

## Preposition Semantic Classification via TREEBANK and FRAMENET

**Tom O'Hara**

Department of Computer Science  
New Mexico State University  
Las Cruces, NM 88003  
tomohara@cs.nmsu.edu

**Janyce Wiebe**

Department of Computer Science  
University of Pittsburgh  
Pittsburgh, PA 15260  
wiebe@cs.pitt.edu

### Abstract

This paper reports on experiments in classifying the semantic role annotations assigned to prepositional phrases in both PENN TREEBANK (version II) and FRAMENET (version 0.75). In both cases, experiments are done to see how the prepositions can be classified given the dataset's role inventory, using standard word-sense disambiguation features, such as the parts of speech of surrounding words, and collocations indicative of the particular roles. In addition to using traditional word collocations, the experiments incorporate class-based collocations in the form of WordNet hypernyms. Separate classifiers are produced for each preposition. For TreeBank, the word-collocations achieve slightly better performance: 78.5% versus 77.4%. However, for FrameNet, the combined collocations achieve better performance: 70.3% versus 68.5%. Furthermore when classifying all the TreeBanks prepositions together, the combined yields a noticeable gain at 85.8% accuracy versus 81.3% for word-only collocations.

### 1 Introduction

English prepositions convey important relations in text. When used as verbal adjuncts, they are the principle means of conveying semantic roles for the sup-

porting entities described by the predicate. Prepositions are highly ambiguous. A typical collegiate dictionary has dozens of senses for each of the common prepositions. These senses tend to be closely related, in contrast to the other parts of speech where there might a variety of distinct senses.

Given the recent advances in word-sense disambiguation, due in part to the SENSEVAL competitions (Kilgarriff, 1998; Edmonds and Cotton, 2001), it would seem natural to apply the same basic approach to handling the disambiguation of prepositions. Of course, given the current state of the art, it is difficult to disambiguate prepositions at the granularity present in collegiate dictionaries, as illustrated later. Nonetheless, in certain cases this is feasible as shown later.

We also present results disambiguating prepositions at a coarse granularity more typical of earlier work in computational linguistics, such as the role inventory proposed by Fillmore (1968), including high level roles such as *instrument* and *location*. Recently, systems have incorporated fine-grain roles, often specific to particular domains. For example, in the Cyc KB, there are close to 200 different types of semantic roles (Lehmann, 1996). These range from high-level roles (e.g., *beneficiaries*) through medium level roles (e.g., *exchanges*) to highly specialized roles (e.g., *catalyst*).<sup>1</sup>

Preposition classification using two different semantic role inventories are investigated in this paper, taking advantage of large annotated corpora. After providing background to the work in Section 2, experiments over the semantic role annotations in

<sup>1</sup>Part of the Cyc KB is freely available at [www.opencyc.org](http://www.opencyc.org).

TREEBANK II (Marcus et al., 1994) are discussed in Section 3.1. These annotations include about a dozen high-level roles similar to Fillmore’s. Next, experiments using the finer-grain semantic role annotations in FRAMENET are discussed in Section 3.2. There are about 100 or so of these roles, approaching but not quite as specialized as Cyc’s. Section 4 then discusses related work.

## 2 Background

### 2.1 Semantic role annotations

#### 2.1.1 PENN TREEBANK

The PENN TREEBANK (Marcus et al., 1993) is one of the most popular resources for corpus-based natural language processing. It is a collection of manually-corrected parse trees for subsets of several commonly-used corpora, such as the Brown corpus and the Wall Street Journal corpus. TREEBANK II (Marcus et al., 1994) added additional syntactic role information, including a few case-style relation annotations, which should be useful for disambiguating prepositions. For example, here is a simple parse tree with the new annotation format:

```
(S (NP-TPC-5 This)
  (NP-SBJ every man)
  (VP contains
    (NP *T*-5)
    (PP-LOC within
      (NP him))))
```

This shows that the prepositional phrase (PP) is providing the location for the state described by verb phrase. Treating this as the preposition sense would yield the following annotation:

This every man contains within<sub>LOC</sub> him

The main semantic relations in TREEBANK are *beneficiary*, *direction*, *spatial extent*, *manner*, *location*, *purpose/reason*, and *temporal*. These tags can be applied to any verb complement but normally occur with clauses, adverbs, and prepositions.

