

# Coverage and Inheritance in The Preposition Project

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

Orin Hargraves  
5130 Band Hall Hill Road  
Westminster, MD 21158  
orinkh@carr.org

## Abstract

In The Preposition Project (TPP), 13 prepositions have now been analyzed and considerable data made available. These prepositions, among the most common words in English, contain 211 senses. By analyzing the coverage of these senses, it is shown that TPP provides potentially greater breadth and depth than other inventories of the range of semantic roles. Specific inheritance mechanisms are developed within the preposition sense inventory and shown to be viable and provide a basis for the rationalization of the range of preposition meaning. In addition, this rationalization can be used for developing a data-driven mapping of a semantic role hierarchy. Based on these findings and methodology, the broad structure of a WordNet-like representation of preposition meaning, with self-contained disambiguation tests, is outlined.

## 1 Introduction

The Preposition Project (TPP, Litkowski & Hargraves, 2005)<sup>1</sup> provides a large amount of data for a small number of prepositions. To date, 13 out of 373 prepositions (among the most frequent in English) have been analyzed. We examined the data for these prepositions to determine (1) their coverage of the semantic space of semantic relations, (2) the extent to which these data could be extrapolated to prepositions not yet covered, and (3) what types of analyses might be useful to fill shortcomings in the data. Examining these issues seems important to determining the extent to which the data in the project can be used in NLP applications.

TPP is designed to provide a comprehensive database of preposition senses, so it is useful to

provide a mechanism for assessing the extent of coverage, not only in comparison with the range of meanings described in traditional grammar, but also in comparison with analyses within the computational linguistics community. Similarly, it seems important to determine how, if at all, the data developed thus far can be leveraged for use with other preposition meanings not yet analyzed, e.g., through mechanisms of inheritance. Finally, through these analyses, it is useful to identify any shortcomings in data being developed in TPP and what further should be undertaken.

In the following sections, we first provide an overview of TPP and extensions to its available data that have occurred since its inception. Next, we examine issues of coverage in relation to the range of preposition meaning contained in Quirk et al. (1985), alongside the ranges in other resources such as the Penn Treebank, FrameNet, and Lexical Conceptual Structures. This analysis also considers accounts of semantic relations that have been presented in literature that has used these other resources. Next, we critically examine claims of the inheritance of preposition meaning as described in Litkowski (2002), including consideration of inheritance mechanisms in FrameNet. This analysis suggests some mechanisms for a data-driven or corpus-based approach to the identification of a semantic relation inventory. Finally, based on these analyses of coverage and inheritance, we identify some next steps TPP needs to take.

## 2 The Preposition Project

The primary objective of TPP is to characterize each of 847 preposition senses for 373 prepositions (including 220 phrasal prepositions with 309 senses) with a semantic role name and the syntactic and semantic properties of its complement and attachment point. The preposition sense inventory is taken from the *Oxford Dictionary of English*

---

<sup>1</sup><http://www.clres.com/prepositions.html>.

(2004).<sup>2</sup> Starting from the senses for a particular preposition, a set of instances of that preposition are extracted from the FrameNet database. A lexicographer then assigns a sense from the inventory to each instance. While engaged in this sense assignment, the lexicographer accumulates an understanding of the behavior of the preposition, assigns a name to each sense (characterizing its semantic type), and characterizes the syntactic and semantic properties of the preposition complement and its point of attachment or head. Each sense is also characterized by its syntactic function and its meaning, identifying the relevant paragraph(s) where it is discussed in Quirk et al.

TPP then makes available the sense analysis (including the lexicographer's overview) and the set of instances for each preposition that is analyzed. In addition, the disambiguated instances are then analyzed to provide the set of FrameNet frames and frame elements associated with each sense. The set of sentences is provided in Senseval format, along with an answer key, for use in development of preposition disambiguation routines (ranging from 300 to over 4000 sentences for 'of'). Finally, using the FrameNet frame and frame element of the tagged instances, syntactic alternation patterns (other syntactic forms in which the semantic role may be realized) are provided for each FrameNet target word; this data constitutes a suitable corpus for use in studying, for example, English verb classes (see Levin, 1993).

