

Pattern Dictionary of English Prepositions

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

Abstract

We present a new lexical resource for the study of preposition behavior, the Pattern Dictionary of English Prepositions (PDEP). This dictionary, which follows principles laid out in Hanks' theory of norms and exploitations, is linked to 81,509 sentences for 304 prepositions, which have been made available under The Preposition Project (TPP). Notably, 47,285 sentences, initially untagged, provide a representative sample of preposition use, unlike the tagged sentences used in previous studies. Each sentence has been parsed with a dependency parser and our system has near-instantaneous access to features developed with this parser to explore and annotate properties of individual senses. The features make extensive use of WordNet. We have extended feature exploration to include lookup of FrameNet lexical units and VerbNet classes for use in characterizing preposition behavior. We have designed our system to allow public access to any of the data available in the system.

1 Introduction

Recent studies (Zapirain et al. (2013); Srikumar and Roth (2011)) have shown the value of prepositional phrases in joint modeling with verbs for semantic role labeling. Although recent studies have shown improved preposition disambiguation, they have received little systematic treatment from a lexicographic perspective. Recently, a new corpus has been made available that promises to be much more representative of preposition behavior. Our initial examination of this corpus has suggested clear indications of senses previously overlooked and reduced prominence

for senses thought to constitute a large role in preposition use.

In section 2, we describe the interface to the Pattern Dictionary of English Prepositions (PDEP), identifying how we are building upon data developed in The Preposition Project (TPP) and investigating its sense inventory with corpora also made available under TPP. Section 3 describes the procedures for tagging a representative corpus drawn from the British National Corpus, including some findings that have emerged in assessing previous studies of preposition disambiguation. Section 4 describes how we are able to investigate the relationship of WordNet, FrameNet, and VerbNet to this effort and how this examination of preposition behavior can be used in working with these resources. Section 5 describes how we can use PDEP for the analysis of semantic role and semantic relation inventories. Section 6 describes how we envision further developments of PDEP and how the data are available for further analysis. In section 7, we present our conclusions for PDEP.

2 The Pattern Dictionary of English Prepositions

Litkowski and Hargraves (2005) and Litkowski and Hargraves (2006) describe The Preposition Project (TPP) as an attempt to describe preposition behavior using a sense inventory made available for public use from the *Oxford Dictionary of English* (Stevenson and Soanes, 2003) by tagging sentences drawn from FrameNet. In TPP, each sense was characterized with its complement and attachment (or governor) properties, its class and semantic relation, substitutable prepositions, its syntactic positions, and any FrameNet frame and frame element usages (where available). The FrameNet sentences were sense-tagged using the sense inventory and were

later used as the basis for a preposition disambiguation task in SemEval 2007 (Litkowski and Hargraves, 2007).

Initial results in SemEval achieved a best accuracy of 69.3 percent (Ye and Baldwin, 2007). The data from SemEval has subsequently been used in several further investigations of preposition disambiguation. Most notably, Tratz (2011) achieved a result of 88.4 percent accuracy and Srikumar and Roth (2013) achieved a similar result. However, Litkowski (2013b) showed that these results did not extend to other corpora, concluding that the FrameNet-based corpus may not have been representative, with a reduction of accuracy to 39.4 percent using a corpus developed by Oxford.

Litkowski (2013a) announced the creation of the TPP corpora in order to develop a more representative account of preposition behavior. The TPP corpora includes three subcorpora: (1) the full SemEval 2007 corpus (drawn from FrameNet data, henceforth FN), (2) sentences taken from the Oxford English Corpus to exemplify preposition senses in the *Oxford Dictionary of English* (henceforth, OEC), and (3) a sample of sentences drawn from the written portion of the British National Corpus (BNC), using the Word Sketch Engine as implemented in the system for the Corpus Pattern Analysis of verbs (henceforth, CPA or TPP).

We have used the TPP data and the TPP corpora to implement an editorial interface, the Pattern Dictionary of English Prepositions (PDEP).¹ This dictionary is intended to identify the prototypical syntagmatic patterns with which prepositions in use are associated, identifying linguistic units used sequentially to make well-formed structures and characterizing the relationship between these units. In the case of prepositions, the units are the complement (object) of the preposition and the governor (point of attachment) of the prepositional phrase. The editorial interface is used to make changes in the underlying databases, as described in the following subsections. Editorial access to make changes is limited, but the system can be explored publicly and the underlying data can be accessed publicly, either in its entirety or through publicly available scripts used in accessing the data during editorial operations.

