Spring April 14, 2015

# Prepositional Phrase Attachment Problem Revisited: How VERBNET Can Help

Dan Bailey, *University of Nebraska at Omaha*
Yuliya Lierler, *University of Nebraska at Omaha*
Benjamin Susman, *University of Nebraska at Omaha*

# Prepositional Phrase Attachment Problem Revisited: How VERBNET Can Help


Dan Bailey
University of Nebraska at Omaha
dbailey@unomaha.edu

Yuliya Lierler
University of Nebraska at Omaha
ylierler@unomaha.edu

Benjamin Susman
University of Nebraska at Omaha
bsusman@unomaha.edu



**Abstract**

Resolving attachment ambiguities is a pervasive problem in syntactic analysis. We propose and investigate an approach to resolving prepositional phrase attachment that centers around the ways of incorporating semantic knowledge derived from the lexico-semantic ontologies such as VERBNET and WORDNET.


## 1 Introduction

Syntactic parsing is a process of uncovering the internal structure of sentences, in particular, articulating what the constituents of a given sentence are and what relationships are between them. Software systems that perform this task are called *(syntactic) parsers*. Parsing technology has seen striking advances. Wide-coverage off-the-shelf parsers are freely available and ready to use. Yet, modern parsers are not at the level of human-expert agreement. One of the notorious problems in parsing technology is determining prepositional phrase attachment. For example, the following phrase by Ray Mooney:

*eat spaghetti with chopsticks.* (1)

is syntactically ambiguous allowing for two syntactic structures: in one, the prepositional phrase *with chopsticks* modifies (attaches to) the verb *eat* and in another, it modifies the noun *spaghetti*. The latter is erroneous as it suggests that *spaghetti with chopsticks* constitutes a meal. The phrase

*eat spaghetti with meatballs* (2)

is syntactically ambiguous in a similar manner. Modern advanced parsers do not produce proper syntactic representations for these phrases: instead they favor the same structure for both statements (Lierler and Schüller, 2013).

These "spaghetti" examples illustrate the necessity to incorporate semantic knowledge into the parsing process, in particular, one has to take into account *selectional restrictions* (Katz and Fodor, 1963) — the semantic, common-sense restrictions that a word imposes on the environment in which it occurs. For instance, the fact that chopsticks are an instrument suggests that *with chopsticks* modifies *eat spaghetti* in phrase (1) as a tool for eating. Current statistical methods, dominant in the field of syntactic analysis, take into account selectional restrictions *implicitly* by assigning the most probable syntactic structure based on observed co-occurrences of words and structures in training corpora. As mentioned, this is often not sufficient. In this work we propose and investigate an approach to the prepositional phrase attachment problem that incorporates explicit semantic knowledge available in the lexico-semantic dataset VERBNET into the decision process for resolving the ambiguity of prepositional statements. Machine learning forms a backbone of the decision procedure that we investigate.

This work targets to bring knowledge representation techniques into syntactic parsing. Indeed, lexical ontologies VERBNET and WORDNET are at heart of this project. Lierler and Schüller (2013) advocated a framework for parsing that results in what they call semantically-coherent syntactic parses. These parses account for selectional restrictions. On the one hand, that work suggests a promising direction. On the other hand, it outlines the need for automatic methods for acquiring lexico-semantic information that relates to parsing a sentence. Present work takes a step in the direction of establishing principles to mine existing lexico-semantic resources and incorporate found information into parsing process.

The importance of taking semantic information, including selectional restrictions, into account during parsing has long been recognized. Ford (1982), Jensen and Binot (1987), Hirst (1988), Dahlgren (1988), and Allen (1995) devised methods for parsing that performed selectional restrictions analysis. These methods assume that a systematic word taxonomy as well as a database of selectional restrictions is available. Developments in the field of Lexical semantics have made such systematic large scale datasets, including WORDNET (Miller et al., 1990) and VERBNET (Kipper et al., 2000), a reality. The WORDNET project provides a taxonomy that organizes words into a coherent representation that reflects some lexical relationships between them. The VERBNET project provides a domain-independent, broad-coverage verb lexicon that includes selectional restriction information. Also recent research illustrates the benefits of lexico-semantic information in tasks closely related to parsing. Zhou et al. (2011) illustrate how web-derived selectional preferences improve dependency parsing. Zapirain et al. (2013) show that selectional preferences obtained by means of WORDNET-based similarity metrics improve semantic role labeling. Srikumar and Roth (2013) illustrate how selectional restrictions posed by prepositions improve relation prediction performance. Agirre et al. (2008, 2011) also suggest the necessity of incorporating semantic information into parsing by providing evidence that word sense information stemming from WORDNET improves parsing and prepositional phrase attachment.

These findings support the importance of developing parsing algorithms that can handle semantic information effectively. We view the decision procedure for resolving prepositional phrase attachment developed in this paper as complementary to above mentioned methods. The main driving vehicle of this work is the VERBNET ontology. To the best of our knowledge no current approach relies on the use of VERBNET in compiling selectional preferences information.

The prepositional phrase attachment problem has received a lot of attention as a stand alone task. Lapata and Keller (2005) provide a summary of systems attempting to solve this problem. All of the reported systems have centered on machine learning approaches such as a maximum entropy model (Ratnaparkhi et al., 1994), a back-off model (Collins and Brooks, 1995), a transformation based approach (Brill and Resnik, 1994), a memory-based learning approach (Zavrel et al., 1997), and unsupervised approaches (Ratnaparkhi, 1998) or (Pantel and Lin, 2000). The reported accuracy of the systems ranged from 81.60% to 88.10%. The average human precision on the task is reported to be 88.20%.

