Corpus Pattern Analysis for Prepositions

Ken Litkowski
CL Research
9208 Gue Road
Damascus, MD 20872, USA
ken@clres.com

Abstract

We introduce a new corpus of 48,000 unlabeled preposition instances to serve as the basis for further analysis of preposition behavior. We describe techniques used for a corpus pattern analysis for verbs and draw upon these techniques in outlining steps that can be used for the preposition analysis. We show how data developed in the preposition project can be used with state-of-the-art techniques used in preposition disambiguation. Included in these techniques, we indicate where data from FrameNet and VerbNet might usefully be employed in characterizing preposition behavior. Finally, we present techniques for validating preposition characterizations developed with these techniques, with the objective of developing simplified representations for disambiguating prepositions.

1 Introduction

The analysis of preposition behavior requires an understanding of a considerable number of interrelated factors. While previous work on preposition disambiguation has uncovered many of these factors, many additional needs have become more obvious.

In section 2, we provide an overview of corpus pattern analysis, which will be used as a framework for analyzing preposition corpus instances. Section 3 describes a new corpus of 48,000 sentences for 272 prepositions to serve as the basis for further analysis of preposition behavior. In section 4, we describe data available in The Preposition Project and how it can guide the analysis of preposition concordances. Section 5 lays out the procedures for performing a corpus pattern analysis for prepositions, showing how we will use the pattern analysis for verbs as a model. In section 6, we present a framework for validating the patterns, particularly identifying necessary interannotator agreement studies and use of existing disambiguated corpora as a test suite. In the last section, we draw some conclusions about this effort.

2 Corpus Pattern Analysis

Hanks (2004a) describes methods for a Corpus Pattern Analysis (CPA) of verbs. It is based on the Theory of Norms and Exploitations (TNE, Hanks (2004b) and Hanks (forthcoming)). The focus of the analysis is identifying patterns (prototypes) with which verbs in actual use are associated. The patterns consist not only of the argument or valency structure, but also of semantic values for each of the arguments.

For example, for the verb *file*, one pattern is **[[Human = Plaintiff]] file [[Document = Lawsuit]]**. Associated with each pattern is an implicature, e.g., "[[Human = Plaintiff]] officially presents [[Document = Lawsuit]] to a court of law in order to start legal proceedings." In the pattern, **Human** and **Document** are semantic values, i.e., we expect the semantic type of the subject and object to fall into these ontological categories. At the same time, **Plaintiff** and **Lawsuit** constitute semantic roles for these semantic types (not to be confused with an argument's semantic role, such as **Agent** for the subject of the verb).

In CPA, no attempt is made to identify the meaning of a verb. Instead, meanings are associated with prototypical sentence contexts, i.e., concordance lines are tagged with pattern numbers. The meaning of a pattern is characterized in its implicature. As may be expected, the development of patterns for verbs is quite complex. CPA is viewed as a pilot project, where over 700 verbs were analyzed.[1] This project is being replaced by a

[1] This pilot project was performed at Masaryk University, Brno, CZ (http://nlp.fi.muni.cz/projects/cpa/).

new project, Disambiguation of Verbs by Collocation (DVC), where the objective is to identify Typical Usage Patterns (TUP).[2]

DVC will characterize verb uses following the principles of TNE, which "says, in essence, that a language consists of two interlinked systems of rules governing word use: a set of rules for the normal uses of words and a second-order set of rules governing the ways in which normal patterns are exploited." In addition to the patterns, DVC will also provide links to FrameNet and VerbNet, where available. Further, individual patterns will be linked to other patterns, showing relationships among the senses of a single verb and between the senses of other verbs.

3 A New Preposition Corpus

The Preposition Project (TPP) (Litkowski & Hargraves (2005), Litkowski & Hargraves (2006)) has provided two sense-tagged corpora suitable for preposition disambiguation.[3] The first corpus, used in SemEval 2007 (Litkowski & Hargraves (2007)), was drawn from FrameNet instances which explicitly referenced prepositions. The 27,000 instances for 57 prepositions were tagged by a professional lexicographer using the TPP sense inventory. The second corpus was built from the OUP sentence dictionary, and consists of 7,650 sentences covering 635 senses for 259 prepositions, with a maximum of 20 sentences for a sense. Neither of these corpora is representative of preposition usage. For the FrameNet-based corpus, many senses of some prepositions have no instances. For the OUP corpus, although all senses are covered, the relative frequency of the senses is not present.

