STRUCTURAL AMBIGUITY AND LEXICAL RELATIONS

Donald Hindle and Mats Rooth
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, NJ 07974

Abstract

We propose that ambiguous prepositional phrase attachment can be resolved on the basis of the relative strength of association of the preposition with noun and verb, estimated on the basis of word distribution in a large corpus. This work suggests that a distributional approach can be effective in resolving parsing problems that apparently call for complex reasoning.

Introduction

Prepositional phrase attachment is the canonical case of structural ambiguity, as in the time worn example,

(1) I saw the man with the telescope

The existence of such ambiguity raises problems for understanding and for language models. It looks like it might require extremely complex computation to determine what attaches to what. Indeed, one recent proposal suggests that resolving attachment ambiguity requires the construction of a discourse model in which the entities referred to in a text must be reasoned about (Altmann and Steedman 1988). Of course, if attachment ambiguity demands reference to semantics and discourse models, there is little hope in the near term of building computational models for unrestricted text to resolve the ambiguity.

Structure based ambiguity resolution

There have been several structure-based proposals about ambiguity resolution in the literature; they are particularly attractive because they are simple and don't demand calculations in the semantic or discourse domains. The two main ones are:

- Right Association - a constituent tends to attach to another constituent immediately to its right (Kimball 1973).

- Minimal Attachment - a constituent tends to attach so as to involve the fewest additional syntactic nodes (Frazier 1978).

For the particular case we are concerned with, attachment of a prepositional phrase in a verb + object context as in sentence (1), these two principles – at least in the version of syntax that Frazier assumes – make opposite predictions: Right Association predicts noun attachment, while Minimal Attachment predicts verb attachment.

Psycholinguistic work on structure-based strategies is primarily concerned with modeling the time course of parsing and disambiguation, and proponents of this approach explicitly acknowledge that other information enters into determining a final parse. Still, one can ask what information is relevant to determining a final parse, and it seems that in this domain structure-based disambiguation is not a very good predictor. A recent study of attachment of prepositional phrases in a sample of written responses to a "Wizard of Oz" travel information experiment shows that neither Right Association nor Minimal Attachment account for more than 55% of the cases (Whittemore et al. 1990). And experiments by Taraban and McClelland (1988) show that the structural models are not in fact good predictors of people's behavior in resolving ambiguity.

Resolving ambiguity through lexical associations

Whittemore et al. (1990) found lexical preferences to be the key to resolving attachment ambiguity. Similarly, Taraban and McClelland found lexical content was key in explaining people's behavior. Various previous proposals for guiding attachment disambiguation by the lexical content of specific

words have appeared (e.g. Ford, Bresnan, and Kaplan 1982; Marcus 1980). Unfortunately, it is not clear where the necessary information about lexical preferences is to be found. In the Whittemore et al. study, the judgement of attachment preferences had to be made by hand for exactly the cases that their study covered; no precompiled list of lexical preferences was available. Thus, we are posed with the problem: how can we get a good list of lexical preferences.

Our proposal is to use cooccurrence of with prepositions in text as an indicator of lexical preference. Thus, for example, the preposition *to* occurs frequently in the context *send NP* —, i.e., after the object of the verb *send*, and this is evidence of a lexical association of the verb *send* with *to*. Similarly, *from* occurs frequently in the context *withdrawal* —, and this is evidence of a lexical association of the noun *withdrawal* with the preposition *from*. Of course, this kind of association is, unlike lexical selection, a symmetric notion. Cooccurrence provides no indication of whether the verb is selecting the preposition or vice versa. We will treat the association as a property of the pair of words. It is a separate matter, which we unfortunately cannot pursue here, to assign the association to a particular linguistic licensing relation. The suggestion which we want to explore is that the association revealed by textual distribution – whether its source is a complementation relation, a modification relation, or something else – gives us information needed to resolve the prepositional attachment.

Discovering Lexical Association in Text

A 13 million word sample of Associated Press new stories from 1989 were automatically parsed by the Fidditch parser (Hindle 1983), using Church's part of speech analyzer as a preprocessor (Church 1988). From the syntactic analysis provided by the parser for each sentence, we extracted a table containing all the heads of all noun phrases. For each noun phrase head, we recorded the following preposition if any occurred (ignoring whether or not the parser attached the preposition to the noun phrase), and the preceding verb if the noun phrase was the object of that verb. Thus, we generated a table with entries including those shown in Table 1.

In Table 1, example (a) represents a passivized instance of the verb *blame* followed by the prepo-

	VERB	HEAD NOUN	PREP
(a)	blame	PASSIVE	for
(b)		money	for
(c)		development	
(d)	control	government	
(e)		military	
(f)		accord	
(g)	enrage	radical	
(h)		WHPRO	
(i)	spare	it	from
(j)	grant	concession	to
(k)	determine	flaw	in

Table 1: A sample of the Verb-Noun-Preposition table.

sition *for*. Example (b) is an instance of a noun phrase whose head is *money*; this noun phrase is not an object of any verb, but is followed by the preposition *for*. Example (c) represents an instance of a noun phrase with head noun *development* which neither has a following preposition nor is the object of a verb. Example (d) is an instance of a noun phrase with head *government*, which is the object of the verb *control* but is followed by no preposition. Example (j) represents an instance of the ambiguity we are concerned with resolving: a noun phrase (head is *concession*), which is the object of a verb (*grant*), followed by a preposition (*to*).

From the 13 million word sample, 2,661,872 noun phrases were identified. Of these, 467,920 were recognized as the object of a verb, and 753,843 were followed by a preposition. Of the noun phrase objects identified, 223,666 were ambiguous verb-noun-preposition triples.

Estimating attachment preferences

Of course, the table of verbs, nouns and prepositions does not directly tell us what the strength lexical associations are. There are three potential sources of noise in the model. First, the parser in some cases gives us false analyses. Second, when a preposition follows a noun phrase (or verb), it may or may not be structurally related to that noun phrase (or verb). (In our terms, it may attach to that noun phrase or it may attach somewhere else). And finally, even if we get accurate attachment information, it may be that fre-

quency of cooccurrence is not a good indication of strength of attachment. We will proceed to build the model of lexical association strength, aware of these sources of noise.

We want to use the verb-noun-preposition table to derive a table of bigrams, where the first term is a noun or verb, and the second term is an associated preposition (or no preposition). To do this we need to try to assign each preposition that occurs either to the noun or to the verb that it occurs with. In some cases it is fairly certain that the preposition attaches to the noun or the verb; in other cases, it is far less certain. Our approach is to assign the clear cases first, then to use these to decide the unclear cases that can be decided, and finally to arbitrarily assign the remaining cases. The procedure for assigning prepositions in our sample to noun or verb is as follows:

1. No Preposition - if there is no preposition, the noun or verb is simply counted with the null preposition. (cases (c-h) in Table 1).

2. Sure Verb Attach 1 - preposition is attached to the verb if the noun phrase head is a pronoun. (i in Table 1)

3. Sure Verb Attach 2 - preposition is attached to the verb if the verb is passivized (unless the preposition is *by*. The instances of *by* following a passive verb were left unassigned.) (a in Table 1)

4. Sure Noun Attach - preposition is attached to the noun, if the noun phrase occurs in a context where no verb could license the prepositional phrase (i.e., the noun phrase is in subject or pre-verbal position.) (b, if pre-verbal)

5. Ambiguous Attach 1 - Using the table of attachment so far, if a t-score for the ambiguity (see below) is greater than 2.1 or less than -2.1, then assign the preposition according to the t-score. Iterate through the ambiguous triples until all such attachments are done. (j and k may be assigned)

6. Ambiguous Attach 2 - for the remaining ambiguous triples, split the attachment between the noun and the verb, assigning .5 to the noun and .5 to the verb. (j and k may be assigned)

7. Unsure Attach - for the remaining pairs (all of which are either attached to the preceding noun or to some unknown element), assign them to the noun. (b, if following a verb)

This procedure gives us a table of bigrams representing our guess about what prepositions associate with what nouns or verbs, made on the basis of the distribution of verbs nouns and prepositions in our corpus.

The procedure for guessing attachment

Given the table of bigrams, derived as described above, we can define a simple procedure for determining the attachment for an instance of verb-noun-preposition ambiguity. Consider the example of sentence (2), where we have to choose the attachment given verb *send*, noun *soldier*, and preposition *into*.

(2) Moscow sent more than 100,000 soldiers into Afganistan ...

The idea is to contrast the probability with which *into* occurs with the noun *soldier* (P(into | soldier)) with the probability with which *into* occurs with the verb *send* (P(into | send)). A t-score is an appropriate way to make this contrast (see Church et al. to appear). In general, we want to calculate the contrast between the conditional probability of seeing a particular preposition given a noun with the conditional probability of seeing that preposition given a verb.

$$ t \equiv \frac{P(prep \mid noun) - P(prep \mid verb)}{\sqrt{\sigma^2(P(prep \mid noun)) + \sigma^2(P(prep \mid verb))}} $$

We use the "Expected Likelihood Estimate" (Church et al., to appear) to estimate the probabilities, in order to adjust for small frequencies; that is, given a noun and verb, we simply add 1/2 to all bigram frequency counts involving a preposition that occurs with either the noun or the verb, and then recompute the unigram frequencies. This method leaves the order of t-scores nearly intact, though their magnitude is inflated by about 30%. To compensate for this, the 1.65 threshold for significance at the 95% level should be adjusted up to about 2.15.

Consider how we determine attachment for sentence (2). We use a t-score derived from the adjusted frequencies in our corpus to decide whether the prepositional phrase *into Afganistan* is attached to the verb (root) *send/V* or to the noun (root) *soldier/N*. In our corpus, *soldier/N* has an adjusted frequency of 1488.5, and *send/V* has an adjusted frequency of 1706.5; *soldier/N* occurred in 32 distinct preposition contexts, and *send/V* in

60 distinct preposition contexts; $f(send/V\ into) = 84$, $f(soldier/N\ into) = 1.5$.

From this we calculate the t-score as follows:[1]

$$t \equiv \frac{P(w|soldier/N) - P(w|send/V)}{\sqrt{\sigma^2(P(w|soldier/N)) + \sigma^2(P(w|send/V))}}$$

$$\approx \frac{\frac{f(soldier/N\ into)+1/2}{f(soldier/N)+V/2} - \frac{f(send/V\ into)+1/2}{f(send/V)+V/2}}{\sqrt{\frac{f(soldier/N\ into)+1/2}{(f(soldier/N)+V/2)^2} + \frac{f(send/V\ into)+1/2}{(f(send/V)+V/2)^2}}}$$

$$= \frac{\frac{1.5+1/2}{1488.5+70/2} - \frac{84+1/2}{1706.5+70/2}}{\sqrt{\frac{1.5+1/2}{1488.5+70/2)^2} + \frac{84+1/2}{1706.5+70/2)^2}}} \approx -8.81$$

This figure of -8.81 represents a significant association of the preposition *into* with the verb *send*, and on this basis, the procedure would (correctly) decide that *into* should attach to *send* rather than to *soldier*. Of the 84 *send/V into* bigrams, 10 were assigned by steps 2 and 3 ('sure attachements').

Testing Attachment Preference

To evaluate the performance of this procedure, first the two authors graded a set of verb-noun-preposition triples as follows. From the AP new stories, we randomly selected 1000 test sentences in which the parser identified an ambiguous verb-noun-preposition triple. (These sentences were selected from stories included in the 13 million word sample, but the particular sentences were excluded from the calculation of lexical associations.) For every such triple , each author made a judgement of the correct attachment on the basis of the three words alone (forced choice - preposition attaches to noun or verb). This task is in essence the one that we will give the computer - i.e., to judge the attachment without any more information than the preposition and the head of the two possible attachment sites, the noun and the verb. This gave us two sets of judgements to compare the algorithm's performance to.

---

[1]V is the number of distinct preposition contexts for either *soldier/N* or *send/V*; in this case V = 70. Since 70 bigram frequencies f(*soldier/N p*) are incremented by 1/2, the unigram frequency for *soldier/N* is incremented by 70/2.

Judging correct attachment

We also wanted a standard of correctness for these test sentences. To derive this standard, we together judged the attachment for the 1000 triples a second time, this time using the full sentence context.

It turned out to be a surprisingly difficult task to assign attachment preferences for the test sample. Of course, many decisions were straightforward; sometimes it is clear that a prepositional phrase is and argument of a noun or verb. But more than 10% of the sentences seemed problematic to at least one author. There are several kinds of constructions where the attachment decision is not clear theoretically. These include idioms (3-4), light verb constructions (5), small clauses (6).

(3) But over time , misery has *given way to* mending .

(4) The meeting will *take place in* Quantico

(5) Bush has said he would not *make cuts in* Social Security

(6) Sides said Francke *kept* a .38-caliber *revolver in* his car 's glove compartment

We chose always to assign light verb constructions to noun attachment and small clauses to verb attachment.

Another source of difficulty arose from cases where there seemed to be a systematic ambiguity in attachment.

(7) ...known to *frequent* the same *bars in* one neighborhood.

(8) Inaugural officials reportedly were trying to *arrange* a *reunion for* Bush and his old submarine buddies ...

(9) We have not *signed* a settlement *agreement with* them

Sentence (7) shows a systematic locative ambiguity: if you frequent a bar and the bar is in a place, the frequenting event is arguably in the same place. Sentence (8) shows a systematic benefactive ambiguity: if you arrange something for someone, then the thing arranged is also for them. The ambiguity in (9) arises from the fact that if someone is one of the joint agents in the signing of an agreement, that person is likely to be a party to the agreement. In general, we call an attachment systematically ambiguous when, given our understanding of the semantics, situations which

make the interpretation of one of the attachments true always (or at least usually) also validate the interpretation of the other attachment.

It seems to us that this difficulty in assigning attachment decisions is an important fact that deserves further exploration. If it is difficult to decide what licenses a prepositional phrase a significant proportion of the time, then we need to develop language models that appropriately capture this vagueness. For our present purpose, we decided to force an attachment choice in all cases, in some cases making the choice on the bases of an unanalyzed intuition.

In addition to the problematic cases, a significant number (120) of the 1000 triples identified automatically as instances of the verb-object-preposition configuration turned out in fact to be other constructions. These misidentifications were mostly due to parsing errors, and in part due to our underspecifying for the parser exactly what configuration to identify. Examples of these misidentifications include: identifying the subject of the complement clause of *say* as its object, as in (10), which was identified as (*say ministers from*); misparsing two constituents as a single object noun phrase, as in (11), which was identified as (*make subject to*); and counting non-object noun phrases as the object as in (12), identified as (*get hell out_of*).

> (10) Ortega also *said* deputy foreign *ministers from* the five governments would meet Tuesday in Managua, ...
>
> (11) Congress made a deliberate choice to *make this commission subject to* the open meeting requirements ...
>
> (12) Student Union, *get the hell out of* China!

Of course these errors are folded into the calculation of associations. No doubt our bigram model would be better if we could eliminate these items, but many of them represent parsing errors that cannot readily be identified by the parser, so we proceed with these errors included in the bigrams.

After agreeing on the 'correct' attachment for the sample of 1000 triples, we are left with 880 verb-noun-preposition triples (having discarded the 120 parsing errors). Of these, 586 are noun attachments and 294 verb attachments.

Evaluating performance

First, consider how the simple structural attachment preference schemas perform at predicting the

	choice		% correct		
	N	V	N	V	total
Judge 1	575	305	91.7	80.7	87.8
Judge 2	511	369	94.3	71.8	84.9
LA	557	323	85.4	65.9	78.3

Table 2: Performance on the test sentences for 2 human judges and the lexical association procedure (LA).

outcome in our test set. *Right Association*, which predicts noun attachment, does better, since in our sample there are more noun attachments, but it still has an error rate of 33%. *Minimal Attachment*, interpreted to mean verb attachment, has the complementary error rate of 67%. Obviously, neither of these procedures is particularly impressive.

Now consider the performance of our attachment procedure for the 880 standard test sentences. Table 2 shows the performance for the two human judges and for the lexical association attachment procedure.

First, we note that the task of judging attachment on the basis of verb, noun and preposition alone is not easy. The human judges had overall error rates of 10-15%. (Of course this is considerably better than always choosing noun attachment.) The lexical association procedure based on t-scores is somewhat worse than the human judges, with an error rate of 22%, but this also is an improvement over simply choosing the nearest attachment site.

If we restrict the lexical association procedure to choose attachment only in cases where its confidence is greater than about 95% (i.e., where t is greater than 2.1), we get attachment judgements on 607 of the 880 test sentences, with an overall error rate of 15% (Table 3). On these same sentences, the human judges also showed slight improvement.

Underlying Relations

Our model takes frequency of cooccurrence as evidence of an underlying relationship, but makes no attempt to determine what sort of relationship is involved. It is interesting to see what kinds of relationships the model is identifying. To investigate this we categorized the 880 triples ac-

	choice		% correct		
	N	V	N	V	total
Judge 1	433	174	94.2	81.0	90.4
Judge 2	394	213	95.6	69.9	86.7
LA	403	204	93.0	67.6	84.5

Table 3: Performance on the test sentences for 2 human judges and the lexical association procedure (LA) for test triples where $t > 2.1$

relation	count	%correct
argument noun	375	88.5
argument verb	103	86.4
adjunct noun	91	72.5
adjunct verb	101	61.3
light verb	19	63.1
small clause	13	84.6
idiom	20	65.0
locative ambiguity	37	75.7
systematic ambiguity	37	64.8
other	84	61.9

Table 4: Performance of the Lexical attachment procedure by underlying relationship

cording to the nature of the relationship underlying the attachment. In many cases, the decision was difficult. Even the argument/adjunct distinction showed many gray cases between clear participants in an action (arguments) and clear temporal modifiers (adjuncts). We made rough best guesses to partition the cases into the following categories: argument, adjunct, idiom, small clause, locative ambiguity, systematic ambiguity, light verb. With this set of categories, 84 of the 880 cases remained so problematic that we assigned them to category *other*.

Table 4 shows the performance of the lexical attachment procedure for these classes of relations. Even granting the roughness of the categorization, some clear patterns emerge. Our approach is quite successful at attaching arguments correctly; this represents some confirmation that the associations derived from the AP sample are indeed the kind of associations previous research has suggested are relevant to determining attachment. The procedure does better on arguments than on adjuncts, and in fact performs rather poorly on adjuncts of verbs (chiefly time and manner phrases). The remaining cases are all hard in some way, and the performance tends to be worse on these cases, showing clearly for a more elaborated model.

Sense Conflations

The initial steps of our procedure constructed a table of frequencies with entries f(x,p), where x is a noun or verb root string, and p is a preposition string. These primitives might be too coarse, in that they do not distinguish different senses of a preposition, noun, or verb. For instance, the temporal use of *in* in the phrase *in December* is identified with a locative use *in Teheran*. As a result, the procedure LA necessarily makes the same attach-

ment prediction for *in December* and *in Teheran* occurring in the same context. For instance, LA identifies the tuple *reopen embassy in* as an NP attachment (t-score 5.02). This is certainly incorrect for (13), though not for (14).[2]

(13) Britain *reopened* the *embassy in* December

(14) Britain *reopened* its *embassy in* Teheran

Similarly, the scalar sense of *drop* exemplified in (15) sponsors a preposition *to*, while the sense represented in *drop the idea* does not. Identifying the two senses may be the reason that LA makes no attachment choice for *drop resistance to* (derived from (16)), where the score is -0.18.

(15) exports are expected to *drop* a further 1.5 *percent to* 810,000

(16) persuade Israeli leaders to *drop* their *resistance to* talks with the PLO

We experimented with the first problem by substituting an abstract preposition *in%MONTH* for all occurrences of *in* with a month name as an object. While the tuple *reopen embassy in%MONTH* was correctly pushed in the direction of a verb attachment (-1.34), in other cases errors were introduced, and there was no compelling general improvement in performance. In tuples of the form *drop/grow/increase percent in%MONTH* , derived from examples such as (16), the preposition was incorrectly attached to the noun *percent*.

---

[2](13) is a phrase from our corpus, while (14) is a constructed example.

234

(16) Output at mines and oil wells dropped 1.8 percent in February

(17) *1.8 percent was dropped by output at mines and oil wells

We suspect that this reveals a problem with our estimation procedure, not for instance a paucity of data. Part of the problem may be the fact that adverbial noun phrase headed by *percent* in (16) does not passivize or pronominalize, so that there are no sure verb attachment cases directly corresponding to these uses of scalar motion verbs.

Comparison with a Dictionary

The idea that lexical preference is a key factor in resolving structural ambiguity leads us naturally to ask whether existing dictionaries can provide useful information for disambiguation. There are reasons to anticipate difficulties in this regard. Typically, dictionaries have concentrated on the 'interesting' phenomena of English, tending to ignore mundane lexical associations. However, the Collins Cobuild English Language Dictionary (Sinclair et al. 1987) seems particularly appropriate for comparing with the AP sample for several reasons: it was compiled on the basis of a large text corpus, and thus may be less subject to idiosyncrasy than more arbitrarily constructed works; and it provides, in a separate field, a direct indication of prepositions typically associated with many nouns and verbs. Nevertheless, even for Cobuild, we expect to find more concentration on, for example, idioms and closely bound arguments, and less attention to the adjunct relations which play a significant role in determining attachment preferences.

From a machine-readable version of the dictionary, we extracted a list of 1535 nouns associated with a particular preposition, and of 1193 verbs associated with a particular preposition after an object noun phrase. These 2728 associations are many fewer than the number of associations found in the AP sample. (see Table 5.)

Of course, most of the preposition association pairs from the AP sample end up being nonsignificant; of the 88,860 pairs, fewer than half (40,869) occur with a frequency greater than 1, and only 8337 have a t-score greater than 1.65. So our sample gives about three times as many significant preposition associations as the COBUILD dictionary. Note however, as Table 5 shows, the overlap is remarkably good, considering the large space of possible bigrams. (In our bigram table

Source	Total	NOUN	VERB
COBUILD	2728	1535	1193
AP sample	88,860	64,629	24,231
AP sample ($f > 1$)	40,869	31,241	9,628
AP sample ($t > 1.65$)	8,337	6,307	2,030
COBUILD ∩ AP	1,931	1,147	784
COBUILD ∩ AP ($t > 1.65$)	1,040	656	384

Table 5: Count of noun and verb associations for COBUILD and the AP sample

there are over 20,000 nouns, over 5000 verbs, and over 90 prepositions.) On the other hand, the lack of overlap for so many cases – assuming that the dictionary and the significant bigrams actually record important preposition associations – indicates that 1) our sample is too small, and 2) the dictionary coverage is widely scattered.

First, we note that the dictionary chooses attachments in 182 cases of the 880 test sentences. Seven of these are cases where the dictionary finds an association between the preposition and both the noun and the verb. In these cases, of course, the dictionary provides no information to help in choosing the correct attachment.