The frequencies for the most frequent prepositions that have occurred in the prepositional phrase annotations are shown later in Table 4. The table is ordered by entropy, which measures the inherent ambiguity in the classes as given by the annotations

(Jurafsky and Martin, 2000). Note that the *Baseline* column is the probability of the most frequent sense, which is an estimate of the lower bound. There are several cases at the bottom of the table for which a single role has been annotated, giving a maximal probability of 1.0 and an entropy value of 0.0.

#### 2.1.2 FRAMENET

The Berkeley FRAMENET (Fillmore et al., 2001) project provides the most recent large-scale annotation of semantic roles. These are at a much finer granularity than those in TREEBANK II, so they should prove quite useful for applications which learn semantics from corpus. In all, there are over 140 roles annotated with over 117,000 tagged instances.

FRAMENET annotations occur at the phrase level instead of the grammatical constituent level as in TREEBANK. The cases that involve prepositional phrases can be determined by the phrase type attribute of the annotation. For example, consider the following annotation.

```
<S TPOS="56879338">
<T TYPE="sense2"></T>
ItpnP hadvhD aat0 sharpaj0
·pun pointedaj0 facenn1 andcjC
<C FE="BodP" PT="NP" GF="Ext">
aat0 featheryaj0 tailnn1 thatcjt
</C>
<C TARGET="y"> archedvvd </C>
<C FE="Path" PT="PP" GF="Comp">
overavp-prp itsdps backnn1
</C>
·pun
</S>
```

The constituent (C) tags identify the phrases that have been annotated. The frame element (FE) attribute indicates the semantic roles, and the phrase type (PT) attribute indicates the grammatical function of the phrase. For the work here, the prepositional phrase annotation is isolated and treated as the sense of the preposition. This yields the following derived annotation:

It had a sharp, pointed face and a feathery tail that arched over<sub>Path</sub> its back.

The annotation frequencies for the 31 most frequent prepositions are shown later in Table 5, again

ordered by entropy. This illustrates that the role distributions are more complicated, yielding higher entropy values on average. In all, there are over 100 prepositions with annotations, 65 with ten or more instances each.

## 2.2 Word-sense disambiguation

The task of selecting the semantic role for the prepositions can be framed as a type of word-sense disambiguation (WSD), where the semantic roles define high-level senses for the prepositions.

Since the TREEBANK roles are more general than those above, the disambiguation in the first set of experiments address a coarse form of sense distinction. For ‘for’ there are 6 distinctions (or 4 with low-frequency pruning). In contrast, since the FRAMENET distinctions are quite specific, the disambiguation in the second set of experiments address fine-grained sense distinctions. For ‘for’, there are 41 distinctions (or 18 with low-frequency pruning). For the final set of experiments, a coarse-grained set of distinctions is again used since the FRAMENET roles are converted into TREEBANK roles.

## 3 Classification experiments

A straight-forward approach for preposition disambiguation would be to use standard WSD features, such as the parts-of-speech of surrounding words, and, more importantly, collocations (e.g., lexical associations). Although this can be highly accurate, it will likely overfit the data and generalize poorly. To overcome these problems, a class-based approach is used for the collocations, using WordNet high-level synsets as the source of the word classes. Therefore, in addition to using collocations in the form of other words, this uses collocations in the form of semantic categories.

A supervised approach for word-sense disambiguation is used following Bruce and Wiebe (1999). The results described here were obtained using the settings in Figure 1. These are similar to the settings used by O’Hara et al. (2000) in the first SENSEVAL competition. This shows that for the hypernym associations, only those words that occur within 5 words of the target prepositions are con-

| Relation | Probability | Example                               |
|----------|-------------|---------------------------------------|
| locative | 0.73        | workers <i>at</i> a factory           |
| temporal | 0.24        | expired <i>at</i> midnight<br>Tuesday |

Table 1: Prior probabilities of semantic relations for ‘at’ in TREEBANK

sidered.<sup>2</sup>

The main difference from that of a standard WSD approach is that, during the determination of collocations, each word token is replaced by synset tokens for its hypernyms in WordNet, several of which might occur more than once. This introduces noise due to ambiguity, but given the conditional-independence selection scheme, the preference for hypernym synsets that occur for different words will compensate somewhat. O’Hara and Wiebe (2003) provide more details on the these types of collocations.