An important next step for TPP is the use of these disambiguated instances to refine the characterization of the syntactic and semantic properties of the complement and the point of attachment. As the lexicographer has analyzed the sense inventory for a preposition, the question of its use in relation to other words is continually raised. In particular, the question is whether a sense stands alone or is selected for by a verb or other word (most frequently, an adjective).<sup>3</sup> The lexicographer has observed that selection might be occurring. The extent to which this occurs will be examined when

---

<sup>2</sup>TPP does not include particle senses of such words as *in* or *over* (or any other particles) used with verbs to make phrasal verbs. In this context, phrasal verbs are to be distinguished from verbs that select a preposition (such as *on* in *rely on*), which may be characterized as a collocation. We are grateful to an anonymous reviewer for raising this issue.

<sup>3</sup>We are grateful to an anonymous reviewer for this characterization.

an attempt is made, for example, to develop decision lists for disambiguating among a preposition's senses.<sup>4</sup> We hope, as a result, that the number of instances available for disambiguation will permit a more definitive characterization of selection.

Since Litkowski & Hargraves (2005), several additions have been made to the data and analyses available under TPP. First, Oxford University Press has granted permission to provide the definitions and examples of the senses for each definition from the *Oxford Dictionary of English* (ODE, 2003) (and its predecessor, the *New Oxford Dictionary of English* (NODE, 1997)). Second, a summary file of all senses has been prepared from the individual preposition sense analyses, facilitating overview analysis of the full sense inventory (e.g., sorting the table on different columns). Third, the lexicographer has disambiguated the ending preposition of definitions as those prepositions are analyzed (e.g., in sense 1 of **about**, *on the subject of*, identifying the applicable sense of **of**); 451 prepositions have been so tagged.

At present, the following 13 prepositions have been analyzed (with the initial number of senses in parentheses): **about** (6), **against** (10), **at** (12), **by** (22), **for** (14), **from** (14), **in** (11), **of** (18), **on** (23), **over** (16), **through** (13), **to** (17), and **with** (16). The number of senses has changed based on changes from NODE to ODE and based on evidence developed in the project (adding 19 senses that are attested with the FrameNet data). These prepositions include the most frequent in English (see Boonthum et al., 2006 for the top 10 based on the Brown corpus). In summary, the 13 prepositions (out of 373 identified in Litkowski, 2002) have 210 senses (19 have been added during the course of TPP) out of the original 847 senses.

It is noteworthy also that in moving from NODE to ODE, 60 prepositions have been removed. Some of these prepositions are variant spellings (e.g. **abaht** for **about**). Most are phrasal prepositions, e.g., **to the accompaniment of**. In

---

<sup>4</sup>The anonymous reviewer asked whether TPP excludes senses that are selected for. This prompted an examination of whether this might be the case. Although it is the intent that such senses be included, an examination of how FrameNet instances are generated raises the possibility that such instances may have excluded. Procedures are currently being developed to ensure that such instances are not excluded.

NODE, the definitions constitute a lexicographic statement that the meaning of the phrase has an idiomatic status, i.e., is not solely recoverable based on an understanding of the meanings of its constituents. In ODE, such phrases are identified as having collocative status and thereby rendered in example usages with italics, but not given a definition. Such phrases will be retained in TPP.

Litkowski & Hargraves (2005) provides more details on the methodology used in TPP and the databases that are available.

### 3 Semantic Coverage of TPP

Although only a small percentage of the prepositions have as yet been analyzed, approximately 25 percent of the total number of senses are included in the 13 prepositions. This percentage is sufficient to assess their coverage of the semantic space of prepositional meaning.

#### 3.1 Assessing the Broad Spectrum of Semantic Space

To assess the coverage, the first question is what inventory should be used. The linguistics and computational linguistics literatures are replete with introspective lists of semantic roles. Gildea & Jurafsky (2002) present a list of 18 that may be viewed as reasonably well-accepted. O'Hara (2005) provides several compilations based on Penn Treebank annotations, FrameNet, OpenCyc, and Factotum. Boonthum et al. (2006) includes an assessment of semantic roles in Jackendoff, Dorr's Lexical Conceptual Structures preposition database, and Barker's analysis of preposition meaning; she posits a list of 7 overarching semantic roles (although specifically intended for use in paraphrase analysis). Without going into a detailed analysis of each of these lists, all of which are relatively small in number, the semantic relations included in TPP clearly cover each of the lists. However, since the semantic relations in these lists are relatively coarse-grained, this assessment is not sufficient.

Quirk et al. (1985) is arguably the most comprehensive introspective compilation of the range of preposition meaning. As indicated above, in analyzing the senses for a preposition, the lexicographer includes a reference to a section in Quirk et al (specifically in Chapter 9). Quirk et al. describe the meanings of prepositions in 50

sections, with the majority of discussion devoted to spatial and temporal prepositions. By comparing the references in the spreadsheets for each preposition (i.e., a data-driven approach), we find that only 4 sections are not yet mentioned. These are 9.21 (between), 9.56 (concession), 9.58 (exception and addition), and 9.59 (negative condition). In general, then, TPP broadly covers the full range of meanings expressed by prepositions as described in Quirk et al.

However, for almost half of the senses analyzed in TPP (100 of 210), the lexicographer was unable to assign a Quirk paragraph in Chapter 9 or elsewhere. This raises the question of whether Quirk et al. can be viewed as comprehensive. A preliminary examination of the semantic relations assigned by the lexicographer and not assigned a Quirk paragraph indicates that the range of prepositional meaning is more extensive than what is provided in Quirk et al.

Two major categories of missing semantic relations emerge from this analysis. Of the 100 senses without a Quirk paragraph, 28 involve prepositional usages pertaining to quantities. These include the semantic relations like **Age** (“*at* six he contracted measles”), **ScaleValue** (“an increase *of* 5%”), **RatioDenominator** (“ten miles *to* the gallon”), **Exponent** (“10 *to* the fourth power”), **ValueBasis** (“a tax *on* tea”), **Price** (“copies are available *for* \$5”), and **UnitSize** (“billing is *by* the minute”). Another 32 involve prepositions used to establish a point of reference, similar to the **Standard** in Quirk (section 9.62), except indicating a much broader set. These include semantic relations like **FormerState** (“wakened *from* a dream”), **KnowledgeSource** (“information *from* books”), **NameUsed** (“call him *by* his last name”), **ParentName** (“a child *by* her first husband”), **Experiencer** (“a terrible time *for* us”), and **Comparator** (“that’s nothing compared *to* this”). The remaining 40 semantic relations, such as **MusicalKey** (“*in* F minor”), **Drug** (“*on* dope”), and **ProfessionAspect** (“a job *in* publishing”), appear to represent finer-grained points of prepositional meaning.

This assessment of coverage suggests that TPP currently not only covers the broad range of semantic space, but also identifies gaps that have not received adequate treatment in the linguistic literature. Perhaps such gaps may be viewed as “beneath the radar” and not warranting elaborate treatment. However, it is highly likely that these

**Table 1. Frequency of “at” FrameNet Instances in The Preposition Project**

Semantic Relation	Frequency	Definitions	Examples
Location	0.404	expressing location or arrival in a particular place or position	crouched <b>at</b> the edge of the track
Temporal	0.072	expressing the time when an event takes place	avoid confusion <b>at</b> this late stage
Level	0.039	denoting a particular point or segment on a scale	charged <b>at</b> two percent
Skill	0.038	expressing a particular state or condition, or a relationship between an individual and a skill	brilliant <b>at</b> the job
ActionObject	0.276	expressing the object of a look, gesture, thought, action, or plan	moaned <b>at</b> him
Stimulus	0.171	expressing the means by which something is done or the cause of an action or reaction	boiled <b>at</b> his lack of thought

senses occur with considerable frequency and should be treated.

It is somewhat premature to perform a comprehensive analysis of coverage that provides a full characterization of the semantic space of preposition meaning based on the 25 percent of senses that have been analyzed thus far. However, the available data are sufficient to begin such an effort; this issue is further discussed below.

### 3.2 Assessing Finer-Grained Spectra of Prepositional Meaning

While examining the broad coverage of preposition meaning, several issues affecting the treatment of individual prepositions in the computational linguistics literature emerged. These issues also provide a perspective on the potential value of the analyses being performed in TPP.

O’Hara (2005), in attempting to create a framework for analysis and identification of semantic relations, examined the utility of Penn Treebank II annotations and FrameNet frame elements. He examined sentences containing **at** in both corpora. In Treebank, he noted that there were four senses: locative (0.732), temporal (0.239), manner (0.020), and direction (0.006). In FrameNet, with some combination of frame elements, he identified five major senses: addressee (0.315), other (0.092), phenomenon (0.086), goal (0.079), and content (0.051).

Table 1 provides a coarse-grained analysis of **at** developed in TPP (6 additional subsenses are not shown). Although frequencies are shown in the table, they should not be taken seriously, since the

FrameNet instances on which they are based makes no claim to be representative. In particular, FrameNet seldom annotates temporal references since they are usually viewed as peripheral frame elements that may occur with virtually all frames. Nonetheless, the frequencies in the FrameNet instances does indicate that each of the **at** senses is likely to occur at levels that should not be ignored or glossed over.

In comparing TPP results with Penn Treebank characterizations, it seems that, not only might the corpus be unrepresentative, but that the linguistic introspection does not capture the more natural array of senses. Thus, by combining corpus evidence (from FrameNet) with a lexicographic perspective for carving out sense distinctions, an improved balance results. It should also be noted that in Table 1, the final sense for **Stimulus** emerged from the FrameNet data and from Quirk and was not identified in the ODE sense inventory.

Comparing TPP results with O’Hara’s aggregation of FrameNet frame elements indicates the difficulty of working directly with the large number of frame elements (currently over 700). As Gildea & Jurafsky noted, it is difficult to map these frame elements into higher level semantic roles. Some assistance is available from the FrameNet inheritance hierarchy, but this is still not well-developed. This issue is taken up further below in describing how TPP’s data-driven approach may facilitate this kind of mapping.

In summary, the methodology being followed in TPP arguably provides a more natural and a more assuredly complete coverage of the fine-grained senses associated with an individual preposition.

## 4 Inheritance Within the Preposition Sense Inventory

The preceding discussion provides some assurance that TPP provides broad coverage of the range of prepositional meaning and fine-grained analysis of the behavior of individual prepositions. However, the large number of preposition senses requires some additional work to manage these broad and fine-grained spectra. Litkowski (2002) provided a graph-theoretical analysis that arranged prepositions into a hierarchy. However, that analysis treated individual prepositions as aggregations, i.e., all senses were combined into nodes in a digraph. With the finer-grained analysis now available in TPP data, a more in-depth examination of inheritance within the preposition sense inventory is possible.

### 4.1 Initial Considerations for Mapping Out the Inheritance Hierarchy

Of the 847 senses described in Litkowski (2002), and used as the starting point for the analysis in TPP, most follow the prototypical form of a prepositional phrase followed by a terminal (dangling) preposition, e.g., for sense 1 of **about**, *on the subject of*. Litkowski viewed the terminal preposition as a hypernym. However, 62 senses do not have terminal prepositions (but rather usually verbs) and an additional 164 senses are usage notes describing behavior (such as the senses of **at** shown in Table 1). These 226 senses were viewed as being primitive, while the remaining 621 were viewed as being derived in some way dependent on the putative hypernym.

Among the 13 prepositions that have been analyzed thus far, 11 senses having a non-preposition hypernym and 100 senses with usage notes have been characterized. Thus, only about half of the so-called primitives have been assigned a semantic relation type. Further analysis of the range of meaning of these primitives should await a more complete coverage of these senses. The kind of analysis envisioned among these senses is determining how they group together and what range of semantic meaning they express. This will be discussed further below.

Of the 621 senses with a preposition hypernym, 411 end in one of the 13 prepositions that have been analyzed, with 175 ending in **of** and 74 in **to**. The remaining 210 senses end in prepositions with at

most a few cases of the same preposition. Most of these remaining senses, in fact, are the ones that gave rise to the definitional cycles and hierarchical analysis of the digraph described in Litkowski (2002). As a result, senses with a preposition hypernym form a set sufficient in size for a more detailed analysis of inheritance within the preposition inventory.

### 4.2 The Meaning of an Inheritance Hierarchy for Prepositions

The assumption underlying an inheritance analysis of preposition definitions with a terminal preposition is that such definitions are substitutable for the preposition that is defined. For example, in *a book about ancient Greece*, **about** can be replaced by its definition to obtain *a book on the subject of ancient Greece*. This sense of **about** has been labeled **SubjectConsidered** (or equivalently, Topic or Subject) by the lexicographer. In the inheritance analysis, this definition of **about** is said to have **of** as its hypernym.

Clearly, the hypernymic ascription for prepositions is by analogy only. To say that **about** *isa of* makes little sense. In TPP, the lexicographer develops three pieces of information about each sense: a semantic relation name, the properties of the prepositional object, and the properties of the word to which the prepositional phrase is attached. In analyzing the definition for **about**, **of** is attached to the word *subject*. Thus, nothing about the attachment properties of **of** can be inherited into saying anything about the attachment properties of **about**. At best, then, the semantic relation name and complement properties of the applicable sense of **of** can be inherited. Indeed, this can be put into the form of a hypothesis: *the semantic relation name and the complement properties of an inherited sense are more general than those of the inheriting sense*.

As mentioned above, the lexicographer has disambiguated the terminal preposition in senses that use one of the 13 prepositions that have been analyzed. This has been done for 451 definitions in the 411 senses. It is noteworthy that in only 29 cases did the lexicographer assign multiple senses (i.e., viewing the applicable sense as ambiguous). In other words, despite the fact that most of these definitions contained only 4 or 5 words, sufficient context enabled resolution to a specific sense of the hypernym. In 8 cases, the multiple inheritance was

**Table 2. Inheritance of Semantic Relations and Complement Properties**

Semantic Relation	Preposition	Complement Properties	Definition	Hypernym Semantic Relation	Hypernym Complement Properties
Opposing Force	against	sth actively resisted	in resistance to; as protection <b>from</b>	Thing Prevented	participle or noun denoting thing prevented
Thing Surmounted	over	a physical entity that can have sth above it	extending directly upwards <b>from</b>	Space Origin	point in space or abstraction identified as origin
Thing Bored	through	permeable or breakable physical object	so as to make a hole or opening <b>in</b> (a physical object)	Thing Entered	sth capable of being entered or of incorporating or enveloping input
Beneficiary	for	usually a person; otherwise, sth capable of benefitting	on behalf of or to the benefit <b>of</b> (someone or something)	Recipient	noun representing the obj. of action denoted in the POA
Feature Backdrop	on	background on which the POA is located	forming a distinctive or marked part <b>of</b> (the surface of something)	Whole	object of which the POA is a part, piece, or sample
Downside	against	downside; the con in a pro/con situation	in conceptual contrast <b>to</b>	Comparator	second term of a comparison

for all senses, as in the case of **frae**, a Scottish dialectal form of **from**.

In making the sense assignments, 175 of which (39 percent) involved **of**, the lexicographer noted that a large number of cases (132 of 373) involved phrasal prepositions that ended in **of**, e.g., **into the arms of** and **in the name of**. In these cases, the definition (as developed by Oxford lexicographers) merely substituted one phrase ending in **of** for the phrase being defined (*into the possession or control of* and *for the sake of* for the two examples). This observation was a major reason for requiring that any hypernymic ascription within the preposition inventory could not be based on the prototypical *isa* hierarchy applicable to nouns.

Among the 411 senses for which the terminal preposition had been disambiguated, 48 senses occurred as definitions of the 13 prepositions that have been analyzed in TPP. For these 48 senses, each of which was fully characterized, the characterization of the terminal prepositions was also available, thus enabling us to test the hypothesis about what could be inherited. Table 2 shows the results for 6 of these senses, giving first the semantic relation assigned to the sense by the lexicographer, the preposition, the characterization of the complement properties for that sense, the definition (with the hypernymic preposition in bold), the semantic relation of the sense that the lexicographer judged to be the appropriate sense of

the hypernymic preposition, and the complement properties of that sense.

The examples in Table 2 support the hypothesis about inheritance. The other 42 cases are similar, although for some, the hypernymic semantic relation or hypernymic complement properties are not as close to the preposition sense being examined. In a few cases, for example, the complement properties are as general as “any noun.” In such cases, what gets inherited may not provide much in the way of specificity to aid in analyzing the behavior of the inheriting preposition. However, viewed from the perspective of the digraph analysis performed in Litkowski (2002), this inheritance analysis provides confidence that there is an ordering relationship within the preposition sense inventory that can be exploited.

In the digraph analysis in Litkowski (2002), where the prepositions were analyzed as aggregated nodes, the inheritance mechanism provides the basis for splitting nodes based on the specific sense assignments that can now be made. In particular, in Table 3, showing one node of the preposition digraph that was characterized as a single strong component (number 12) containing 33 prepositions, the sense-specific assignments will permit the disaggregation of these prepositions into smaller groups that are closely related.

Table 3. Strong Components	
	Entries
12	in favour of, along with, with respect to, in proportion to, in relation to, in connection with, with reference to, in respect of, as regards, concerning, about, with, in place of, instead of, in support of, except, other than, apart from, in addition to, behind, beside, next to, following, past, beyond, after, to, before, in front of, ahead of, for, by, according to

In considering the type of analysis described by Table 2, it is important to note that the results followed from the reliance on a data-driven approach. Characterizations of individual senses are made locally with respect to observed behavior of a single preposition. It is only after these analyses that results from several tables and spreadsheets can be conjoined to produce something like Table 2.

It is also important to note that the results in Table 2 must be viewed as preliminary. Although it is expected that the central hypothesis about inheritance will remain valid, it is expected that the characterizations of the complement properties will undergo considerable refinement. One of the primary goals of TPP is to develop a data-driven set of disambiguation criteria for distinguishing among preposition senses. Methods such as those developed by O'Hara (2005) and Boonthum et al. (2006) suggest that refined characterizations will emerge. The large instance sets (in Senseval format) will provide an ample data set for this analysis.

Finally, it is expected that the semantic relation names will also undergo some additional revisions. Again, since these names are developed locally with respect to single prepositions, they do not reflect what may be a final set when they are analyzed together. This is discussed in the next section.

## 5 Next Steps for The Preposition Project

The analyses of issues concerning coverage and inheritance within the preposition sense inventory suggest at least two major new goals for TPP. One is the rationalization of the semantic relation types and the other is the aggregation of characterizations about the senses into a convenient and usable data structure, perhaps following WordNet.

### 5.1 Rationalization of Semantic Relation Types

The semantic relation types that have been

developed thus far in TPP have been extremely useful in assessing the current coverage of the semantic space of prepositions and in examining the possibilities of an inheritance structure for the senses. However, the analyses have shown that there are some gaps in broad coverage and some that will affect fine-grained characterizations of the semantic space.

In performing the analyses of the 13 prepositions and their 211 senses, the names for the semantic relations for an individual preposition have been developed without regard to those from other prepositions or the linguistic literature, based on the individual definitions in ODE and the instances from FrameNet that have been tagged. Although frame element names are available to the lexicographer when examining FrameNet instances, they are only in the background. As a result, these names provide a data-driven basis for characterizing the semantic space of prepositions.

Given the importance of these names for the types of analyses described above, it is valuable to complete the assignment of names, even without the full-scale analysis of sentence instances. Completion of this task would represent only a preliminary assignment, modifiable when instances are more fully analyzed.

With a relatively complete set of semantic relation types, "rationalization" of the set, i.e., reorganization in such a way to make it more logical and consistent can be performed. At present, among the 211 semantic relation types, there are 36 duplicate names, some appearing multiple times, e.g., **AgentName** appears 5 times. Some names are only slight variants of one another, such as **Timeframe** and **TimePeriod**. Many names can be grouped together for analysis. For example, in the time space, such semantic relations as **ActivePeriod**, **ClockHour**, **CreationDate**, **FutureTime**, **Hour**, **PeriodBisected**, **PointInTime**, **TargetTime**, **TimeOrigin**, and **TimePeriod** would be examined together.

In pursuing this rationalization, outside resources can be used more efficiently. In particular, the FrameNet naming conventions and inheritance hierarchy can be examined in more detail (as well as critiqued). In addition, it will be possible to take into account other treatments of particular prepositions or fine-grained areas of semantic space more easily.

Rationalization not only will ensure consistency in naming, but provide a vehicle for appropriate

data-driven mapping. This will provide a basis for or against conventional groupings that have been posited in the linguistics and computational linguistics literature. It is not expected that this rationalization will produce anything unexpected, but it will provide an underlying support for characterizing the range of prepositional meaning.

## 5.2 Towards a WordNet Representation of Prepositional Meaning

The amount of data generated in TPP has been prodigious and is difficult to comprehend and exploit. With a firmer basis established in section 4 above for inheritance mechanisms, combined with the digraph analysis described in Litkowski (2002), it seems possible to move toward a representation that is similar to WordNet.

By following the inheritance structure, based on the analyses described in section 4, combined with a rationalization of semantic relation names, it seems likely that there will be a relatively small number of primitive concepts. The digraph analysis yields synsets in the manner of WordNet, so we can visualize that nodes in a WordNet preposition network will consist of preposition names and preposition glosses (i.e., definitions). In addition, the objective will be to provide an improved characterization of complement and attachment properties that will accompany each node. Thus, such a WordNet-like preposition network will represent not only meanings, but also provide the capability for disambiguation.

## 6 Conclusions

Although only a small number of prepositions have been analyzed in The Preposition Project, the data that has been generated has proved sufficient for a broad assessment of the range of preposition meaning. Not only has it been possible to demonstrate that the project currently provides a comparable broad coverage, but also that it reveals potential gaps in previous analyses of coverage. The data has also proved sufficient for the articulation of appropriate inheritance mechanisms within the preposition sense inventory. These results have permitted the development of procedures that can be used for mapping out the space of semantic roles. In addition, with these results, it is possible to lay out steps toward a WordNet-like representation of prepositions and their behavior.

## References

- Bonnie Dorr. 1996. Lexical Conceptual Structures for Prepositions (<http://www.umiacs.umd.edu/~bonnie/AZ-preps-English.lcs>)
- Chutima Boonthum, Shunichi Toida, & Irwin Levinstein. 2006. Preposition Senses: Generalized Disambiguation Model. Conference on Intelligent Text Processing and Computational Linguistics (CICLING-2006). Mexico City.
- Daniel Gildea and Daniel Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28 (3), 245-288.
- Kenneth C. Litkowski. 2002. Digraph Analysis of Dictionary Preposition Definitions. Word Sense Disambiguation: Recent Success and Future Directions. Philadelphia, PA: Association for Computational Linguistics.
- Kenneth C. Litkowski & Orin Hargraves. 2005. The Preposition Project. ACL-SIGSEM Workshop on "The Linguistic Dimensions of Prepositions and their Use in Computational Linguistic Formalisms and Applications", University of Essex - Colchester, United Kingdom. 171-179.
- The New Oxford Dictionary of English*. 1998. (J. Pearsall, Ed.). Oxford: Clarendon Press.
- Thomas P. O'Hara. 2005. Empirical Acquisition of Conceptual Distinctions via Dictionary Definitions. Ph.D. Thesis. New Mexico State University.
- The Oxford Dictionary of English*. 2003. (A. Stevenson and C. Soanes, Eds.). Oxford: Clarendon Press.
- Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, & Jan Svartik. (1985). *A comprehensive grammar of the English language*. London: Longman.