Standard dictionaries include definitions of prepositions, but only loosely characterize the syntagmatic patterns associated with each sense.

¹ <http://www.clres.com/db/TPPEditor.html>

PDEP takes this a step further, looking for prototypical sentence contexts to characterize the patterns. PDEP is modeled on the principles of Corpus Pattern Analysis (CPA), developed to characterize syntagmatic patterns for verbs.² These principles are described more fully in Hanks (2013). Currently, CPA is being used in the project Disambiguation of Verbs by Collocation to develop a Pattern Dictionary of English Verbs (PDEV).³ PDEP is closely related to PDEV, since most syntagmatic patterns for prepositions are related to the main verb in a clause. PDEP is viewed as subordinate to PDEV, sufficiently so that PDEP employs significant portions of code being used in PDEV, with appropriate modifications as necessary to capture the syntagmatic patterns for prepositions.⁴

2.1 The Preposition Inventory

After a start page for entry into PDEP, a table of all prepositions in the sense inventory is displayed. Figure 1 contains a truncated snapshot of this table. The table has a row for each of 304 prepositions as identified in TPP. The second column indicates the number of patterns (senses) for each preposition. The next two columns show the number of TPP (CPA) instances that have been tagged and the total number of TPP instances that have been obtained as the sample from the total number of instances in the BNC.

Preposition ↑	Patterns	TPP Tagged	TPP Insts	BNC Freq
because of	1	250	250	9346
before	4	125	250	20821
behind	9	0	250	6207
below	4	16	250	2646
beneath	6	0	250	1277
beside	3	0	250	932
besides	1	0	250	468
between	9	8	250	58865

Figure 1. Preposition Inventory

Additional columns not shown in Figure 1 show (1) the status of the analysis for the preposition, (2) the number of instances from FrameNet (i.e., FN Insts, as developed for SemEval 2007), and (3) the number of instances from the Oxford English Corpus (i.e., OEC Insts). The number of prepositions with

² See <http://nlp.fi.muni.cz/projects/cpa/>.

³ See <http://clg.wlv.ac.uk/projects/DVC>

⁴ PDEP is implemented as a combination of HTML and Javascript. Within the Javascript code, calls are made to PHP scripts to retrieve data from MySQL database tables and from additional files (described below).

FrameNet instances is 57 (larger than the 34 prepositions used in SemEval). There are no OEC instances for 57 prepositions. There are no TPP instances for 41 prepositions. Notwithstanding the lack of instances, there are TPP characterizations for all 304 prepositions.

The BNC frequency shown in Figure 1 provides a basis for extrapolating results from PDEP to the totality of prepositions. In total, the number of instances in the BNC is 5,391,042, which can be used as the denominator when examining the relative frequency of any preposition (e.g., *between* has a frequency of 0.0109, 58,865/5,391,042).⁵

In general, the target sample size was 250 CPA instances. If the number available was less than 250, all instances were used. The TPP CPA corpus contains 250 instances for 170 prepositions. Where the number of senses for a preposition was large (about 15 or more), larger samples of 750 (*of, to, on, and with*) or 500 (*in, for, by, from, at, into, over, like, and through*) were drawn.

2.2 Preposition Patterns

When a row in Figure 1 is clicked, the preposition is selected and a new page is opened to show the patterns for that preposition. Figure 2 shows the four patterns for *below*. Each pattern is presented as an instance of the template **[[Governor]] prep [[Complement]]**, followed by its primary implicature, where the current definition is substituted for the preposition.

Patterns for below				
Add pattern Annotate All Corpus				
Number of Tagged Instances: FN 34 OEC 72 TPP 16 Untagged: TPP 234 Status: WIP				
#	FN	OEC	TPP	Pattern & primary implicature
1(1)	28	13	null	[[Governor]] below [[Complement]] [[Governor]] at a lower level or layer than [[Complement]]
2(1a)	null	20	1	[[Governor]] below [[Complement]] [[Governor]] lower in grade or rank than [[Complement]]
3(1b)	6	20	15	[[Governor]] below [[Complement]] [[Governor]] lower than (a specified amount or standard) [[Complement]]
4(2)	null	19	null	[[Governor]] below [[Complement]] [[Governor]] extending underneath [[Complement]]

Figure 2. Preposition Pattern List

The display in Figure 2 provides an overview for each preposition, with the top line showing the number of tagged instances available from

⁵ The total number of instances for *of* and *in* in this estimate is 1,000,000. As a result, the relative frequency calculation should not be construed as completely accurate.

each corpus. For the TPP instances, this identifies the number of instances that have been tagged and the number that remain to be tagged. In the body of the table, the first column shows the TPP sense number. The next three columns show the number of instances that have been tagged with this sense. Note that the top line of the pattern list includes a menu option for adding a pattern, for the case when we find that a new sense is required by the corpus evidence.