A new untagged corpus of 48,000 sentences from the BNC has been constructed using the CPA mechanism for constructing samples. This mechanism uses the Word Sketch Engine for drawing the samples. The BNC corpus for CPA is called BNC50, since it includes only written utterances, excluding the spoken portion of BNC. We had to employ two methods for drawing the samples, one for single-word prepositions and one for phrasal prepositions. For single-word prepositions, we restricted the search to the use of the lemma as a preposition. For phrasal prepositions, we used a phrasal search query. In general, our target was to obtain 250 sentences for each preposition. If fewer than 250 instances were available, all sentences were retrieved. If more than 250 instances were available, we used the sketch engine random sampling mechanism to obtain 250 instances. For prepositions with a high number of senses (up to 20) and a large number of available instances, we drew samples of 500 or 750 instances. We were able to construct an instance set for 272 prepositions (out of 303 in TPP), with 250 or more instances for 140 prepositions. We believe this corpus to be balanced and representative.

Several interesting observations emerged during the construction of the corpus. Prepositions for which no instances were found are generally dialectic (*fornent*, *frae*) or tributary (*agin*), and hence unlikely to be encountered very often. Instances for single-word prepositions occasionally may be tagged incorrectly as prepositions, when they are adverbs or particles, e.g., *down* in *the sun has gone* ***down*** *day after day* and *across* in *he came* ***across*** *them*. Phrasal prepositions may have more diverse concordance sets. In some cases, the phrase is not a preposition instance, but rather consists of the individual words taken apart. In other cases, the phrase needs to be construed literally and not in the idiomatic phrasal preposition sense (e.g., *in the pay of*). The corpus may thus be able to serve as a source for examining infelicitous instances that are tagged as prepositions.

The general procedure followed in CPA is for the lexicographer to examine the concordance and to develop patterns. Each pattern is given a number and all relevant instances in the concordance that adhere to that pattern are tagged with this number. An essential part of CPA is that all concordance lines must be tagged. In other words, a pattern must be developed to cover each instance. Patterns may be renumbered (perhaps reflecting some organization of the patterns), with the renumbering then propagated through the concordance.

In the CPA for verbs, no relations are currently made from one verb to another, although the lexicographer may examine patterns under different verbs as a model for new patterns. In TPP, by contrast, several data elements have been used to comprise an integrated view of prepositions. These are described in the next section as a prelude to a

[2] DVC began on October 1, 2012 at the University of Wolverhampton (http://clg.wlv.ac.uk/projects/DVC/).

[3] Both corpora are available on request from CL Research, where an online version of TPP is also available.

description of how a CPA for prepositions might proceed.

4 The Preposition Project Data

The basic data in TPP consists of (1) a comprehensive sense inventory with definitions and examples, and for each sense, (2) identification of the preposition class, (3) semantic relation characterizations, (4) complement properties, (5) attachment properties, (6) permissible syntactic positions, (7) FrameNet frame and frame element characterizations, (8) synonymous prepositions, and (9) sense relations. For major prepositions, a lexicographic "treatment" is available to provide insights into the behavior of the preposition along with identification of idioms that use the preposition. These data, including a DIMAP dictionary and a MySQL database, can be downloaded from Online TPP.[4] As noted above, TPP data also includes two sense-tagged corpora, one of which was used for preposition disambiguation in SemEval. These data can be used to facilitate the corpus pattern analysis of the new corpus.

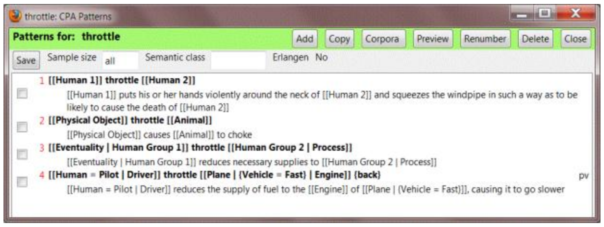
To begin, we can use the sense inventory as corresponding to the pattern set for a preposition. However, there are two difficulties. First, assuming a pre-existing sense inventory is contrary to the data-driven approach in CPA, which requires that the concordance be used to determine the set of patterns. Second, when the lexicographer assigned sense tags to the FrameNet-based corpus instances, he found it necessary to add new senses to the original OUP sense inventory, approximately a 10 percent increase, mostly subsenses. With a more representative sample, the new corpus may lead to further additions. Thus, we can use the existing senses as a candidate set, keeping these two factors in mind and using CPA's renumbering mechanism where necessary.

