters from sparse data. In *Proceedings of the Workshop on Pattern Recognition in Practice*, Amsterdam, The Netherlands: North-Holland, May 1980.
- [Jensen and Binot, 1987] Karen Jensen and Jean-Louis Binot. Disambiguating prepositional phrase attachments by using on-line dictionary definitions. *American Journal of Computational Linguistics*, 13(3):251–260, 1987.
- [Katz, 1987] Slava M. Katz. Estimation of probabilities from sparse data for the language model component of a speech recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-35(3):400–401, March 1987.
- [Kimball, 1973] John Kimball. Seven principles of surface structure parsing in natural language. *Cognition*, 2:15–47, 1973.
- [Miller, 1990] George Miller. Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4), 1990. (Special Issue).
- [Resnik, 1992] Philip Resnik. WordNet and distributional analysis: A class-based approach to lexical discovery. In *AAAI Workshop on Statistically-based NLP Techniques*, San Jose, California, July 1992.
- [Resnik, 1993] Philip Resnik. Semantic classes and syntactic ambiguity. ARPA Workshop on Human Language Technology, March 1993. Princeton.
- [Schubert, 1984] Lenhart Schubert. On parsing preferences. In *COLING-84*, 1984.
- [Weischedel *et al.*, 1989] Ralph Weischedel, Marie Meteer, Richard Schwartz, and Jeff Palmucci. Coping with ambiguity and unknown words through probabilistic models. *ms.*, 1989.
- [Whittemore *et al.*, 1990] Greg Whittemore, Kathleen Ferrara, and Hans Brunner. Empirical study of predictive powers of simple attachment schemes for post-modifier prepositional phrases. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 23–30, 1990. Pittsburgh, Pennsylvania.
- [Wilks *et al.*, 1985] Yorick Wilks, Xiuming Huang, and Dan Fass. Syntax, preference and right attachment. In *IJCAI-85*, pages 779–784, 1985.
- [Woods *et al.*, 1986] Anthony Woods, Paul Fletcher, and Arthur Hughes. *Statistics in Language Studies*. Cambridge Textbooks in Linguistics. Cambridge University Press: Cambridge, England, 1986.
- [Yarowsky, 1993] David Yarowsky. One sense per collocation. ARPA Workshop on Human Language Technology, March 1993. Princeton.