A Hybrid Disambiguation Model for Prepositional Phrase Attachment

HAODONG WU and TEIJI FURUGORI
The University of Electro-Communications, Tokyo, Japan

Abstract

Prepositional Phrase (PP) attachment is a major cause of structural ambiguity in natural language. Many proposals have increasingly relied on large-scale corpus to resolve this problem. However, this approach encounters the notorious sparse-data problem that produces poor results on disambiguation. We in this paper offer a hybrid method which integrates corpus-based approach with knowledge-based techniques for PP attachment disambiguation. It explores a wide-variety of information, including co-occurrence frequencies from annotated corpora, conceptual relationships and conceptual features from a machine-readable dictionary, and syntactic clues from our linguistic observations. We use dictionary definitions and human knowledge to overcome the sparse-data problem. An experiment shows an accuracy rate of 87.7% of our method over 3043 sentences in real English text that contain ambiguous PPs. This result is better than those of any existing methods.

1. Introduction

The resolution of prepositional phrase attachment ambiguity is a difficult problem in NLP. There have been many proposals to attack this problem. Traditional proposals are mainly based on knowledge-based techniques which heavily depend on empirical knowledge encoded in handcrafted rules and domain knowledge in knowledge base: they are therefore not scalable. Recent work has turned to corpus-based or statistical approaches (e.g. Hindle and Rooth, 1993; Ratnaparkhi, Reynar and Roukos; 1994; Brill and Resnik, 1994; Collins and Brooks, 1995). Unlike traditional proposals, corpus-based approaches do not need to prepare a large amount of handcrafted rules: they have therefore the merit of being scalable or easy to transfer to new domains. However, corpus-based approaches suffer from the notorious sparse-data problem: information from sparse data is very unreliable and often it becomes a source for poor performances in disambiguation. To cope with this problem, Brill and Resnik (1994) use word classes from Word-Net noun hierarchy to cluster words into semantic classes. Collins and Brooks (1995), on the other hand, use morphological analysis both on test and training data. Unfortunately, all these smoothing methods are not efficient enough to make a significant improvement on performance.

Instead of using pure statistical approaches, we in this paper propose a hybrid approach to attack the PP attachment problem. We employ corpus-based likelihood estimation to predict most-likely attachments. Where the occurrence frequency is too low to make a reliable choice, we employ a rule-based approach which uses conceptual information from a machine-readable dictionary to make a decision on PP attachments.[1]

In what follows we first outline the idea of using hybrid information to supply preferences for resolving ambiguous PP attachment. We then describe how this information is used in disambiguating PP attachment. After putting the process in an algorithm, we finally show a disambiguation experiment and discuss its results.

2. Using Multiple Information Sources in Disambiguation

Like other work, we use the combination of the following four head words to make a decision on PP attachment: main verb (v), head noun (n1) ahead of the preposition (p), and head noun (n2) of the object of the preposition. We henceforth use the quadruple (v, n1, p, n2) for these four head words.

In our linguistic observation, we found that a wide variety of information supplies important clues for disambiguation. The clues come from presuppositions, syntactic and lexical cues, collocations, syntactic and semantic restrictions, features of head words, conceptual relationships, and world knowledge. We explore in our disambiguation model those clues that are general and reliable so that they make the computation efficient and extensible. They include:

(1) *Syntactic or lexical cues*
If n1 is the same as n2, for example, often n1 + PP is a fixed phrase such as *step by step*.

(2) *Co-occurrences of head words*
Co-occurrences of head words in annotated corpora (See Section 4)

(3) *Syntactic and semantic features*
Features of y of n1 or n2 sometimes indicate the 'correct' attachment. For example, if v is for a motion, p is *to* and n2 is a place or direction, the PP tends to be attached to the verb.

(4) *Conceptual relations*
Conceptual relations between v and n2, or between n1 and n2 reflect the role-exceptions of prepositions, which supply important clues for disambiguation. For example, in the sentence *Tom broke the window with a stone*, we are sure that the PP *with a stone* is attached to *broke*/v with an understanding that *stone*/n2 is an instrument for *broke*.v.

Correspondence: Teiji Furugori, Department of Computer Science, The University of Electro-Communications, Chofu, Tokyo, Japan.