We can next exploit existing preposition disambiguation efforts to make an initial assignment of senses (see Ye & Baldwin (2007), Yuret (2007), Popescu et al. (2007), Tratz & Hovy (2009), Hovy et al. (2010), and Hovy et al. (2011)). Several of these efforts used maximum entropy modeling. Simple use of the models against the new corpus can provide an initial sense assignment. In-depth analysis of the results can perhaps identify preposition-specific features important for disambiguation. Many of the important features point to the preposition complement. In Hovy et al. (2011), using unsupervised techniques, the attachment point was found to be important. These studies also indicated the (small) contribution of syntactic positioning in disambiguation. These studies used the SemEval-2007 data, so their results are limited to 34 (major) prepositions. A key question is whether these results can be extended to other single-word and phrasal prepositions.

Yuret (2007) presented a somewhat different approach to disambiguation, attempting to determine the extent to which substitutable prepositions could be used (a technique found useful for disambiguation of content words). While achieving reasonable results, he noted that a drawback in preposition disambiguation was the small number of synonymous prepositions. The new corpus may provide a way in which substitutability may be exploited in reverse. Many of the prepositions in the new corpus, particularly phrasal prepositions, are monosemous. In TPP, the lexicographer frequently identified one or two substitutable prepositions for these entries, frequently a sense of a highly polysemous preposition. For example, the preposition *on the stroke of* (the **Temporal** class) has the substitutable preposition *at*, which has two temporal senses. By examining the properties in the concordance for *on the stroke of*, some characteristics of its behavior can perhaps be used to infer characteristics for the *at* senses. Similarly, the preposition *in the person of* (the **Agent** class) has substitutable prepositions *by* and *from*, both of which have agent senses.

As suggested, properties of the preposition complement and the point of attachment are important features of preposition disambiguation. We wish to characterize these features in the simplest way possible. Toward this end, we have begun simple part-of-speech tagging of the corpus instances and developing Perl patterns to recognize the preposition objects and the possible points of attachment. This approach not only enables some syntactic characterizations of these two key elements, but also the identification of the specific lemmas involved. A key next step is to attempt to characterize some semantic characteristics of the elements, using a shallow ontology similar to the one provided in CPA.

[4] Online TPP: http://www.clres.com/cgi-bin/onlineTPP/find_prep.cgi.



Patterns for: throttle

Sample size all Semantic class Erlangen No

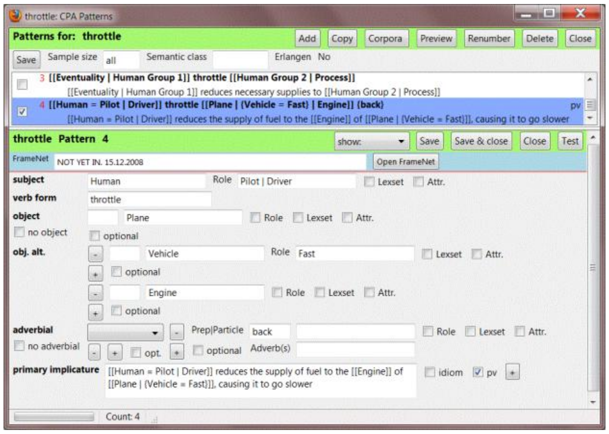
□ 1 **[[Human 1]] throttle [[Human 2]]**
[[Human 1]] puts his or her hands violently around the neck of [[Human 2]] and squeezes the windpipe in such a way as to be likely to cause the death of [[Human 2]]

□ 2 **[[Physical Object]] throttle [[Animal]]**
[[Physical Object]] causes [[Animal]] to choke

□ 3 **[[Eventuality | Human Group 1]] throttle [[Human Group 2 | Process]]**
[[Eventuality | Human Group 1]] reduces necessary supplies to [[Human Group 2 | Process]]

□ 4 **[[Human = Pilot | Driver]] throttle [[Plane | {Vehicle = Fast} | Engine]] {back}** pv
[[Human = Pilot | Driver]] reduces the supply of fuel to the [[Engine]] of [[Plane | {Vehicle = Fast}]], causing it to go slower

As Hanks pointed out (see above), it is necessary to examine both intrinsic and extrinsic (i.e., semantic role) characteristics of the lexical items. Also, since each preposition instance is a frame element (see Litkowski (2012)), it is worthwhile attempting to identify the frame in which the prepositional phrase is participating.