The outline of the paper follows. We begin by introducing the relevant resources and concepts, in particular, VERBNET and WORDNET along with the concept of selectional restriction. Following that, we introduce the problem of prepositional phrase attachment. Once these foundations have been laid, we provide the details of a machine-learning based algorithm that makes use of the VERBNET and WORDNET resources in a systematic way. We then evaluate a system that implements the outlined algorithm and discuss our plans for its future.

## 2 The VERBNET, WORDNET Lexicons and Selectional Restrictions

Levin classes (Levin, 1993) are groups of verbs that share usage patterns and have semantic similarity. For instance, the Levin class for the verb *hit* includes the words *bang*, *bash*, *click* and *thwack*. These words can be used alike and suggest similar sentence structures. Organizing verbs into groups according to the similarity of their syntactic behavior is the basis of Levin classes. It is supported by an extensive study suggesting that similar syntactic behavior translates into common semantic features of verbs (Levin, 1993). The VERBNET ontology (Kipper et al., 2000) is an English-language verb lexicon that collects verbs into extended Levin classes and provides information about the sentence structure that these classes share.

The VERBNET dataset is composed of basic structures, called *frame syntax*. For example, a frame syntax for a *hit*-verb class follows:

$$\texttt{AGENT}_{intControl}\ \texttt{V}\ \texttt{PATIENT}\ \{\texttt{with}\}\ \texttt{INSTRUMENT}_{concrete} \tag{3}$$

This frame syntax suggests that one possible structure for the use of a verb in the *hit*-class is to have an AGENT followed by the verb itself, then a PATIENT, the preposition *with* and an INSTRUMENT. The AGENT, PATIENT and INSTRUMENT are called *thematic roles*. VERBNET allows 23 such roles including THEME, RECIPIENT, SOURCE.[1] The thematic roles in VERBNET are augmented further by *restrictions*. VERBNET maintains a hierarchy of restrictions based on the top-level entries in EuroWordNet (Kipper-Schuler, 2005, Section 3.1.2) consisting of 37 entries. This hierarchy allows VERBNET to specify that the AGENT thematic role for the verbs in class *hit* is of the type *intelligent control* (*intControl*) and the INSTRUMENT role in *hit* is *concrete*. In other words, an entity that serves an INSTRUMENT role of *hit* possesses a property of being *concrete* – some concrete physical object.

The WORDNET system is a comprehensive manually developed lexical database from Princeton University (Miller et al., 1990). In WORDNET, nouns, verbs, and adjectives are organized into synonym sets each representing one underlying lexical concept. Several semantic relations among words are incorporated into WORDNET as links between the synonym sets. These semantic relations include super/subordinate relations — *hypernymy, hyponymy or ISA relation*; and part-whole relation — *meronymy*. Thus we can investigate relationships between various concepts by following links within WORDNET. For instance, by following the ISA links, one may easily establish that synonym set containing a noun *boy* is in ISA relation with a synonym set for the *intelligent control* concept. The WORDNET lexicon has been extensively used for developing metrics and procedures for determining the relatedness/similarity of lexical concepts. The task of identifying whether and how given words are related has many applications in natural language processing (NLP) (e.g., word sense disambiguation, information extraction). Budanitsky and Hirst (2006) present a comprehensive study that compares five different measures of relatedness based on WORDNET including a measure by Leacock and Chodorow (1998). In this work we also use WORDNET for similar purposes. For example, with the help of WORDNET we define what it means that a noun "matches" a restriction or a thematic role. Section 4 presents the definition of matching.

Selectional restrictions (Katz and Fodor, 1963) are the semantic, common-sense restrictions that a word imposes on the environment in which it occurs. A *selectional construct* is a tuple $[w, t, r, p]$ where (i) $w$ is a word, (ii) $t$ is a thematic role that the word $w$ allows, (iii) $r$ is a restriction on the thematic role $t$ with respect to the word $w$ (by the empty set we denote no restrictions), (iv) $p$ is a set of prepositions that can be used to realize the thematic role $t$ of the word $w$ (this set may be empty suggesting that no preposition is necessary to realize this thematic role). Selectional construct is meant to capture the selectional restrictions (sometimes we use these terms interchangeably). The VERBNET lexicon can be viewed as a systematic, wide-coverage knowledge base about selectional restrictions of verbs. Recall a frame syntax (3) for the verb *hit*. We now present three selectional constructs that follow from the frame:

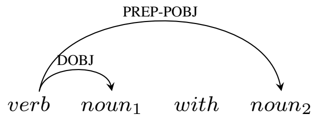
$$(hit, \text{AGENT}, intControl, \emptyset),\quad (hit, \text{PATIENT}, \emptyset, \emptyset),\quad (hit, \text{INSTRUMENT}, concrete, \{with\}).$$

## 3 Prepositional Phrase Attachment

Resolving prepositional phrase (PP) attachment ambiguities is a pervasive problem in NLP exemplified by phrases (1) and (2). They look "identical" modulo one word, yet the proper syntactic analyzer will process (1) differently from (2). Indeed, the "instrumental" use of the preposition *with* — as in phrase (1)

[1]Table 2 of the VERBNET website `http://verbs.colorado.edu/~mpalmer/projects/verbnet.html` lists the thematic roles and brief explanations.

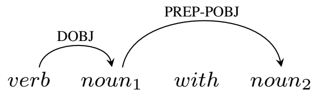
should be parsed into dependency structure of the form:




(4)

This parse structure reveals the prepositional phrase attachment describes the action being undertaken.

"Comitative" use of *with* such as in phrase (2) leads to a structure of the form




(5)

We call (4) and (5) $\mathcal{P}$-parse structures. We call a phrase, $\mathcal{P}$*-phrase*, when it has the form

verb noun-phrase preposition noun-phrase.

In the introduction, we argued that selectional restrictions provide sufficient information to disambiguate many PP attachments. We now incorporate selectional restrictions into syntactic parsing of $\mathcal{P}$-phrases. We say that a selectional construct *justifies* an edge annotated by PREP-POBJ from $w$ to $n$ if it has the form $(w, t, r, p)$, where thematic role $t$ and restriction $r$ on $t$ are "matched" by word $n$. Section 4 presents the definition of matching for different thematic roles and restrictions. A $\mathcal{P}$-parse structure is *semantically coherent* if its edge annotated by PREP-POBJ is justified by some selectional construct triggered by the words occurring in a $\mathcal{P}$-parse structure.

For example, $\mathcal{P}$-phrase (1) triggers the selectional construct

$$(eat, \text{INSTRUMENT}, concrete, \{with\}). \tag{6}$$

Intuitively, this construct justifies the PREP-POBJ edge between $spaghetti$ and $chopsticks$. $\mathcal{P}$-parse of the form (4) is thus semantically coherent.

We can view selectional restrictions as conditions that must be satisfied in the process of parsing. In other words, semantically coherent parse structures are the ones that satisfy these conditions. It is clear that at times more than one parse structure is semantically coherent for a phrase. Similarly, more than one selectional construct may justify a PREP-POBJ edge.

# 4 PP-attachment Selection Algorithm

The dataset by Ratnaparkhi et al. (1994) is often used to devise and evaluate PP-attachment resolution systems. We use it in this work also. For the rest of the paper we refer to the Ratnaparkhi et al. dataset as $\mathcal{R}$. The $\mathcal{R}$ dataset is a collection of $\mathcal{P}$-phrases stemming from Penn Treebank. Each data entry in $\mathcal{R}$ is a tuple of the form

$$(verb,\ noun_1,\ prep,\ noun_2). \tag{7}$$

Intuitively, each tuple corresponds to a $\mathcal{P}$-phrase. Figure 1 presents the basic statistics on ten most occurring prepositions in the dataset. The second row titled *Total* gives the number of tuples contained in $\mathcal{R}$ that mention the respective preposition. The third row presents the ratio of the number of occurrences of a preposition (the second row) over the size of the $\mathcal{R}$ dataset. Overall $\mathcal{R}$ contains $23898$ $\mathcal{P}$-phrases. The last row represents the frequencies of the verb attachment for the respective prepositions. We note how by far the most frequent preposition $of$ is also very bias in a sense that $99\%$ of the time it triggers the attachment to a noun. This is why we present the column named *All-of* that gives the statistics for all tuples that do not contain the preposition $of$.

Machine learning methods are commonly used for implementing decision/classification procedures called classifiers. In supervised learning, the classifier is first trained on a set of labeled data (training data) that is representative of the domain of interest. Typically labeled data consists of pairs of

| Preposition | All | All - of | of | in | to | for | on | from | with | at | as | by |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Total | 23898 | 17395 | 6503 | 3973 | 3005 | 2522 | 1421 | 1059 | 1049 | 780 | 564 | 526 |
| % of $\mathcal{R}$ | 100 | 72.8 | 27.2 | 16.6 | 12.6 | 10.6 | 5.9 | 4.4 | 4.4 | 3.3 | 2.4 | 2.2 |
| % Verb Attachment | 46.9 | 64.2 | 0.9 | 54.6 | 80.1 | 51.2 | 53.8 | 68.6 | 64.4 | 80.4 | 81.2 | 72.2 |

Figure 1: Basic statistics on the Ratnaparkhi et al. (1994) dataset.

input objects and a desired output. An input object is often summarized by so called feature vector. A classifier of choice analyzes the training data and produces a model that can be used to evaluate unlabeled input objects. In this work we rely on Logistic Regression classification algorithm as the vehicle for implementing the decision procedure for the PP-attachment selection problem. To implement this procedure we used the Logistic Regression classifier with a ridge estimator of $10^{-7}$ available in Weka[2] (Hall et al., 2009) – software by University of Waikato which contains tools for data preprocessing, classification, and clustering. We call our system PPATTACH, which is available at `http://www.unomaha.edu/nlpkr/software/ppattach/`.

In our settings we used the $\mathcal{R}$ dataset to produce training data. Each $\mathcal{P}$-phrase (7) in $\mathcal{R}$ is mapped to a feature vector composed of five elements:

- the preposition $prep$ of the tuple (7);
- the VERBNET verb class of the $verb$ in (7). If the verb class is unavailable in VERBNET then lemmatized $verb$ serves the role of a feature itself. We call this feature *Verbclass*;
- Features named VERBNET$[noun_1, noun_2]$, VERBNET$[noun_2]$, and *Nominalization*, which encode information that some selectional constructs stemming from VERBNET are "applicable" to the tuple (7).

We now speak about the rationale behind choosing these features. Figure 1 clearly indicates that prepositions are bias to one or another attachment decision. Lapata and Keller (2005) present a generic baseline for the prepositional phrase attachment decision by choosing noun attachment in all cases, which achieves correct attachment 56.9% of the time. They further present that this baseline can be improved simply by choosing the most likely attachment decision for each preposition reaching 72.2% accuracy. These observations provide strong evidence for the necessity of the first feature. As discussed verbs impose selectional restrictions. The second feature in combination with the first one allows us to use the $\mathcal{R}$ dataset to collect the statistical information about verb classes and their usage. The last three features are based on the information stemming from VERBNET. These features allow us to incorporate explicit information on selectional restrictions available in VERBNET into the decision procedure for the PP-attachment selection problem. We now proceed towards the description of how these three VERBNET-based features are computed.

To describe the computation of the VERBNET-based features precisely, we define a concept of matching. We say that a noun *matches* a thematic role (a restriction) listed in Figure 2 if one of its WORDNET senses has a path in WORDNET justified by the ISA links to a corresponding lexical concept depicted in Figure 2. We also say that a noun *matches* the thematic role INSTRUMENT if the definition (gloss) of one of the noun's WORDNET senses contains a string "used". Likewise, we establish a *match* with the restriction $pointy$ by finding the string, "sharp" within the definition for one of the noun's senses. Accounting for parts of word's definition stems from the work by Jensen and Binot (1987). Figure 2 contains all 23 thematic roles of the VERBNET dataset.

Descriptions of the computation procedures of VERBNET-based features follow. Each procedure is given a tuple of the form (7) as its input. These features are binary, their default values are 0.

**Feature** VERBNET$[noun_1, noun_2]$: We start by searching for all verb-classes that include $verb$ from tuple (7). Frame syntax structures of the form

$$\text{THEMROLE verb THEMROLE}_{1\,restriction_1}\ \{\text{prep}\}\ \text{THEMROLE}_{2\,restriction_2} \qquad (8)$$

[2] `http://www.cs.waikato.ac.nz/ml/weka/`

| Thematic Role | WORDNET Concept |
|---|---|
| THEME, PATIENT, RECIPIENT, OBLIQUE, DESTINATION, EXPERIENCER, SOURCE, BENEFICIARY, AGENT, PRODUCT, MATERIAL, TOPIC, PREDICATE, ASSET, EXTENT, PROPOSITION, CAUSE, VALUE | entity.n.01 |
| ACTOR | causal_agent.n.01 |
| LOCATION | location.n.01, location.n.03 |
| INSTRUMENT | instrumentality.n.03, act.n.02, communication.n.02, body_part.n.01 |
| ATTRIBUTE | attribute.n.02 |
| STIMULUS | stimulation.n.02 |

| Restriction | WORDNET Concept |
|---|---|
| *abstract* | abstraction.n.06 |
| *communication* | communication.n.02 |
| *body_part* | body_part.n.01 |
| *force* | entity.n.01 |
| *pointy*, *concrete*, *refl*, *solid* | physical_entity.n.01 |
| *organization* | group.n.01 |
| *region* | region.n.01 |

| Restriction | WORDNET Concept |
|---|---|
| *location* | location.n.01, location.n.03 |
| *animal* | animal.n.01 |
| *animate* | causal_agent.n.01, living_thing.n.01 |
| *currency* | currency.n.01 |
| *machine* | machine.n.01 |
| *scalar* | scalar.n.01 |
| *comestible* | comestible.n.01 |

Figure 2: WORDNET ISA-Parent

are extracted from these classes. Frame syntax (8) translates into selectional constructs that include

$$(verb, \text{THEMROLE}_1, \textit{restriction}_1, \emptyset) \tag{9}$$

$$(verb, \text{THEMROLE}_2, \textit{restriction}_2, \{prep\}) \tag{10}$$

For each frame syntax, we (i) verify whether $noun_1$ matches the thematic role $\text{THEMROLE}_1$ as well as the restriction $\textit{restriction}_1$, which suggests that selectional construct (9) justifies an edge between $verb$ and $noun_1$, and (ii) verify whether $noun_2$ matches $\text{THEMROLE}_2$ as well as $\textit{restriction}_2$, which suggests that selectional construct (10) justifies a PREP-POBJ edge between $verb$ and $noun_2$. If this test is positive for at least one frame syntax we assign value VERB to the feature VERBNET$[\textit{noun}_1, \textit{noun}_2]$.

**Feature** VERBNET$[\textit{noun}_2]$: This procedure is similar to the previous method. Frame syntax structures of the form

$$\text{THEMROLE v \{prep\} THEMROLE}_{2\,restriction_2} \tag{11}$$

are extracted from the verb-classes in VERBNET that include $verb$ from tuple (7). The frame syntax (11) translates into selectional constructs that include restriction (10). We then verify whether $noun_2$ matches $\text{THEMROLE}_2$ as well as $\textit{restriction}_2$, which suggests that selectional construct (10) justifies an edge between $verb$ and $noun_2$. Subsequently we assign value VERB to the feature VERBNET$[\textit{noun}_2]$.

**Feature *Nominalization***: *Nominalization* is the use of a verb, an adjective, or an adverb as a noun, with or without morphological transformation. In this work, we are especially interested in nouns derived from verbs. Such nouns typically behave as nouns grammatically, yet semantically they carry information of a respective verb. For example, a noun "conversation" is derived from a verb "to converse", which informally suggests at least two participants in the event of conversation. Given tuple (7), the *Nominalization* method starts by identifying whether $noun_1$ is derived from a verb. The WORDNET lexicon contains edges between nouns and verbs that are called *derivationally related forms*. We search WORDNET for connections via these edges between $noun_1$ and some verb. We require that the root word remains the same between $noun_1$ and a found verb. If such verb exists we consider $noun_1$ to be a nominalization. If $noun_1$ is derived from some verb, the VERBNET lexicon is searched for all verb-classes that include this verb. Frame syntax structures of the form (11) are extracted. We then verify whether $noun_2$ matches $\text{THEMROLE}_2$ as well as $\textit{restriction}_2$, which suggests that selectional construct (10) justifies a NOUN assignment for the feature.

## 5 Evaluation

We use various metrics to gauge the overall performance of the PPATTACH system. First we consider a baseline which consists of the most likely attachment on a per preposition basis. We also construct a PPATTACH- system by dropping the *Verbclass* feature from PPATTACH. We construct a GENERIC system by dropping the VERBNET-based features from the PPATTACH system.

We train and test each system on the whole $\mathcal{R}$ dataset and subsets of $\mathcal{R}$ on a preposition-by-preposition basis. Given that each resultant dataset is of limited size (see the second row in Figure 1), we use 10-fold cross-validation to evaluate the methods. The cross-validation was done in Weka. The main idea is to randomly select instances that constitute the test set. Subsequently we train a classifier on the remaining instances and evaluate the model on the selected test set. This is conducted ten times (with different test-training set pairs). Figure 3 summarizes the classification accuracy (the number of correct classifications over the number of classifications) of the system using Logistic Regression classifier.

| Preposition | All | All - of | of | in | to | for | on | from | with | at | as | by |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Baseline | 74.6 | 65.4 | 99.1 | 54.6 | 80.1 | 51.2 | 53.8 | 68.6 | 64.4 | 80.4 | 81.2 | 72.2 |
| PPATTACH- | 79.3 | 72.7 | 99.0 | 64.6 | 87.8 | 66.6 | 68.5 | 75.5 | 70.9 | 81.8 | 79.8 | 80.0 |
| GENERIC | 79.0 | 72.3 | 99.0 | 64.7 | 87.8 | 67.0 | 68.2 | 76.3 | 69.7 | 82.9 | 79.8 | 82.3 |
| PPATTACH | 79.3 | 72.5 | 99.0 | 64.7 | 88.0 | 66.9 | 69.6 | 75.4 | 70.7 | 81.9 | 78.5 | 81.7 |

Figure 3: Evaluation Data on PPATTACH using Logistic Regression.

We see substantial improvements from Baseline across most prepositions. Figure 4 presents data that can be used to explain this. For each VERBNET-based feature, this figure presents two rows. The row named *Recall* gives a percentage that describes the frequency at which the feature is assigned a value different from default; the row named *Precision* gives a percentage of relevant instances such that the feature assignment agrees with the correct attachment decision. For six out of ten prepositions the precision for the VERBNET$[noun_1, noun_2]$ feature is at least 83.6%. For five out of these prepositions the recall ranges from 10.3% to 37.6%. There are two prepositions $at$ and $as$ that have high precision, yet the performance of PPATTACH is comparable to that of Baseline. We also find that in this case the verb class does not play a role in improving the classification accuracy (Baseline and GENERIC behave practically identical). Figure 1 illustrates that the prepositions $at$ and $as$ have strong attachment bias for verb. Most of the features in PPATTACH also favor such attachment. Gaining evidence for the other decision shall improve the situation.

| VERBNET-based features | | of | in | to | for | on | from | with | at | as | by |
|---|---|---|---|---|---|---|---|---|---|---|---|
| VERBNET$[noun_1, noun_2]$ | Recall | 4.1 | 12.9 | 37.6 | 25.5 | 9.4 | 25.3 | 15.7 | 10.3 | 22.3 | 0.8 |
| | Precision | 6.0 | 66.0 | 91.5 | 59.6 | 71.6 | 83.6 | 92.7 | 93.8 | 98.4 | 100.0 |
| VERBNET$[noun_2]$ | Recall | 1.2 | 7.5 | 27.5 | 3.8 | 10.6 | 7.0 | 9.2 | 1.7 | 0.0 | 5.5 |
| | Precision | 0.0 | 59.8 | 89.7 | 60.4 | 85.3 | 82.4 | 88.5 | 100 | N/A | 96.6 |
| *Nominalization* | Recall | 1.5 | 12.8 | 9.8 | 3.9 | 3.7 | 2.8 | 10.3 | 0.3 | 0.0 | 1.0 |
| | Precision | 100.0 | 70.2 | 27.8 | 66.3 | 64.2 | 56.7 | 66.7 | 50.0 | N/A | 60.0 |

Figure 4: Features Evaluation for PPATTACH.

We now note on the difference that changing the classification algorithm can make to PPATTACH performance. Figure 5 summarizes the classification accuracy of PPATTACH using the Naïve Bayes classifier of Weka. In this case, PPATTACH- markedly lags behind GENERIC and PPATTACH, indicating the importance of the classification algorithm selection.

The PPATTACH system lags behind its peers (see Introduction). The top performing system for disambiguating prepositional attachment on $\mathcal{R}$ by Stetina and Nagao (1997) reported in (Lapata and Keller, 2005) incorporates manual word sense disambiguation. Also, let us take a closer look at several samples from $\mathcal{R}$. Consider tuples $(held, talks, with, parties)$ and $(establish, relation, with, institution)$.

They were annotated in Penn Treebank as having verb attachment suggesting errors in this corpus[3].

| Preposition | All | All - of | of | in | to | for | on | from | with | at | as | by |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PPATTACH- | 74.5 | 67.7 | 99.1 | 59.8 | 80.1 | 54.5 | 57.0 | 69.2 | 67.9 | 80.3 | 81.2 | 71.9 |
| GENERIC | 79.0 | 71.3 | 99.1 | 64.5 | 86.2 | 66.5 | 68.3 | 76.1 | 69.9 | 80.4 | 81.0 | 80.0 |
| PPATTACH | 78.9 | 71.2 | 99.1 | 65.3 | 87.9 | 66.5 | 68.5 | 76.5 | 72.8 | 80.5 | 81.0 | 80.0 |

Figure 5: Evaluation Data on PPATTACH using Naïve Bayes.

## 6 Beyond VERBNET: Preposition *with* Case-Study

This section focuses on a specific preposition, *with*. We investigate whether and how *with*-specific features improve classification accuracy. We start by noting that VERBNET often omits information. Consider sentence (1). There is nothing in VERBNET that suggests the selectional construct (6). This construct is intuitively triggered by the preposition *with* itself. Indeed, there are three main uses of *with*: *instrumental*, *adverbial*, and *comitative*. The instrumental use of *with* indicates that the prepositional phrase conveys details in which the object serves the role of an instrument while executing the action suggested by the verb. Phrase (1) illustrates the instrumental use of *with*. In contrast, the phrase *eat spaghetti with enthusiasm* illustrates an adverbial use of *with*. Here, the prepositional phrase answers the question of *how* the action was undertaken. To accommodate for common instrumental and adverbial uses of *with* we propose the following generic selectional constructs

$$(v, \text{INSTRUMENT}, \mathit{concrete}, \{\mathit{with}\}) \tag{12}$$
$$(v, \text{MANNER}, \emptyset, \{\mathit{with}\}), \tag{13}$$

where $v$ is a variable that can be substituted by *any* verb including *eat* or *hit*.

The grammatical case *comitative* denotes accompaniment. In English this case is often realized by *with* and captures the sense of *together with* or *in company with*. Expressions *spaghetti with meatballs* and *boat with an engine* illustrate the comitative case. In the former example, words *spaghetti* and *meatballs* are closely *related* to each other as they both denote food entities. In the later example, *boat* and *engine* are in lexical relation meronymy. We propose two selectional constructs that account for such examples

$$(w, \text{COMPANION}, \mathit{related}(w), \{\mathit{with}\}) \tag{14}$$
$$(w, \text{COMPANION}, \mathit{meronym}(w), \{\mathit{with}\}) \tag{15}$$

where $w$ is a variable that can be substituted by *any* word, e.g., *spaghetti* or *boat*; $\mathit{related}(w)$ stands for any word $w'$ such that $w$ and $w'$ are related (according to a certain metric); $\mathit{meronym}(w)$ stands for any word $w'$ such that $w'$ is a meronym of $w$.

In describing selectional constructs (14) and (15) we identified the need not only for a metric to establish relatedness between words, but also for a wide-coverage meronym relation database. The WORDNET lexicon records meronymy relations between synonym sets. However, it is not flawless and questions arise when attempting automatic methods for identifying meronymy. For example, in WORDNET *arm* is listed as a direct meronym of *human*, but *leg* is not. Thus to identify that *leg* is a meronym of *human*, deeper mining of WORDNET becomes a necessity. Later in this section we describe an algorithm that we devised for this purpose. To establish relatedness between words we rely on WORDNET and the metric developed by Leacock and Chodorow (1998).

Below we present features that capture the aforementioned reasoning as well as several other observations. We use these features to augment the PPATTACH system to construct the system PPATTACH+.

[3] *Held talks* represents the case of light verb construction; *establish* is an aspectual verb: both cases hint a noun attachment.

**Feature *Instrumentality***: This method accounts for "instrument" selectional construct (12). We verify whether $noun_2$ matches the thematic role INSTRUMENT, which suggests that selectional construct (12) justifies an edge between $verb$ and $noun_2$. We assign VERB to the feature.

**Feature *Adverbial Use***: We proposed to characterize an adverbial use of $with$ by selectional construct (13). We say that a noun *matches* the thematic role MANNER if there exists an adverbial derivation from a noun to some verb in WORDNET. The "derivationally related forms" edges of WORDNET are used to establish an adverbial derivation. If $noun_2$ in the given tuple (7) matches the thematic role MANNER then selectional construct (13) justifies an edge between $verb$ and $noun_2$. We assign VERB to the feature.

**Feature *Similarity***: This procedure accounts for "related" selectional construct (14). We verify whether $noun_1$ and $noun_2$ of the given tuple (7) are related using the Leacock-Chodorow algorithm (1998).If the value produced by the Leacock-Chodorow procedure exceeds 2, we assume that the nouns are related. This translates into the fact that selectional construct (12) justifies an edge between $noun_1$ and $noun_2$. We assign a value NOUN to the feature.

**Feature *Meronymy***: This procedure accounts for "meronymy" selectional construct (15). For a given tuple (7), we verify whether $noun_2$ is a meronym of $noun_1$ using a WORDNET-based method that we propose. First, we take the $noun_1$ and construct a set containing its full hypernymy and hyponymy tree for all of its WORDNET synsets. Second, we construct a set consisting of the full hyponymy for $noun_2$ for all of its WORDNET synsets. If an element from the set for $noun_2$ is a meronym, as defined by WORDNET, of an element in the set for $noun_1$, then we conclude that $noun_2$ is a meronym of $noun_1$. If the meronymy selectional construct justifies an edge between $noun_1$ and $noun_2$, the feature is assigned NOUN.

**Feature *Relational Noun***: Phrases such as $developed\ a\ relationship\ with\ people$ contain a relational noun $relationship$. Relational nouns suggest that there is a possessive relation between "individuals" participating in an utterance. To accommodate for relational nouns we propose the following generic selectional construct $(n, \text{POSSESSOR}, \emptyset, \{with\})$, where $n$ is a relational noun. Given tuple (7), the *Relational Noun* method identifies whether $noun_1$ is a relational noun by observing if one of its WORDNET senses has a path justified by the ISA links in WORDNET to a lexical concept *relation*. Currently, we assume that any noun matches the thematic role POSSESSOR. We assign the feature NOUN if we establish that $noun_1$ is relational.

**Feature *Idiom***: Some verb/noun combinations represent an idiomatic use, such as "make hay". The WORDNET lexicon contains entries representing idioms. We verify whether $verb$ and $noun_1$ of the given tuple (7) form an idiom by means of WORDNET. If this is the case, we assign VERB to the feature.

We analyzed performance of each described feature. Figure 6 presents the data in a similar fashion as Figure 4. On the left, we list higher precision features. We note the high recall of *Instrumentality* and rather reliable precision. This is a positive indication that we may address the limitations encountered for VERBNET and to generally improve classification. On the right, we list the lower precision features. The "right" results suggest that the ways to refine algorithms for implementing low-precision features should be sought out. Also, it is possible that the semantic information carried by the verb outweighs the information available from *Similarity* and *Meronymy*. In the future we plan to investigate these possibilities. The classification accuracy of the PPATTACH+ system is 71.2% for $with$. Due to the

| | | |
|---|---|---|
| *Instrumentality* | Recall | 48.0 |
| | Precision | 70.6 |
| *Relational Noun* | Recall | 5.8 |
| | Precision | 75.4 |
| *Adverbial Use* | Recall | 2.3 |
| | Precision | 75.0 |

| | | |
|---|---|---|
| *Similarity* | Recall | 15.7 |
| | Precision | 38.8 |
| *Meronymy* | Recall | 3.1 |
| | Precision | 48.5 |
| *Idiom* | Recall | 1.6 |
| | Precision | 35.3 |

Figure 6: Features Evaluation for PPATTACH+.

poor precision we witnessed for the "right" features, we retested the PPATTACH+ after removing these features. We subsequently achieve a classification accuracy of 72.0%, outperforming the PPATTACH accuracy of 70.7%. Overall, the results appear to be promising, suggesting that preposition-specific selectional constructs will lead to better classification as a whole.

## 7 Discussion and Future Work

In this work we proposed a principled method for incorporating wide-coverage lexical resources VERBNET and WORDNET into decision making for the task of resolving prepositional phrase attachment. Our preliminary system PPATTACH illustrates the feasibility and promise of the approach.

The proposed method relies on a number of features that are suggestive of why a particular attachment is reasonable. For instance, consider the feature *Instrumentality*. In cases when the value of this feature is VERB, we are urged to believe that the second noun of a given $\mathcal{P}$-phrase tuple can be labeled as an instrument of the action indicated by the verb of the tuple (following from the fact that the "instrument" selectional construct is applicable to this $\mathcal{P}$-phrase tuple). A long-term goal of this project is to incorporate elements of the proposed decision procedure into modern parsing technology and, in particular, into semantic role labeling methods. Work by Zhou et al. (2011), Srikumar and Roth (2013), Agirre et al. (2008, 2011), and Belinkov et al. (2014) encourages research in this direction.

As we continue development of this project, we hope to improve the presented method in several ways. We will use WORDNET to a greater extent to determine selectional restrictions on nouns. For example, the current method does not draw any distinction between AGENT and ASSET. We also intend to incorporate a semantic ontology called NOMLEX (Macleod et al., 1998) that incorporated noun-based selectional restrictions. Figure 1 illustrates that all but one preposition *of* prefer verb attachment. Most of the features we investigated also favor such attachment. Gaining evidence for the other decision will be helpful. We illustrated how we improve on preposition *with* by augmenting available lexico-semantic ontologies with knowledge specific to this preposition. We will pursue similar effort for other prepositions in the future. We also intend to go beyond the development and evaluation geared by the $\mathcal{R}$ dataset. Our discussion in the Evaluation section suggests such necessity.

## Acknowledgments

Thanks to A. Artsymenia, J. Beavers, C. Bonial, M. Chernyshevich, A. Harrison, V. Lifschitz, J. Michael, M. Palmer, D. Postanogov, P. Schüller, Q. Zhu, and the anonymous reviewers for useful discussions and comments. Daniel Bailey was supported by a FIRE-2013: Fund for Investing in the Research Enterprise Grant of the University of Nebraska.

## References

Agirre, E., T. Baldwin, and D. Martínez (2008). Improving parsing and pp attachment performance with sense information. In *46th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 317–325.

Agirre, E., K. Bengoetxea, K. Gojenola, and J. Nivre (2011). Improving dependency parsing with semantic classes. In *49th Annual Meeting of the Association for Computational Linguistics (ACL, Short Papers)*, pp. 699–703.

Allen, J. (1995). *Natural Language Understanding (2Nd Ed.)*. Redwood City, CA, USA: Benjamin-Cummings Publishing Co., Inc.

Belinkov, Y., T. Lei, R. Barzilay, and A. Globerson (2014). Exploring Compositional Architectures and Word Vector Representations for Prepositional Phrase Attachment. *Transactions of the Association for Computational Linguistics 2*, 561–572.

Brill, E. and P. Resnik (1994). A rule-based approach to prepositional phrase attachment disambiguation. In *15th conference on Computational linguistics-Volume 2*, pp. 1198–1204.

Budanitsky, A. and G. Hirst (2006, March). Evaluating wordnet-based measures of lexical semantic relatedness. *Computional Linguistics 32*(1), 13–47.

Collins, M. and J. Brooks (1995). Prepositional phrase attachment through a backed-off model. In *Proceedings of the Third Workshop on Very Large Corpora*, pp. 27–38.

Dahlgren, K. (1988). *Naive Semantics for Natural Language Understanding*. Norwell, MA, USA: Kluwer Academic Publishers.

Ford, Bresnan, K. (1982). A competence-based theory of syntatic closure. In Bresnan (Ed.), *The Mental Representation of Grammatical Relations*, pp. 727–796. The MIT Press.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten (2009, November). The weka data mining software: An update. *SIGKDD Explor. Newsl. 11*(1), 10–18.

Hirst, G. (1988, March). Semantic interpretation and ambiguity. *Artificial intelligence 34*(2), 131–177.

Jensen, K. and J.-L. Binot (1987). Disambiguating prepositional phrase attachments by using on-line dictionary definitions. *Computational Linguistics 13*(3-4), 251–260.

Katz, J. J. and J. A. Fodor (1963). The structure of a semantic theory. *Language 39*(2), pp. 170–210.

Kipper, K., H. T. Dang, and M. Palmer (2000). Class-based construction of a verb lexicon. In *7th National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, pp. 691–696. AAAI Press / The MIT Press.

Kipper-Schuler, K. (2005). *VerbNet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph. D. thesis, University of Pennsylvania.

Lapata, M. and F. Keller (2005). Web-based models for natural language processing. *ACM Transactions on Speech and Language Processing (TSLP) 2*(1), 3.

Leacock, C. and M. Chodorow (1998). Combining local context and wordnet similarity for word sense identification. In C. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*, Cambridge, MA, USA, pp. 265–283. The MIT Press.

Levin, B. (1993). *English verb classes and alternations : a preliminary investigation*. University Of Chicago Press.

Lierler, Y. and P. Schüller (2013). Towards a tight integration of syntactic parsing with semantic disambiguation by means of declarative programming. In *10th International Conference on Computational Semantics (IWCS)*.

Macleod, C., R. Grishman, A. Meyers, L. Barrett, and R. Reeves (1998). Nomlex: A lexicon of nominalizations. *Proceedings og EURALEX 98*, 187–193.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, and K. Miller (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography 3*, 235–244.

Pantel, P. and D. Lin (2000). An unsupervised approach to prepositional phrase attachment using contextually similar words. In *38th Annual Meeting on Association for Computational Linguistics (ACL)*, Stroudsburg, PA, USA, pp. 101–108.

Ratnaparkhi, A. (1998). Statistical models for unsupervised prepositional phrase attachment. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2 (ACL)*, Stroudsburg, PA, USA, pp. 1079–1085.

Ratnaparkhi, A., J. Reynar, and S. Roukos (1994). A maximum entropy model for prepositional phrase attachment. In *Workshop on Human Language Technology*, pp. 250–255.

Srikumar, V. and D. Roth (2013). Modeling semantic relations expressed by prepositions. *TACL 1*, 231–242.

Stetina, J. and M. Nagao (1997). Corpus based pp attachment ambiguity resolution with a semantic dictionary. In *5th Workshop pn Very Large Corpora*, pp. 66–80.

Zapirain, B., E. Agirre, L. Màrquez, and M. Surdeanu (2013). Selectional preferences for semantic role classification. *Computational Linguistics 39*(3), 631–663.

Zavrel, J., W. Daelemans, J. Veenstra, et al. (1997). Resolving PP attachment ambiguities with memory-based learning. In *Workshop on Computational Language Learning (CoNLL'97), ACL, Madrid*.

Zhou, G., J. Zhao, K. Liu, and L. Cai (2011). Exploiting web-derived selectional preference to improve statistical dependency parsing. In *49th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1556–1565.