We use (2) in corpus-based disambiguation and (1), (3) and (4) in rule-based disambiguation.

3. Likelihood Estimation Based on Corpora

We consider two kinds of PP attachment in our corpus-based approach, namely, attachment to verb phrase (VP attachment) and to noun phrase (NP attachment). Here, we use two annotated corpora: EDR English Corpus (EEC) and Susanne Corpus (SC) as the training data.

EEC, compiled by Japan Electronic Dictionary Research Institute, Ltd., contains 160,000 sentences with annotated morphological, syntactic and semantic information. SC, compiled by Geoffrey Sampson, is an annotated corpus comprising about 130,000 words of written American English text. EEC and SC contain about 228,000 PPs. Each sentence in them is syntactically tagged. This means that a given PP is attached to an unique phrase. For example, in the sentence *A girl I had never met slipped a letter into my hand and left*, there is a quadruple of (slip/v, letter/n1, into/p, hand/n2), where the PP *into my hand* is assigned to the verb *slipped*.

We use RA (v,n1,p,n2) score to estimate the likelihood of attachment for a certain (v,n1,p,n2). RA score is defined in (1) as a value of counts of VP attachments divided by the total occurrences of (v,n1,p,n2) in the training data.

$$\mathrm{RA}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) = \frac{f(vp|v, n1, p, n2)}{f(v, n1, p, n2)}$$

$$\simeq \frac{f(vp|v, n1, p, n2)}{f(vp|v, n1, p, n2) + f(np|v, n1, p, n2)} \qquad (1)$$

In (1), the symbol *f* denotes frequency of a particular quadruple in the training data. For example, f(vp|share, apartment, with, friend) is the number of times the quadruple (share, apartment, with, friend) is seen with a VP attachment. The RA score ranges from 0 to 1, and we could choose an attachment according to the score. If RA > 0.5, choose VP attachment. Otherwise, choose NP attachment.

Quadruples in test data are seldom found in the training data due to the data sparseness, however. We thus turn to collect triplets of (v, p, n2), (n1, p, n2), (v, n1, p) and pairs of (v, p), (n1, p), (p, n2) like Collins and Brooks (1995) did,[2] and compute RA score using either (2) or (3).

$$\mathrm{RA}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) =$$

$$\frac{f(vp|v, p, n2) + f(vp|n1, p, n2) + f(vp|v, n1, p1)}{f(v, p, n2) + f(n1, p, n2) + f(v, n1, p)} \qquad (2)$$

$$\mathrm{RA}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) =$$

$$\frac{f(vp|v, p) + f(vp|n1, p) + f(vp|p, n2)}{f(v, p) + f(n1, p) + f(p, n2)} \qquad (3)$$

To avoid using very low frequencies, we set two thresholds for each one above. For (2), the condition is:

$\mathrm{f}_{\mathrm{triplet}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) \geqslant 2$, and
$|2 * \mathrm{RA}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) - 1| * \log(\mathrm{f}_{\mathrm{triplet}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2})) < 0.5$

where $\mathrm{f}_{\mathrm{triplet}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2})$ is defined as:

$\mathrm{f}_{\mathrm{triplet}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2})\ 5\ \mathrm{f}(\mathrm{v}, \mathrm{p}, \mathrm{n2})\ 1\ \mathrm{f}(\mathrm{n1}, \mathrm{p}, \mathrm{n2})\ 1\ \mathrm{f}(\mathrm{v}, \mathrm{n1}, \mathrm{p})$

For (3), the condition is:

$\mathrm{f}_{\mathrm{pair}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) \geqslant 4$, and
$|2 * \mathrm{RA}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) - 1| * \log(\mathrm{f}_{\mathrm{pair}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2})) < 0.5$

where $\mathrm{f}_{\mathrm{pair}}$ is defined as:

$\mathrm{f}_{\mathrm{pair}}(\mathrm{v}, \mathrm{n1}, \mathrm{p}, \mathrm{n2}) = \mathrm{f}(\mathrm{v}, \mathrm{p}) + \mathrm{f}(\mathrm{n1}, \mathrm{p}) + \mathrm{f}(\mathrm{p}, \mathrm{n2})$

We can avoid using low frequency tuples with the first threshold in each case; the second one in each case allows us to throw away the RA score which is close to 0.5 as this value is rather unstable.

4. Conceptual Information and Preference Rules

As we try to use only 'reliable' data from corpora to make a decision on PP attachment based on RA score, there are many instances in which many PPs' attachments may be left undetermined due to the data sparseness. We deal with these undetermined PPs with a rule-based approach. Unlike traditional rule-based approaches which depend on a large amount of hand-crafted knowledge, we employ a machine-readable dictionary, EDR Electronic Dictionary, as a knowledge source that supplies conceptual information for disambiguation. This information is incorporated into preference rules for determining semantically possible attachments.

EDR Electronic Dictionary is a machine-readable dictionary that catalogues the lexical knowledge of Japanese and English (the Word Dictionary, the Bilingual Dictionary, and the Co-occurrence Dictionary). It also contains the Concept Dictionary (CD) in which concepts are classified into thesaurus-like structures.

4.1 *Conceptual Information and Concept Dictionary*

The CD consists of about 400,000 concepts, where a concept corresponds to a word sense. CD is divided into three parts. The first part contains the definition of each concept; the second part is a property inheritance network in which concepts are arranged in superclass-subclass relation; the third part contains concept pairs that may hold one of a few dozen semantic or co-occurrence relations.

In PP attachment disambiguation, we use two kinds of conceptual information derived from CD: *concept class* and *conceptual relation*. A concept class is a set of concepts which share the same semantic features or syntactic function to be used for disambiguation. An example of the concept class is *animal* whose members may be such concepts as *dog*, *tiger*, *fish*. We derive concept classes that are useful for disambiguation task from the superclass-subclass relation in CD. A conceptual relation is a semantic relation between two concepts which is to be used in preference rules for predicting PP attachment. The conceptual relations we use are *implement*, *a-object*, *possessor*, etc. Here, *implement* is the relation seen as *implement* (cut, knife) with the meaning *knife* is an *instrument* for the action *cut*;

a-object is the relation seen as *a-object (red, apple) with the meaning* red color is an *attribute* of *apple*; *possessor* is the relation seen as *possessor* (father, coat) with the meaning the *owner* of the *coat* is *father*.

4.2 *Preference Rules*

We use preference rules to encode syntactic and lexical clues, as well as clues from conceptual information to determine PP attachments. We divide the rules into two categories: a rule applicable to any preposition in general is called a *global rule*; a rule applicable only to a particular preposition is called a *local rule*. Four global rules used in our disambiguation model are listed below.

1. lexical (passivized(v) + PP) AND prep ≠ 'by' → vp_attach(PP)
   The PP is attached to the VP if the verb is passivized.
2. n1 = n2 → vp_attach(n1 + PP)
   If n2 repeats n1 (e.g. *loss on loss*), n1 + PP is a fixed phrase to modify the verb.
3. (prep ≠ 'of' AND prep ≠ 'for') AND (time(n2) OR date(n2)) → vp_attach(PP)
   The PP is attached to the VP if n2 is a time or a date (e.g. *next week*) and the preposition is not **of** or **for**.
4. lexical(Adjective + PP) → adjp_attach(PP)
   If the PP comes after an adjective (including participial adjective), it is attached to the adjective as its complement.

Local rules use conceptual information to determine PP attachment. Here, we show two examples of local rules for preposition *with* and *to*. Local rules for other prepositions are found in Wu, Ito and Furugori (1995).

with-rules:
implement (v, n2) → vp_attach(PP)
(a-object (n1, n2) OR possessor (n1, n2)) AND NOT (implement (v, n2)) → np_attach(PP)
Default → vp_attach(PP)
to-rules:
a-object (n1, n2) → np_attach(PP)
motion (n1) AND direction (n2) → np_attach(PP)
state (n1) AND degree (n2) → np_attach(PP)
goal (v, n2) → vp_attach (PP)
Default → vp_attach(PP)

On the left hand of each rule, a one-atom predicate presents a concept class (e.g., time[n2]]), and a two-atom predicate describes a conceptual relation between two atoms (e.g., implement[v, n2]).

Before applying local rules, we must project each of v, n1 and n2 used by local rules into the concept which represents the 'correct' word sense. This process is described as follows.

1. For each v, n1, n2 used in local rules, search the EDR word dictionary for its definition, pick up the concepts which coincide with the word classes and syntactic restrictions (e.g. transitive or intransitive verb) and add them to candidate set.
2. If more than two candidate concepts remain for a word, we want to know which one is better to go with other words in the sentence. We use the mutual information (Church and Hanks, 1990) as a measurement to test in ECC how often each concept co-occurs with each unambiguous word in the sentence. Concepts with low values are then dropped out.

4.3 *Concept-based Disambiguation with Preference Rule*

We now illustrate how to use preference rule and CD for determining PP attachment. Consider the sentence:

(1) The man killed a girl with a gun.

We first try to use global rules for this sentence. As no global rules apply to it, we turn to check local rules. The first **with**-rule consults CD to see if *gun* is an instrument or implement of the verb *kill*. It succeeded since implement (kill, gun) is found in CD, and since the second **with**-rule fails to apply, the PP *with a gun* is attached to the verb *killed*.

5. Disambiguation Model

The disambiguation process thus far is put in an algorithm form:

```
Phase 1. (disambiguation using global rules):
  Try global rules one by one. If a rule succeeds, use it to decide the attachment, and exit.
Phase 2. (statistics-based disambiguation):
  Set the initial valve of −1 for RA(v, n1, p, n2) = −1;

  Define:
    $f_{triplet}(v, n1, p, n2)$ as $f(v, p, n2) + f(n1, p, n2) + f(v, n1, p)$
    $f_{pair}(v, n1, p, n2)$ as $f(v, p) + f(n1, p) + f(p, n2)$
  if $f_{triplet}(v, n1, p, n2) \geqslant 2$, then
    $RA(v, n1, p, n2) = \frac{f(vp|v, p, n2) + f(vp|n1, p, n2) + f(vp|v, n1, p)}{f(v, p, n2) + f(n1, p, n2) + f(v, n1, p)}$
  if $|2 * RA(v, n1, p, n2) - 1| * \log(f_{triplet}(v, n1, p, n2))$ , 0.5
    then $RA(v, n1, p, n2) = -1$
  if $RA(v, n1, p, n2) < 0$ and $f_{pair}(v, n1, p, n2) \geqslant 4$, then
    $RA(v, n1, p, n2) = \frac{f(vp|v, p) + f(vp|n1, p) + f(vp|p, n2)}{f(v, p) + f(n1, p) + f(p, n2)}$
  if $|2 * RA(v, n1, p, n2) - 1| * \log(f_{pair}(v, n1, p, n2)) < 0.5$
    then $RA(v, n1, p, n2) = -1$
  if $RA(v, n1, p, n2) \geqslant 0$, then [
    if $RA(v, n1, p, n2) < 0.5$, then choose NP attachment; otherwise choose VP attachment.
    exit.]
```

```
Phase 3. (concept-based disambiguation):
    1. Project each of v, n1, n2 into its concept sets.
    2. Try the rules related to the preposition; if only
       one rule is applicable, use it to decide the
       attachment, and then exit.
Phase 4. (attachment by default):
    if f(p) > 0, then [
        if $\frac{f(vp|p)}{f(p)} < 0.5$, then choose NP attachment;
        otherwise choose VP attachment]
    otherwise choose NP attachment.
```

When two rules succeed and make different decisions on attachment in Phase 3, it may mean that the attachment is semantically undetermined as shown in example 2, or that conceptual information from CD is not adequate for making the decision. In such a case, the attachment is determined in Phase 4.

(2) Peter sliced the cheese on the table.

Phase 4 is for the default attachment process. The symbol f(p) here denotes the number of occurrences of a preposition in the corpora.

This algorithm differs from the one described in Wu, Ito and Furugori (1995) where preference rules were applied before statistical computing. We have changed the order partly because we understand that using the data of triplets and pairs with high occurrences results in a high success rate, and partly because the statistical model has a grounded mathematical basis.

6. Experiment on Testing Attachment

To evaluate the performance of the disambiguation algorithm, 3,043 English sentences which contain ambiguous PPs were randomly selected from a computer manual, a grammar book and *Japan Times*. We extracted quadruples of head words, applied the algorithm to make judgment on PP attachments, and evaluated performance by comparing the results with human being's judgments.

Table 1 shows the performance on the test sentences using the algorithm. Here, the second column shows the number of PPs tested in each phase, the third column the cases successfully processed, the final column the success rate. As is shown here, we have achieved an overall success rate of 86.9%.

We use a procedure similar to that of Collins and Brooks (1995) to improve the performance further. It is applied for processing head words both in training data and in test data by reducing the sparse data problem and dealing with undefined words in the dictionary. The procedure includes the following processes:

1. All 4-digit numbers are replaced with 'date'.
2. All verbs are replaced with their stems in lower cases.
3. Nouns starting with a capital letter are replaced with 'name'.
4. Personal pronouns in the n2 field are replaced with 'person'.
5. Prepositions like synonyms and antonyms are clustered into groups as follows and replaced by a representative preposition (e.g. *till* and *pending* are replaced by *until*.) as are shown below:

group 1:	*with*, without
group 2:	*after*, before
group 3:	*beside*, opposite
group 4:	*inside*, outside
group 5:	*off*, onto
group 6:	*above*, below, beneath
group 7:	*among*, amongst, amid, amidst
group 8:	*despite*, notwithstanding
group 9:	*until*, till, pending
group 10:	*through*, throughout
group 11:	*toward*, towards
group 12:	*under*, underneath
group 13:	*like*, unlike
group 14:	*on*, upon
group 15:	*but*, except

The results with this modification are shown in Table 2. Now the accuracy rate has become 87.7%, an improvement of 0.8% over the result in Table 1. It is also shown that the number of PPs resolved by using triplets and pairs has been raised, and the number of PPs resolved by using local rules and by default has been decreased.

6.1 *Evaluating Performance*

Hindle and Rooth (1995) have reported that an 'average' person's success rate for judging PP attachments is 88.2% when he or she used the four head words alone. Table 2 shows that the success rate of our method is comparable to the performance of human beings. We attribute the good results to the hybrid method in which the sound clues are used to less reliable ones in the disambiguation process. The two thresholds are also of help in improving the result. When we do not use the thresholds at all, the success rate becomes 89.1% in triplet-combinations, and 81.2% in pair-combinations. They are 2.1% and 3.7% lower than the results attained with using the thresholds. Using local rules to tackle unattached PPs is also of value in improving the overall success rate since local rules in Phase 3 work much better than the default decision in Phase 4.

Table 1 Result of the test in PP attachment

Phase	Number Tested	Number Correct	Success Rate
Global rules	507	487	96 1%
Triplets	564	518	91 8%
Pairs	1093	931	85.3%
Local rules	662	557	84 1%
Default	217	151	69 6%
Total	3043	2644	86.9%

Table 2 Result with processing head words

Phase	Number Tested	Number Correct	Success Rate
Global rules	507	487	96.1%
Triplets	708	646	91 2%
Pairs	1142	941	84 9%
Local rules	592	498	84 1%
Default	94	66	70 2%
Total	3043	2668	87.7%

Our method is not perfect, however. Some failures are found mainly in the following:

a. *Idiomatic phrases or fixed phrases.* Examples: *in vain, brother in blood, at last.* Other examples may be found in phrases with literal and metaphorical senses such as *in deep water*, which metaphorically means of 'in trouble'.
b. *Misprojecting of words to concepts.*
c. *Quadruple (v, n1, p, n2) being insufficient to predicting PP attachment.* In an example, sentences *I kept the bicycle in the garage* and *I kept the spare bicycle in the garage well oiled* share the same quadruple of (keep/v, bicycle/n1, in/p, garage/n2), the PP in the former modifies the verb *kept* whereas the PP in the latter acts as the complement of the NP headed by *bicycle*.

6.2 *Comparison with Related Work*

Our hybrid disambiguation method employs annotated corpora, machine-readable dictionary and some handcrafted rules. On the surface it looks somewhat similar to those of rule-based (e.g. Dahlgren and McDowell, 1986), dictionary-based (Jensen and Binot, 1987), and corpus-based (e.g. Collins and Brooks, 1995) methods.

Conventional rule-based approach depends on a large amount of handcrafted knowledge, so it suffers from the problem of not being scalable. Our method, on the other hand, uses an on-line dictionary as knowledge source, so only a limited number of rules are needed. In addition these rules are almost domain-free.

Jenson and Binot (1987) offered a method of extracting semantic information from on-line dictionary definitions (Websters's Seventh New Collegiate, W7). They used it in heuristics to choose attachments with high certainties. However, a general dictionary like W7 is not suitable for use in a disambiguation task. For example, we can hardly find information in W7 on the semantic relation between *murder* and *gun*. The CD we used, on the other hand, contains rich conceptual information fit for predicting PP attachment.

In corpus-based proposals, ours is close to that of Collins and Brooks (1995). They use a backed-off estimate to decide PP attachments. Our method differs from them in use of threshold values in one aspect, in another we use preference rules and a machine-readable dictionary, whereas they use neither of them. To make a comparison of efficiency, we have tried their method both on our training data and test data, and got a success rate of 84.2%, a result 3.5% lower than that of our method.

7. Conclusions

Pure statistical models for disambiguation tasks suffer from the sparse-data problem. It is difficult to avoid making poor estimations on low occurrences in corpora even when applying smooth techniques such as semantic similarity or clustering. On-line dictionaries which contain rich semantic or conceptual information may be of help in improving the performance. Our experiment has proven that the hybrid approach we employed is both effective and applicable in practice.

The hybrid method, on the other hand, has some weaknesses. We find that the use of CD in disambiguation endangers over-generation and mis-generation in selecting a concept for a polysemous head word. Although full-sentence context and full-text context are helpful for determining PP attachment, we do not use them in the disambiguation process. Further improvement may be attained by using larger training data and other linguistic resources such as lexical databases, and perhaps an incremental disambiguation process consulting a wider context.

We hope the hybrid method may be of use for resolving other ambiguities such as word sense disambiguation, relative-clause attachment and anaphor disambiguation.

Acknowledgements

We are very grateful for the anonymous referees. Criticisms and comments made by them have helped us improve the context and the presentation. We also owe much to Mr G. Dixon, the Editor-in-Chief of L&LC, in the process of editing the paper.

Notes

[1] We intend to use this disambiguation method to build a disambiguation module in the PFTE (Parser for Free Text of English) system. The PFTE system is a versatile parsing system in development which covers a wide range of phenomena in lexical, syntactic and semantic dimensions of English. It is designed as a linguistic tool for applications in text understanding, data-base generation from text and computer-based language learning.

[2] As Collins and Brooks pointed out, triplets and pairs not containing prepositions do not work well in predicting PP attachments.

References

Brill, E. and Resnik, P. (1994) A rule-based approach to prepositional phrase attachment disambiguation. *Proc. of the 15th COLING*, 1198–1204.

Charniak, E. (1993). *Statistical language learning*. The MIT Press.

Charch, K. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16: 22–29.

Collins, M. and Brooks, J. (1995). Prepositional phrase attachment through a backed-off model. http://xxx.lanl.gov/cmp-lg/9506021.

Dahlgren, K. and McDowell, J. (1986). Using commonsense knowledge to disambiguate prepositional phrase modifiers. *Proc. of the 5th AAAI*, 589–593.

Fukumoto, F. (1995). Disambiguating prepositional phrase attachments by using statistical information about word triplets. *Journal of Natural Language Processing*, 2: 67–74. (in Japanese)

Hindle, D. and Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, 19: 103–120.

Japan Electronic Dictionary Research Institute, Ltd (1993). *EDR Electronic Dictionary Specifications Guide.*

Jensen, K. and Binot, J. (1987). Disambiguating prepositional phrase attachments by using on-line dictionary definition. *Computational Linguistics*, 13: 251–260.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, 2: 15–47.

Luk, A K. (1995). Statistical sense disambiguation with relatively small corpora using dictionary definitions. *Proc. of the 33rd ACL Meeting*, 181–188.

Whittemore, G. et al. (1990). Empirical study of predictive powers of simple attachment schemes for post-modifiers prepositional phrases. *Proc. of the 28th ACL Meeting*, 23–30.

Wu, H., Ito, T., and Furugori, T. (1995). A preferential approach for disambiguating prepositional phrase modifiers. *Proc. of the 3rd Natural Language Processing Pacific Rim Symposium*, 745–751.