Patterns for: throttle

Sample size all Semantic class Erlangen No

□ 3 **[[Eventuality | Human Group 1]] throttle [[Human Group 2 | Process]]**
[[Eventuality | Human Group 1]] reduces necessary supplies to [[Human Group 2 | Process]]

☑ 4 **[[Human = Pilot | Driver]] throttle [[Plane | {Vehicle = Fast} | Engine]] {back}** pv
[[Human = Pilot | Driver]] reduces the supply of fuel to the [[Engine]] of [[Plane | {Vehicle = Fast}]], causing it to go slower

throttle Pattern 4 show:

FrameNet NOT YET IN. 15.12.2008

subject Human Role Pilot | Driver □ Lexset □ Attr.
verb form throttle
object Plane □ Role □ Lexset □ Attr.
□ no object □ optional
obj. alt. - Vehicle Role Fast □ Lexset □ Attr.
+ □ optional
- Engine □ Role □ Lexset □ Attr.
+ □ optional
adverbial - Prep|Particle back □ Role □ Lexset □ Attr.
□ no adverbial - + □ opt. + □ optional Adverb(s)
primary implicature [[Human = Pilot | Driver]] reduces the supply of fuel to the [[Engine]] of [[Plane | {Vehicle = Fast}]], causing it to go slower □ idiom ☑ pv +

Count: 4

5 Pattern Analysis for Prepositions

To develop patterns for prepositions, we use the CPA mechanisms as a model. In Figure 1, we show an example of a completed entry (for the verb *throttle*), with four patterns. Figure 1 presents a summary of the patterns that were developed, identifying the subjects and objects of the verb and providing the implicature. This is the result of the pattern development, where the **Add** button was used in the development of each pattern.

Figure 2 shows the details entered in specifying the pattern. Initially, this pattern detail was empty and various options were used to specify more and more detail. Thus, for example, initially there was only one slot for an object; the form allowed the addition of two additional alternatives. Note that the first object only has checkboxes for **Role** and **Lexset**. When these boxes are checked, these fields change to edit boxes that allow the entry of specific text. When a pattern is deemed complete, the **Save** button is used to add the pattern to the entry and summarize it in Figure 1.

Clearly, verb behavior is much more complicated than preposition behavior. Taking Figure 1 as an example, we may expect preposition behavior to be summarized by the general template **[[Attachment]] preposition [[Complement]]**, where characteristics of the attachment point and the complement would be detailed. The extent of the detail would depend on the specific preposition sense. It is worth noting that standard dictionaries usually provide little information about the attachment point, whereas we will attempt to give as much information as possible, reflecting the importance given to this item in the various disambiguation studies.

It is also important to note that the general template does not reflect the morphosyntactic behavior that may be realized in actual usage. Again, we may analogize to verb behavior, where the patterns in Figure 1 do not specifically mention the possibility that the verb may appear in the passive voice.

For prepositions, position may be more important, e.g., where Hovy et al. (2010) noted that fronting (i.e., the prepositional phrase at the beginning of a sentence) was significant in some disambiguation. In TPP, the item "permissible syntactic positions" refers to an identification by the lexicographer of where a prepositional phrase could occur. The analysis used for this item was based on the classification used by Quirk et al. (1986). These are (1) noun postmodifier, (2) adverbial, (a) adjunct, (b) subjunct, (c) disjunct, (d) conjunct, (3) complementation, (a) of a verb, and (b) of an adjective. This suggests that the detail form for prepositions should include an item that specifies such positions.

Specification of the preposition complement is somewhat easier than the attachment point. The CPA specification of a verb object can be used for preposition objects as well. The main field is the identification of an ontological category, where a default value could be "Anything", but where evidence might allow us to be more specific. An important issue here is the availability of a suitable ontology. In some cases, we may be able to identify specific lexical items that fill the slot, possibly with some attributes (such as "no determiner" or "possessive determiner"). We also need to allow for the possibility of a clausal object. In CPA, several types of clauses may occur; for TPP, only two seem likely (a gerundial or a WH-clause). CPA also allows for some semantic characterization of clauses (usually defaulting to an **Event**. The value for a **Role** field can be filled with a description that characterizes an object. For example, in *this animal always hunts **by night***, *night* is a **PeriodOfActivity**. Below, we will consider whether such a description comports with a FrameNet frame element specification.

Characterization of the point of attachment is somewhat more difficult. For this field, the permissible syntactic positions will constrain the types of information that may be required. For noun postmodifier phrases, the specification of the modified noun can follow the characterization used for the complement. Subjunct, disjunct, and conjunct adverbials may perhaps all be instances of fronting. Quirk et al. (1986) suggest that complement phrases are verb- and adjective-specific. Adverbial adjuncts may have similar behavior and may be licensed by FrameNet frames. These latter two cases may be problematic.

Quirk et al. identify a class of "prepositional verbs" in which the preposition is so closely tied to the verb as to lose its separate meaning as a preposition (*of* in *They accused him **of** favoritism*). Ordinary dictionaries will identify these nearly idiomatic forms in verb entries; it will be difficult to discern a distinct preposition sense for these cases.[5] The same is true of adjective complements (*for* in *he felt sorry **for** her*). CPA has a mechanism (see Figure 2) for marking a verb pattern as idiomatic or as a phrasal verb. This mechanism can be used here, as well as covering idiomatic prepositional phrases (e.g., *across the board*). One potential difficulty in attempting to use this mechanism is that there may be a considerable number of lexical items that have this near-idiomatic status.

Adverbial adjuncts are likely to be the most prevalent type of prepositional phrase. They are most likely to be licensed by a verb or a verbal noun, i.e., they fill a semantic role or frame element. In general, we may expect such adjuncts to be closely related to a FrameNet frame or a VerbNet class. As shown in Figure 2, CPA has a field for entering a FrameNet frame (which, in the case of *throttle*, has been identified as not yet being present) and some discussions have taken place about including a VerbNet verb class in the CPA pattern specification.

The CPA for prepositions should also include fields for both FrameNet frames and VerbNet verb classes. In addition, the pattern should also include fields for specifying the FrameNet frame element and the VerbNet thematic roles (usually identified in the verb class frames with **PP.role**). A potential difficulty in specifying values for these fields is the large number that may be needed. In TPP, for sentences used in SemEval-2007 (i.e., the most common prepositions), (Frame, FrameElement) pairs were captured. Some senses had as many as 50 such pairs. Some pairs occurred many times. Also, frequently, the frame element name in the pairs for a sense occurred many times, but with different frame names, for example, (*Motion_directional, Source*) and (*Cause_motion, Source*) for one sense of *for*.

None of the disambiguation studies used the TPP FrameNet data directly. However, many attempted to identify and use semantic role labels as disambiguation features, without significant benefit. The large number of frame element names and the difficulty of the semantic role labeling task may explain the lack of significance. In studies that also attempted to use nearby verbs as disambiguation features, the difficulty of specifying verb classes (such as the tops of WordNet verb classes) may also be significant.

While the preceding discussion has focused on building preposition patterns using outside resources, these resources can be examined themselves for preposition specifications. CPA itself records adverbial patterns (see Figure 2), frequently identifying and characterizing significant prepositions associated with verb patterns. The adverbial specifications can include not only the prepositions and a characterization of their complements, but also can specify the major type of the prepositional phrase (*Direction, Location, Manner, Time, Completive,* and *Privative*). FrameNet and VerbNet specify individual prepositions along with a characterization of type (semantic role or frame element and thematic role, respectively). These databases can be inverted to identify all cases that specify an individual preposition.

Clearly, this aspect of specifying preposition patterns will be the most difficult. On the other hand, attempts to do so may provide a basis for a wider analysis. Adherence to the CPA model re-

[5] These cases may be the ones that give the most difficulty for non-native speakers precisely because the preposition has no separate meaning.

quires analysis of individual concordance instances. As the CPA for prepositions proceeds, perhaps initially focusing on monosemous prepositions, it becomes possible to examine the database for generalizations across many prepositions.

6 Verifying Preposition Patterns

The CPA database for verbs, the Pattern Dictionary of English Verbs (PDEV), is largely the result of efforts by one lexicographer. Since it currently covers only slightly over 10 percent of English verbs, several questions about PDEV have emerged. Cinková et al. (2012) asks whether pattern development can be learned and whether these patterns can be used in NLP tasks. They indicate that these questions are complex, deferring them, and then go on to investigate narrower questions as first steps, among which two are immediately relevant to a CPA of prepositions, namely, whether the pattern structure is well-designed and whether the patterns can be applied to corpus instances with reasonable interannotator agreement. Hovy et al. (2011), after describing their results with unsupervised training methods for coarse-grained disambiguation, also suggested annotating the data to produce a test set for use with Amazon's Mechanical Turk to measure label accuracy for preposition arguments.

Cinková et al. developed a pilot resource of 30 verbs from CPA. They developed annotation guidelines, cleaned up the CPA data to facilitate interannotator agreement, ran several rounds of annotation, and analyzed annotator behavior. They found it useful to include additional information in the pattern specification, particularly identifying syntactic dependencies and further syntactic features and other changes specific to verb patterns, which included ensuring the semantic distinctiveness of implicatures. They were able to show clear improvements in interannotator agreements during the several rounds, but also found it necessary to increase the number of patterns. However, they were unable to achieve the 90 percent agreement that was the goal of Ontonotes (Hovy et al., 2006).

These studies underscore the need for a validation phase of CPA-derived preposition patterns. They suggest an initial phase, perhaps using Turkers, where major difficulties in writing preposition patterns may be identified. Subsequent phases can look more closely at how well the characterizations perform in preposition disambiguation.

7 Conclusions

We have introduced a new unlabeled preposition corpus of over 48,000 sentences, randomly drawn from the British National Corpus to provide a representative set of instances. This process followed procedures used in creating samples of verb instances for their corpus pattern analysis, which provides a model for a comprehensive analysis of preposition usage. We have provided some initial characterizations of this corpus and shown how we can use data from the preposition project to assist in the analysis of the concordances. We have provided a framework for analyzing this data, with a particular objective of integrating links to FrameNet and VerbNet. Finally, we presented a set of steps for validating sense assignments, using interannotator agreement studies and use of existing corpora that have previously been disambiguated.

References

Silvie Cinková, Martin Holub, Adam Rambousek, and Lenka Smejkalová. 2012. A Database of Semantic Clusters of Verb Usages. In *Proceedings of Language Resources and Evaluation (LREC 2012),* Istanbul, Turkey, pp. 3176-3183.

Patrick Hanks. 2004a. Corpus Pattern Analysis. In *EURALEX Proceedings.* Vol. I, pp. 87-98. Lorient, France: Université de Bretagne-Sud.

Patrick Hanks. 2004b. The Syntagmatics of Metaphor and Idioms. *International Journal of Lexicography*, 17(3):245-74.

Patrick Hanks. Forthcoming. *Lexical Analysis: Norms and Exploitations*. MIT Press.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and and Ralph Weishedel. 2006. OntoNotes: The 90% Solution. In *Proceedings of Human Language Technologies: The 2011 Annual Conference of the North American Chapter of the ACL,* New York, June. Association for Computational Linguistics.

Dirk Hovy, Stephen Tratz, and Eduard Hovy. 2010. What's in a Preposition? Dimensions of Sense Disambiguation for an Interesting Word Class. In *Coling 2010: Posters*, pages 454–462, Beijing, China, August. Coling 2010 Organizing Committee.

Dirk Hovy, Ashish Vaswani, Stephen Tratz, David Chiang, and Eduard Hovy. 2011. Models and Training for Unsupervised Preposition Disambiguation. In

Proceedings of Human Language Technologies: The 2011 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Portland, Oregon, June. Association for Computational Linguistics.

Ken Litkowski and Orin Hargraves. 2005. The preposition project. *ACL-SIGSEM Workshop on "The Linguistic Dimensions of Prepositions and Their Use in Computational Linguistic Formalisms and Applications"*, pages 171–179.

Ken Litkowski and Orin Hargraves. 2006. Coverage and Inheritance in The Preposition Project. In: *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*. Trento, Italy.ACL. 89-94.

Ken Litkowski and Orin Hargraves. 2007. SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic.

Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. A Comprehensive Grammar of the English Language. Longman, New York.

Stephen Tratz and Dirk Hovy. 2009. Disambiguation of Preposition Sense Using Linguistically Motivated Features. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Companion Volume: Student Research Workshop and Doctoral Consortium*, pages 96–100, Boulder, Colorado, June. Association for Computational Linguistics.

Patrick Ye and Timothy Baldwin. 2007. MELB-YB: Preposition Sense Disambiguation Using Rich Semantic Features. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic.

Deniz Yuret. 2007. KU: Word Sense Disambiguation by Substitution. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic.