STRUCTURAL AMBIGUITY AND LEXICAL RELATIONS

Donald Hindle and Mats Rooth

AT&T Bell Laboratories

Abstract

We propose that many ambiguous prepositional phrase attachments can be resolved on the basis of the relative strength of association of the preposition with verbal and nominal heads, estimated on the basis of distribution in an automatically parsed corpus. This suggests that a distributional approach can provide an approximate solution to parsing problems that, in the worst case, 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

An analysis where the prepositional phrase $[_{\text{PP}}$ with the telescope] is part of the object noun phrase has the semantics "the man who had the telescope"; an analysis where the PP has a higher attachment (perhaps as daughter of VP) is associated with a semantics where

[0]We thank Bill Gale, Ken Church and David Yarowsky for many helpful discussions of this work, and are grateful to four reviewers and Christian Rohrer for their comments on an earlier version. The new affiliation of the second author is: Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.

the seeing is achieved by means of a telescope. The existence of such ambiguity raises problems 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 are represented and reasoned about (Altmann and Steedman 1988). We take this argument to show that reasoning essentially involving reference in a discourse model is implicated in resolving attachment ambiguities in a certain class of cases. If this phenomenon is typical, there is little hope in the near term for building computational models capable of resolving such ambiguities in unrestricted text.

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 acknowledges 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 that 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 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.

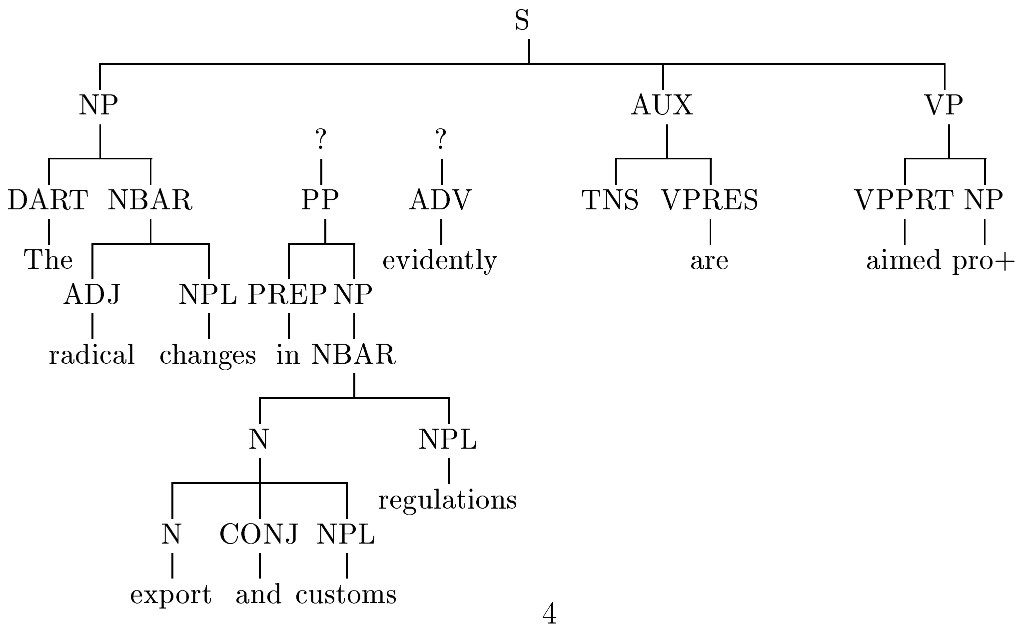
Our proposal is to use co-occurrence of verbs and nouns with prepositions in a large body of text as an indicator of lexical preference. Thus, for example, the preposition *to* occurs frequently in the context *send NP* —, that is after the object of the verb *send*. This is evidence of a lexical association 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*. This kind of association is a symmetric notion: it provides no indication of whether the preposition is selecting the verbal or nominal head, or vice versa. We will treat the association as a property of the pair of words. It is a separate issue, which will not be concerned with in the initial part of this paper, 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 prepositional attachment in the majority of cases.