Clicking on any row in the pattern list opens the details for that pattern, with a pattern box entitled with the preposition and the pattern number, as shown in Figure 3. The pattern box contains data developed in TPP and several new fields intended to capture our enhancements.

TPP data include the fields for the **Complement**, the **Governor**, the **TPP Class**, the **TPP Relation**, the **Substitutable Prepositions**, the **Syntactic Position**, the **Quirk Reference**, the **Sense Relation**, and the **Comment**. We have added the checkboxes for complement type (common nouns, proper nouns, WH-phrases, and -ing phrases), as well as a field to identify a particular lexical item (lexset) if the sense is an idiomatic usage. We have added the **Selector** fields for the complement and the governor. For the complement, we have a field **Category** to hold its ontological category (using the shallow ontology being developed for verbs in the DVC project mentioned above).⁶ We also provided a field for the **Semantic Class** of the governor; this field has not yet been implemented.

We have added two **Cluster/Relation** fields. The **Cluster** field is based on data available from Tratz (2011), where senses in the SemEval 2007 data have been put into 34 clusters. The **Relation** field is based on data available from Srikumar and Roth (2013), where senses in the SemEval 2007 data have been put into 32 classes. A key element of Srikumar and Roth was the use of these classes to model semantic relations across prepositions (e.g., grouping all the Temporal senses of the SemEval prepositions). In the pattern box, each of these two fields has a drop-down list of the clusters and relations, enabling us to categorize the senses of other prepositions with these classes. Below, we describe how we are able to use the TPP classes and relations along with the Tratz clusters and Srikumar relations in an analysis of these classes across the

⁶ This ontology is an evolution of the Brandeis Semantic Ontology (Pustejovsky et al., 2006).

Pattern below 3(1b) Corpus Instances Save Save&Close Close

Complement: NPs denoting quantity or a point on a scale NN NNP WH -ING Lexset

Selectors: hr.pos.od; hr.l.standard,average Category

Governor: no pattern. Noun Verb Adj

Selectors: Sem Class

Semantics: TPP Class: Scalar TPP Relation: QuantityAbove

Cluster/Relation: Cluster: Scalar Relation:

Substitutable Prepositions: under, underneath

Syntactic Position: 1 2a 2b 2c 2d 3a 3b Quirk Reference: 9.32

Sense Relation: 1: extension of space to r

Primary implicature: idiom

[[Governor]] lower than (a specified amount or standard) [[Complement]]

Figure 3. Preposition Pattern Details

full set of prepositions, instead of just those used in SemEval.

Any number of pattern boxes may be opened at one time. The data in any of the fields may be altered (with the menu bar changing color to red) and then saved to the underlying databases. An individual pattern box may then be closed.

The drop-down box labeled **Corpus Instances** in the menu bar is used to open the set of corpus instances for the given sense. As shown in Figure 2, this sense has 6 FN instances, 20 OEC instances, and 15 TPP instances. The drop-down box has an option for each of these sets, along with an option for all TPP instances that have not yet been tagged. When one of these options is selected, the corresponding set of instances is opened in a new tab, discussed in the next section.

2.3 Preposition Corpus Instances

As indicated, selecting an instance set from the pattern box opens this set in a separate tab, as shown in Figure 4. This tab, labeled **Annotation: below (3(1b))**, identifies the preposition and the sense, if any, associated with the instance set (the sense will be identified as **unk** if the set has not yet been tagged. The instance set is displayed, identifying the corpus, the instance identifier, the

TPP sense (if identified, or “unk” if not), the location in the sentence of the target preposition, and the sentence, with the preposition in bold.

This tab is where the annotation takes place. Any set of sentences may be selected; each selected sentence is highlighted in yellow (as shown in Figure 6). The sense value may be changed using the drop-down box labeled **Tag Instances** in the menu bar. This drop-down box contains all the current senses for the preposition, along with possible tags **x** (to indicate that the instance is invalid for the preposition) and **unk** (to indicate that a tagging decision has not yet been made). The sense tags in Figure 4 were originally untagged in the CPA (TPP) corpus and were tagged in this manner.

