

Notes on Barbecued Opakapaka: Ontology in Preposition Patterns

Ken Litkowski

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

Abstract

The development of patterns in the Pattern Dictionary of English Prepositions (PDEP) is a fraught process, since it is not clear just what should be included. At a minimum, preposition behavior requires a specification of the complement (the preposition object) and the governor (the point of attachment). In general, it would seem that the complement can be characterized by an ontological category. This is less the case for the governor, since in addition to modifying a noun, a prepositional phrase can be attached to verb or adjectives. PDEP has assembled a corpus of 47,285 sentences for 304 single-word and phrasal prepositions and has tagged the instances with 1015 senses. These sentences have been parsed with a dependency parser and extracted about 3500 features for each sentence. SVM models have been developed for the polysemous prepositions and the features are being analyzed in depth. This paper describes many of the factors under analysis.

1 Introduction

In theory, ontologies can play an important role in natural language processing. In practice, various ontologies have performed less than adequately. While ontological theory has made great strides, filling in the nuts and bolts, that is, the concepts in the ontology, has been considerably more difficult. I believe this is largely due to attempting to construct ontologies from the top down, rather than from the bottom up. That is, I advocate a data-driven approach, one that builds on lexicographic principles.

El Maarouf et al. describes technique focused on ontology population. My efforts will not be directed toward that end, but rather will focus on the use of the ontology in characterizing preposition

behavior. However, the methodologies I attempt to develop may have some benefit for the ontology population task. Perhaps importantly, PDEP has the same basic methodology of a data-driven approach and focusing on the patterns of usage, rather than the analysis of word patterns in isolation. For prepositions, the task is more difficult since the appropriate patterns do not have a single structure.

TPP was designed to provide a well-defined framework for examining preposition behavior (Litkowski & Hargraves (2005); Litkowski & Hargraves (2006)). Recognizing that one of the most important lessons of word-sense disambiguation (WSD) studies is the need for well-defined sense inventory, we were able to obtain data from the Oxford Dictionary of English (ODE; Stevenson & Soanes (2003)) for use in TPP. In addition to an appropriate sense inventory, WSD also requires a set of instances tagged with senses from the sense inventory.

Like PDEV, PDEP has slots, one for the preposition complement and one for the governor of the prepositional phrase. In general, the complement will specify semantic preferences, for which the PDEV ontology should be applicable. In many cases, the governor may be a noun (where the prepositional phrase is a post-modifier), in which case the PDEV ontology may likewise apply. When the governor is a verb or an adjective, it may be more difficult to provide an ontological semantic type. A possible option for verbs is to identify a verb class (for which VerbNet may be explored). There may also be some analog that can be applied for adjectives. Similar problems with the PDEP corpus may apply here (250 instances, semantic alternations, and potential exploitations from the norms).

While the PDEP methodology is essentially similar to PDEV, there is one important difference. PDEP begins with a sense inventory (taken from ODE) and attempts to match the instances to this

inventory. However, the inventory is not viewed as completely accurate. Senses have been added (perhaps as many as 20 percent). In addition, in the original development and application of the sense inventory to a set of nearly 30,000 instances from FrameNet, the lexicographer frequently questioned its organization and completeness. A crucial goal of PDEP is to develop methodologies that can be used to study and hopefully improve the sense inventory. Several methodologies are under investigation.

In describing the CPA ontology, a discussion of the suitability of WordNet identifies several problems. These are applicable to PDEP as well. The analysis described herein begins with the existence of a well-designed machine learning model using SVM techniques. In parsing the PDEP sentences, WordNet is the basis for many features. As will be described, while WordNet has provided many insights, they are not conclusive. As a result, several additional lexical resources will be investigated for the development of other features. Included in these resources is the CPA ontology.

El Maarouf et al. describe unambiguity in PDEV (primarily in service of ontology population). While PDEP includes many polysemous prepositions, there are many that are monosemous and that may provide the ability to characterize individual senses of polysemous prepositions. This is similar in spirit to the efforts of Srikumar & Roth (2013) of cross-preposition analysis for the benefit of semantic role labeling.

The PDEP corpus used for developing preposition patterns mirrors that used for PDEV, using the same guidelines and the same word sketch interface. Specifically, 250 random instances were drawn from BNC50, when available. For prepositions with fewer than 250 instances, all instances were used. For 13 polysemous prepositions, either 500 or 750 instances were drawn. Since each sample identified the number of instances from which the samples were drawn, we are able to extrapolate to a normalized frequency of different preposition classes. (Further details are available in Litkowski (2013) and Litkowski (2014).)