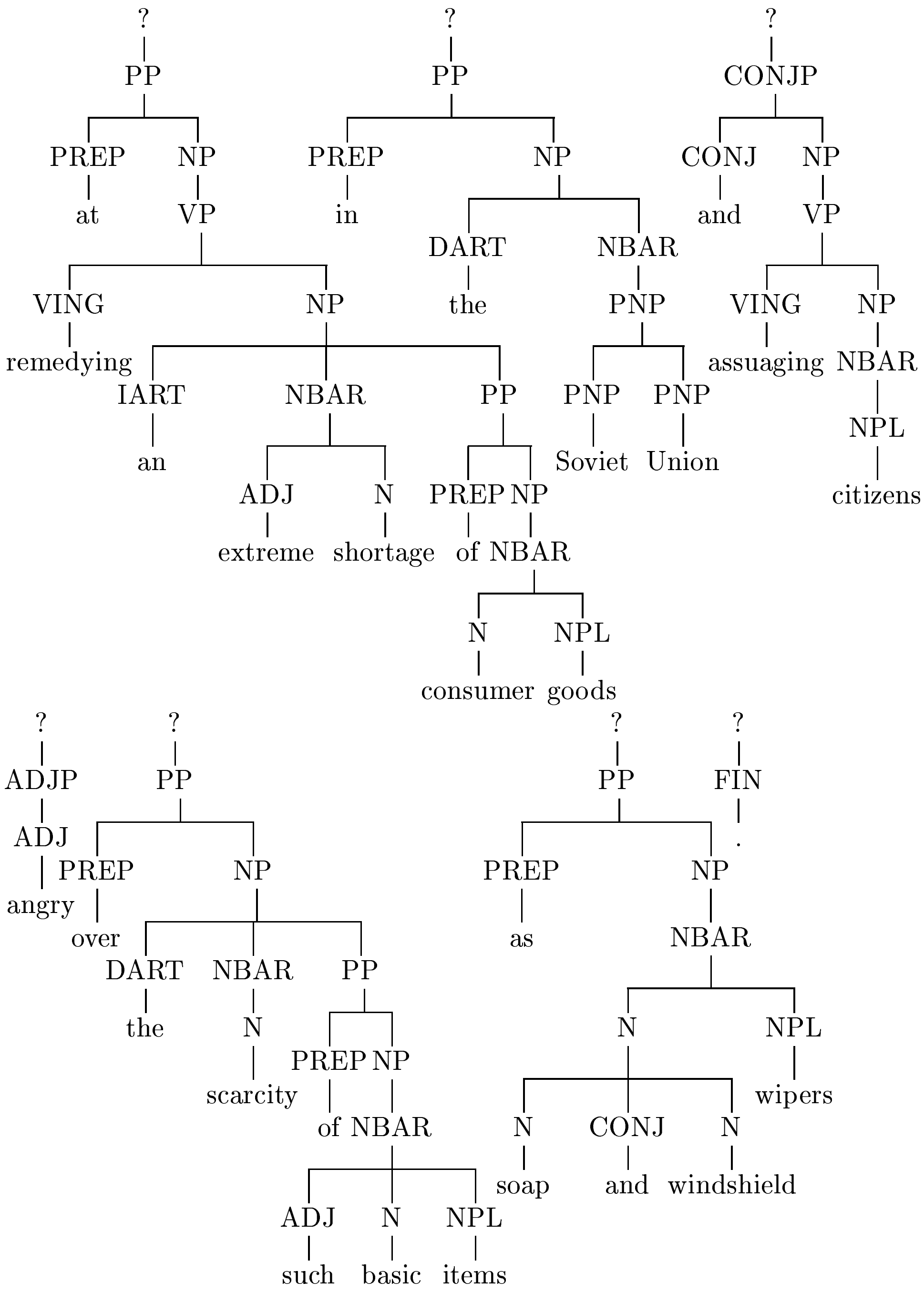
Discovering Lexical Association in Text

A 13 million word sample of Associated Press news stories from 1989 were automatically parsed by the Fidditch parser (Hindle 1983, to appear), using Church's part of speech analyzer as a preprocessor (Church 1988), a combination which we will call simply "the parser". The parser produces a single partial syntactic description of a sentence. Consider the sentence (2), and its parsed representation (3). The information in the tree representation is partial in the sense that some attachment information is missing: the nodes dominated by "?" have not been integrated into the syntactic representation. Note in particular that many PPs have not been attached. This is a symptom of the fact that the parser does not (in many cases) have the kind of lexical information which we have just claimed is required in resolving PP attachment.

(2) The radical changes in export and customs regulations evidently are aimed at remedying an extreme shortage of consumer goods in the Soviet Union and assuaging citizens angry over the scarcity of such basic items as soap and windshield wipers.

(3)





From the syntactic analysis provided by the parser, we extracted a table containing 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 had attached the preposition to the noun

	Verb	Noun	Prep	Syntax
a.		change	in	-V
b.		regulation		
c.	aim	PRO-+	at	
d.	remedy	shortage	of	
e.		good	in	
f.		DART-PNP		
g.	assuage	citizen		
h.		scarcity	of	
i.		item	as	
j.		wiper		
k.		VING		
l.		VING		

Table 1: A sample of NP heads, preceding verbs, and following prepositions derived from the parsed corpus.

phrase), and the preceding verb if the noun phrase was the object of that verb. The entries in Table 1 are those generated from the text above. Each noun phrase in (3) is associated with an entry in the Noun column of the table. Usually this is simply the root of the head of the noun phrase: *good* is the root of the head of *consumer goods*. Noun phrases with no head, or where the head is not a common noun, are coded in a special way: *DART-PNP* represents a noun phrase beginning with a definite article and headed by a proper noun, and *VING* represents a gerundive noun phrase. *PRO-+* represents the empty category which, in the syntactic theory underlying the parser, is assumed to be the object of the passive verb *aimed*. In cases where a prepositional phrase follows the noun phrase, the head preposition appears in the Prep column; attached and unattached prepositional phases generate the same kinds of entries. If the noun phrase is an object, the root of the governing verb appears in the

Verb column: *aim* is the root of *aimed*, the verb governing the empty category [$_{\text{PRO}}$ +]. The last column in the table, labeled Syntax, marks with the symbol -*V* all cases where there is no preceding verb that might license the preposition: the initial subject of sentence (2) is such a case.

In 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 object noun phrases identified, 223,666 were ambiguous verb-noun-preposition triples.