In general, sense-tagging follows standard lexicographic principles, where an attempt is made to group instances that appear to represent distinct senses. PDEP provides an enhanced environment for this process. Firstly, we can make use of the current TPP sense inventory to tag sentences. Since the pattern sets (definitions) are based on the *Oxford Dictionary of English*, the likelihood that the coverage and accuracy of the sense distinctions is quite high. However, since prepositions have not generally received the close attention of words in other parts of speech,

Figure 4. Preposition Corpus Instance Annotation

Sentences for below Sense 3(1b) Corpus TPP Examine WFRs FERs FN Select Tag Instances Select All Unselect Save

Corpus Instance	Sense	Location	Sentence
CPA 1	3(1b)	60	Fifty per cent of all people in Britain are educationally below average . '
CPA 5	3(1b)	44	Selling pressure also sent sterling briefly below DM2.95 for the first time since August 1987.
CPA 27	3(1b)	196	How come we can send a spacecraft to Venus or write the Bible on a microchip yet remain incapable of operating a sophisticated transportation network when the temperature drops just a few degrees below zero ?
CPA 36	3(1b)	220	It was they who caused a reluctant minister to accept , in 1963 , the introduction of an examination at 15 + other than O level (the CSE examination) specifically designed for the next 40 per cent of the ability range below the 20 per cent for whom O level was thought to be appropriate .
CPA 37	3(1b)	36	Grade G is aligned to CSE grade 5 (below average) and there is an unclassified grade below G.
CPA 62	3(1b)	42	Having said that , my playing never sinks below a certain standard , although sometimes it 's hard to tell if you 're actually improving , because you 're so close to it .
CPA 93	3(1b)	182	What the movies taught Welford Beaton was ` we are not interested in average things , whether animate or inanimate -- we are interested in anything in the degree that it is above or below the average ' .

PDEP is intended to ensure the coverage and accuracy. During the tagging of the SemEval instances, the lexicographer found it necessary to increase the number of senses by about 10 percent. Since the lack of coverage of FrameNet is well-recognized, the representative sample developed for the TPP corpus should provide the basis for ensuring the coverage and accuracy.

In addition to adhering to standard lexicographic principles, the availability of the tagged FN and OEC instances can be used as the basis for tagging decisions. Where available, these tagged instances can be opened in separate tabs and used as examples for tagging the unknown TPP instances.

3 Tagging the TPP Corpus

3.1 Examining Corpus Instances

The main contribution of the present work is the ability to interactively examine characteristics of the context surrounding the target preposition in the corpus instances. In the menu bar shown in Figure 4, there is an **Examine** item. Next to it are two drop-down boxes, one labeled **WFRs** (word-finding rules) and one labeled **FERs** (feature extraction rules). These rules are taken from the system described in Tratz and Hovy (2011) and Tratz (2011).⁷ The TPP corpora described in Litkowski (2013a) includes full dependency parses and feature files for all sentences. Each sentence may have as many as 1500 features describing the context of the target preposition. We have made the feature files for these sentences (1309 MB) available for exploration in PDEP.

In our system, we make available seven word-finding rules and nine feature extraction rules. The word-finding rules fall into two groups: words pertaining to the governor and words pertaining to the complement. The five governor word-finding rules are (1) verb or head to the left (**l**), (2) head to the left (**hl**), (3) verb to the left (**vl**), (4) word to the left (**wl**), and (5) governor (**h**). The two complement word-finding rules are (1) syntactic preposition complement (**c**) and (2) heuristic preposition complement (**hr**). The feature extraction rules are (1) word class (**wc**), (2) part of speech (**pos**), (3) lemma (**l**), (4) word (**w**), (5) WordNet lexical name (**ln**), (6) WordNet synonyms (**s**), (7) WordNet hypernyms (**h**), (8) whether the word is capitalized (**c**), and (9) affixes (**af**). Thus, we are able to examine any of 63

WFR FER combinations for whatever corpus set happens to be open.

In addition to these features, we are able to determine the extent to which prepositions associated with FrameNet lexical units and VerbNet classes occur in a given corpus set. In Figure 4, there is a checkbox labeled FN next to the FERs drop-down list to examine FrameNet lexical units. There is a similar checkbox labeled VN to examine members of VerbNet classes. These boxes appear only when either of these resources has identified the given preposition as part of its frame (75 for FrameNet and 31 for VerbNet).

When a particular WFR-FER combination is selected and the **Examine** menu item is clicked, a new tab is opened showing the values for those features for the given corpus set, as shown in Figure 5. The tab shows the WFR and FER that were used, the number of features for which the value was found in the feature data, the values, and the count for each feature. The description column is used when displaying results for the part of speech, the affix type, FrameNet frame elements, and VerbNet classes, since the value column for these hits are not self-explanatory. The example in Figure 5 is showing the lemma, which requires no further explanation.