Looking at the 175 cases where the dictionary finds one and only one association for the preposition, we can ask how well it does in predicting the correct attachment. Here the results are no better than our human judges or than our bigram procedure. Of the 175 cases, in 25 cases the dictionary finds a verb association when the correct association is with the noun. In 3 cases, the dictionary finds a noun association when the correct association is with the verb. Thus, overall, the dictionary is 86% correct.

It is somewhat unfair to use a dictionary as a source of disambiguation information; there is no reason to expect that a dictionary to provide information on all significant associations; it may record only associations that are interesting for some reason (perhaps because they are semantically unpredictable.) Table 6 shows a small sample of verb-preposition associations from the AP sam-

235

	AP sample	COBUILD
approach	about (4.1)	
	with (2.4)	
appropriate		for
approve	for (2.5)	
approximate		to
arbitrate		between
argue		with
arm	with (2.5)	with
arraign	as (3.2)	
	in (2.4)	
	on (4.1)	on
arrange		for
	through (5.9)	
array		in
arrest	after (3.4)	
	along_with (6.1)	
	during (3.1)	
		for
	on (2.8)	
	while (3.9)	
arrogate		to
ascribe		to
ask	about (4.3)	about
assassinate	in (2.4)	
assemble	at (3.8)	
assert	over (5.8)	
assign	to (5.1)	to
assist	in (2.4)	in
		with
associate	with (6.4)	with

Table 6: Verb-(NP)-Preposition associations in AP sample and COBUILD.

ple and from Cobuild. The overlap is considerable, but each source of information provides intuitively important associations that are missing from the other.

Conclusion

Our attempt to use lexical associations derived from distribution of lexical items in text shows promising results. Despite the errors in parsing introduced by automatically analyzing text, we are able to extract a good list of associations with prepositions, overlapping significantly with an existing dictionary. This information could easily be incorporated into an automatic parser, and additional sorts of lexical associations could similarly be derived from text. The particular approach to deciding attachment by t-score gives results nearly

as good as human judges given the same information. Thus, we conclude that it may not be necessary to resort to a complete semantics or to discourse models to resolve many pernicious cases of attachment ambiguity.

It is clear however, that the simple model of attachment preference that we have proposed, based only on the verb, noun and preposition, is too weak to make correct attachments in many cases. We need to explore ways to enter more complex calculations into the procedure.

References

Altmann, Gerry, and Mark Steedman. 1988. Interaction with context during human sentence processing. *Cognition*, 30, 191-238.

Church, Kenneth W. 1988. A stochastic parts program and noun phrase parser for unrestricted text, *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas.

Church, Kenneth W., William A. Gale, Patrick Hanks, and Donald Hindle. (to appear). Using statistics in lexical analysis. in Zernik (ed.) *Lexical acquisition: using on-line resources to build a lexicon*.

Ford, Marilyn, Joan Bresnan and Ronald M. Kaplan. 1982. A competence based theory of syntactic closure, in Bresnan, J. (ed.) *The Mental Representation of Grammatical Relations*. MIT Press.

Frazier, L. 1978. *On comprehending sentences: Syntactic parsing strategies*. PhD. dissertation, University of Connecticut.

Hindle, Donald. 1983. User manual for fidditch, a deterministic parser. Naval Research Laboratory Technical Memorandum 7590-142.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language, *Cognition*, 2, 15-47.

Marcus, Mitchell P. 1980. *A theory of syntactic recognition for natural language*. MIT Press.

Sinclair, J., P. Hanks, G. Fox, R. Moon, P. Stock, et al. 1987. *Collins Cobuild English Language Dictionary*. Collins, London and Glasgow.

Taraban, Roman and James L. McClelland. 1988. Constituent attachment and thematic role assignment in sentence processing: influences of content-based expectations, *Journal of Memory and Language*, 27, 597-632.

Whittemore, Greg, Kathleen Ferrara and Hans Brunner. 1990. Empirical study of predictive powers of simple attachment schemes for post-modifier prepositional phrases. *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, 23-30.