Estimating Associations

The table of verbs, nouns and prepositions is in several respects an imperfect source of information about lexical associations. First, the parser gives us incorrect analyses in some cases. For instance, in the analysis partially described in (4a), the parser incorrectly classified *probes* as a verb, resulting in a table entry *probe lightning in.* Similarly, in (4b), the infinitival marker *to* has been misidentified as a preposition.

(4) a. [$_{\text{NP}}$The space] [$_{\text{VPRES}}$ probes] [$_{\text{NP}}$detected lightning] [$_{\text{PP}}$ in Jupiter's upper atmosphere] and observed auroral emissions like Earth 's northern lights in the Jovian polar regions.

b. The Bush administration told Congress on Tuesday it wants to [$_{\text{V}}$ preserve] [$_{\text{NP}}$the right] [$_{\text{PP}}$ [$_{\text{P}}$ to] control entry] to the United States of anyone who was ever a Communist .

Second, a preposition in an entry might be structurally related to neither the noun of the entry nor the verb (if there is one), even if the entry is derived from a correct parse.[1] For instance, the phrase headed by the preposition might have a higher locus of attachment:

[1]For present purposes, we can consider a parse correct if it contains no incorrect information in the relevant area. Provided the PPs in (5) are unattached, the parses would be correct in this sense. The incorrect information is added by our table construction step, which (given our interpretation of the table) assumes that a preposition following an object NP modifies either the NP or its governing verb.

(5) a. The Supreme Court today agreed to consider reinstating the murder conviction of a New York City man who confessed to $[_{\text{VING}}$ killing] $[_{\text{NP}}$his former girlfriend] $[_{\text{P}}$ after] police illegally arrested him at his home.

b. NBC was so afraid of hostile advocacy groups and unnerving advertisers that it shot its dramatization of the landmark court case that $[_{\text{VPAST}}$ legalized] $[_{\text{NP}}$abortion] $[_{\text{PP}}$ under two phony script titles]

The temporal phrase headed by *after* modifies *confess*, but given the procedure described above, (5a) results in a tuple *kill girlfriend after*. In the second example, a tuple *legalize abortion under* is extracted, although the PP headed by *under* modifies the higher verb *shot*.

Finally, entries of the form *verb noun preposition* do not tell us whether to induce a lexical association between *verb* and *preposition*, or between *noun* and *preposition*. We will view the first two problems as noise which we do not have the means to eliminate, and partially address the third problem in a procedure we will now describe. We want to use the verb-noun-preposition table to derive a table of bigrams counts, where a bigram is a pair consisting of a noun or verb and 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 whether the preposition attaches to the noun or the verb; in other cases, this 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 divide the data in the remaining unresolved cases between the two hypotheses (verb and noun attachment). The procedure for assigning prepositions is as follows:

1. No Preposition - if there is no preposition, the noun or verb is simply entered with a special symbol *NULL*, conceived of as the null preposition. (Cases b, f, g, and j-l in Table 1 are assigned).

2. Sure Verb Attach 1 - the preposition is attached to the verb if the noun phrase head is a pronoun.

3. Sure Verb Attach 2 - the 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. (Item c in Table 1 is assigned).

4. Sure Noun Attach - the preposition is attached to the noun, if the noun phrase occurs in a context where no verb could license the prepositional phrase, specifically if the noun phrase is in a subject or other pre-verbal position. The required syntactic information is present in the last column of the table derived from the parse. (Item a in Table 1 is assigned.)

5. Ambiguous Attach 1 - Using the table of attachments computed so far, if the LA-score for the ambiguity (a score that compares the probability of noun versus verb attachment, as described below) is greater than 2.0 or less than -2.0, then assign the preposition according to the LA-score. Iterate until this step produces no new attachments. (Item d in Table 1 may be assigned.)

6. Ambiguous Attach 2 - for the remaining ambiguous triples, split the datum between the noun and the verb, assigning a count of .5 to the noun-preposition pair and .5 to the verb-preposition pair.. (Item d in Table 1 is assigned, if not assigned in the previous step.)

7. Unsure Attach - assign remaining pairs to the noun. (Items e, h and i in Table 1 are assigned.)

This procedure gives us bigram counts representing the frequency with which a given noun occurs associated with an immediately following preposition (or no preposition), or a given verb occurs in a transitive use and is associated with a preposition immediately following the object of the verb. We use the following notation: $f(w,p)$ is the frequency count for the pair consisting of the verb or noun w and the preposition p. The unigram frequency count for the word w (either a verb, noun, or preposition) can be viewed as a sum of bigram frequencies, and is written $f(w)$. For instance, if p is a preposition, $f(p) = \sum_w f(w,p)$.

The procedure for guessing attachment

Our object is to develop a procedure to guess whether a preposition is attached to the verb or its object when a verb and its object are followed by a preposition. We assume that in each case of attachment ambiguity, there is a forced choice between two outcomes: the preposition attaches either to the verb or to the noun.[2] For example, in sentence 6, we want to choose between two possibilities: either *into* is attached to the verb *send* or it is attached to the noun *soldier*.

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

In particular, we want to choose between two structures:

(7) a. verb_attach structure: $[_{\mathrm{VP}} \textit{send}\ [_{\mathrm{NP}} ..\ \textit{soldier NULL}]\ [_{\mathrm{PP}}\ \textit{into}\ ..]\ ..]$

b. noun_attach structure: $[_{\mathrm{VP}} \textit{send}\ [_{\mathrm{NP}} ..\ \textit{soldier}\ [_{\mathrm{PP}}\ \textit{into}\ ..]]\ ..\ ]$

For the verb_attach case, we require not only that the preposition attach to the verb *send* but also that the noun *soldier* have no following prepositional phrase attached: since *into* directly follows the head of the object noun phrase, there is no room for any post-modifier of the noun *soldier*. We use the notation *NULL* to emphasize that in order for a preposition licensed by the verb to be in the immediately post-nominal position, the noun must have no following complements (or adjuncts). For the case of noun attachment, the verb may or may not have additional prepositional complements following the prepositional phrase associated with the noun.

Since we have a forced choice between two outcomes, it is appropriate to use a likelihood ratio to compare the attachment probabilities (cf. Mosteller and Wallace 1964).[3] In partic-

[2] Thus we are ignoring the fact that the preposition may in fact be licensed by neither the verb nor the noun, as in sentences 5a-b.

[3] In earlier versions of this paper we used a t-test for deciding attachment and a different procedure for estimating the probabilities. The current procedure has several advantages. Unlike the t-test used previously, it is sensitive to the magnitude of the difference between the two probabilities, not to our confidence in our ability to estimate those probabilities accurately. And our estimation procedure has the property that it defaults (in case of novel words) to the average behavior for nouns or verbs, for instance reflecting a default

ular, we look at the log of the ratio of the probability of verb_attach to the probability of noun_ attach. We will call this log likelihood ratio the LA (lexical association) score.

$$LA(v,n,p) = log_2 \frac{P(verb\_attach\ p \mid v,n)}{P(noun\_attach\ p \mid v,n)}$$

For the current example,

$$P(verb\_attach\ into \mid send_{\mathrm{V}}, soldier_{\mathrm{N}}) \approx P(into|send_{\mathrm{V}}) * P(NULL|soldier_{\mathrm{N}})$$

and

$$P(noun\_attach\ into \mid send_{\mathrm{V}}, soldier_{\mathrm{N}}) \approx P(into|soldier_{\mathrm{N}})$$

Again, the probability of noun attachment does not involve a term indicating that the verb sponsors no (additional) complement; when we observe a prepositonal phrase which is in fact attached to the object NP, the verb might or might not have a complement or adjunct following the object phrase.

We can estimate these probabilities from the table of co-occurrence counts as:[4]

$$P(into|send_{\mathrm{V}}) \approx \frac{f(send_{\mathrm{V}}, into)}{f(send_{\mathrm{V}})} = \frac{86}{1742.5} = .049$$

$$P(NULL|soldier_{\mathrm{N}}) \approx \frac{f(soldier_{\mathrm{N}}, NULL)}{f(soldier_{\mathrm{N}})} = \frac{1182}{1478} = .800$$

$$P(into|soldier_{\mathrm{N}}) \approx \frac{f(soldier_{\mathrm{N}}, into)}{f(soldier_{\mathrm{N}})} = \frac{1}{1478} = .0007$$

Thus, the LA score for this example is:

$$LA(send_{\mathrm{V}}, soldier_{\mathrm{N}}, into) = log_2 \frac{.049 * .800}{.0007} \approx 5.81$$

---

preference with *of* for noun attachment.

[4]The non-integral count for *send* is a consequence of the data-splitting step Ambiguous Attach 2, and the definition of unigram frequencies as a sum of bigram frequencies.

The LA score has several useful properties. The sign indicates which possibility, verb attachment or noun attachment, is more likely; an LA score of zero means they are equally likely. The magnitude of the score indicates how much more probable one outcome is than the other. For example, if the LA score is 2.0, then the probability of verb attachment is 4 times greater than noun attachment. Depending on the task, we can require a certain threshold of LA score magnitude before making a decision.[5]

As usual, in dealing with counts from corpora we must confront the problem of how to estimate probabilities when counts are small. The maximum likelihood estimate described above is not very good when frequencies are small, and when frequencies are zero, the formula will not work at all. We use a crude adjustment to observed frequencies which has the right general properties, though it is not likely to be a very good estimate when frequencies are small. For our purposes however – exploring in general the relation of distribution in a corpus to attachment disambiguation – we believe it is sufficient. Other approaches to adjusting small frequencies are discussed in (Church et al. 1991) and in (Gale et al. 1992).

---

[5]An advantage of the likelihood ratio approach is that we can use it in a *Bayesian discrimination* framework to take into account other factors that might influence our decision about attachment (see Gale et al. (1992) for a discussion of this approach). We know of course that other information has a bearing on the attachment decision. For example, we have observed that if the noun phrase object includes a superlative adjective as a premodifier, then noun attachment is certain (for a small sample of 16 cases). We could easily take this into account by setting the prior odds ratio to heavily favor noun attachment: let's suppose that if there is a superlative in the object noun phrase, then noun attachment is say 1000 times more probable than verb attachment; otherwise, they are equally probable. Then following Mosteller and Wallace (1964), we assume that

$$Final\ attachment\ odds = log_2(initial\ odds) + LA$$

In case there is no superlative in the object, the initial log odds will be zero (verb and noun attachment are equally probable), and the final odds will equal our LA score. If there is a superlative,

$$Final\ attachment\ odds = log_2 \frac{1}{1000} + LA(v, n, p)$$

The idea is to use the typical association rates of nouns and verbs to interpolate our probabilities. Where $f(N,p) = \sum_n f(n,p)$, $f(V,p) = \sum_v f(v,p)$, $f(N) = \sum_n f(n)$ and $f(V) = \sum_v f(v)$, we redefine our probability estimates in the following way:

$$P(p \mid n) = \frac{f(n,p) + \frac{f(N,p)}{f(N)}}{f(n) + 1}$$

$$P(p \mid v) = \frac{f(v,p) + \frac{f(V,p)}{f(V)}}{f(v) + 1}$$

When $f(n)$ is zero, the estimate for $P(p \mid n)$ is the average $(\frac{f(N,p)}{f(N)})$ across all nouns, and similarly for verbs. When $f(n,p)$ is zero, the estimate used is proportional to this average. If we have seen only one case of a noun and it occurred with a preposition p (that is $f(n,p) = 1$ and $f(n) = 1$), then our estimate is nearly cut in half. This is the kind of effect we want, since under these circumstances we are not very confident in 1 as an estimate of $P(p \mid n)$. When $f(n,p)$ is large, the adjustment factor does not make much difference. In general, this interpolation procedure adjusts small counts in the right direction and has little effect when counts are large.

For our current example, this estimation procedure changes the LA score little:

$$LA(send_{\mathrm{V}}, soldier_{\mathrm{N}}, into) = log_2 \frac{\frac{f(\mathrm{send_V,into}) + \frac{f(\mathrm{V,into})}{f(\mathrm{V})}}{f(\mathrm{send_V}) + 1} \; \frac{f(\mathrm{soldier_N,NULL}) + \frac{f(\mathrm{N,NULL})}{f(\mathrm{N})}}{f(\mathrm{soldier_N}) + 1}}{\frac{f(\mathrm{soldier_N,into}) + \frac{f(\mathrm{N,into})}{f(\mathrm{N})}}{f(\mathrm{soldier_N}) + 1}}$$

$$= log_2 \frac{\frac{86 + \frac{2292}{385435}}{1742.5+1} \; \frac{1182 + \frac{2047311}{2656594}}{1478+1}}{\frac{1 + \frac{2434}{2656594}}{1478+1}} = 5.87$$

The LA score of 5.87 for this example is positive and therefore indicates verb attachment; the magnitude is large enough to suggest a strong preference for verb attachment.

This method of calculating the LA score was used both to decide unsure cases in building the bigram tables as described in Ambiguous Attach 1, and to make the attachment decisions in novel ambiguous cases, as discussed in the sections following.

Testing Attachment

To evaluate the performance of the procedure, 1000 test sentences in which the parser identified an ambiguous verb-noun-preposition triple were randomly selected from the AP news stories. 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. The two authors first guessed attachments on the verb-noun-preposition triples, making a judgement on the basis of the three head words alone. The judges were required to make a choice in each instance. This task is in essence the one that we will give the computer – to judge the attachment without any more information than the preposition and the heads of the two possible attachment sites.

This initial step provides a rough indication of what we might expect to be achievable based on the information our procedure is using. We also wanted a standard of correctness for the test sentences. We again judged the attachment for the 1000 triples, this time using the full sentence context, first grading the test sentences separately, and then discussing examples on which there was disagreement. Disambiguating the test sample turned out to be a surprisingly difficult task. While many decisions were straightforward, 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 as in (8, 9), light verb constructions (10), and small clauses (11).

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

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

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

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

In the case of idioms, we made the assignment of the basis of a guess about the syntactic structure of the idiom, though this was sometimes difficult to judge. We chose always to assign light verb constructions to noun attachment, based on the fact that the noun supplies the lexical information about what prepositions are possible, and small clauses to verb attachment, based on the fact that this is a predicative construction lexically licensed by the verb.

Another difficulty arose with cases where there seemed to be a systematic semantically based indeterminacy about the attachment. In the situation described by (12a), the bar and the described event or events are presumably in the same location, and so there is no semantic reason to decide on one attachment. Sentence (12b) shows a systematic benefactive indeterminacy: if you arrange something for someone, then the thing arranged is also for them. The problem in (12c) is that signing an agreement usually involves two participants who are also parties to the agreement. (13) gives some further examples drawn from another test sample.

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

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

c. We have not *signed* a settlement *agreement with* them

(13)a. It said the rebels *issued* a *challenge to* soldiers in the area and fought with them for 30 minutes .

b. The worst such attack came Nov. 11 when a death squad firing submachine guns *killed* 43 *people in* the northwest town of Segovia .

c. Another charge raised at the Contra news conference was that the Sandinistas have *mined roads along* the Honduran and Costa Rican borders .

d. Buckner said Control Data is in the process of *negotiating* a new lending *agreement with* its banks .

e. ... which would require ailing banks and S&Ls to *obtain* advance *permission from* regulators before raising high-cost deposits through money brokers.

f. She said the organization has *opened* a cleaning *center in* Seward and she was going to Kodiak to open another.

In general, we can say that an attachment is semantically indeterminate if situations which verify the meaning associated with one attachment also make the meaning associated with the other attachment true. Even a substantial overlap (as opposed to identity) between the classes of situations verifying the two meanings makes an attachment choice difficult.

The problems in determining attachments are heterogeneous. The idiom, light verb, and small clause constructions represent cases where the simple distinction between noun attachment and verb attachment perhaps does not make sense, or is very theory-dependent. It seems to us that the phenomenon of semantically based indeterminacy deserves further exploration. If it is often difficult to decide what licenses a prepositional phrase, we need to develop language models that appropriately capture this. For our present purpose, we decided to make an attachment choice in all cases, in some cases relying on controversial theoretical considerations, or relatively unanalyzed intuitions.

In addition to the problematic cases, 120 of the 1000 triples identified automatically as instances of the verb-object-preposition configuration turned out in fact to be other constructions, often as the result of parsing errors. Examples of this kind were given above, in the context of our description of the construction of the verb-noun-preposition table. Some further misidentifications which showed up in the test sample are: identifying the subject of the complement clause of *say* as its object, as in (10), which was identified as (*say ministers from*), and misparsing two constituents as a single object noun phrase, as in (11), which was identified as (*make subject to*).

(14)a. Ortega also *said* deputy foreign *ministers from* the five governments would meet Tuesday in Managua, ...

b. Congress made a deliberate choice to *make this commission subject to* the open meeting requirements ...

After agreeing on the 'correct' attachment for the test sample, we were left with 880 disambiguated verb-noun-preposition triples, having discarded the examples which were not instances of the relevant construction. Of these, 586 are noun attachments and 294 are verb attachments.

LA	actual N	actual V		precision	recall
N guess	496	89	N	.848	.846
V guess	90	205	V	.695	.697
neither	0	0	combined	.797	.797
Judge 1	actual N	actual V		precision	recall
N guess	527	48	N	.917	.899
V guess	59	246	V	.807	.837
neither	0	0	combined	.878	.878
Judge 2	actual N	actual V		precision	recall
N guess	482	29	N	.943	.823
V guess	104	265	V	.718	.901
neither	0	0	combined	.849	.849

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

Evaluating performance

First, consider how the simple structural attachment preference schemas perform at predicting the outcome in our test set. Right Association predicts noun attachment and does better, since in our sample there are more noun attachments, but it still has an error rate of 33%. Minimal Attachment, interpreted as entailing verb attachment, has the complementary error rate of 67%. Obviously, neither of these procedures is particularly impressive.

Now consider the performance of our lexical association (LA) 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 figures in the entry labeled "combined precision" indicate that the human judges had overall error rates of 12-15%.[6] The lexical association procedure is somewhat worse than the human judges, with an error rate of 20%, but this is an improvement over the structural strategies.

The table also gives results broken down according to N vs. V attachment. The *precision* figures indicate the proportion of test items assigned to a given category which actually belong to the category. For instance, N precision is the fraction of cases which the procedure identified a s N attachments which actually were N attachments. The *recall* figures indicate the proportion of test items actually belonging to a given category which were assigned to that category: N precision is the fraction of actual N attachments which were identified as N attachments. The LA procedure recognized about 85% of the 586 actual noun attachment examples as noun attachments, and about 70% of the actual verb attachments as verb attachments.

If we restrict the lexical association procedure to choose attachment only in cases where the absolute value of the LA score is greater than 2.0 (an arbitrary threshold indicating that the probability of one attachment is four times greater than the other), we get attachment judgements on 621 of the 880 test sentences, with overall precision of about 89%. On these same examples, the judges also showed improvement, as evident in table Table 3.[7]

The fact that an LA score threshold improves precision indicates that the LA score gives information about how confident we can be about an attachment choice. In some applications, this information is useful. For instance, suppose that we wanted to incorporate the PP attachment procedure in a parser such as Fidditch. It might be preferable to achieve increased precision in PP attachment, in return for leaving some PPs unattached. For this purpose, a threshold could be used. Table 4 shows the combined precision and recall levels

[6] Combined precision is the proportion of assigned items which were correctly assigned. Combined recall is the proportion of all items which were correctly assigned. These differ if some items are left unassigned, as in Table 3.

[7] The recall figures for the judges do not have the usual interpretation, since the judges did not have the option of passing on test items. The ratios given in parentheses indicate the proportion of items in the relevant category which were both correctly labeled by the judge, and had a likelihood ratio score with absolute value greater than 2.0. Thus these figures are of interest only in relation to the recall figures for LA.

LA	actual N	actual V		precision	recall
N guess	416	31	N	.931	.710
V guess	39	135	V	.776	.459
neither	131	128	combined	.887	.626
Judge 1	actual N	actual V		precision	recall
N guess	424	20	N	.955	(.724)
V guess	31	146	V	.825	(.497)
neither	131	128	combined	.918	(.648)
Judge 2	actual N	actual V		precision	recall
N guess	402	12	N	.971	(.686)
V guess	53	154	V	.744	(.524)
neither	131	128	combined	.895	(.632)

Table 3: Performance for the LA procedure with a cutoff of 2 ($\mid LA \mid > 2.0$), and for the human judges on the sentences classified by LA

at various LA thresholds. It is clear that the LA score can be used effectively to trade off precision and recall, with a floor for the forced choice at about 80%.

A comparison of table 3 with table 2 indicates, however, that the decline in recall is severe for V attachment. And in general, the performance of the LA procedure is worse on V attachment examples than on N attachments, according to both precision and recall criteria. The next section is concerned with a classification of the test examples which gives insight into why performance on V attachments is worse.

Underlying relations

Our model takes frequency of co-occurrence as evidence of an underlying relationship, but makes no attempt to determine what sort of relationship is involved. It is interesting to

LA score	cases covered	precision	recall
16.0	103 (11.7%)	0.990	0.116
8.0	298 (33.9)	0.966	0.327
4.0	478 (54.3)	0.923	0.501
3.0	530 (60.2)	0.917	0.552
2.0	621 (70.6)	0.887	0.626
1.5	666 (75.7)	0.871	0.659
1.0	718 (81.6)	0.852	0.695
0.5	796 (90.5)	0.823	0.744
0.25	840 (95.5)	0.807	0.770
0	880 (100.0)	0.797	0.797

Table 4: Precision recall tradeoff for the LA procedure

see what kinds of relationships are responsible for the associations the model is identifying. To investigate this we categorized the 880 triples according to the nature of the relationship underlying the attachment. In many cases, the decision was difficult. The argument/adjunct distinction showed many gray cases between clear participants in an action and clear adjuncts, such as temporal modifiers. We made rough best guesses to partition the cases into the following categories: argument, adjunct, idiom, small clause, systematic locative indeterminacy, other systematic indeterminacy, and light verb. With this set of categories, 78 of the 880 cases remained so problematic that we assigned them to the category *other*.

Table 5 shows the proportion of items in a given category which were assigned the correct attachment by the lexical association procedure. Even granting the roughness of the categorization, some clear patterns emerge. Our approach is most successful at attaching arguments correctly. Notice that the 378 noun arguments constitute 65% of the total 586 noun attachments, while the 104 verb arguments amount to only 35% of the 294 verb attachments. Furthermore, performance with verb adjuncts is worse than with noun adjuncts.

relation	count	%correct
argument noun	378	92.1
argument verb	104	84.6
adjunct noun	91	74.7
adjunct verb	101	64.4
light verb	19	57.9
small clause	13	76.9
idiom	19	63.2
locative indeterminacy	42	69.0
systematic indeterminacy	35	62.9
other	78	61.5

Table 5: Performance of the Lexical attachment procedure by underlying relationship

Thus much of the problem with V attachments noted in the previous section appears to be attributable to a problem with adjuncts, particularly verbal ones. Performance on verbal arguments remains worse than performance on nominal ones, however.

The remaining cases are all complex in some way, and the performance is poor on these classes, showing clearly the need for a more elaborated model of the syntactic structure which is being identified.

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 information relevant to disambiguation. The Collins COBUILD English Language Dictionary (Sinclair et al. 1987) is useful for a comparison 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 other works, and it provides, in a separate field, a direct indication of prepositions typically associated with many nouns and verbs. From a machine-readable version of the dictionary, we extracted a list of 1942 nouns associated with a particular preposition, and of 2291 verbs associated with a particular preposition after an object noun phrase.[8]

These 4233 pairs are many fewer than the number of associations in the AP sample (see Table 6), even if we ignore the most infrequent pairs. Of the total 76,597 pairs, 20,005 have a frequency greater than 3, and 7,822 have a frequency which is greater than 3 and more than 4 times what one would predict on the basis of the unigram frequencies of the noun or verb and the preposition.[9]

We can use the fixed lexicon of noun-preposition and verb-preposition associations derived from COBUILD to choose attachment in our test set. The COBUILD dictionary has information on 257 of the 880 test verb-noun-preposition triples. In 241 of those cases, there is information only on noun or only on verb association. In these cases, we can use the dictionary to choose the attachment according to the association indicated. In the remaining 16 cases, associations between the preposition and both the noun and the verb are recorded in the dictionary. For these, we select noun attachment, since it is the more probable outcome in general. For the remaining cases, we assume that the dictionary makes no decision. Table 7 gives the results obtained by this attachment procedure. The precision figure is similar to that obtained by the lexical association procedure with a threshold of zero, but the recall is far lower: the dictionary provides insufficient information in most cases.

Like the lexicon derived from the COBUILD dictionary, the fixed lexicon of 7,822 corpus-derived associations derived from our bigram table as described above (that is, all bigrams where $f(w,p) > 3$ and $I(w,p) > 2$) contains categorical information about associations. Using it for disambiguation in the way the COBUILD dictionary was used gives the results

[8]We included particles in these associations, since these were also included in our corpus of bigrams.

[9]The latter condition is spelled out as $\frac{f(w,p)}{N} > 4\frac{f(w)}{N}\frac{f(p)}{N}$, where U is $\sum_{w,p} f(w,p)$, the total number of token bigrams. It is equivalent to w and p having a mutual information (defined as $I(w,p) = log_2 \frac{\frac{f(w,p)}{U}}{\frac{f(w)}{U}\frac{f(p)}{U}}$) greater than 2. This threshold of 2, of course, is an arbitrary cutoff.

Source	Total	NOUN	VERB
COBUILD	4,233	1,942	2,291
AP sample	76,597	56,368	20,229
AP sample ($f > 3$)	20,005	14,968	5,037
AP sample ($f > 3$ *and* $I > 2.0$)	7,822	5,286	2,536
COBUILD $\cap$ AP	2,703	1,415	1,288
COBUILD $\cap$ AP ($f > 3$ *and* $I > 2.0$)	1,235	671	564

Table 6: Count of noun and verb associations for COBUILD and the AP sample

indicated in Table 7. The precision is similar to that which was achieved with the LA procedure with a threshold of 2, although the recall is lower. This suggests that while overall coverage of association pairs is important, the information about the relative strengths of associations contributing to the LA score is also significant.

It must be noted that the dictionary information we derived from COBUILD was composed for people to use in printed form. It seems likely that associations were left out because they did not serve this purpose in one way or another. For instance, listing many infrequent or semantically predictable associations might be confusing. Furthermore, our procedure undoubtedly gained advantage from the fact that the test items are drawn from the same body of text as the training corpus. Nevertheless, the results of this comparison suggest that for the purpose of concern in this paper, a partially parsed corpus is a better source of

COBUILD	actual N	actual V		precision	recall
N guess	132	17	n	.886	.225
V guess	41	67	v	.620	.228
neither	413	210	combined	.774	.226

AP Fixed	actual N	actual V		precision	recall
N guess	283	31	n	.901	.483
V guess	27	119	v	.815	.405
neither	276	144	combined	.874	.457

Table 7: Performance on the test sentences for a fixed lexicon extracted from the COBUILD dictionary, and a fixed lexicon extracted from the AP sample.

information than a dictionary. This conclusion should not be overstated, however. Table 6 showed that most of the associations in each lexicon are not found in the other. Table 8 is a sample of a verb-preposition association dictionary obtained by merging information from the AP sample and from COBUILD, illustrating both the common ground and the differences between the two lexicons. Each source of information provides intuitively important associations that are missing from the other.

Discussion

In our judgment, the results of the lexical association procedure are good enough to make it useful for some purposes, in particular for inclusion in a parser such as Fidditch. The fact that the LA score provides a measure of confidence increases this usefulness, since in some applications (such as exploratory linguistic analysis of text corpora) it is advantageous to be able to achieve increased precision in exchange for discarding a proportion of the data.

From another perspective, our results are less good than what might be demanded. The performance of the human judges with access just to the verb-noun-preposition triple is a standard of what is possible based on this information, and the lexical association procedure falls somewhat short of this standard. The analysis of underlying relations indicated some particular areas in which the procedure did not do well, and where there is therefore room for improvement. In particular, performance on adjuncts was poor. A number of classes of adjuncts, such as temporal ones, are fairly easy to identify once information about the object of the preposition is taken into account. Beginning with such an identification step (which could be conceived of as adding a feature such as [+temporal] to individual prepositions, or replacing individual token prepositions with an abstract temporal preposition) might yield a lexical association procedure which would do better with adjuncts. But it is also possible that a procedure which evaluates associations with individual nouns and verbs in simply inappropriate for adjuncts. This is an area for further investigation.

This experiment was deliberately limited to one kind of attachment ambiguity. However, we expect that the method will be extendable to other instances of PP attachment ambiguity, such the ambiguity which arises when several prepositional phrases follow a subject NP, and to ambiguities involving other phrases, especially phrases such as infinitives which have syntactic markers analogous to a preposition.

We began this paper by alluding to several approaches to PP attachment, specifically work assuming the construction of discourse models, approaches based on structural attachment preferences, and work indicating a dominant role for lexical preference. Our results tend to confirm the importance of lexical preference. However, we can draw no firm conclusions about the other approaches. Since our method yielded incorrect results on roughly 20% of the cases, its coverage is far from complete. This leaves a lot of work to be done, within both psycholinguistic and computational approaches. Furthermore, as we noted above, contemporary psycholinguistic work is concerned with modeling the time course of parsing. Our experiment gives no information about how lexical preference information is exploited at this level of detail, or the importance of such information compared to other factors such as structural preferences at a given temporal stage of the human parsing process. However, the numerical estimates of lexical association we have obtained may be relevant to a psycholinguistic investigation of this issue.

	AP sample	COBUILD
approach	about as at with	
appropriate	**for**	**for**
approve	because by for under	
approximate		to
arbitrate		between
argue	before	out with
arm	**with**	**with**
arouse	in	
arraign	as in **on**	**on**
arrange	**for** so through with	**for**
array		in
arrest	after along_with at during near on outside while	for
arrogate		to
ascribe		to
ask	**about** for	**about**
assassinate	in	
assault	at	
assemble	at	
assert	over	
assign	**to**	**to**
assist	**in**	**in** with
associate	**with**	**with**

Table 8: Verb-(NP)-Preposition associations in the COBUILD dictionary and in the AP sample (with $f(v,p) > 3$ and $I(v,p) > 2.0$).

References

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

[2] 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.

[3] Church, Kenneth, William Gale, Patrick Hanks, and Donald Hindle. 1991. Using statistics in lexical analysis. In Zernik (ed.) *Lexical acquisition: exploiting on-line resources to build a lexicon.*

[4] 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.

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

[6] Gale, William A., Kenneth W. Church, and David Yarowsky. (to appear). A method for disambiguating word senses in a large corpus. In Computers and Humanities.

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

[8] Hindle, Donald. A parser for text corpora. To appear in B.T.S. Atkins and A. Zampolli, editors, *Computational Approaches to the Lexicon.* Oxford University Press, Oxford.

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

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

[11] Mosteller, Frederick and David L. Wallace. 1964. *Inference and disputed authorship: The Federalist.* Addison-Wesley.

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

[13] 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.

[14] 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.