In section 2, we describe the format used for all corpora. In section 3, we describe how the sentences were parsed. In section 4, we discuss the features used for examining preposition behavior. In Section 5, we describe, in broad outlines, how we are attempting to analyze the features, with some

discussion of our attempts to focus on semantic characteristics, including ontological investigations. In section 6, we describe investigations of the properties of preposition semantic classes, enabling us to examine features across prepositions. Section 7 describes the planned inclusion of several other lexical resources in feature generation, beyond the emphasis on WordNet in the present system, including PDEV, FrameNet, and VerbNet. Section 8 identifies several methodological insights from El Maarouf et al. that should be included in further analysis of PDEP data.

2 The SemEval Format

The format for the PDEP corpus follows the standard lexical sample format used in Senseval and SemEval, i.e., when the objective is to disambiguate individual words. A similar format is used for two other corpora, a FrameNet corpus used in SemEval 2007 and an OEC corpus developed by Oxford. Each preposition has its own XML file (e.g., underneath.xml). Each file contains a number of instances, as shown in Figure 1 (only one instance is shown in the example). The first line identifies the lexical item and its part of speech (always "prep" in these corpora). Each instance is given an identifying number and a document source. These are FN (FrameNet), OEC (Oxford English Corpus), and CPA (Corpus Pattern Analysis). The next line gives the answer for the instance, identifying the instance number and the TPP sense identifier. These answers are given for the SemEval and OEC corpora, but not the CPA corpus. For the CPA corpus, separate "key" files have been generated for each preposition. The next line gives the sentence (the context), with the target preposition surrounded by a "head" tag. Each sentence has been tokenized using TreeTagger,¹ that is, separated into space-separated strings, so that, for example, an apostrophe and the letter s forms a possessive token ('s) and the terminal period is separated from the preceding word.

3 Parsing the Corpora

The tokenized sentences of each corpus have been further processed with a lemmatizer, part-of-speech tagger, and dependency parser, using an updated version of the system described in Tratz &

¹<http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

```

<lexelt item="underneath" pos="prep">
  <instance id="underneath.p.fn.635810" docsrc="FN">
    <answer instance="underneath.p.fn.635810" senseid="1(1)"/>
    <context>
      He always used to tuck it <head>underneath</head> the water butt .
    </context>
  </instance>
</lexelt>

```

Figure 1. Example of a Senseval/SemEval lexical sample instance.

Hovy (2011).² These parses are provided in an expanded CoNLL-X format.³ The Tratz system includes a module specifically designed to process files that use the lexical sample format shown in Figure 1.

The Tratz system creates 14 tab-separated items, compared with 10 items in the original CoNLL-X format. However, in producing these files, only 6 items are included: (1) the token counter (item 1), (2) the word form (item 2), (3) the lemma (item 3), (4) the fine-grained part of speech tag (item 5), (5), the head of the current token, i.e., the token number of its head or 0 for the ROOT of the sentence (item 7), and (6) the dependency relation of each item to its head (item 8).

4 Features for Analyzing Preposition Behavior

The Tratz system has been used for developing SVM models for each polysemous preposition. These models examine the context of the preposition and develop a large number of features for as many as seven words in the context (e.g., the preposition complement and the governor of the preposition phrase). Many of the features make use of WordNet, identifying the WordNet lexical file name, synonyms, immediate hypernyms, and all hypernyms of a word. We are in the process of extending these features using other lexical resources. For example, when the governor is a word in FrameNet, we have added a feature identifying the FrameNet frame.⁴

Initial examination of the effect of adding the FrameNet frame suggests that a considerable amount of exploration of just what features might be valuable will be necessary. Moreover, in exam-

ining other lexical resources, the primary objective will involve assessing their value from an ontological perspective. In particular, this will require identification of the appropriate granularity for each resource.

As indicated above, one feature generated by the Tratz system is the WordNet lexical file name. In fact, a single lexical item may lead to a large number of such features, since all possible values are retained. For example, for one set of 20 instances assigned to one sense of *within*, 48 lexical file names are generated for the 20 identified complements and 69 for the 20 governors. This indicates that there is some ambiguity in assigning the correct label.

McCarthy et al. (2015) have extended the sketch engine to provide semantic word sketches using the 41 WordNet lexicographer classes, characterizing the use of these classes as supersense tagging. By fining the predominant classes that serve as arguments in syntactic relationships, they are able to group together dominant categories filling argument slots.

El Maarouf et al. develop many patterns for extracting potential ontological items.^{5,6} PDEP is not focusing on the same kind of surface pattern matching. Instead, the patterns are hypothetically implicit in the features that have been generated. It

⁵ It is perhaps worth noting that many of these patterns were originally developed as a result of the TREC question answering competitions over several years. It is possible that similar pattern matching has occurred in subsequent competitions involving knowledge base population.

⁶ It is also worth noting that the [Membership](#) class developed in PDEP may prove useful in developing further patterns for ontology population. This class identifies 37 preposition senses in PDEP that deal with the identification of groups and species. It is not known how many of these might be useful in ontology population, but it is estimated that this class of prepositions is the most frequent of all classes, accounting for almost 18 percent of all preposition uses.

² Available at <http://sourceforge.net/projects/miacp/>

³ Described in detail at <http://ilk.uvt.nl/conll/>,

⁴ <https://framenet.icsi.berkeley.edu/fndrupal/>

is hoped that the application of machine learning techniques to word-sense classification will uncover these patterns.

5 Analysis Approach

At this time, only a broad outline of an approach for analyzing the features has been developed. As described in Litkowski (2014), classification models have been developed for the 117 polysemous prepositions in PDEP using the CPA-derived instances. These models have been applied to the other TPP corpora. The results have shown that the PDEP representative corpus is not a panacea for preposition disambiguation. Generally, classification on the two test sets is about 50 percent. These results are an improvement on results described in Litkowski (2013), of about 40 percent accuracy.

The Tratz-Hovy system generates a considerable number of features for use in developing the SVM models. Most polysemous prepositions have 250 instances, leading to 75,000 features. Since there are no general principles for interpreting the coefficients of the SVM models, examining these features is quite challenging.

A first step in taming the feature set is the use of a process of recursive feature elimination, as described in Guyon et al. (2002). This process recursively eliminates features by powers of two, starting with the full set, eliminating down to the number of features with the highest power of two less than the full number, and then eliminating half in each further step down to the last standing feature. This process makes use of the coefficients of the SVM models, applying some criterion for the elimination. Three criteria are being investigated: (1) using the sum of the absolute values, (2) using the sum of the squares, (3) using the sum of the absolute values of the “impact” of each feature (as defined in Yano, et al. (2012)), and (4) using the chi-square of each feature (used in constructing the original SVM models). Preliminary results using these tests indicates that the models using the full feature set are not optimum and that as many as 90 percent of the features can be eliminated while increasing the accuracy on the test sets. At the optimums, with the second test as the best, the SVM models yield an accuracy 5 percent higher on both test sets. In running these tests, the features eliminated on each iteration were printed out along with their scores, permitting a more detailed analy-

sis of which features are deemed the most important.

The first two and fourth elimination tests do not allow a detailed examination of the feature importance by sense, whereas the third test does. The impact of a feature is computed as the weight times the relative frequency of the feature in the training set. That is, the impact measures how important the feature is in making a decision for the sense. This is a frequentist interpretation of the weights. This is perhaps the most lexicographically salient way of assessing the features.

Another way of assessing the features is simply to determine their relative frequency. There are several fields in characterizing preposition patterns in PDEP. The relative frequencies can assist in identifying the type of complement, where the prepositional phrase is attached, the semantic category of the complement, and dominant selectors of the complement and the governor.

With smaller feature sets, it is possible that other machine learning techniques may provide the ability for more intuitive assessments of features. A desirable objective, from a lexicographic perspective, would be the generation of decision lists that can be applied to disambiguating the prepositions.⁷

6 Class Analyses

The analysis described in the previous section focuses features associated with individual preposition senses. Srikumar & Roth (2013) investigated the possibility of modeling semantic relations expressed by prepositions. A starting point for their work was preposition classes identified in TPP. These classes are recorded as a data element in the pattern for each sense in PDEP. During the course of tagging the PDEP CPA corpus, each sense in PDEP was placed into a description of each class.⁸ As a result of these efforts, the number of classes was reduced to 10 major classes, from an original 21.

⁷ We expect to make use of the [WEKA](#) machine learning environment to obtain different perspectives on the data.

⁸ These classes are described at <http://www.cres.com/db/classes/ClassAnalysis.php> with links to each class and its subclasses, including a list of all senses in the class and an estimate of the frequency of each sense in the BNC.

The baseline models developed from the PDEP corpus are attempting to disambiguate prepositions at the fine grain. The classes constitute a coarser level of granularity. We have examined the accuracy of using these semantic classes, and found an improvement of 18 percentage points over the accuracy obtained at the fine-grained level. In addition, we have examined the accuracy of the models in predicting the classes of an independently developed sense inventory, known as preposition supersense tags (Schneider et al., 2015). On this set of corpus instances, known as the Reviews corpus, we obtained an accuracy of over 55%, compared to their baseline accuracy of 43%. We have added a supersense field in describing the preposition patterns and have maintained a synchronization with their project.

In addition to providing the basis for improved disambiguation into broad classes, the class analyses may provide additional insights into the behavior of senses across prepositions. Thus, for example, the Temporal class identifies 93 senses under 75 prepositions. We can examine the features associated with these senses in an attempt to provide a more precise characterization of the subclasses in this class (Simple Time Point, Time Preceding, Time Following, and Time Periods), following the methods described in Srikumar & Roth, particularly since those methods have been refined. We expect that these analyses might benefit characterization of ontological categories.

7 Additional Lexical Resources

As mentioned earlier, a FrameNet feature was added to the feature set in a preliminary investigation. This feature is added only when the governor of the prepositional phrase is in the FrameNet dictionary and has a frame element realized with the preposition under analysis. The added feature is only the frame name. Although the analysis of this feature has only been preliminary, it appears as if it eliminated fairly early in the recursive feature elimination. That is, it does not appear to be significant. Several factors may be at work here. First, FrameNet coverage of lexical items is known to be somewhat limited. Second, it is possible that use of the immediate frame may be too granular; it may be useful to consider the FrameNet frame hierarchy. Third, it is possible that frame features may be more significant for individual senses, rather than

for the preposition as a whole. Fourth, it may be useful to create a consolidated feature that merely identifies whether a FrameNet item is found (this is simply identifying whether the rule fired). Fifth, it may be useful to identify the frame element name as a feature; once again, since these names are frame-specific, they may be quite granular, and consideration of the frame element hierarchy may be warranted. Each of these points may be relevant to consideration of other lexical resources that may be added as features.

In addition to FrameNet, several other lexical resources may usefully be examined and are planned. The PDEV dictionary itself can be used, particularly when a verb is the governor and a PDEV pattern identifies a specific preposition. Importantly in this case, when a specific preposition is identified, the ontological category included in the pattern may be useful. This would be particularly useful in characterizing the category for the preposition complement in PDEP patterns. Use of PDEV for PDEP feature analysis would also have to include consideration of coverage issues.⁹

At least three other lexical resources may be examined for feature analysis. These are: (1) VerbNet, (2) the noun hierarchy embodied in the superconcept field of the Oxford Dictionary of English, and (3) the category system included in the USAS tagger.

8 Additional Lexical Resources

Sections 6 and 7 of El Maarouf et al. provide a number of important methodological considerations that can guide the further analysis of PDEP data. The bootstrapping algorithm described in section 6.1 may be useful in examining disambiguation predictions from the Tratz system, where each prediction is associated with a score; the strength of the score may perhaps be usefully examined. This may be done with some modification of equations (1) and (2) in that section, particularly in ranking the importance of “selectors” used in specifying the properties of the complement and governor in PDEP patterns.

Section 7 describes various manual evaluations applied to the ontology population. A similar pro-

⁹ It is possible that the Tratz parser could be used on samples drawn for PDEV analysis for verbs not yet started, to provide a potential identification of prepositional phrases that might be associated with these verbs.

cess seems desirable for PDEP data, but the questions to be addressed are identifying just what should be examined. One possibility is using the confusion matrix as the starting point. Another is systematic exclusion of particular feature groups; an initial analysis of excluding such groups has shown little effect on overall performance, perhaps because there is considerable redundancy in the features. The analysis in section 7 questions the coverage of WordNet hyponyms. Since the Tratz system includes many WordNet features, this is something that a detailed analysis of the features should include. The PDEP data strongly suggest that there is a considerable amount of “exploitation” going on, e.g., characterizing something as a Location that would not be identified as such in any lexical resources.

References

- Ismail El Maarouf, Georgiana Marsci, and Constantin Orasan. 2015. Barbecued Opakapaka: Using Semantic Preferences for Ontology Population. In *RANLP Proceedings*. Hissar, Bulgaria.
- 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. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.
- Ken Litkowski. 2013. *Corpus Pattern Analysis of Prepositions*. Technical Report 12-02. Damascus, MD: CL Research.
- 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.
- Diana McCarthy, Adam Kilgarriff, Milos Jakubicek, and Siva Reddy. 2015. *Semantic Word Sketches. Corpus Linguistics 2015*. Lancaster University.
- Nathan Schneider, Vivek Srikumar, Jena D. Hwang, and Martha Palmer. 2015. A Hierarchy with, of, and for Prepositions. In *Proceedings of the 9th Linguistic Annotation Workshop*, pages 112-123 Denver, Colorado.
- Vivek Srikumar and Dan Roth. 2013. Modeling Semantic Relations Expressed by Prepositions. *Transactions of the Association for Computational Linguistics*, 1”231-42.
- Angus Stevenson and Catherine Soanes (Eds.). 2003. *The Oxford Dictionary of English*. Oxford: Clarendon Press.
- Stephen Tratz and Eduard Hovy. 2011. A Fast, Accurate, Non-Projective, Semantically-Enriched Parser. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. Edinburgh, Scotland, UK.
- Tae Yano, Noah A. Smith, and John D. Wilkerson. 2012. Textual Predictors of Bill Survival in Congressional Committees. In: *2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Montreal, Canada. 793-802.