### 3.1 TREEBANK

To see how these conceptual associations are derived, consider the differences in the prior versus class-based conditional probabilities for the semantic roles of the preposition ‘at’ in TREEBANK. Table 1 shows the global probabilities for the roles assigned to ‘at’, excluding cases below 5%. Table 2 shows the conditional probabilities for these roles given that certain high-level WordNet categories occur in the context. These category probability estimates were derived by tabulating the occurrences of the synsets for the words occurring within a 5-word window of the target preposition. In a context with a concrete concept (ENTITY#1), the difference in the probability distributions shows that the *locative* interpretation is even more likely. In contrast, in a context with an abstract concept, the difference in the probability distributions shows that the *temporal* interpretation becomes more likely. Therefore, these class-based lexical associations reflect the intuitive use of the prepositions.

<sup>2</sup>This window size was chosen after estimating that on average the prepositional objects occur within 2.35 +/- 1.26 words of the preposition and that the average attachment site is within 3.0 +/- 2.98 words. These figures were produced by analyzing the parse trees for the semantic role annotations in the PENN TREEBANK.

**Features:**

- POS -2            part-of-speech 2 words to left  
 POS -1:           part-of-speech 1 word to left  
 POS:              part-of-speech of preposition  
 POS +1:          part-of-speech 1 word to right  
 POS +2:          part-of-speech 2 words to right  
 Prep                preposition being classified  
 WordColl<sub>i</sub>:      context includes word collocation for role i  
 HypernymColl<sub>i</sub>: context includes hypernym collocation for role i

**Collocation Context:**

- Word:              anywhere in the sentence  
 Hypernyms:      within 5 words of target preposition

**Collocation selection:**

- Frequency constraint:  $f(word) > 1$   
 Conditional independence threshold:  $\frac{p(c|coll)-p(c)}{p(c)} \geq 0.2$   
 Organization:      per-class-binary grouping (Wiebe et al., 1998)

**Model selection:**

- overall classifier: Decision tree (via Weka’s J48 classifier (Witten and Frank, 1999))  
 individual classifiers: Naive Bayes  
 10-fold cross-validation

Figure 1: Feature settings used in the preposition classification experiments

| Category      | Relation | P(R C) |
|---------------|----------|--------|
| entity#1      | locative | 0.86   |
| entity#1      | temporal | 0.12   |
| abstraction#6 | locative | 0.51   |
| abstraction#6 | temporal | 0.46   |

Table 2: Sample conditional probabilities of semantic relations for ‘at’ in TREEBANK

Classifying the prepositions in the PENN TREEBANK shows that this approach is very effective. Table 3 shows the results of classification when all of the prepositions are classified together. Unlike the general case for WSD, the sense inventory is the same for all words here, so that a single classifier can be produced rather than individual classifiers. This has the advantage of allowing for more training data to be used in the derivation of the collocations indicative of each semantic role. High accuracy is achieved when just using standard word collocations. The table also shows that further improvements are generally achieved using both types of collocations. For the combined case, the accuracy is 85.8%, using Weka’s

# Instances: 26616  
 # Classes: 7  
 Entropy: 1.917  
 Baseline: 0.480

| Experiment | Accuracy | Stdev |
|------------|----------|-------|
| Word       | 81.3     | 0.50  |
| Combined   | 85.8     | 0.60  |

Table 3: Overall results for preposition disambiguation with TREEBANK semantic roles

J48 classifier (Witten and Frank, 1999), which is an implementation of Quilan’s C4.5 (Quinlan, 1993) decision tree learner. For comparison, Table 4 shows the results for individual classifiers created for each preposition (using Naive Bayes). In this case, the word-only collocations perform slightly better: 78.5% versus 77.4% accuracy.

**3.2 FRAMENET**

It is illustrative to compare the prior probabilities for FRAMENET to those seen earlier for ‘at’. See Table 6 for the most frequent roles out of the 40 cases that were assigned to it. This highlights a dif-

| Prep       | Freq | Entropy | Baseline | Word Only | Both  |
|------------|------|---------|----------|-----------|-------|
| through    | 332  | 1.668   | 0.438    | 0.598     | 0.634 |
| as         | 224  | 1.647   | 0.399    | 0.820     | 0.879 |
| by         | 1043 | 1.551   | 0.501    | 0.867     | 0.860 |
| between    | 83   | 1.506   | 0.483    | 0.733     | 0.751 |
| of         | 30   | 1.325   | 0.567    | 0.800     | 0.814 |
| out        | 76   | 1.247   | 0.711    | 0.788     | 0.764 |
| for        | 1406 | 1.223   | 0.655    | 0.805     | 0.796 |
| on         | 1927 | 1.184   | 0.699    | 0.856     | 0.855 |
| throughout | 61   | 0.998   | 0.525    | 0.603     | 0.584 |
| across     | 78   | 0.706   | 0.808    | 0.858     | 0.748 |
| from       | 1521 | 0.517   | 0.917    | 0.912     | 0.882 |
| total      | 6781 | 1.233   | 0.609    | 0.785     | 0.778 |

Table 4: Per-word results for preposition disambiguation with TREEBANK semantic roles

| Prep    | Freq  | Entropy | Baseline | Word Only | Both  |
|---------|-------|---------|----------|-----------|-------|
| between | 286   | 3.258   | 0.490    | 0.325     | 0.537 |
| against | 210   | 2.998   | 0.481    | 0.310     | 0.586 |
| under   | 125   | 2.977   | 0.385    | 0.448     | 0.440 |
| as      | 593   | 2.827   | 0.521    | 0.388     | 0.598 |
| over    | 620   | 2.802   | 0.505    | 0.408     | 0.526 |
| behind  | 144   | 2.400   | 0.520    | 0.340     | 0.473 |
| back    | 540   | 1.814   | 0.544    | 0.465     | 0.567 |
| around  | 489   | 1.813   | 0.596    | 0.607     | 0.560 |
| round   | 273   | 1.770   | 0.464    | 0.513     | 0.533 |
| into    | 844   | 1.747   | 0.722    | 0.759     | 0.754 |
| about   | 1359  | 1.720   | 0.682    | 0.706     | 0.778 |
| through | 673   | 1.571   | 0.755    | 0.780     | 0.779 |
| up      | 488   | 1.462   | 0.736    | 0.736     | 0.713 |
| towards | 308   | 1.324   | 0.758    | 0.786     | 0.740 |
| away    | 346   | 1.231   | 0.786    | 0.803     | 0.824 |
| like    | 219   | 1.136   | 0.777    | 0.694     | 0.803 |
| down    | 592   | 1.131   | 0.764    | 0.764     | 0.746 |
| across  | 544   | 1.128   | 0.824    | 0.820     | 0.827 |
| off     | 435   | 0.763   | 0.892    | 0.904     | 0.899 |
| along   | 469   | 0.538   | 0.912    | 0.932     | 0.915 |
| onto    | 107   | 0.393   | 0.926    | 0.944     | 0.939 |
| past    | 166   | 0.357   | 0.925    | 0.940     | 0.938 |
| total   | 10432 | 1.684   | 0.657    | 0.685     | 0.703 |

Table 5: Per-word results for preposition disambiguation with FRAMENET semantic roles

| Relation   | RelFreq | Example                                    |
|------------|---------|--|
| addressee  | 0.315   | growled <i>at</i> the attendant            |
| other      | 0.092   | chuckled heartily <i>at</i> this admission |
| phenomenon | 0.086   | gazed <i>at</i> him with disgust           |
| goal       | 0.079   | stationed a policeman <i>at</i> the gate   |
| content    | 0.051   | angry <i>at</i> her stubbornness           |

Table 6: Prior probabilities of semantic relations for ‘at’ in FRAMENET

| Category      | Relation   | $P(R C)$ |
|---------------|------------|----------|
| entity#1      | addressee  | 0.28     |
| entity#1      | goal       | 0.11     |
| entity#1      | phenomenon | 0.10     |
| entity#1      | other      | 0.09     |
| entity#1      | content    | 0.03     |
| abstraction#6 | addressee  | 0.22     |
| abstraction#6 | other      | 0.14     |
| abstraction#6 | goal       | 0.12     |
| abstraction#6 | phenomenon | 0.08     |
| abstraction#6 | content    | 0.05     |

Table 7: Sample conditional probabilities of semantic relations for ‘at’ in FRAMENET

ference between the two sets of annotations. The common *temporal* role from TREEBANK is not directly represented in FRAMENET, and it is not subsumed by another specific role. Similarly, there is no direct role corresponding to *locative*, but it is partly subsumed by *goal*. This reflects the bias of FRAMENET towards roles that are an integral part of the frame under consideration: location and time apply to all frames, so these cases are not generally annotated.

The overall classifier for the FRAMENET prepositions was not produced due to system limitations related to the larger number of roles. However, the results over FRAMENET using individual classifiers shows that the combined collocations produce better results: 70.3% versus 68.5%. Table 5 shows these results

## 4 Related work

Until recently, there has not been much work specifically on preposition classification, outside of early work by Halliday (1956). Preposition classification is indirectly addressed in machine translation. In some cases, the issue is avoided by translating the preposition into a corresponding foreign function word without regard to the preposition’s underlying meaning (i.e., direct transfer). Other times an internal representation is helpful. Trujillo (1992) discusses these issues in depth. Japkowicz and Wiebe (1991) illustrate the deep meaning approach in using conceptual structures to account for the differences in how prepositions are used to conceptualize objects.

There is currently more interest in this type of classification. Litkowski (2002) presents manually-derived rules for disambiguating prepositions, in particular for ‘of’. Srihari et al. (2001) present manually-derived rules for disambiguating prepositions used in named entities.

The work by Gildea and Jurafsky (2002) and by Blaheta and Charniak (2000) address the more general problem of assigning semantic roles to arbitrary constituents of a sentence. Their approaches use more syntactically oriented features since many of the constituents reflect syntactic dependencies.

Gildea and Jurafsky (2002) classify semantic role assignments using all the annotations in FRAMENET. They use several features derived from the output of a parser, such as the constituent type of the phrase (e.g., NP) and the grammatical function (e.g., subject). They include lexical features for the headword of the phrase and the predicating word for the entire annotated frame. To account for sparsity of the data, especially when dealing with the lexical features, a backoff model is applied with more specific feature combinations being used only when present in the training data. Note that by using a bag-of-words approach to the collocations, we avoid the need for smoothing. They report an accuracy of 76.9% with a baseline of 40.6% over the same FRAMENET data versus our 70.3% accuracy with a baseline of 65.7%. However, due to the conditioning the classification on the predicating word, the range of roles for a particular classification is more limited than in our case.

Blaheta and Charniak (2000) classify semantic role assignments using all the annotations in TREEBANK. They also include a few parser-derived features, such as the constituent labels for nearby nodes and part-of-speech for parent and grandparent nodes. They also include lexical features for the head and alternative head (since prepositions are considered as the head by their parser). To account for data sparsity, smoothing is employed based on a tree organization of the features. They report an accuracy of 77.6% over the form/function tags from the PENN TREEBANK with a baseline of 37.8%,<sup>3</sup> which is below our 85.8% accuracy with 48.0% baseline. The tasks are somewhat different, since they address all adjuncts, not just prepositions, hence their lower baseline. In addition, they tag for the *nominal* and *adverbial* roles, which are syntactic and presumably more predictable than the others in this group.

## 5 Conclusion

Classifying prepositions according to the PENN TREEBANK annotations can be accurate (78.5% individually and 85.8% together), while retaining ability to generalize via class-based lexical associations. Thus, these annotations can be used for default classification of prepositions in case more fine-grained semantic role information cannot be determined. For the fine-grained FRAMENET roles, the performance is less accurate but still respectable (70.3%). In both cases, the best accuracy is achieved using a combination of standard word collocations along with class collocations in the form of WordNet hypernyms.

Future work will address cross-dataset experiments. In particular, we will see whether the word and hypernym associations learned over FrameNet can be carried over into TreeBank, giving a mapping of the fine-grained FrameNet roles into the coarse-grain TreeBank ones. Such a mapping would be similar the one discussed in (Gildea and Jurafsky, 2002).

<sup>3</sup>They target all of the TREEBANK function tags but give performance figures broken down by the groupings defined in the TREEBANK tagging guidelines. The baseline figure shown above is their recall figure for the 'baseline 2' performance.

## Acknowledgements

The first author is supported by a generous GAANN fellowship from the Department of Education. Some of the work used computing resources at NSU made possible through MII Grants EIA-9810732 and EIA-0220590.

## References

- Don Blaheta and Eugene Charniak. 2000. Assigning function tags to parsed text. In *Proc. NAACL-00*.
- Rebecca Bruce and Janyce Wiebe. 1999. Decomposable modeling in natural language processing. *Computational Linguistics*, 25 (2):195–208.
- P. Edmonds and S. Cotton, editors. 2001. *Proceedings of the SENSEVAL 2 Workshop*. Association for Computational Linguistics.
- Charles J. Fillmore, Charles Wooters, and Collin F. Baker. 2001. Building a large lexical databank which provides deep semantics. In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong.
- C. Fillmore. 1968. The case for case. In Emmon Bach and Robert T. Harms, editors, *Universals in Linguistic Theory*. Holt, Rinehart and Winston, New York.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
- M.A.K. Halliday. 1956. The linguistic basis of a mechanical thesaurus, and its application to English preposition classification. *Mechanical Translation*, 3.3:81–88.
- Nathalie Japkowicz and Janyce Wiebe. 1991. Translating spatial prepositions using conceptual information. In *Proc. 29th Annual Meeting of the Assoc. for Computational Linguistics (ACL-91)*, pages 153–160.
- Daniel Jurafsky and James H. Martin. 2000. *Speech and Language Processing*. Prentice Hall, Upper Saddle River, New Jersey.
- A. Kilgarriff. 1998. SENSEVAL: An exercise in evaluating word sense disambiguation programs. In *Proc. First International Conference on Language Resources and Evaluation*, pages 581–588. Granada, Spain.
- Fritz Lehmann. 1996. Big posets of participations and thematic roles. In P. Eklund, G. Ellis, and G. Mann, editors, *Conceptual Structures: Knowledge Representation as Interlingua*, pages 50–74, Berlin. Springer-Verlag.

- K. C. Litkowski. 2002. Digraph analysis of dictionary preposition definitions. In *Proceedings of the Association for Computational Linguistics Special Interest Group on the Lexicon*. July 11, Philadelphia, PA.
- M. Marcus, Santorini, B., and M. Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Mitchell Marcus, Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating predicate argument structure. In *ARPA Human Language Technology Workshop*.
- Tom O’Hara and Janyce Wiebe. 2003. Classifying functional relations in Factotum via WordNet hypernym associations. In *Proc. Fourth International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2003)*.
- Tom O’Hara, Janyce Wiebe, and Rebecca F. Bruce. 2000. Selecting decomposable models for word-sense disambiguation: The grling-sdm system. *Computers and the Humanities*, 34 (1-2):159–164.
- J. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Mateo, California.
- Rohini Srihari, Cheng Niu, and Wei Li. 2001. A hybrid approach for named entity and sub-type tagging. In *Proc. 6th Applied Natural Language Processing Conference*.
- Arturo Trujillo. 1992. Locations in the machine translation of prepositional phrases. In *Proc. TMI-92*, pages 13–20.
- Janyce Wiebe, Kenneth McKeever, and Rebecca Bruce. 1998. Mapping collocational properties into machine learning features. In *Proc. 6th Workshop on Very Large Corpora (WVLC-98)*, pages 225–233, Montreal, Quebec, Canada. Association for Computational Linguistics SIGDAT.
- Ian H. Witten and Eibe Frank. 1999. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann.