

Using Parsed Corpora for Structural Disambiguation in the TRAINS Domain

Mark Core

Department of Computer Science
University of Rochester
Rochester, New York 14627
mcore@cs.rochester.edu

Abstract

This paper describes a prototype disambiguation module, KANKEI, which was tested on two corpora of the TRAINS project. In ambiguous verb phrases of form V ... NP PP or V ... NP adverb(s), the two corpora have very different PP and adverb attachment patterns; in the first, the correct attachment is to the VP 88.7% of the time, while in the second, the correct attachment is to the NP 73.5% of the time. KANKEI uses various n-gram patterns of the phrase heads around these ambiguities, and assigns parse trees (with these ambiguities) a score based on a linear combination of the frequencies with which these patterns appear with NP and VP attachments in the TRAINS corpora. Unlike previous statistical disambiguation systems, this technique thus combines evidence from bigrams, trigrams, and the 4-gram around an ambiguous attachment. In the current experiments, equal weights are used for simplicity but results are still good on the TRAINS corpora (92.2% and 92.4% accuracy). Despite the large statistical differences in attachment preferences in the two corpora, training on the first corpus and testing on the second gives an accuracy of 90.9%.

1 Introduction

The goal of the TRAINS project is to build a computerized planning assistant that can interact conversationally with its user. The current version of this planning assistant, TRAINS 95, is described in (Allen et al., 1995); it passes speech input onto a parser whose chart is used by the dialog manager and other higher-level reasoning components. The planning problems handled involve moving several trains from given starting locations to specified destinations on a map display (showing a network of rail lines in the eastern United States). The 95 dialogs

are a corpus of people's utterances to the TRAINS 95 system; they contain 773 instances of PP or adverb postmodifiers that can attach to either NPs or VPs. Many of these cases were unambiguous, as there was no NP following the VP, or the NP did not follow a verb. Only 275 utterances contained ambiguous constructions and in 73.5% of these, the correct PP/adverb attachment was to the NP.

One goal of the TRAINS project is to enhance the TRAINS 95 system sufficiently to handle the more complex TRAINS 91-93 dialogs. This corpus was created between 1991 and 1993 from discussions between humans on transportation problems involving trains. The dialogs deal with time constraints and the added complexity of using engines to pick up boxcars and commodities to accomplish delivery goals. This corpus contains 3201 instances of PP or adverb postmodifiers that can attach to either NPs or VPs. 1573 of these examples contained both an NP and a VP to which the postmodifier could attach. The postmodifier attached to the VP in 88.7% of these examples. On average, a postmodifier attachment ambiguity appears in the 91-93 dialogs after about 54 words, which is more frequent than the 74 word average of the 95 dialogs. This suggests that a disambiguation module is going to become necessary for the TRAINS system. This is especially true since some of the methods used by TRAINS 95 to recover from parse errors will not work in a more complex domain. For instance in the 95 dialogs, a PP of form *at city-name* can be assumed to give the current location of an engine that is to be moved. However, this is not true of the 91-93 dialogs where actions such as *load* often take *at city-name* as adjuncts.

2 Methodology

KANKEI¹ is a first attempt at a TRAINS disambiguation module. Like the systems in (Hindle and Rooth, 1993) and (Collins and Brooks, 1995), KANKEI records attachment statistics on informa-

¹From the Japanese word, *kankei*, meaning "relation."

tion extracted from a corpus. This information consists of phrase head patterns around the possible locations of PP/adverb attachments. Figure 1 shows how the format of these patterns allows for combinations including a verb, NP-head (rightmost NP before the postmodifier), and either the preposition and head noun in the PP, or one or more adverbs.² These patterns are similar to ones used by the disambiguation system in (Collins and Brooks, 1995) and (Brill and Resnik, 1994) except that Brill and Resnik form rules from these patterns while KANKEI and the system of Collins and Brooks use the attachment statistics of multiple patterns. While KANKEI combines the statistics of multiple patterns to make a disambiguation decision, Collins and Brooks' model is a backed-off model that uses 4-gram statistics where possible, 3-gram statistics where possible if no 4-gram statistics are available, and bigram statistics otherwise.

verb NP-head (preposition obj-head | adverb1 adverb2)

Figure 1: Format of an attachment pattern

Most items in this specification are optional. The only requirement is that patterns have at least two items: a preposition or adverb and a verb or NP-head. The singular forms of nouns and the base forms of verbs are used. These patterns (with hyphens separating the items) form keys to two hash tables; one records attachments to NPs while the other records attachments to VPs. Numbers are stored under these keys to record how often such a pattern was seen in a not necessarily ambiguous VP or NP attachment. Sentence 1 instantiates the longest possible pattern, a 4-gram that here consists of *need, orange, in, and Elmira*.

1) I need the oranges in Elmira.

The TRAINS corpora are much too small for KANKEI to rely only on the full pattern of phrase heads around an ambiguous attachment. While searching for attachment statistics for sentence 1, KANKEI will check its hash tables for the key *need-orange-in-Elmira*. If it relied entirely on full patterns, then if the pattern had not been seen, KANKEI would have to randomly guess the attachment. Such a technique will be referred to as *full matching*. Normally KANKEI will do *partial matching*, i.e., if it cannot find a pattern such as *need-orange-in-Elmira*, it will look for smaller partial patterns which here would be: *need-in, orange-in, orange-in-Elmira, need-in-Elmira, and need-orange-in*. The frequency with which NP and VP attachment occurs for these patterns is totaled to see if one attachment is preferred. Currently, we count partial patterns equally, but in future refinements we would

²Examples of trailing adverb pairs are *first off* and *right now*.

like to choose weights more judiciously. For instance, we would expect shorter patterns such as *need-in* to carry less weight than longer ones. The need to choose weights is a drawback of the approach. However, the intuition is that one source of evidence is insufficient for proper disambiguation. Future work needs to further test this hypothesis.

The statistics used by KANKEI for partial or full matching can be obtained in various ways. One is to use the same kinds of full and partial pattern matching in training as are used in disambiguation. This is called *comprehensive training*. Another method, called *raw training*, is to record only full patterns for ambiguous and unambiguous attachments in the corpus. (Note that full patterns can be as small as bigrams, such as when an adverb follows an NP acting as a subject.) Although raw training only collects a subset of the data collected by comprehensive training, it still gives KANKEI some flexibility when disambiguating phrases. If the full pattern of an ambiguity has not been seen, KANKEI can test whether a partial pattern of this ambiguous attachment occurred as an unambiguous attachment in the training corpus.

Like the disambiguation system of (Brill and Resnik, 1994), KANKEI can also use word classes for some of the words appearing in its patterns. The rudimentary set of noun word classes used in this project is composed of **city** and **commodity** classes and a **train** class including cars and engines.

3 Measure of Success

One hope of this project is to make generalizations across corpora of different domains. Thus, experiments included trials where the 91-93 dialogs were used to predict the 95 dialogs³ and vice versa. Experiments on the effect of training and testing KANKEI on the same set of dialogs used cross validation; several trials were run with a different part of the corpus being held out each time. In all these cases, the use of partial patterns and word classes was varied in an attempt to determine their effect.

Word Classes	Yes	Yes	Yes	Yes
Raw Training	Yes	Yes	No	Yes
P. Matching	Yes	Yes	No	No
Default Guess	VP	NP	NP	NP
% by Default	7.6	3.6	42.3	43.0
% Accuracy	90.9	87.6	86.9	85.5

Table 1: Results of training with the 93 dialogs and testing on the 95 dialogs

Tables 1, 2, and 3 show the results for the best parameter settings from these experiments.

³91-93 dialogs were used for training and the 95 dialogs for testing.

Word Classes	Yes	Yes	No	Yes	No
Raw Training	No	No	No	Yes	No
P. Matching	Yes	Yes	Yes	Yes	Yes
Default Guess	NP	VP	VP	VP	NP
% by Default	3.9	3.6	5.1	9.5	4.0
% Accuracy	92.4	92.0	92.0	91.6	90.9

Table 2: Results of training and testing on the 95 dialogs

Word Classes	Yes	No	Yes	Yes
Raw Training	No	No	Yes	No
P. Matching	Yes	Yes	No	No
Default Guess	VP	VP	VP	VP
% by Default	7.2	8.8	42.9	42.7
% Accuracy	92.2	91.3	91.0	91.0

Table 3: Results of training and testing on the 93 dialogs

The rows labeled *% by Default* give the portion of the total success rate (last row) accounted for by KANKEI's default guess. The results of training on the 95 data and testing on the 93 data are not shown because the best results were no better than always attaching to the VP. Notice that all of these results involve either word classes or partial patterns. There is a difference of at least 30 attachments (1.9% accuracy) between the best results in these tables and the results that did not use word classes or partial patterns. Thus, it appears that at least one of these methods of generalization is needed for this high-dimensional space. The 93 dialogs predicted attachments in the 95 test data with a success rate of 90.9% which suggests that KANKEI is capable of making generalizations that are independent of the corpus from which they were drawn. The overall accuracy is high: the 95 data was able to predict itself with an accuracy of 92.2%, while the 93 data predicted itself with an accuracy of 92.4%.

4 Discussion and Future Work

The results for the TRAINS corpora are encouraging. We would also like to explore how KANKEI performs in a more general domain such as the Wall Street Journal corpus from the Penn Treebank. We could then compare results with Collins and Brooks' disambiguation system which was also tested using the Penn Treebank's Wall Street Journal corpus.

Weighting the n-grams in a nonuniform manner should improve accuracy on the TRAINS corpora as well as in more general domains. (Alshawi and Carter, 1994) address a related problem, weighting scores from different disambiguation systems to obtain a single rating for a parse tree. They achieved good results using a hill climbing technique to ex-

plore the space of possible weights. Another possible technique for combining evidence is the maximum-entropy technique of (Wu, 1993). We are also considering using logical forms (instead of word and word classes) in collocation patterns.

The integration of KANKEI with the TRAINS parser needs to be completed. As a first attempt, when the TRAINS parser tries to extend the arcs associated with the rules: VP \rightarrow VP (PP|ADV) and NP \rightarrow NP (PP|ADV), KANKEI will adjust the probabilities of these arcs based on attachment statistics.⁴ Ultimately, the TRAINS disambiguation module will contain functions measuring rule habituation and distance effects. Then it will become necessary to weight the scores of each disambiguation technique according to its effectiveness. The ability to adjust probabilities based on evidence seen is an advantage over rule-based approaches. This advantage is obtained at the expense of storing all the patterns seen.

Acknowledgments

This work was supported in part by National Science Foundation grant IRI-95033312. Len Schubert's supervision and many helpful suggestions are gratefully acknowledged. Thanks also to James Allen for his helpful comments.

References

- James Allen, George Ferguson, Bradford Miller, and Eric Ringger. 1995. Spoken dialogue and interactive planning. In *Proc. of the ARPA Spoken Language Technology Workshop*, Austin, TX.
- Hiyan Alshawi and David Carter. 1994. Training and scaling preference functions. *Computational Linguistics*, 20(4):635-648.
- Eric Brill and Philip Resnik. 1994. A rule-based approach to prepositional phrase attachment disambiguation. In *Proc. of 15th International Conference on Computational Linguistics*, Kyoto, Japan.
- Michael Collins and James Brooks. 1995. Prepositional phrase attachment through a backed-off model. In *Proc. of the 3rd Workshop on Very Large Corpora*, pages 27-38, Boston, MA.
- Donald Hindle and Mats Rooth. 1993. Structural ambiguity and lexical relations. *Computational Linguistics*, 19(1):103-120.
- Dekai Wu. 1993. Estimating probability distributions over hypotheses with variable unification. In *Proc. of the 11th National Conference on Artificial Intelligence*, pages 790-795, Washington D.C.

⁴The TRAINS parser is probabilistic although the probabilities are parse scores not formal probabilities.