Feature Examination Results for below		
Word-Finding Rule: <i>Preposition Complement (Heuristic)</i>		
Feature Extraction Rule: <i>Lemma</i>		
Number of hits: 13 out of 15 instances		
Value	Count	Description
5	1	
20	1	
average	9	
standard	1	
zero	1	

Figure 5. Feature Examination Results

For most features (e.g., lemma or part of speech), the number of possible values is relatively small, limited by the number of instances in the corpus set. For features such as the WordNet lexical name, synonyms and hypernyms, the number of values may be much larger. For FrameNet and VerbNet, the feature examination is limited to the combination of the WFR for the governor (**h**) and the FER lemma (**l**), both of which will generally identify verbs in the value column.

The general objective of examining features is to identify those that are diagnostic of specific senses. When applied to the full untagged TPP corpus set, this process is akin to developing

⁷ An updated version of this system is available at <http://sourceforge.net/projects/miacp/>.

Sentences for below Sense 3(1b) Corpus TPP			
Examine		hr	I
FN		<input checked="" type="checkbox"/>	Select
average		Tag Instances	
Select All		Unselect	
Save			
Corpus Instance	Sense	Location	Sentence
CPA 1	3(1b)	60	' Fifty per cent of all people in Britain are educationally below average . '
CPA 5	3(1b)	44	Selling pressure also sent sterling briefly below DM2.95 for the first time since August 1987.
CPA 27	3(1b)	196	How come we can send a spacecraft to Venus or write the Bible on a microchip yet remain incapable of operating a sophisticated transportation network when the temperature drops just a few degrees below zero ?
CPA 36	3(1b)	220	It was they who caused a reluctant minister to accept , in 1963 , the introduction of an examination at 15 + other than O level (the CSE examination) specifically designed for the next 40 per cent of the ability range below the 20 per cent for whom O level was thought to be appropriate .
CPA 37	3(1b)	36	Grade G is aligned to CSE grade 5 (below average) and there is an unclassified grade below G.
CPA 62	3(1b)	42	Having said that , my playing never sinks below a certain standard , although sometimes it 's hard to tell if you 're actually improving , because you 're so close to it .
CPA 93	3(1b)	182	What the movies taught Welford Beaton was ` we are not interested in average things , whether animate or inanimate -- we are interested in anything in the degree that it is above or below the average ' .

Figure 6. Selected Corpus Instances

word sketches for prepositions (Kilgarriff et al., 2004). However, since we have tagged corpus sets for most preposition senses, we can begin our efforts looking at these sets. The hypothesis is that the tagged corpora will show patterns which can then be used for tagging instances in the TPP corpus.⁸

The first step in examining features generally is to look at the word classes and parts of speech for the complement and the governor.⁹ These are useful for filling in their checkboxes in Figure 3. Another useful feature is word to the left (**wl**), which can be used to verify the syntactic position checkboxes, particularly the adverbial positions (adjunct, subjunct, disjunct, and conjunct). These first steps provide a general overview of a sense's behavior.

The next step of feature examination delves more into the semantic characteristics of the complement and the governor. Tratz (2011) reported that the use of heuristics provided a more accurate identification of the preposition complement; this is the WFR **hr** in our system. After getting some idea of the word class and the part of speech, we next examine the WordNet lexical name of the complement to determine its broad semantic grouping. As mentioned, this feature may return a number of values larger than the size of the corpus set, since WordNet senses for a given lexeme may be polysemous. Notwithstanding, this feature examination generally shows the dominant categories and can be used to charac-

terize and act as a selector for the complement in the pattern details. Similar procedures are used for characterizing the governor selection criteria.

In the example in Figure 3, for *below*, sense 3(1b), our preliminary analysis shows **hr:pos:cd** (i.e., a cardinal number) and **hr:l:average, standard** (i.e., the lemmas *average* and *standard*) are particularly useful for identifying this sense.

3.2 Selecting Corpus Instances

In addition to enabling feature examination, PDEP also facilitates selection of corpus instances. We can use the specifications for any WFR - FER combination, along with one of the values (as shown in Figure 5), to select the corpus instances having that feature. Figure 6 shows, in part, the result of the WFR **hr** and FER **I** with the value *average*, against the instances in the open corpus set.

As shown in the menu bar in Figure 6, we can select all instances and unselect all selections. Based on any selections, we can then tag such instances with one of the options that appear in the **Tag Instances** drop-down box. In the specific example, we could change all the selected instances to some other sense, if we have decided that the current assignment is not the best.

The selection mechanism is not used absolutely. For example, in examining the untagged instances for *over*, we used the specification **hr:ln:noun.time** (looking for instances with the heuristic complement having the WordNet lexical name **noun.time**). Out of 500 instances, we found 122 with this property. We then scrolled through the selected items, deselecting instances that did not provide a time period, and then tagged 99 instances with the sense 14(5), with the meaning *expressing duration*. Once we have made such a tagging, we can look at just those instances the next time we examine this sense. In this case, we might decide, pace the TPP lexicographer's comment, that the instances should be

⁸ Currently, 21.5 percent of the TPP instances (10347 of 47,285) have been tagged.

⁹ Accurate identification of the complement and governor is likely improved with the reliance on the Tratz dependency parser. Moreover, this is likely to improve the word sketches in PDEP. Ambati et al. (2012) report that dependency parses provide improved word sketches over purpose-built finite-state grammars. Their findings provide additional support for the methods presented here.

broken down into those which express a time period and those which describe “accompanying circumstances” (e.g., *over coffee*).

3.3 Accuracy of Features

PDEP uses the output from Tratz’ system (2011), which is of high quality, but which is not always correct. In addition, the TPP corpus also has some shortcomings, which are revealed in examining the instances. The TPP corpus has not been cleaned in the same manner as the FN and the OEC corpora. As a result, we see many cases which are more difficult to parse and hence, from which to generate feature sets. We believe this provides a truer real-world picture of the complexities of preposition behavior. As a result, in the **Tag Instances** drop-down box, we have included an option to tag a sentence as **x**, to indicate that it is not a valid instance.

A small percentage of the TPP instances are ill-formed, i.e., incomplete sentences; these are marked as **x**. For some prepositions, e.g., *down*, a substantial number of instances are not prepositions, but rather adverbs or particles. For some phrasal prepositions, such as *on the strength of*, the phrase is literal, rather than the preposition idiom; in this case, 20 of 124 instances were marked as **x**. The occurrence of these invalid instances provides an opportunity for improving taggers, parsers, and semantic role labelers.

4 Assessment of Lexical Resources

Since the PDEP system enables exploration of features from WordNet, FrameNet, and VerbNet, we are able to make some assessment of these resources.

WordNet played a statistically significant role in the systems developed by Tratz (2011) and Srikumar and Roth (2013). This includes the WordNet lexicographer’s file name (e.g., **noun.time**), synsets, and hypernyms. We make extensive use of the file name, but less so from the synsets and hypernyms. However, in general, we find that the file names are too coarse-grained and the synsets and hypernyms too fine-grained for generalizations on the selectors for the complements and the governors. The issue of granularity also affects the use of the DVC ontology. We discuss this issue further in section 6, on investigations of suitable categorization schemes for PDEP.

In using FrameNet, our results illustrate the unbalanced corpus used in SemEval 2007 (as suggested in Litkowski (2013b)). For the sense

of *of*, “used to indicate the contents of a container”, we first examined the FrameNet corpus set for that sense, which contains 278 instances (out of 4482, or 6.2 percent). Using PDEP, we found that FrameNet feature values for the governor accounted for 264 of these instances (95 percent), all of which were related to the frame elements *Contents* or *Stuff*. However, in the TPP corpus, only 3 out of 750 instances were identified for this sense (0.4 percent). Thus, while FrameNet culled a large number of instances which had these frame element realizations, these instances do not appear to be representative of their occurrence in a random sample of *of* uses. We have seen similar patterns for the other SemEval prepositions.

A similar situation exists for *Cause* senses of major prepositions: *for* (385 in FrameNet, 5/500 in TPP), *from* (71 in FrameNet, 16/500 in TPP), *of* (68 in FrameNet, 0/750 in TPP), and *with* (127 in FrameNet, 8/750 in TPP). Each of these cases further emphasizes how the SemEval 2007 instances are not representative and thus degrade the ability to apply existing preposition disambiguation results beyond these instances. (We discuss *Cause* senses further in the wider context of all PDEP prepositions in the next section on class analyses.)

As indicated earlier, VerbNet identifies fewer prepositions in its frames than FrameNet. We believe this is the case since VerbNet prepositions are generally arguments, rather than adjuncts. Many of the FrameNet prepositions are evoking peripheral and extra-thematic frame elements, so the number of prepositions is correspondingly higher. Also, VerbNet contains fewer members in its verb classes. As a result, the number of hits when using VerbNet is somewhat smaller, although some use of VerbNet classes is possible with the governor selectors.

PDEP provides a vehicle for expanding the items in all these resources. While prepositions are not central to these resources, their supporting role provides additional information that might be useful in developing and using these other resources.

5 Class Analyses

In SemEval 2007, Yuret (2007) investigated the possibility of using the substitutable prepositions as the basis for disambiguation (as part of more general lexical sample substitution). Although his methodology yielded significant gains over the baseline, his best results were only 54.7 per-

cent accuracy, concluding that preposition use is highly idiosyncratic. Srikumar and Roth (2013) broadened this perspective by considering a class-based approach by collapsing semantically-related senses across prepositions, thereby deriving a semantic relation inventory. While their emphasis was on modeling semantic relations, they achieved an accuracy of 83.53 percent for preposition disambiguation.

As mentioned above, PDEP has a field for the Srikumar semantic relation, initially populated for the SemEval prepositions, and being extended to cover all other prepositions. For example, Srikumar and Roth identified 21 temporal senses across 14 SemEval prepositions, while we have thus far identified 62 senses across 50 prepositions. Similar increases in the sizes of other classes occur as well. For causal senses, Srikumar and Roth identified 11 senses over 7 prepositions, while PDEP has 27 senses under 25 prepositions.

PDEP enables an in-depth analysis of TPP classes, Tratz clusters, and Srikumar semantic realations. First, we query the database underlying Figure 3 to identify all senses with a particular class. We then examine each sense on each list in detail.

We follow the procedures laid out above for examining the features to add information about selectors, complement types, and categories. We use this information to tag the TPP instances, conservatively assuring the tagging, e.g., leaving untagged questionable instances. Finally, we carefully place each sense into a preposition class or subclass, grouping senses together and making annotations that attempt to capture any nuance of meaning that distinguishes the sense from other members of the class.

To build a description of the class and its subclasses, we make use of the Quirk reference in Figure 3 (i.e., the relevant discussions in Quirk et al. (1985)). We build the description of a class as a separate web page and make this available as a menu item in Figure 3 (not shown for the Scalar class when that screenshot was made). The description provides an overview of the class, making use of the TPP data and the Quirk discussion, and indicating the number of senses and the number of prepositions. Next, the description provides a list of the categories within the class, characterizing the complements of the category and then listing each sense in the category, with any nuance of meaning as necessary. Finally, we attempt to summarize the selection criteria that have been used across all the senses in the class.

The process of building a class description reveals inconsistencies in each of the class fields. When we place a preposition sense into the class, we may find it necessary to make changes in the underlying data.

At the top level, these class analyses in effect constitute a coarse-grained sense inventory. As the subclasses are developed, a finer-grained analysis of a particular area is available. We believe these analyses may provide a comprehensive characterization of particular semantic roles that can be used for various NLP applications.

6 Availability of PDEP Data and Potential for Further Enhancements

As indicated above, each of the tables shown in the figures is generated in Javascript through a system call to a PHP script. Each of these scripts is described in detail at the PDEP web site. Each script returns data in Javascript Object Notation (JSON), enabling users to obtain whatever data is of interest to them and perhaps using this data dynamically.

While PDEP provides access to a large amount of data, the architecture is very flexible and easy to extend. For this, we are grateful for the Tratz parser and the DVC code.

In building PDEP, we found it necessary to reprocess the SemEval 2007 data of the full 28,052 sentences that were available through TPP, rather than just those that were used in the SemEval task itself. Tagging, parsing, and creating feature files for these sentences took less than 10 minutes, with an equal time to upload the feature files. We would be able to add or substitute new corpora to the PDEP databases with relatively little effort.

Similarly, we can add new elements or modify existing elements that describe preposition patterns. This would require easily-made modifications to the underlying MySQL database tables. The PHP scripts that access these tables are also easily developed or modified. Most of these scripts use less than 100 lines of code.

In developing PDEP, we have added various resources incrementally. This applies to such resources as the DVC ontology, FrameNet, and VerbNet. Each of these resources required relatively little effort to integrate into PDEP. We will continue to investigate the utility of other resources that will assist in characterizing preposition behavior. We have begun to look at the noun clusters used in Srikumar and Roth (2013) for better characterizing complements. We are also

examining an Oxford noun hierarchy as another alternative for complement analysis. We are examining the WordNet detour to FrameNet, as described in Burchardt et al. (2005), particularly for use in further characterizing the governors.

We recognize that an important element of PDEP will be in its utility for preposition disambiguation. While we have not yet begun the necessary experimentation and evaluation, we believe the representativeness and sample sizes of the TPP corpus (mostly with 250 or more sentences per preposition) should provide a basis for constructing the needed studies. We expect that this will follow techniques used by Cinkova et al. (2012), in examining the Pattern Dictionary of English Verbs developed as the precursor to DVC.

We expect that interaction with the NLP community will help PDEP evolve into a useful resource, not only for characterizing preposition behavior, but also for assisting in the development of other lexical resources.

7 Conclusion and Future Plans

We have described the Pattern Dictionary of English Prepositions (PDEP) as a new lexical resource for examining and recording preposition behavior. PDEP does not introduce any ideas that have not already been explored in the investigation of other parts of speech. However, by bringing together work from these disparate sources, we have shown that it is possible to analyze preposition behavior in a manner equivalent to the major parts of speech. Since dictionary publishers have not previously devoted much effort in analyzing preposition behavior, we believe PDEP may serve an important role, particularly for various NLP applications in which semantic role labeling is important.

On the other hand, PDEP as described in this paper is only in its initial stages. In following the principles laid out for verbs in PDEV, a main goal is to provide a sufficient characterization of how frequently different preposition patterns (senses) occur, with some idea of a statistical characterization of the probability of the conjunction of a preposition, its complement, and its governor. Better development of a desired syntagmatic characterization of preposition behavior, consistent with the principles of TNE, is still needed. Since preposition behavior is strongly linked to verb behavior, further effort is needed to link PDEP to PDEV.

The resource will benefit from further experimentation and evaluation stages. We expect that desired improvements will come from usage in various NLP tasks, particularly word-sense disambiguation and semantic role labeling. In particular, we anticipate that interaction with the NLP community will identify further enhancements, developments, and hints from usage.

Acknowledgments

Stephen Tratz (and Dirk Hovy) provided considerable assistance in using the Tratz parser. Vivek Srikumar graciously provided his data on preposition classes. Vitek Baisa similarly helped with the adaptation of the PDEV Javascript modules. Orin Hargraves, Patrick Hanks, and Eduard Hovy continued to provide valuable insights. Reviewer comments helped sharpen the draft version of the paper.

References

- Bharat Ram Ambati, Siva Reddy, and Adam Kilgarriff. 2012. Word Sketches for Turkish. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*. Istanbul, 2945-2950.
- Aljoscha Burchardt, Katrin Erk, and Anette Frank. 2005. A WordNet Detour to FrameNet. *Proceedings of GLDV workshop GermaNet II*. Bonn.
- Silvie Cinkova, Martin Holub, Adam Rambousek, and Lenka Smejkalova. 2012. A database of semantic clusters of verb usages. *Lexical Resources and Evaluation Conference*. Istanbul, 3176-83.
- Patrick Hanks. 2004. Corpus Pattern Analysis. In *EURALEX Proceedings*. Vol. I, pp. 87-98. Lorient, France: Université de Bretagne-Sud.
- Patrick Hanks. 2013. *Lexical Analysis: Norms and Exploitations*. MIT Press.
- Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. The Sketch Engine. *Proceedings of EURALEX*. Lorient, France, pp. 105-16.
- Ken Litkowski. 2013a. *The Preposition Project Corpora*. Technical Report 13-01. Damascus, MD: CL Research.
- Ken Litkowski. 2013b. *Preposition Disambiguation: Still a Problem*. Technical Report 13-02. 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.
- James Pustejovsky, Catherine Havasi, Jessica Littman, Anna Rumshisky, and Marc Verhagen. 2006. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. *5th Edition of the International Conference on Lexical Resources and Evaluation.*, 1702-5.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman Inc.
- Vivek Srikumar and Dan Roth. 2011. A Joint Model for Extended Semantic Role Labeling. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. ACL, 129-139.
- Vivek Srikumar and Dan Roth. 2013. Modeling Semantic Relations Expressed by Prepositions. *Transactions of the Association for Computational Linguistics*, 1.
- Angus Stevenson and Catherine Soanes (Eds.). 2003. *The Oxford Dictionary of English*. Oxford: Clarendon Press.
- Stephen Tratz. 2011. *Semantically-Enriched Parsing for Natural Language Understanding*. PhD Thesis, University of Southern California.
- 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.
- Deniz Yuret. 2007. KU: Word Sense Disambiguation by Substitution. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, Prague, Czech Republic.
- Zapirain, B., E. Agirre, L. Marquez, and M. Surdeanu. 2013. Selectional Preferences for Semantic Role Classification. *Computational Linguistics*, 39:3.