Chapter 4

Preposition Sense Disambiguation

4.1 Introduction

Like most other parts of speech, prepositions can exhibit of a variety of different meanings. For example, the word *by* can have several senses including *means* (e.g., 'melt it by cooking it'), *spatial proximity* (e.g., 'it is by the car'), and *magnitude* (e.g., 'fell by six points'). Determining which meaning a particular word takes in a given context, word sense disambiguation (WSD), is a valuable preprocessing step for many natural language processing (NLP) tasks including machine translation (Carpuat and Wu, 2007; Chan et al., 2007), sentiment analysis (Rentoumi et al., 2009), subcategorization acquisition (Korhonen and Preiss, 2003), and summarization (Plaza et al., 2010). Since prepositions 1) can encode many important functional relationships, 2) can be highly polysemous, with an average of over 9 senses per preposition for the 34 most frequent prepositions according to The Preposition Project (Litkowski and Hargraves, 2005), and 3) are quite common,

comprising approximately 9.4% of the tokens in the Penn Treebank (Marcus et al., 1993), it is important to be able to accurately disambiguate them.

To date, most word sense disambiguation work has focused on assigning content words, especially nouns and verbs, to their appropriate WordNet (Fellbaum, 1998) senses while ignoring the disambiguation of other important word classes such as prepositions. Some of the earliest work on preposition sense disambiguation (PSD) was performed by O'Hara and Wiebe (2003), who trained classifiers to annotate prepositions with the coarse function tags provided in the Penn Treebank (e.g., TMP - temporal, LOC - locative). Later, Litkowski and Hargraves (2005) created The Preposition Project (TPP), which consists of a corpus of preposition instances annotated according to a fine-grained sense inventory based upon Oxford English Dictionary definitions. The TPP dataset was used as the basis for the first fine-grained preposition sense disambiguation evaluation, SemEval 2007 Task #6 (Litkowski and Hargraves, 2007), a task for which participants built classifiers using a training set of 16,557 annotated prepositions covering the 34 most frequent English prepositions and evaluated against a separate, held-out test collection of 8,096 instances. The best participating system in the SemEval task achieved 69.3% accuracy, a result that was later improved upon by Tratz and Hovy (2009) and Hovy et al. (2010), who reported 75.1% and 84.8% accuracy, respectively.

In this chapter, I describe the current version of the state-of-the-art preposition sense disambiguation system previously described by Tratz and Hovy (2009) and Hovy et al. (2010); work done to refine the sense inventory created by The Preposition Project, the

reasons for doing so, and the results of related inter-annotator agreement work; experimental results using both the original TPP data and the refined data; and mappings of the refined senses to the work on noun compounds (Chapter 3) and possessives (Chapter 5). The sense refinement work produced a set of 278 preposition senses for the 34 most frequent English prepositions—54 fewer senses than were originally used. In most cases, the agreement between annotators was in excess of 80%. The automatic preposition sense disambiguation system (Section 4.3) achieves 85.7% on the SemEval 2007 Task #6 data, which is over 16% more accurate than the highest performing SemEval participant and 1.2% more accurate than an earlier version of the system described by Hovy et al. (2010). With the same training/test split but using the new, refined definitions and annotations instead, the system achieves 88.1% accuracy.

4.2 Sense Inventory Refinement and Alignment

4.2.1 Motivation

Unfortunately, even though most of the definitions used by The Preposition Project are relatively good, there are various opportunities for improvement, including cases of vague definitions or overly fine sense distinctions. Furthermore, all the TPP annotation work was performed by a single individual and, as such, no inter-annotator agreement study was conducted, leaving open the question of how well humans can agree. Though the annotation was fairly consistent, it is hard for anyone not to make mistakes, especially

when there are over 24,000 examples being annotated and some of the definitions are in need of improvement.

An example of a particularly vague definition is one for the word *of—indicating an association between two entities, typically one of belonging, in which the first is the head of the phrase and the second is something associated with it*. This definition is assigned to a wide variety of examples including those that express location (e.g., 'the bogs of Ireland', 'the loess plateau of northern China'), ownership (e.g., 'locker of club skipper Cyril Proby', 'the Derbyshire mill of Elias Needham'), part-of-whole (e.g., 'the belly of the whale', 'beards of the elders'), logical grammatical subject (e.g., 'duel of two astute minds', 'clash of pocket-sized titans') and de-adjectival noun complement (e.g., 'proficiency of Russian students', 'the susceptibility of a crop'). Clearly, this is a case where some revisions could be made, both with the definition and the data annotation, especially since there are already separate senses for part-of-whole and logical grammatical subject.

One definition for *of* that is more reasonable is *expressing the relationship between a general category or type and the thing being specified which belongs to such a category*. However, even though a significant portion of the data annotated with this sense seems reasonable, including 'sacrament of baptism' and 'nickname of fungus-breath', there are many instances that appear as though they belong elsewhere, including cases of time (e.g., 'Olson's splendid essay of 1949', 'Napoleon's expedition of 1798'), location (e.g., 'political struggles of South America', 'charities ... of the area'), and de-adjectival noun complement (e.g., 'probability of such an event', 'absurdity of Ceau's methods').

In addition to cases where the definition is vague or the annotation could be improved, there are some overly fine or subtle distinctions, including some for the word *above*. Two of the senses for *above* are *in extended space over and not touching* and *at a higher level or layer than*. Annotated examples for the former include 'the water was bubbling above the metal floor grille' and 'a kite above the beach' and, for the latter, 'swam above her' and 'live above the shop.' Arguably, the intended distinction, or its value, isn't very obvious, and the definitions could be combined.

Given the aforementioned examples, it should be clear that some additional refinement of the definitions and annotations is warranted.

4.2.2 Refinement Process

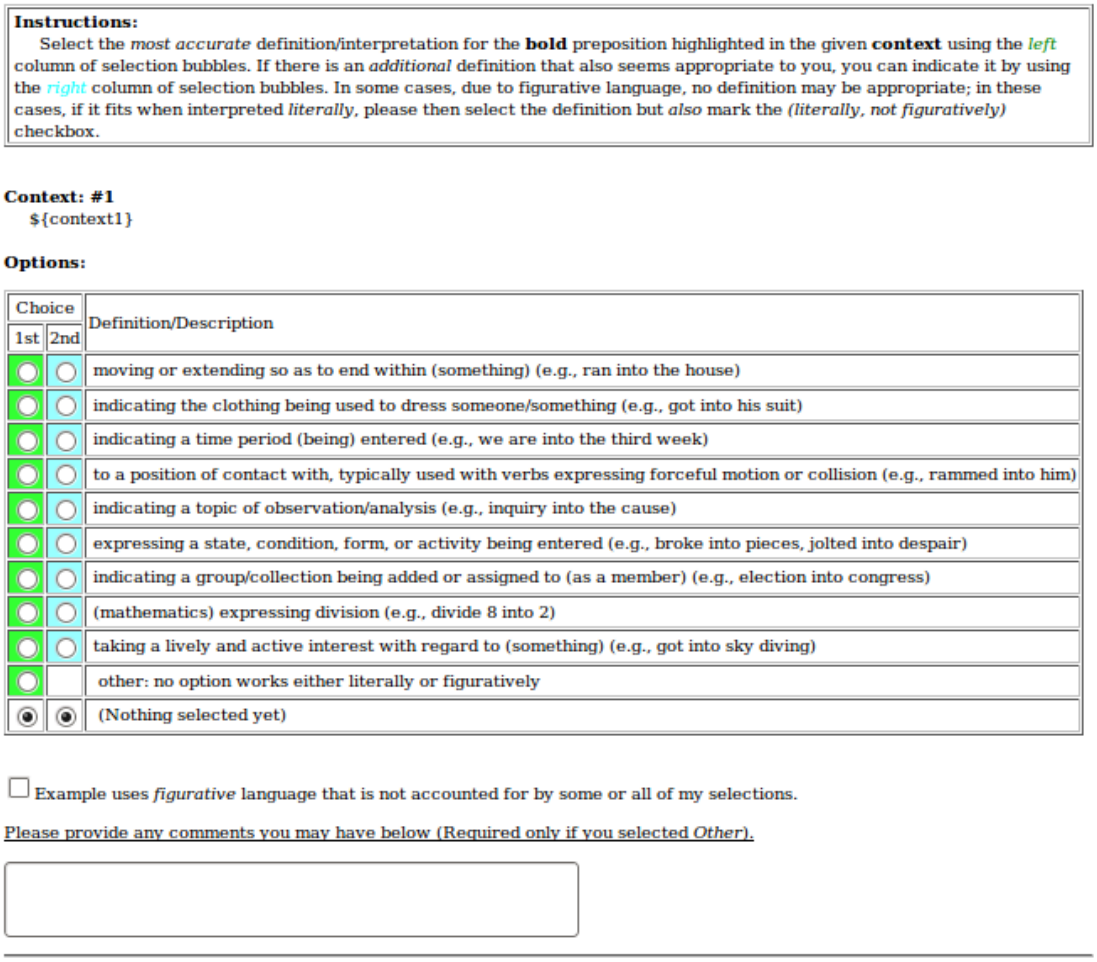
The refinement process consisted of two main phases. The first phase was a largely independent exercise in which I (Annotator A) examined the annotated preposition data in order to create an initial sense inventory. The second phase was an iterative process involving an additional annotator who would annotate portions of the data and provide feedback on the definitions so that I could compare his annotations with mine, calculate inter-annotator agreement, and, ultimately, revise the sense inventory and, if necessary, the annotations. After the refinement was completed, a final batch of data was selected to calculate final inter-annotator agreement figures.

The first phase of the refinement process was begun by sorting the data for each preposition based upon the original annotation given by The Preposition Project and then automatically extracting the governor and prepositional complement/object using the MALTPARSER (Nivre et al., 2006) [1]. Examining a list of governor and complement/object pairs is a quick way to understand the scope of a particular preposition sense. During the first pass over the data, the senses were broken apart as finely as possible, with the data grouped based upon particular characteristics of the governor and/or complement, such as part-of-speech or some other general category (e.g., emotion, communication, competition). Ultimately, this process resulted in a much large number of potential senses than were originally present in The Preposition Project. During the second pass over the data, the various sub-senses were merged together to form an initial sense inventory. The preposition *around* and its British variant *round* were combined for the purpose of creating the new sense inventory due to their similarity. Finally, the actual definitions for the senses were written. Many of defintions were taken directly from the original The Preposition Project inventory, though a significant number were altered to make them more appropriate to the actual data or more clear. By the end of the whole multi-month process, many of original senses that seemed incohesive were split apart and many of sense distinctions that had seemed somewhat arbitrary had disappeared. Of course, such judgments are subjective, and it is important that multiple people can agree on any sort of data annotation.

1. The parser described in Chapter 7 could have been used, but it was still under very active development at the time and support for using the MALTPARSER already existed in the PSD code base.

In the second phase, an additional annotator (Annotator B) was given batches of data to annotate, about thirty instances at a time, along with the list of possible definitions. An image depicting the annotating interface for the preposition *into* is presented in Figure 4.1. After each batch had been annotated, I examined the annotations and compared them to my own in order to find cases where the definitions were poorly defined, overly general, or overly specific so that they could be improved. The number of batches varied by preposition, depending both on the overall level of disagreement as well as individual disagreements. For each batch, the entropy of the annotations, the level of simple agreement, and Fleiss' Kappa (1971) [2] were calculated—these calculations are provided in Table 4.1. For these calculations, if, while inspecting the data points for which disagreement, I concluded that Annotator B's were correct, I corrected my own annotations before performing the calculations. The number of iterations performed varied by preposition. No objective measure, such as agreement level, was used to determine when to end the annotation process for a given preposition, and, indeed, it would not have been reasonable to expect the same level of agreement to be possible for all prepositions given the substantial variation in their level of polysemy. In general, the process continued until agreement was near 90% and/or no significant areas for improvement were apparent. The full list of definitions are presented in Appendix A.

2. Fleiss' Kappa was preferred over Cohen's Kappa because it can be applied to the annotation of more than two annotators. Both are intended to produce agreement figures that account for chance agreement but vary in how the probability of chance agreement is calculated. The difference between Fleiss' Kappa and Cohen's Kappa was rarely more than 0.01.



Instructions:
Select the *most accurate* definition/interpretation for the **bold** preposition highlighted in the given **context** using the *left* column of selection bubbles. If there is an *additional* definition that also seems appropriate to you, you can indicate it by using the *right* column of selection bubbles. In some cases, due to figurative language, no definition may be appropriate; in these cases, if it fits when interpreted *literally*, please then select the definition but *also* mark the *(literally, not figuratively)* checkbox.

Context: #1
${context1}

Options:

Choice 1st	Choice 2nd	Definition/Description
○	○	moving or extending so as to end within (something) (e.g., ran into the house)
○	○	indicating the clothing being used to dress someone/something (e.g., got into his suit)
○	○	indicating a time period (being) entered (e.g., we are into the third week)
○	○	to a position of contact with, typically used with verbs expressing forceful motion or collision (e.g., rammed into him)
○	○	indicating a topic of observation/analysis (e.g., inquiry into the cause)
○	○	expressing a state, condition, form, or activity being entered (e.g., broke into pieces, jolted into despair)
○	○	indicating a group/collection being added or assigned to (as a member) (e.g., election into congress)
○	○	(mathematics) expressing division (e.g., divide 8 into 2)
○	○	taking a lively and active interest with regard to (something) (e.g., got into sky diving)
○		other: no option works either literally or figuratively
◉	◉	(Nothing selected yet)

☐ Example uses *figurative* language that is not accounted for by some or all of my selections.

Please provide any comments you may have below (Required only if you selected *Other*).

Figure 4.1: Screen shot of the HTML template page used for annotating the preposition *into*.

Prep	Agree	κ	E1	E2	Prep	Agree	κ	E1	E2
about	1.00	1.00	0.57	0.57	inside	0.62	0.39	1.55	1.18
above	1.00	1.00	0.96	0.96		0.90	0.84	1.43	1.42
across	0.75	-0.13	0.00	0.78	into	0.93	0.84	0.98	1.05
	0.97	0.84	0.56	0.67		0.87	0.74	1.52	1.25
after	0.90	0.86	1.85	2.10		0.90	0.80	1.21	1.43
	1.00	1.00	1.64	1.64	like	0.87	0.69	0.97	1.37
against	0.84	0.71	0.95	1.50		0.87	0.44	0.42	0.91
	0.97	0.94	1.00	1.18	of	0.67	0.60	2.79	2.81
along	1.00	*	0.00	0.00		0.63	0.57	2.99	2.94
among	0.77	0.66	1.77	1.77		0.77	0.74	3.67	3.51
	0.90	0.85	1.82	1.91		0.83	0.82	3.59	3.51
a(round)	0.37	-0.14	1.16	1.40		0.77	0.74	3.49	3.22
	0.93	0.86	1.16	1.16		0.77	0.74	3.54	3.27
	0.67	0.42	1.18	1.54		0.57	0.52	3.75	3.15
	0.73	0.51	1.00	1.37		0.70	0.65	3.13	3.04
as	1.00	1.00	0.39	0.39	off	0.85	0.44	0.50	0.87
at	0.74	0.65	2.21	2.34		0.80	0.17	0.57	0.77
	0.93	0.92	2.48	2.56		0.90	0.53	0.42	0.88
before	0.90	0.79	0.98	1.22	on	0.76	0.68	2.12	2.73
	0.97	0.94	1.34	1.37		1.00	1.00	2.76	2.76
behind	0.76	0.37	0.58	0.96	onto	1.00	1.00	0.35	0.35
	1.00	*	0.00	0.00	over	0.73	0.64	2.10	2.14
beneath	0.97	0.65	0.21	0.35		0.77	0.72	2.64	2.75
beside	0.97	-0.02	0.00	0.21		0.90	0.87	2.43	2.63
between	0.93	0.91	2.38	2.32		0.90	0.87	2.31	2.17
	0.87	0.78	1.64	1.37		0.80	0.74	2.20	2.42
by	0.93	0.88	1.82	1.80	through	0.86	0.59	0.79	1.24
	0.93	0.85	1.34	1.15		0.90	0.66	0.97	0.88
	0.87	0.82	2.25	2.37		0.93	0.85	1.11	1.41
down	0.90	0.80	1.17	1.11	to	0.97	0.95	2.10	1.95
during	1.00	*	0.00	0.00		0.87	0.83	2.30	2.60
for	0.67	0.60	3.08	2.79		0.87	0.82	2.27	2.25
	0.83	0.79	2.53	2.65		0.97	0.95	2.10	2.17
	0.83	0.81	3.15	3.48	toward	0.63	0.13	0.56	1.32
	0.77	0.73	3.14	3.26		0.93	0.84	1.11	1.09
	0.73	0.70	3.20	3.26	with	0.83	0.78	2.43	2.51
	0.87	0.85	3.24	3.40		0.83	0.79	2.62	2.37
from	0.70	0.64	3.12	3.21		0.90	0.86	2.27	2.14
	0.83	0.80	2.83	3.08		0.83	0.77	2.21	2.29
	0.90	0.88	2.69	2.69		0.90	0.84	1.71	1.94
in	0.83	0.81	3.12	3.12					
	0.83	0.79	2.82	2.97					
	0.73	0.68	3.01	2.95					
	0.90	0.88	3.40	3.33					

Table 4.1: Intermediate results for preposition sense refinement work. Each line shows agreement and entropy for an individual batch. Agree indicates the plain agreement (number of agreements over number of data points) and κ indicates Fleiss' Kappa agreement.

4.2.3 Inter-annotator Agreement

For the final inter-annotator agreement calculation, a third annotator (Annotator C) was brought in. This final round of annotation used batches of fourty-five examples, one for each preposition, for a total of 1,485 annotations. During an initial inspection of the annotator disagreements, it became apparent that a significant portion of the disagreements were likely due to some form of annotator error—that is, if the annotator were to re-examine the particular examples again in detail, he/she would likely make a different selection. To correct for this, annotators were asked to re-annotate all 352 examples for which there was some disagreement. For this annotation revision process, annotators were only shown the definitions that one or more annotators had selected. No indication was given as to who had selected the particular definitions or the number of annotators who had selected them[3]. Annotators were instructed to choose the best definition and not to choose a definition simply because they thought they may or may not have chosen it before. The initial entropy and agreement figures are presented in Table 4.2 and the figures after the revisions are given in Table 4.3.

3. Of course, for the 41 cases for which all three annotators had chosen different definitions, it was obvious that each definition had been selected by exactly one person.

		Agreement (%)			κ				Entropy		
Prep	Senses	A vs B	A vs C	B vs C	A vs B	A vs C	B vs C	All	A	B	C
about	3	0.91	0.98	0.93	0.77	0.93	0.83	0.84	0.72	1.12	0.87
above	4	0.95	0.93	0.98	0.88	0.82	0.94	0.88	1.19	1.13	1.17
across	3	0.91	0.84	0.78	-0.04	-0.06	0.01	-0.01	0.00	0.50	0.87
after	6	1.00	0.93	0.93	1.00	0.89	0.89	0.93	1.43	1.43	1.76
against	5	1.00	0.96	0.96	1.00	0.91	0.91	0.94	1.00	1.00	1.26
along	3	1.00	0.91	0.91	1.00	0.16	0.16	0.31	0.15	0.15	0.43
among	5	0.96	0.87	0.82	0.93	0.80	0.73	0.82	1.67	1.73	1.89
(a)round	5	0.80	0.42	0.56	0.63	-0.06	0.13	0.24	1.26	1.32	1.38
as	2	0.91	0.73	0.73	-0.05	0.09	-0.15	0.01	0.43	0.00	0.84
at	10	0.89	0.67	0.69	0.85	0.56	0.60	0.67	2.35	2.33	2.23
before	4	0.93	0.93	0.91	0.89	0.89	0.85	0.88	1.43	1.45	1.44
behind	5	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
beneath	3	0.96	0.93	0.93	-0.02	0.36	0.36	0.30	0.15	0.15	0.43
beside	3	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
between	7	0.89	0.87	0.91	0.81	0.77	0.84	0.81	1.50	1.47	1.38
by	16	0.93	0.69	0.67	0.91	0.59	0.57	0.69	2.11	2.32	2.38
down	4	0.84	0.80	0.76	0.73	0.63	0.57	0.65	1.29	1.43	1.21
during	1	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
for	17	0.76	0.53	0.56	0.71	0.46	0.49	0.55	2.67	3.12	3.23
from	16	0.89	0.69	0.73	0.86	0.61	0.67	0.72	2.73	3.01	2.82
in	23	0.62	0.49	0.51	0.57	0.42	0.45	0.48	3.44	3.54	3.41
inside	5	0.80	0.84	0.87	0.69	0.76	0.80	0.75	1.58	1.65	1.62
into	9	0.73	0.84	0.64	0.27	0.70	0.23	0.43	1.31	0.93	1.72
like	5	0.76	0.78	0.76	0.32	0.35	0.49	0.41	0.58	1.36	1.25
of	26	0.67	0.53	0.47	0.61	0.44	0.39	0.48	2.99	3.58	3.23
off	6	0.96	0.78	0.76	0.78	0.36	0.30	0.44	0.66	0.61	1.44
on	19	0.73	0.62	0.67	0.67	0.54	0.59	0.60	2.74	2.90	2.94
onto	3	0.98	0.98	0.96	-0.01	-0.01	-0.02	-0.02	0.00	0.15	0.15
over	12	0.87	0.82	0.78	0.83	0.77	0.71	0.77	2.45	2.33	2.51
through	7	0.87	0.84	0.82	0.44	0.36	0.49	0.44	0.41	0.97	1.16
to	14	0.87	0.87	0.84	0.82	0.82	0.79	0.81	2.29	2.36	2.59
towards	7	0.98	0.87	0.87	0.95	0.70	0.69	0.78	1.28	1.21	1.32
with	15	0.87	0.82	0.80	0.82	0.76	0.74	0.77	2.14	2.45	2.41

Table 4.2: Preposition sense refinement agreement results before revisions. Agree indicates the plain agreement and κ indicates Fleiss' kappa agreement.

		Agreement (%)			κ				Entropy		
Prep	Senses	A vs B	A vs C	B vs C	A vs B	A vs C	B vs C	All	A	B	C
about	3	0.98	1.00	0.98	0.93	1.00	0.93	0.96	0.72	0.87	0.72
above	4	0.95	1.00	0.95	0.88	1.00	0.88	0.92	1.19	1.13	1.19
across	3	0.98	0.93	0.96	-0.01	-0.03	0.48	0.23	0.00	0.15	0.41
after	6	1.00	0.98	0.98	1.00	0.96	0.96	0.97	1.43	1.43	1.56
against	5	0.98	0.98	1.00	0.96	0.96	1.00	0.97	1.00	1.13	1.13
along	3	0.93	0.98	0.96	0.37	0.85	0.48	0.61	0.50	0.15	0.41
among	5	0.96	0.96	0.91	0.93	0.93	0.87	0.91	1.70	1.79	1.70
(a)round	5	0.87	0.67	0.71	0.77	0.37	0.44	0.54	1.39	1.46	1.21
as	2	0.89	0.89	0.91	0.23	0.38	0.45	0.36	0.43	0.35	0.50
at	10	0.87	0.93	0.89	0.82	0.91	0.85	0.86	2.19	2.39	2.30
before	4	0.98	0.98	1.00	0.96	0.96	1.00	0.98	1.46	1.44	1.44
behind	5	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
beneath	3	0.96	0.98	0.98	0.48	0.79	0.79	0.70	0.26	0.26	0.35
beside	3	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
between	7	0.96	0.96	0.96	0.93	0.93	0.93	0.93	1.55	1.60	1.54
by	16	0.98	0.93	0.96	0.97	0.91	0.94	0.94	2.23	2.32	2.41
down	4	0.93	0.82	0.84	0.89	0.69	0.73	0.77	1.34	1.39	1.31
during	1	1.00	1.00	1.00	*	*	*	*	0.00	0.00	0.00
for	17	0.89	0.71	0.71	0.87	0.66	0.67	0.73	2.58	2.99	3.22
from	16	0.93	0.80	0.80	0.91	0.75	0.75	0.8	2.69	2.64	2.88
in	23	0.78	0.62	0.60	0.75	0.57	0.55	0.63	3.53	3.53	3.54
inside	5	0.96	0.91	0.91	0.93	0.86	0.86	0.89	1.62	1.62	1.62
into	9	0.84	0.82	0.87	0.59	0.55	0.71	0.62	0.95	1.42	1.50
like	5	0.84	0.78	0.82	0.60	0.44	0.62	0.56	0.93	1.31	1.31
of	26	0.69	0.71	0.69	0.63	0.64	0.63	0.64	3.04	3.53	3.03
off	6	0.98	0.96	0.93	0.91	0.82	0.74	0.83	0.76	0.87	0.76
on	19	0.84	0.82	0.84	0.81	0.78	0.81	0.8	2.76	2.88	2.87
onto	3	0.98	0.98	1.00	-0.01	-0.01	*	-0.01	0.15	0.00	0.00
over	12	0.93	0.87	0.89	0.91	0.83	0.85	0.86	2.41	2.29	2.36
through	7	0.93	0.98	0.91	0.74	0.91	0.68	0.78	0.76	0.76	0.91
to	14	0.84	0.87	0.96	0.79	0.82	0.94	0.85	2.38	2.54	2.47
towards	7	0.93	0.89	0.91	0.85	0.75	0.79	0.80	1.28	1.15	1.27
with	15	0.98	0.93	0.91	0.97	0.91	0.88	0.92	2.14	2.16	2.32

Table 4.3: Preposition sense refinement agreement results after revisions. Agree indicates the plain agreement and κ indicates Fleiss' kappa agreement.

4.2.4 Results Discussion

It is clear from looking at the results that the Kappa scores can differ substantially from the simple agreement scores, particularly for heavily lopsided distributions. For example, Annotator C agreed with Annotators A and B on 91% of the examples for 'along', but the corresponding Kappa score was only 0.16 because there were three disagreements (two for cases where 'along' was actually an adverb, not a preposition). Since A and B each annotated 44 of the 45 examples as sense 1 and Annotator C annotated 42 examples as sense 1, the probability of chance agreement was very high and, as such, the disagreements had a substantially larger effect on the Kappa figure.

Annotators A and B tended to agree more often than A and C or B and C. In fact, in the unrevised results, there are only six cases where the A versus B agreement figure is not the highest or tied for the highest. This perhaps is not be too surprising given that Annotator A used Annotator B's feedback to refine the definitions and Annotator C was only brought in at the end. In the original results, Annotators A and B agreed 1314 out of 1485 times, A and C 1204 out of 1485 times, and B and C 1192 out of 1485 times. In only 41 instances, or 2.8%, did all three annotators disagree.

During annotation revision process, the agreement levels increased substantially. Annotator A changed 66 annotations, Annotator B changed 112 annotations, and Annotator C changed 143 annotations. After these changes, Annotators A and B were in agreement 1374 out of 1485 times, A and C 1333 out of 1485 times, and B and C 1337 out of 1485

times. Moreover, there were only 7 instances for which all three annotators continued to disagree.

In most of the cases, the percentage agreement is over 80%, though the level of agreement varies somewhat from preposition to preposition. Not surprisingly, the prepositions with more senses, including *of* and *in*, generally have lower agreement levels than those with fewer senses. There are a handful of prepositions for which agreement with Annotator C was substantially lower than the agreement level between Annotators A and B, with the greatest drops occurring for *(a)round*, *at*, *by*, *for*, *from*, and *off*. These agreement gaps were substantially reduced during the annotation revision phase, with the gaps for *at*, *by*, and *off* more or less disappearing.

In a few cases, it would be possible to combine certain definitions together to get a higher agreement figure. For example, there are several senses of *by* that cover specific types of *means* (e.g., means of transport) that could be collapsed into the general *means* category, by_2. There are many opportunities for doing this; however, there are almost no cases where merging only two similar senses together would produce any sort of substantial improvement in the agreement figures for any of the prepositions. Conceivably, the two most commonly disagreed upon definitions—*positioned, extending, or moving so as to form a curved path or arc* and *in, to, and/or through many places throughout (an area or space)*, both for the preposition *(a)round*—could be combined into some sort of vague *spatial* category.

One source of disagreement that cut across several prepositions is the *location* vs *destination* distinction. For the example *If an insect walks by, she grabs it and instantly drags it *inside* the tunnel*, one annotator chose *in; situated within the confines of (something)* instead of *into; moving or extending so as to become/end within (something)*, and, for the example *The iodine-131 deposited *on* the ground would decay in a few weeks*, two annotators chose *onto; indicating where (or against what) contact is made* whereas the other chose *physically in contact with and supported by (a surface)*.

4.2.5 Alignment with Noun Compound Relations

Prepositions have sometimes been used to paraphrase the relations between nouns in noun compounds. Work by Levi (1978) and Lauer (1995) are classic examples of this. However, prepositions can be highly ambiguous, which is both one of the motivations for this preposition sense disambiguation work as well as one of the reasons for not using plain prepositions in Chapter 3 for interpreting noun compounds. Ideally, there should be some way to map between the different relations, and Table 4.4 shows mappings from the noun compound relations presented in Chapter 3 to the most relevant fine-grain preposition senses created in this work.

As shown in Table 4.4, there are several preposition senses that are spread across many different noun compound relations, with OF_4 being the most obvious case. OF_4 covers cases where the modifier is the OBJECT of the verb that the head noun is derived from. As discussed in Section 3.2.4, an OBJECT relation tends to be ambiguous with many other

relations (e.g., *soup pot* → *containment* but *soup container* → *containment/object*). Warren (1978) went so far as to completely ignore such examples in her analysis. One problem with excluding an *object* sense is that doing so can lead to a proliferation of senses. Indeed, a few of the noun compound senses such as MODIFY/PROCESS/CHANGE, PROPEL, and TRAVERSE/VISIT probably would not have been created if the OBJECT cases had been excluded or placed in their own category. Similarly, excluding OF_4 from the preposition senses would have resulted in additional senses of *of*.

To allow for a better mapping between prepositions and noun compound relations, I created a revised noun compound relation inventory (shown in Table 4.5) and dataset that maps more directly to the refined preposition senses. The largest change is the introduction of an *object* category that maps directly to OF_4. This change caused several relations including MODIFY/PROCESS/CHANGE, PROTECT/CONSERVE, PROPEL, and TRAVERSE/VISIT to lose most of their annotated examples, and, thus, these relations were removed. Some of the remaining examples for these categories were moved to a new category, FOR (VAGUE PURPOSE). There are also a few new categories introduced to allow for better mappings for a couple of preposition senses, including VARIETY/GENUS-OF, which took its examples from the COLLECTION/CONFIGURATION/SERIES category, JUSTIFICATION, which took its examples from the *cause* portion of CREATOR/PROVIDER/CAUSE OF, and AMOUNT-OF, which mostly took some arguably mislabeled examples from the TOPIC/THING + ATTRIBUTE category. A group of adjective-like

nouns (e.g., core, chief, key, mainstream) were split off from the EQUATIVE category (previously named COREFERENTIAL) to form a new category. Also, the TOPIC/THING + ATTRIBUTE category was folded into a new, somewhat larger category, OTHER 'OF' NOUN COMPLEMENT, which was designed to map to $\text{OF}_1 8$, which includes attribute/property words along with many somewhat idiosyncratic cases.

Category Name	Approximate Mappings
Causal Group	
COMMUNICATOR OF COMMUNICATION	$BY_{1,1.1}$ $FROM_8$ OF_5
PERFORMER OF ACT/ACTIVITY	BY_1 $FROM_8$ OF_5
CREATOR/PROVIDER/CAUSE OF	$BY_{1.1}$ FOR_{14} $FROM_9$
Purpose/Activity Group	
PERFORM/ENGAGE_IN	FOR_3
CREATE/PROVIDE/SELL	FOR_{16} OF_4
OBTAIN/ACCESS/SEEK	FOR_4 OF_4
MODIFY/PROCESS/CHANGE	FOR_{16} OF_4 ON_{11}
MITIGATE/OPPOSE/DESTROY	$AGAINST_1$ FOR_2 OF_4 ON_{11}
ORGANIZE/SUPERVISE/AUTHORITY	$OF_{4,18}$ ON_3 $OVER_{10}$
PROPEL	FOR_{16} OF_4
PROTECT/CONSERVE	FOR_{16} OF_4
TRANSPORT/TRANSFER/TRADE	FOR_{16} OF_4
TRAVERSE/VISIT	OF_4
Ownership, Experience, Employment, Use	
POSSESSOR + OWNED/POSSESSED	OF_{12}
EXPERIENCER + COGINITION/MENTAL	OF_{11}
EMPLOYER + EMPLOYEE/VOLUNTEER	FOR_{11} $IN_{14.1}$ OF_1 $WITH_3$
CONSUMER + CONSUMED	FOR_9
USER/RECIPIENT + USED/RECEIVED	FOR_9 $TO_{4.2}$
OWNED/POSSESSED + POSSESSION	OF_4 $WITH_7$
EXPERIENCE + EXPERIENCER	$OF_{4,18}$ $WITH_{7.2}$
THING CONSUMED + CONSUMER	OF_4
THING/MEANS USED + USER	$BY_{2,2.2,2.3}$ ON_8 $OVER_6$ $THROUGH_4$ $WITH_4$
Temporal Group	
TIME [SPAN] + X	AT_3 $DURING_1$ IN_8 OF_9 ON_5
X + TIME [SPAN]	OF_{18}
Location and Whole+Part/Member of	
LOCATION/GEOGRAPHIC SCOPE OF X	AT_1 BY_6 $FROM_{1,4}$ IN_1 $INSIDE_1$ OF_{10} $ON_{1,1.4}$
WHOLE + PART/MEMBER OF	OF_1
Composition and Containment Group	
SUBST/MATERIAL/INGREDIENT + WHOLE	$FROM_{10}$ $OF_{2.3}$
PART/MEMBER + COLLECT/CONFIG/SERIES	$OF_{2.2,2.4}$
X + SPATIAL CONTAINER/LOC/BOUNDS	FOR_1 $OF_{3,4}$ $WITH_7$
Topic Group	
TOPIC OF COMMUNICATION/IMAGERY/INFO	$ABOUT_1$ OF_4 ON_3 $OVER_8$
TOPIC OF PLAN/DEAL/ARRANGE/RULES	$ABOUT_1$ $OF_{4,18}$ ON_3 $OVER_8$
TOPIC OF OBSERVATION/STUDY/EVAL	$INTO_3$ OF_4
TOPIC OF COGNITION/EMOTION	$ABOUT_1$ AT_5 OF_4 ON_3 $OVER_8$ $TOWARDS_3$ $WITH_{10.1}$
TOPIC OF EXPERT	$ABOUT_1$ AT_6 IN_{15} ON_3 $WITH_{10.3}$
TOPIC OF SITUATION	$ABOUT_1$ OF_4 $OVER_8$ $WITH_{10.3}$
TOPIC OF EVENT/PROCESS	$OF_{4,5}$
Attribute Group	
TOPIC/THING + ATTRIB	$OF_{2.1,18,18.1}$
TOPIC/THING + ATTRIB VALUE CHARAC OF	OF_{18}
Attributive and Coreferential	
COREFERENTIAL	
PARTIAL ATTRIBUTE TRANSFER	$LIKE_4$
MEASURE + WHOLE	BY_3 $FOR_{12.1,12.2}$
Other	
HIGHLY LEXICALIZED / FIXED PAIR	
OTHER	Various

Table 4.4: Mappings of the noun compound relations discussed in Chapter 3 to the most relevant preposition senses.

Category Name	% Example	Approximate Mappings
Objective		
OBJECTIVE	17.26 leaf blower	$\approx$OF$_4$
Doer-Cause-Means		
SUBJECTIVE	3.56 police abuse	$\approx$BY$_1$ $\subset$FROM$_8$ $\approx$OF$_5$
CREATOR/PROVIDER/CAUSE OF	1.54 ad revenue	$\subset$BY$_{1.1}$ $\approx$OF$_{5.1}$ $\approx$FROM$_9$
JUSTIFICATION	0.26 murder arrest	$\approx$FOR$_{14}$
MEANS + USE(R)	1.51 faith healer	$\approx$BY$_{2\cup2.2\cup2.3}$ $\subset$ON$_8$ $\subset$OVER$_6$ $\subset$THROUGH$_4$ $\approx$WITH$_4$
Purpose/Activity Group		
PERFORM/ENGAGE_IN	11.57 cooking pot	$\subset$FOR$_3$
CREATE/PROVIDE/SELL	4.87 nicotine patch	$\supset$FOR$_{16}$
OBTAIN/ACCESS/SEEK	0.87 shrimp boat	$\subset$FOR$_4$
MITIGATE/OPPOSE	0.79 flak jacket	$\approx$AGAINST$_1$ $\approx$FOR$_2$ $\propto$ON$_{11}$
ORGANIZE/SUPERVISE/AUTHORITY	1.61 ethics authority	$\propto$ON$_3$ $\propto$OVER$_{10}$
PURPOSE (VAGUE)	1.96 chicken spit	$\supset$FOR$_{16}$
Ownership, Experience, Employment, Use		
POSSESSOR + OWNED/CONTROLLED	2.14 family estate	$\approx$OF$_{12}$
EXPERIENCER + COGINITION/MENTAL	0.53 family greed	$\approx$OF$_{11}$ $\approx$WITH$_{7.2}$
EMPLOYER + EMPLOYEE/VOLUNTEER	2.34 team doctor	$\approx$FOR$_{11}$ $\propto$IN$_{14.1}$ $\propto$OF$_1$ $\propto$WITH$_3$
USER/RECIPIENT + USED/RECEIVED	1.02 voter pamphlet	$\approx$FOR$_9$ $\subset$TO$_{4.2}$
Temporal Group		
TIME [SPAN] + X	2.19 night work	$\subset$AT$_3$ $\approx$DURING$_1$ $\approx$IN$_8$ $\approx$OF$_9$ $\subset$ON$_5$
X + TIME [SPAN]	0.48 birth date	$\supset$OF$_{18}$
Location and Whole+Part/Member of		
LOCATION/GEOGRAPHIC SCOPE OF X	5.20 hillside home	$\subset$AT$_1$ $\subset$BY$_6$ $\subset$FROM$_{1,4}$ $\subset$IN$_1$ $\subset$INSIDE$_1$ $\subset$OF$_{10}$ $\subset$ON$_{1,1.4}$
WHOLE + PART/MEMBER OF	1.70 robot arm	$\approx$OF$_1$
Composition and Containment Group		
X + SPATIAL CONTAINER/LOC/BOUNDS	1.42 shoe box	$\subset$FOR$_1$ $\subset$OF$_3$ $\propto$WITH$_7$
SUBST/MATERIAL/INGREDIENT + WHOLE	2.66 plastic bag	$\subset$FROM$_{10}$ $\approx$OF$_{2.3}$
PART/MEMBER + COLLECT/CONFIG/SERIES	1.79 truck convoy	$\approx$OF$_{2.2}$
VARIETY/GENUS-OF	0.11 plant species	$\approx$OF$_{2.4}$
AMOUNT-OF	0.88 traffic volume	$\approx$OF$_{2.1}$
Topic Group		
TOPIC (GENERAL)	6.87 travel story	$\propto$ABOUT$_1$ $\propto$ON$_3$ $\propto$OVER$_8$
TOPIC OF COGNITION/EMOTION	0.32 auto fanatic	$\propto$ABOUT$_1$ $\subset$AT$_5$ $\propto$ON$_3$ $\propto$OVER$_8$ $\subset$TOWARDS$_3$ $\subset$WITH$_{10.1}$
TOPIC OF EXPERT	0.68 policy expert	$\propto$ABOUT$_1$ $\subset$AT$_6$ $\propto$IN$_{15}$ $\propto$ON$_3$ $\propto$WITH$_{10.3}$
Other Complements Group		
OTHER 'OF' NOUN COMPLEMENT	5.93 eye shape	$\subset$OF$_{18\cup18.1}$
TOPIC/THING + ATTRIB VALUE CHARAC OF	0.29 earth tone	$\subset$OF$_{18}$
Attributive and Equative		
EQUATIVE	5.47 fighter plane	
ADJ-LIKE NOUN	1.34 core mission	
PARTIAL ATTRIBUTE TRANSFER	0.35 skeleton crew	$\approx$LIKE$_4$
MEASURE + WHOLE	4.25 hour meeting	$\subset$BY$_3$ $\subset$FOR$_{12.1\cup12.2}$
Other		
HIGHLY LEXICALIZED / FIXED PAIR	0.78 pig iron	
OTHER	4.21 contact lense	
OTHER PREP SENSE	1.23 volume leader	Various

Table 4.5: Modified noun compound relations with preposition sense mappings. $\approx$-approximately equivalent; $\supset$/$\subset$-super/sub set; $\propto$-some overlap; $\cup$-union.

4.3 System Description

4.3.1 Models

Unlike the previous version of the system described by Hovy et al. (2010), which used Maximum Entropy models (Berger et al., 1996) trained using the MALLET toolkit (McCallum, 2002), the current system uses linear SVMs built in the one-against-the-rest fashion using the LIBLINEAR toolkit (Fan et al., 2008). The reason for this change is that the linear SVMs train much faster than the Maximum Entropy models but are just as accurate. Because the senses for a given preposition are often quite distinct, the system builds separate models for each preposition.

4.3.2 Feature Generation

The feature generation process used in the current system is very similar to that described by Hovy et al. (2010), but there are a few differences, some of them significant.

The most significant difference is the use the part-of-speech tagger and parser described in Chapters 8 and 7, respectively, instead of the Shen et al. (2007) part-of-speech tagger and the MALTPARSER (Nivre et al., 2006). Since most of the feature generation process relies in some way or another on part-of-speech tagging or parsing, the more accurate these systems are, the better. Early experiments showed that these simple changes accounted for as much as a 0.5% improvement in the overall accuracy of the system.

The most notable difference in the features used is the introduction of features based upon WordNet's (Fellbaum, 1998) 'derivationally-related' links. These links relate words belonging to different parts-of-speech where one word is derived from the other. For example, from the adjective 'aware', there is a link to 'awareness' and, from the verb 'detain', there is a link to 'detention.' Adding these features improves system accuracy by about .3%, as shown later.

There are a handful of other, more minor differences including some updates and bug fixes to some of the heuristics as well as the removal of features that weren't helpful, including some based upon other WordNet link types such a holonym links.

The system uses a two-step process to generate the features for training and testing. In the first step, word selection rules identify words from the surrounding context from which to generate features. The second step involves applying a variety of different feature-generating functions to these words. It is worth noting that some words may be selected by multiple word selection rules. For example, the *First probable governor to the left* and the *First word to the left* rule often independently identify the same word. Since the word selection rule is considered part of the feature identifier, separate features are created for these cases. All the features are boolean (e.g., 'lemma of *word to left* = jump' $\rightarrow$ true/false).

The five different word selection rules are listed below.

Word Selection Rules

- Heuristically-determined complement of the preposition (cf. Algorithm 1)
- First verb to the left
- First probable governor (adjective, number, noun, pronoun, or verb) to the left
- Union of (First verb to the left, First probable govenor to the left)
- First word to the left

The *Heuristically-determined complement of the preposition* is somewhat more complicated than the other rules. A simple first-noun-to-the-right rule would fail on more complicated noun phrases, including, on the more difficult end of the scale, examples like *this/DT very/RB prolific/JJ author/NN 's/POS latest/JJS story/NN*. The rule is described in detail in Algorithm 1.

The feature-generating functions, many of which utilize WordNet (Fellbaum, 1998), are listed below. To conserve space, curly braces are used to represent multiple functions in a single line. The name of each feature is the combination of the word-selection rule, the feature-generating function, and the output from the feature-generating function.

input : w_1 ... w_n, #the sentence
$prepIndex$, #the index of the preposition
output: t #the selected token
t = null;
$finished$ = false;
$index$ = $prepIndex$ + 1;
while $index$ <= n *and* $finished$ == *false* **do**
  pos = `getPartOfSpeech(`w_{index}`);`
  if t *!= null* **then**
    if pos == *POS* **then**
      t := null;
    else if $pos \in$ *{NN,NNS,NNP,NNPS} or (*`getPartOfSpeech(`t`)` $\notin$ *{NN,NNS,NNP,NNPS} and* $pos \in$ *{RB,RBR,RBS,JJ,JJR,JJS,CD})* **then**
      t := w_{index};
    else
      $finished$:= true;
  else
    if $pos \in$ *{VBG,PRP,WP}* **then**
      t := w_{index};
      $finished$:= true;
    else
      $text$:= `getTokenText(`w_{index}`);`
      if $text \in$ *{this,these,those,that,all,some,few,each,another,more,most,less} or* $pos \in$ *{NN,NNS,NNP,NNPS,CD}* **then**
        t := w_{index};
        $tPos$:= pos;
  $index$:= $index$ + 1;
return t

Algorithm 1: Prepositional complement selection heuristic.

WordNet-based Features

- Hypernyms for {1st, all} sense(s) of the word
- Synonyms for 1st sense of the word
- Synonyms and derivationally-related words for all senses of the word
- All terms in the definitions ('glosses') of the word
- Lexicographer file names for the word
- Lists of all link types (e.g., meronym links) associated with the word
- Part-of-speech indicators for the existence of noun/verb/adjective/adverb entries for the word

Other Features

- Indicator that the word-finding rule found a word (rule itself)
- Capitalization indicator
- {Lemma, surface form} of the word
- Part-of-speech tag for the word
- General part-of-speech tag for the word (e.g., NNS $\rightarrow$ NN, VBZ $\rightarrow$ VB)
- The {first, last} {two, three} letters of each word
- Indicators for suffix types (e.g., de-adjectival, de-nominal [non]agentive, de-verbal [non]agentive)
- Indicators for a wide variety of other affixes including those related to degree, number, order, etc[4].

4. See Appendix F for the list of affixes.

4.4 System Evaluation

For evaluation, the data split from the SemEval 2007 task #6 was used. This data contains examples of 34 of the most common English prepositions, including 16,557 training examples and 8,096 testing examples. The C parameter for the SVMs, which controls the tradeoff between margin width and classifier error, was set to 0.005 because this performed best using 5-fold cross validation on the training data. The system was run twice, once with the original TPP sense inventory annotations and once with the refined inventory and annotations. The results for these experiments are presented in Table 4.6.

To assess the impact of the different word selection rules and feature-generating rules, the same experiment was rerun multiple times, each time including or excluding exactly one of the rules. For each variation, the C parameter was retuned using 5-fold cross validation on the training data. The results for these runs are shown in Tables 4.7 and 4.8.

			Original Data			Refined Data		
Preposition	#Train	#Test	Senses	Entropy	Accuracy	Senses	Entropy	Accuracy
about	710	364	6	0.705	0.956	3	0.420	0.984
above	48	23	9	1.651	0.739	4	0.559	0.826
across	319	151	3	0.267	0.967	3	0.057	0.993
after	103	53	11	2.246	0.660	6	1.951	0.830
against	195	92	10	1.596	0.902	5	0.988	0.978
along	365	173	4	0.270	0.954	3	0.836	0.896
among	100	50	4	1.630	0.820	5	1.992	0.780
around	335	155	6	2.112	0.697	5	1.147	0.819
as	174	84	2	0.000	1.000	2	0.685	0.952
at	715	367	12	2.222	0.875	10	2.328	0.926
before	47	20	4	1.513	0.900	4	1.513	0.900
behind	138	68	9	1.393	0.809	5	0.411	0.926
beneath	57	28	6	1.522	0.821	3	0.371	0.929
beside	62	29	3	0.000	1.000	3	0.000	1.000
between	211	102	9	1.888	0.961	7	1.295	0.990
by	510	248	22	2.563	0.871	16	2.130	0.867
down	332	153	5	1.268	0.824	4	1.520	0.791
during	81	39	2	0.961	0.846	1	0.000	1.000
for	951	478	15	2.977	0.822	17	3.363	0.814
from	1206	578	16	2.848	0.889	16	2.871	0.841
in	1397	688	15	2.806	0.783	23	3.346	0.786
inside	67	38	5	1.360	0.711	5	1.620	0.632
into	604	297	10	2.146	0.855	9	1.322	0.902
like	266	125	7	1.207	0.912	5	0.495	0.952
of	3004	1478	20	3.198	0.886	26	3.800	0.887
off	161	76	7	1.039	0.868	6	1.245	0.855
on	872	441	25	3.358	0.825	19	2.948	0.853
onto	117	58	3	0.603	0.914	3	0.126	0.983
over	200	98	17	2.721	0.755	12	2.429	0.878
round	181	82	8	2.270	0.707	5	1.340	0.768
through	441	208	16	2.627	0.490	7	0.575	0.962
to	1183	572	17	2.452	0.907	14	2.461	0.897
towards	214	102	6	0.673	0.990	7	0.868	0.951
with	1191	578	18	3.004	0.879	15	2.697	0.905
Overall	16557	8096	332		0.857	278		0.881

Table 4.6: Preposition sense disambiguation results.

	Original Data		Refined Data	
Word-Finding Rule	LOO	OO	LOO	OO
Head to left + Verb to left	0.852 (0.01)	0.826 (0.02)	0.881 (.005)	0.846 (0.02)
Head to left	0.849 (0.01)	0.788 (0.04)	0.878 (.005)	0.817 (0.01)
Verb to left	0.855 (.005)	0.606 (0.01)	0.877 (.005)	0.628 (0.02)
Preposition complement	0.856 (.005)	0.568 (.005)	0.880 (.005)	0.620 (.005)
Word to left	0.850 (0.01)	0.780 (0.02)	0.878 (.005)	0.809 (0.01)
Head word	0.851 (0.01)	0.804 (0.01)	0.879 (0.01)	0.824 (0.01)
Preposition complement	0.857 (.005)	0.560 (0.01)	0.877 (.005)	0.624 (.005)

Table 4.7: Leave-one-out (LOO) and only-one (OO) results for the word selection rules using the original The Preposition Project annotations. The C parameter used for training the SVMs is shown in parentheses.

	Original Data		Refined Data	
Feature Rule	LOO	OO	LOO	OO
Rule Itself	0.856 (0.01)	0.402 (0.08)	0.880 (.005)	0.470 (0.16)
Lowercase Word	0.856 (0.01)	0.730 (0.64)	0.881 (.005)	0.755 (0.32)
Lemma Form	0.856 (0.01)	0.805 (0.64)	0.880 (.005)	0.828 (0.64)
Part-of-Speech	0.855 (0.01)	0.520 (0.08)	0.880 (0.01)	0.574 (0.16)
General Part-of-Speech	0.857 (0.01)	0.469 (0.16)	0.880 (.005)	0.536 (0.08)
Capitalization	0.856 (0.01)	0.201 (0.08)	0.880 (.005)	0.209 (0.16)
Affixes	0.855 (0.01)	0.424 (0.08)	0.879 (.005)	0.469 (0.08)
Lexicographer Filenames	0.855 (.005)	0.680 (0.04)	0.881 (0.01)	0.730 (0.08)
1st Sense Hypernyms	0.857 (0.01)	0.814 (0.02)	0.880 (.005)	0.840 (0.08)
All Hypernyms	0.856 (0.01)	0.816 (0.02)	0.880 (.005)	0.842 (0.08)
1st Sense Synonyms	0.856 (0.01)	0.810 (0.08)	0.880 (.005)	0.834 (0.08)
All Synonyms	0.855 (0.01)	0.810 (0.08)	0.880 (.005)	0.834 (0.08)
Derivationally-related	0.854 (0.01)	0.781 (0.08)	0.881 (0.01)	0.806 (0.04)
Gloss Terms	0.854 (0.02)	0.820 (0.02)	0.882 (0.01)	0.847 (0.02)

Table 4.8: Leave-one-out (LOO) and only-one (OO) results for the feature-generating rules on the original The Preposition Project annotations. The C parameter used for training the SVMs is shown in parentheses.

4.4.1 Results Discussion

The system achieved 85.7% on the original TPP data and 88.1% on the refined version. The 85.7% figure is a statistically-significant increase over the 84.8% result achieved by an earlier version of the system described by Hovy et al. (2010). For most prepositions, the accuracy figures for the refined version were higher than the those for the original version. This can largely be explained by the fact that the refined version has fewer overall senses overall and, for most prepositions, the entropy of the class distribution of the test examples is lower. Indeed, in ten of the eleven cases where the accuracy was lower for the refined data, the refined data had a higher entropy distribution. There are, of course, a few deviations from this rule; the entropy and the accuracy both increased for *of*, *at*, and *in* and the entropy and accuracy both decreased for *by*.

The *leave-one-out* and *only-one* experiments produced a variety of interesting results. Clearly, the *preposition complement* rules are less important than the rules designed to locate the head/governor of the preposition, with the union of the *head to left* and *verb to left* heuristics appearing to be the single most useful word selection rule overall. The use of the words from the WordNet glosses (definitions) are very effective, about as useful as the hypernyms. And, it appears that they may provide some information that the other rules do not—leaving them out results in a drop of 0.3% accuracy. A similar drop occurs if the 'derivationally-related word' features are removed.

4.5 Conclusion

In this chapter, I described work to refine the sense inventory used by The Preposition Project, the associated inter-annotator agreement study, the current version of the state-of-the-art preposition sense disambiguation system previously described by Tratz and Hovy (2009) and Hovy et al. (2010), and the results of applying the system to the SemEval 2007 Task #6 fine-grained preposition sense disambiguation task. The sense refinement work produced a set of 278 preposition senses that covers the 34 most frequent English prepositions. The state-of-the-art automatic preposition sense disambiguation system achieves 85.7% on the SemEval data, which is over 16% more accurate than the highest performing SemEval participant and 1.2% more accurate than the system described by Hovy et al. (2010). Using the refined sense inventory and annotations results in even higher accuracy, 88.1%.

Bibliography

Adam L. Berger, Vincent J. Della Pietra, and Stephen A. Della Pietra. 1996. A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics*, 22(1):39–71.

Marine Carpuat and Dekai Wu. 2007. Improving Statistical Machine Translation using Word Sense Disambiguation. In *The 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 61–72.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word Sense Disambiguation Improves Statistical Machine Translation. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*.

Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A Library for Large Linear Classification. *Journal of Machine Learning Research*, 9:1871–1874.

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge, MA.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Dirk Hovy, Stephen Tratz, and Eduard Hovy. 2010. What's in a Preposition?—Dimensions of Sense Disambiguation for an Interesting Word Class. In *Proceedings of COLING*.

Anna Korhonen and Judita Preiss. 2003. Improving Subcategorization Acquisition using Word Sense Disambiguation. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*, pages 48–55. ACL.

Mark Lauer. 1995. Corpus Statistics Meet the Noun Compound: Some Empirical Results. In *Proceedings of the 33rd Annual Meeting on Association for Computational Linguistics*, pages 47–54. ACL.

Judith N. Levi. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.

Ken Litkowski and Orin Hargraves. 2005. The Preposition Project. In *Proceedings of the Second ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, pages 171–179. ACL.

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*, pages 24–29. ACL.

Mitchell P. Marcus, Mary A. Marcinkiewicz, and Beatrice Santorini. 1993. Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):330.

Andrew K. McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. *URL http://mallet.cs.umass.edu*.

Joakim Nivre, Johan Hall, and Jens Nilsson. 2006. MaltParser: A Data-Driven Parser-Generator for Dependency Parsing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Tom O'Hara and Janyce Wiebe. 2003. Preposition Semantic Classification via Penn Treebank and FrameNet. In *Proceedings of the 7th Conference on Natural Language Learning at HLT-NAACL 2003*, pages 79–86. ACL.

Laura Plaza, Mark Stevenson, and Alberto Díaz. 2010. Improving Summarization of Biomedical Documents using Word Sense Disambiguation. In *Proceedings of the 2010 Workshop on Biomedical Natural Language Processing*, pages 55–63. ACL.

Vassiliki Rentoumi, George Giannakopoulos, Vangelis Karkaletsis, and George A. Vouros. 2009. Sentiment Analysis of Figurative Language using a Word Sense Disambiguation Approach. In *Proceedings of the International Conference RANLP*, pages 370–375.

Libin Shen, Giorgio Satta, and Aravind K. Joshi. 2007. Guided Learning for Bidirectional Sequence Classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*.

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

Beatrice Warren. 1978. *Semantic patterns of noun-noun compounds*. Acta Universitatis Göthoburgensis, Göteborg, Sweden.

Appendix A

Preposition Definitions

This appendix contains the refined preposition definitions defined by the work described in Chapter 4. Each definition comes with a mapping to the version of The Preposition Project (Litkowski and Hargraves, 2005, TPP) used in the SemEval 2007 Preposition Sense Disambiguation Task (Litkowski and Hargraves, 2007). The following symbols are used are part of the TPP mappings: $\approx$ – approximately equivalent, $\cup$ – union, $\supset$ – superset, and $\subset$ – subset. If a $\subset$ symbol appears as a subscript, it indicates a subset of the relevant sense, as though the $\subset$ symbol in this case was a unary operator. Thus, a mapping like $\approx$ 1(1)$\cup$5(7b)$_{\subset}$ indicates that the sense is a approximately equivalent to the union of what is covered by TPP sense 1(1) and a portion of what is covered by TPP sense 5(7b).

ABOUT

1) on the subject of; concerning (as of communication, thoughts, information, fights, etc.)

TPP mapping: $\approx$ 1(1)

1.1) so as to affect, as in “do something about it”

TPP mapping: ≈ 2(1a)

2) indicating a focal point or region in space; partially, if not fully, encircling

TPP mapping: ≈ 3(2)∪3(2)-1∪4(3)

ABOVE

1) at, in, or to the extended space vertically higher than

TPP mapping: ≈ 1(1)∪2(1a)∪3(1b)∪4(2)

2) higher/greater than along a particular (but, perhaps, unmentioned) scale.

TPP mapping: ≈ 5(2a)∪6(2b)∪9(3)

2.1) higher/greater than along the scale of desirability/preference.

TPP mapping: ≈ 7(2c)

3) in spite of / at a higher pitch or volume than (as in the case of interfering noise)

TPP mapping: ≈ 8(2d)

ACROSS

1) from one side of to another. (e.g., he ran across the bridge)

TPP mapping: ≈ 1(1)

2) indicating a time period through/over which something happens or remains unchanged (e.g., changed across the centuries)

TPP mapping: ≈ 1(1)-1

3) located on/at the other side of (e.g., the house across the river)

TPP mapping: $\approx$ 2(2)

AFTER

1) following in time (and, in some cases, also caused by) (e.g., happen after the wedding)

TPP mapping: $\approx$ 1(1)$\cup$1(1)-1$\cup$3(1b)$\cup$4(1c)

2) moving so as to follow or intercept; in the direction of someone/something that is moving further away (e.g., walked after him)

TPP mapping: $\approx$ 5(2)$\cup$6(2a)$\cup$7(3)$_{\subset}$

3) indicating the subject of a desire or inquiry (e.g., inquired after the stolen diamonds)

TPP mapping: $\supset$ 7(3)

4) indicating a base model (as of design or imitation) or namesake (e.g., designed after the Taj Mahal)

TPP mapping: $\approx$ 9(5)$\cup$10(5a)

5) at a subsequent position in a sequence or ranking (e.g., B is after A)

TPP mapping: $\approx$ 8(4)

6) used in phrases indicating repetition or continuation (e.g., night after night)

TPP mapping: $\approx$ 2(1a)

AGAINST

1) indicating an opposed, mitigated, or disfavored party/thing, including conceptual abstractions (e.g., argued against it)

TPP mapping: ≈ 1(1)∪2(1a)∪3(1b)∪4(2)∪5(2a)∪6(2b)

2) indicating opposition by physical forces, usually involving contact (e.g., the broom against the wall)

TPP mapping: ≈ 10(4)

3) in relation to (an amount of money owed, due, or lent) so as to reduce, cancel, or secure it (e.g., paid money against his debts)

TPP mapping: ≈ 7(2c)

4) *in contrast to*

4.1) in conceptual contrast to (e.g., must compare it against the alternative)

TPP mapping: ≈ 8(3)

4.2) in visual contrast to (e.g., the yellow stood out against the black background)

TPP mapping: ≈ 9(3a)

ALONG

1) positioned, extending, traversing, or moving in parallel with some sort of path

TPP mapping: ≈ 1(1)∪3(2)

2) used to indicate a process that progress is being made on

TPP mapping: ≈ 2(1a)

3) in adherence/accord with

TPP mapping: $\approx$ 4(3)

AMONG

1) situated more or less centrally in relation to (several other things); surrounded by (e.g., walk among the trees)

TPP mapping: $\approx$ 1(1)

2) expressing membership in a group (e.g., foremost among them)

TPP mapping: $\approx$ 2(2)

3) indicating connection(s), interaction(s), or relationship(s) involving two or members of a group or community (e.g., fight among the miners)

TPP mapping: $\supset$ 3(3)

4) shared or possessed by (a group or community) (e.g., thoughts among dockworkers)

TPP mapping: $\supset$ 3(3)

5) indicating a set of two or more people/things in the context of selection, division, and/or differentiation (e.g., choose among them, split it among them)

TPP mapping: $\approx$ 4(4)

(A)ROUND

1) positioned, extending, or moving so as to form a curved path or arc (e.g., he ran around the corner)

TPP mapping: ≈ 1(1)∪4(3)∪5(4) (for 'around')

TPP mapping: ≈ 1(1)∪3(2)∪4(2a)∪5(3) (for 'round')

2) located at a position on the other side of something (e.g., the building around the corner)

TPP mapping: ⊃ 3(2) (for 'around')

TPP mapping: ⊃ 8(4) (for 'round')

3) in, to, and/or through many places throughout (an area or space)

TPP mapping: ⊃ 3(2) (for 'around')

TPP mapping: ⊃ 8(4) (for 'round')

4) about; on the subject of

TPP mapping: see also 2(1a)

TPP mapping: see also 2(1a)

5) used figuratively to indicate a challenge, obstacle, or problem, as with avoidance or circumvention (e.g., get around that problem)

TPP mapping: see also 2(1a)

TPP mapping: see also 2(1a)

AS

1) indicating the function or character someone/something has or is supposed/assigned/believed to have

TPP mapping: $\supset$ 1(1)

2) while acting in (the function/role specified) / during the time while being (the function/role/thing specified)

TPP mapping: $\subset$ 2(2)

AT

1) expressing location or arrival at a particular place or position (e.g., he is at the store)

TPP mapping: $\approx$ 1(1),4(2b)

2) in the direction of (e.g., threw coins at the bottle)

TPP mapping: $\supset$ 9(5)

2.1) in the direction of , specifically in the case of repetitive (and perhaps unsuccessful) actions (e.g., claw at the portrait)

TPP mapping: $\approx$ 10(5a)

3) expressing the time when an event takes place (e.g., happened at night)

TPP mapping: $\approx$ 2(2)$\cup$3(2a)

4) indicating a position on a scale (e.g., room was at 100 degrees)

TPP mapping: $\approx$ 5(3)

4.1) indicating the age of someone/something (e.g., was foolish at 17)

TPP mapping: $\approx$ 6(3a)

5) indicating the subject matter (often one that provokes a mental reaction or an expression indicative thereof) (e.g., angry at Joe)

TPP mapping: $\approx$ 11(6)-1

6) expressing a relationship between an individual and a skill or a level of skill and the skill itself (e.g., good at singing)

TPP mapping: $\approx$ 8(4a)

7) indicating the target or goal of an attempt, action, thought, or plan (e.g., attempts as unification)

TPP mapping: $\supset$ 9(5)

8) expressing a state, condition, or circumstance (often somewhat idiomatic) (e.g., they are at odds)

TPP mapping: $\approx$ 7(4)$\cup$11(6)

BEFORE

1) preceding in time (e.g., snack before bedtime)

TPP mapping: $\approx$ 1(1)

2) in front of (e.g., the food before her)

TPP mapping: $\approx$ 2(2)

2.1) in front of, specifically in situations of judgment or examination (e.g., brought before the judge)

TPP mapping: $\approx$ 3(2a)

3) in an earlier position in a sequence or ranking; in preference to

TPP mapping: $\approx$ 4(3)

BEHIND

1) at or to the far/rear side of

TPP mapping: $\approx$ 1(1)$\cup$3(2)$\cup$4(2a)

2) influencing, responsible for, or underlying (as of reasons or purposes) but not necessarily readily obvious to an observer

TPP mapping: $\approx$ 2(1a)$\cup$6(3a)

3) in support of or giving guidance to (someone/something else)

TPP mapping: $\approx$ 5(3)

4) less advanced than (someone or something) in terms of achievement or development

TPP mapping: $\approx$ 8(5)

5) having a lower score than (another competitor)

TPP mapping: $\approx$ 9(6)

BENEATH

1) at, to, or through an area or location positioned vertically lower than

   TPP mapping: ≈ 1(1)∪2(1a)∪3(2)

2) lower in status or rank than

   TPP mapping: ≈ 4(2a)∪5(2b)

3) not obvious, as if hidden behind an appearance metaphorically

   TPP mapping: ≈ 6(2c)

BESIDE

1) indicating a location at the side of / next to

   TPP mapping: ≈ 1(1)

2) in comparison with , as though next to each other so as to make differences apparent

   TPP mapping: ≈ 2(1a)

3) other than; in addition to

   TPP mapping: ≈ 3(2)

BETWEEN

1) in, into, or across the space separating (two objects or regions) (e.g., between the house and the garage)

   TPP mapping: ≈ 1(1)

2) in the period separating (two points in time) (e.g., between May and July)

TPP mapping: ≈ 2(2)

3) in the interval separating (two points on a scale / in a series) (e.g., between N and Z)

TPP mapping: ≈ 3(3)

4) indicating a connection, interaction, or relationship involving two or more things (e.g., discussion between the two factions)

TPP mapping: ≈ 4(4)∪5(4a)

4.1) with reference to decision-making processes involving two or more things being considered together/against each other, as with choices or the discernment of differences (e.g., choose between them)

TPP mapping: ≈ 7(4c)

4.2) with reference to a contrast or failure to correspond (e.g., the difference between them)

TPP mapping: ≈ 6(4b)

5) by combining (or summing over as in arithmetic) the resources or actions of (two or more things) (e.g., between the two of them, they own 50% of the city)

TPP mapping: ≈ 8(5)∪9(5a)

BY

1) indicating the logical subject (that is, the word that would be the subject in an active construction), typically some form of agent (e.g. provided by Joe)

TPP mapping: ≈ 1(1)∪2(1a)∪3(1b)

1.1) indicating the creator/author of (e.g., plays by Shakespeare)

TPP mapping: ≈ 4(1c)

2) indicating the means of achieving something (e.g., melt it by cooking it)

TPP mapping: ≈ 5(2)

2.1) indicating a term or name to which some meaning or referent is assigned (e.g., what is meant by 'fair')

TPP mapping: ≈ 6(2a)∪7(2b)

2.2) indicating the means/medium/route of transport (e.g., travel by car)

TPP mapping: ≈ 8(2c)

2.3) indicating the means/criteria/basis of judgment (e.g., judge by looks)

TPP mapping: ≈ 15(3c)

2.4) indicating the other parent of someone's/something's child[ren] (e.g., his children by his ex-wife)

TPP mapping: ≈ 9(2c)∪9(2d)

3) indicating the size or amount, as of a margin (e.g., increase by 7%)

TPP mapping: ≈ 12(3),13(3a)

4) as in phrases where something is happening repeatedly or progressively, indicating the unit of time (e.g., month by month)

TPP mapping: ≈ 14(3b)

5) (mathematics) expressing multiplication (e.g., 6 feet by 7 feet)

TPP mapping: ≈ 16(3d)

6) near; close to in space (e.g., the tree by the house)

TPP mapping: ≈ 18(5)

6.1) near but then past and beyond, as of movement (e.g., drove by the house)

TPP mapping: ≈ 19(5a)

7) indicating a deadline or the end of a particular time period (e.g., be done by Friday)

TPP mapping: ≈ 17(4)

8) indicating the period in which something happens (e.g., 'by day')

TPP mapping: ≈ 20(6)

9) according to; concerning (e.g., did well by him)

TPP mapping: ≈ 21(7)

10) as used in oaths/swearing (e.g., swore by God)

TPP mapping: ≈ 22(8)

DOWN

1) at or to a lower point of something (e.g., he went down the hill)

TPP mapping: $\approx$ 1(1)

2) moving toward or located at a point further along a path (e.g., he went down the road)

TPP mapping: $\approx$ 3(1b)

2.1) moving toward or located at a point lower in elevation along a water course (e.g., a river)

TPP mapping: $\approx$ 2(1a)

3) throughout a period of time (e.g., down the ages)

TPP mapping: $\approx$ 5(2)

DURING

1) throughout, within, or at a particular point in the course or duration of (a period of time) (e.g., ate a hamburger during the match)

TPP mapping: $\approx$ 1(1)$\cup$2(1a)

FOR

1) used to or intended to contain (e.g., jar for cookies)

TPP mapping: $\supset$ 5(4)

2) with the purpose of mitigating, preventing, or reducing (e.g., cure for cancer)

TPP mapping: $\supset$ 5(4)

3) with the purpose of performing or engaging in (an activity) (e.g., hammer for building)

TPP mapping: $\supset$ 5(4)

4) with the purpose of obtaining, accessing, experiencing, or detecting (e.g., asked for the book, paid $5 for it)

TPP mapping: $\supset$ 5(4)

5) indicating something given or required in exchange (e.g., bought it for $5)

TPP mapping: $\approx$ 10(8a)

6) indicating the topic of desire (e.g., longed for peace)

TPP mapping: $\supset$ 6(5)

7) indicating the focus of an emotion (other than desire) (e.g., disdain for chocolate)

TPP mapping: $\approx 2(2)_{\subset} \cup 3(3)_{\subset}$

8) indicating the destination of movement (e.g., headed for the door)

TPP mapping: $\approx$ 7(6)

9) indicating the intended/expected recipient (e.g., bought it for her)

TPP mapping: $\supset$ 3(3)

10) benefiting; in favor of (e.g., ruled for the defense)

TPP mapping: $\approx 1(1)_{\subset} \cup 3(3)$

11) indicating the employer (e.g., worked for the bank)

TPP mapping: $\approx$ 4(3a)

12) *indicating an extent*

12.1) indicating the temporal extent (e.g., for 16 minutes)

TPP mapping: $\approx$ 12(10)

12.2) indicating the distance/extent (e.g., for 7 miles)

TPP mapping: $\approx$ 13(11)

13) indicating the ordinal position within a series (e.g., for the 5th time)

TPP mapping: $\approx$ 14(12)

14) indicating the justification (and, perhaps also the topic) (e.g., criticize for losing)

TPP mapping: $\supset$ 6(5)

15) relative to the expected norm (e.g., soft for a turtle shell)

TPP mapping: $\approx$ 11(9)

16) regarding; in respect/relation to; with purpose(s) related to (e.g., a name for it)

TPP mapping: $\approx$ 2(2)$\cup$2(2)-1$_{\subset}\cup$5(4)$_{\subset}$

FROM

1) indicating the physical starting point, as of movement or extension (e.g., travel from Maine, a rope from the front to the back)

TPP mapping: $\approx$ 1(1)$\cup$2(1a)$\cup$10(7)$_{\subset}$

2) indicating the vantage point (e.g., see from the tower)

TPP mapping: $\approx$ 8(5)

3) indicating the location where someone/something who is communicating is at (e.g., he called from the lobby)

TPP mapping: $\supset$ 4(3)

4) indicating the origin of an entity or object, as of where it/they came into existence or developed (e.g., people from Britain)

TPP mapping: $\supset$ 4(3)

5) indicating the location associated with someone/something, as of where someone/something was originally or is generally encountered (e.g., friends from school)

TPP mapping: $\supset$ 4(3)

6) so as to be or become (non-physically) separated, unassociated with, or not in accordance with (e.g., defect from the party)

TPP mapping: $\supset$ 10(7)

7) indicating a previous state or condition (e.g., recovered from illness)

TPP mapping: $\approx 10(7)_{\subset} \cup 7(4a) \cup 10(7)\text{-}1$

8) indicating the creator or performer of something (e.g., criticism from Joe)

TPP mapping: $\approx$ 12(9)-1

9) indicating the cause of something (e.g., was cold from the wind)

TPP mapping: ≈ 12(9)

10) indicating an input used in a creative process (e.g., made from recycled paper)

TPP mapping: ⊃ 9(6)

11) indicating a source of knowledge or basis for judgment (e.g., concluded from his remarks)

TPP mapping: ≈ 13(10)

12) indicating the source of obtainment (e.g., I bought it from Joe)

TPP mapping: ⊃ 10(7)

13) indicating the starting point along a particular scale or range (e.g., rose from $5)

TPP mapping: ⊃ 6(4)

14) indicating the starting point in time (e.g., from January onward)

TPP mapping: ≈ 3(2)∪5(3a)

15) indicating something against which a distinction/difference is made or perceived (e.g., distinguish old from new)

TPP mapping: ≈ 14(11)

16) indicating something that is avoided (e.g., abstain from drugs)

TPP mapping: ≈ 11(8)

IN

1) indicating a location that surrounds (or can be viewed as surrounding) something else (e.g., the cow in the meadow)

TPP mapping: $\approx$ 1(1)$\cup$7(5)

2) indicating the (perhaps abstract) place where information is stored, written, or perceived (e.g., stored in the database)

TPP mapping: $\supset$ 7(5)

3) indicating the location (typically a body part/region) where contact is made, typically some form of striking, AND in may not be replaced by into in this context (e.g., hit him in the jaw)

TPP mapping: $\supset$ 2(1a)

4) indicating clothing being worn; wearing (e.g., was in a blue dress)

TPP mapping: $\approx$ 1(1)-1

5) indicating a state/condition/form, often a mental/emotional one that is being experienced (it may in addition be the trigger of some response) (e.g., be in agony)

TPP mapping: $\approx$ 5(4)

6) (into) expressing motion or extension to a position within or surrounded by (e.g., jumped in the pit)

TPP mapping: $\approx$ 2(1a)

7) (into) indicating a state/condition/form that is entered (e.g., tie in a knot)

TPP mapping: $\supset$ 5(4)

8) indicating a time period during which something happens or continues (e.g., in the morning)

TPP mapping: $\approx$ 3(2)

9) indicating an age range (e.g., in their 20s)

TPP mapping: $\supset$ 5(4)

10) indicating the substance/material something is made of (usually somewhat restricted) (e.g., sweater in wool)

11) indicating the language, medium, or means of encoding (e.g., spoke in German)

TPP mapping: $\approx$ 7(5)$_{\subset}\cup$9(7)

12) indicating the color or pattern of (e.g., clothes in blue)

TPP mapping: $\approx$ 9(7)-1

13) indicating the key of a piece of music (e.g., in B flat)

TPP mapping: $\approx$ 10(7a)

14) indicating a group as a context (e.g., tension in his squadron)

TPP mapping: $\supset$ 7(5)

14.1) indicating a group/organization that someone/something is a member/employee of (e.g., he is in the navy)

TPP mapping: $\supset$ 7(5)

15) indicating an academic or occupational subfield or other activity (e.g., he works in astrophysics)

TPP mapping: $\approx 7(5)_{\subset} \cup 8(6)$

16) indicating a manner that something happens or is done, often somewhat idiomatic (e.g., done in a neat way)

TPP mapping: $\approx$ 6(4a)-1

17) during; due to exposure to or involvement with (e.g., died in the earthquake)

TPP mapping: $\supset$ 7(5)

18) expressing participation with an activity (e.g., he was involved in the planning)

TPP mapping: $\supset$ 7(5)

19) indicating the aspect with respect to which a change occurs (e.g., change in its color)

TPP mapping: $\supset$ 6(4a)

20) indicating the quality or aspect with respect to which an assertion or judgment is made; similar to the idiom "in terms of" (e.g., brilliant in its style)

TPP mapping: $\approx 6(4a)_{\subset} \cup 7(5)_{\subset} \cup 11(8)_{\subset}$

21) used in conjunction with 'delight' or 'interest'

TPP mapping: $\supset$ 7(5)

22) expressing the length of time before a future event is expected to happen (e.g., will happen in 1 hour)

TPP mapping: $\approx$ 4(3)

INSIDE

1) in; situated within the confines of (something)

TPP mapping: $\approx$ 1(1)

1.1) into; moving or extending so as to become/end within (something)

TPP mapping: $\approx$ 2(1a)

1.2) metaphorically within (a person's body or mind), as of emotions and sensations (i.e., the anger burned inside him)

TPP mapping: $\approx$ 3(1b)

1.3) (in soccer, rugby, and other sports) closer to the centre of the field than (another player)

TPP mapping: $\approx$ 4(1c)

2) in less than (the period of time specified)

TPP mapping: $\approx$ 5(2)

INTO

1) moving or extending so as to end within (something) (e.g., ran into the house)

TPP mapping: $\approx 1(1)_{\subset} \cup 3(3)$

1.1) indicating the clothing being used to dress someone/something (e.g., got into his suit)

TPP mapping: $\supset$ 1(1)

1.2) indicating a time period (being) entered (e.g., we are into the third week)

TPP mapping: $\approx$ 1(1)-1

2) to a position of contact with, typically used with verbs expressing forceful motion or collision (e.g., rammed into him)

TPP mapping: $\approx$ 2(2)$\cup$4(4)

3) indicating a topic of observation/analysis (e.g., inquiry into the cause)

TPP mapping: $\approx$ 5(5)

4) expressing a state, condition, form, or activity being entered (e.g., broke into pieces, jolted into despair)

TPP mapping: $\approx$ 6(6)$\cup$7(7)$_{\subset}$

4.1) indicating a group/collection being added or assigned to (as a member) (e.g., election into congress)

TPP mapping: $\supset$ 7(7)

5) (mathematics) expressing division (e.g., divide 8 into 2)

TPP mapping: $\approx$ 8(8)

6) taking a lively and active interest with regard to (something) (e.g., got into sky diving)

TPP mapping: $\approx$ 9(9)

LIKE

1) (specific to verbs / actions) in the manner of; as if done by; as if the complement of the preposition were the subject of an active voice sentence using the same verb or predicative adjective (e.g., screamed like a banshee [screams], smells like a dog [smells])

TPP mapping: $\approx$ 2(1a)$\cup$4(1c)$\cup$5(1d)$\cup$1(1)$_{\subset}$

2) (specific to verbs / actions) in a manner appropriate to; as if done to; as if the complement of the preposition were the object of an active voice sentence using the same verb (e.g., eating it like [one eats] candy, petting it like [one pets] a cat)

TPP mapping: $\approx$ 3(1b)

3) (specific to verbs / actions) in the manner of (another event or action), often used with demonstratives (i.e., this/that)

TPP mapping: $\approx$ 6(2)

4) (specific to noun-like-noun) used to express similarity or resemblance between two nouns

TPP mapping: $\supset$ 1(1)

5) such as; for example

TPP mapping: $\approx$ 7(3)

OF

1) (part/member-of-whole) indicating the whole in a part-of-whole relationship (e.g., the roots of the tree)

TPP mapping: $\approx$ 1(1)$\cup$6(3)$_{\subset}$

2) *connecting a number, quantifier, or partitive noun with the thing(s)/substance that compose(s) it*

2.1) (measure/number/amount -of- something(s)) connecting a formal measure term or count with what is being measured/counted (e.g., six of them)

TPP mapping: $\supset$ 3(1b)

2.2) (collection/configuration/series -of- members/parts) connecting a collection/configuration term with the member(s)/part(s) that constitute it (e.g., group of people)

TPP mapping: $\supset$ 3(1b)

2.3) (substance/material/ingredient) indicating a/the substance/material/ingredient constituting something (e.g., ball of paste)

TPP mapping: $\approx$ 3(1b)$_{\subset}\cup$17(8)

2.4) expressing a (species/variety -of- specimens) relationship (e.g., genus of plant)

TPP mapping: $\approx$ 10(5a)

3) (container -of- contents) connecting container(s) (or containing area(s)) with their contents (e.g., basket of apples)

TPP mapping: $\approx$ 3(1b)-1

4) indicating the direct object, theme, or patient (the thing modified) associated with the governing verb-derived/verb-like noun (e.g., modification of cars)

TPP mapping: $\approx$ 11(6)$\cup$12(6a)

5) indicating the agent or logical grammatical subject of a verb-like noun (e.g., conclusions of the report)

TPP mapping: $\approx$ 10(5a)

5.1) expressing the relationship between a creator/producer and their product(s), most often for artistic works (e.g., books of J.K. Rowling)

TPP mapping: $\approx$ 7(3a)

6) indicating a secondary argument not paraphrasable using 'about' (limited to a small set of verbs) (e.g., drained it of money, convict him of theft)

TPP mapping: $\supset$ 14(7)

7) indicating the subject/topic matter; about (e.g., spoke of the trouble)

TPP mapping: $\approx$ 15(7a)

7.1) expressing and/or motivated by (e.g., grin of happiness)

TPP mapping: $\supset$ 6(3)-1

8) expressing an equative relationship, often where the second of the preposition's arguments is a specific instance of the class of things specified by the first (e.g., sacrament of baptism, city of Perth)

TPP mapping: $\approx$ 9(5)

8.1) expressing the relationship between a scale or measure and a value (e.g., a loss of 5%)

TPP mapping: $\approx$ 4(2)

9) indicating the time when something occurred, existed, or came into being (e.g., the famine of 1745)

TPP mapping: $\supset$ 6(3)

10) indicating the location where something is at/from (e.g., people of Toledo)

TPP mapping: $\supset$ 6(3)

11) (mental experience -of- experiencer) indicating the experiencer who experiences the thought(s)/emotion(s) (e.g., fury of voters)

TPP mapping: $\supset$ 11(6)

12) (possession -of- possessor) indicating the possessor who owns or otherwise holds rights towards something (e.g., the house of the Johnsons)

TPP mapping: $\supset$ 6(3)

13) connecting an attribute with the person/thing that exhibits it (e.g., that was smart of him)

TPP mapping: $\supset$ 13(6b)

14) caused by (used mostly with death, perceptions, and feelings) (e.g., died of cancer)

TPP mapping: $\supset$ 16(7b)

15) (with verbs of smell/taste) indicating the thing that is detected or resembled (e.g., stinks of beer)

TPP mapping: $\supset$ 16(7b)

16) indicating a possessed attribute such as age or color (e.g., cloak of blue)

TPP mapping: $\approx$ 5(2a)$\cup$6(3)-$1_{\subset}$

17) indicating an action/activity something is involved with or is used for (e.g., rod of smiting)

TPP mapping: $\supset$ 9(5)

18) indicating the complement of a relational noun (e.g., son of, price of)

TPP mapping: $\supset$ 6(3)

18.1) indicating the complement of a measurable noun such as de-adjectival nouns (e.g., adequacy, compatibility, width)

19) indicating the complement of a predicative adjective (e.g., symptomatic of cancer)

TPP mapping: $\supset$ 13(6b)

OFF

1) moving away from so as to no longer be located at or in contact with (e.g., walk off the stage)

TPP mapping: $\approx$ 1(1)$\cup$4(3)

2) abutting and extending away from (e.g., a road off the main boulevard)

TPP mapping: $\approx$ 2(2)

3) in or at a position separated from (e.g., the bird was off its perch)

TPP mapping: $\approx$ 1(1)

3.1) in or at a position in a body of water separated from (e.g., the boats off the coast)

TPP mapping: $\approx$ 3(2a)

4) indicating a substance (often a drug or something that can be viewed as such) being abstained from (e.g., stay off drugs)

TPP mapping: $\approx$ 6(3b)$\cup$7(4)

5) indicating an activity (typically work) avoided or not engaged in (e.g., he took time off work)

TPP mapping: $\approx$ 5(3a)

ON

1) physically in contact with and supported by (a surface) (e.g., the book on the table)

TPP mapping: $\approx$ 1(1)$\cup$2(1a)$\cup$

1.1) indicating a surface that, due to contact with, serves as the cause or means of causing something (e.g., cooked on the open flame)

seealso 3(1b)

1.2) onto; indicating where (or against what) contact is made (e.g., threw spaghetti on the wall)

TPP mapping: $\approx$ 5(1d)$\cup$11(5)-1$_{\subset}\cup$16(7b)$_{\subset}$

1.3) indicating the part(s) of the body supporting the rest of the body (e.g., stood on his feet)

TPP mapping: $\approx$ 4(1c)

1.4) indicating the vehicle or mode of transportation (e.g., flew on a plane)

TPP mapping: $\approx$ 15(7a)

1.5) in the possession of; being carried by (e.g., a gun on him)

TPP mapping: $\approx$ 6(1e)

2) forming a distinctive or marked part of the surface of (e.g., a smile on his lips)

TPP mapping: $\approx$ 7(2)

3) having (the thing mentioned) as a topic; about (e.g., knowledge on sailing)

TPP mapping: $\approx$ 8(3)$\cup$11(5)-1$_{\subset}$

4.1) having (the thing mentioned) as criteria used in judgment or evaluation or as a/the justification for an action (e.g., judged on his actions)

TPP mapping: $\supset$ 9(3a)

4.2) having (the thing mentioned) as a/the basis for a monetary action (e.g., was charged on his loan)

TPP mapping: $\approx$ 9(3a)-1

4.3) having (the thing mentioned) as a/the basis being adhered to or system of rules governing an action or situation (e.g., employed on a contract)

TPP mapping: $\supset$ 9(3a)

5) indicating a time; at the time of (e.g., happened on Oct. 22nd)

TPP mapping: $\approx$ 17(8)$\cup$18(8a)

6) indicating some foodstuff that is being consumed or otherwise forms the basis of a diet/meal (e.g., dined on fine wine)

TPP mapping: $\supset$ 9(3a)

7) indicating some substance (e.g., a drug or medicine) one is currently taking or under the influence of (e.g., he is on drugs)

TPP mapping: $\approx$ 20(10)

8) indicating the means/medium used for storing or transmitting information (e.g., talked on the phone)

TPP mapping: $\approx$ 12(6)$\cup$13(6a)

9) in the course of (a journey) (e.g., it happened on the way to the store)

TPP mapping: $\approx$ 14(7)

10) indicating an activity engaged in (e.g., he was on a mission)

TPP mapping: $\approx$ 19(9)

11) indicating the target of some action or behavior; metaphorically onto or against (e.g., get revenge on Sam)

TPP mapping: $\approx 11(5)_{\subset} \cup 11(5)\text{-}1_{\subset}$

12) denoting a particular level or setting along a scale or range (e.g., set the dryer on Permanent Press)

TPP mapping: $\approx$ 23(13)

ONTO

1) moving (or extending) to a location on the surface of

TPP mapping: $\approx$ 1(1)

1.1) moving aboard a vehicle of transportation

TPP mapping: $\approx$ 2(2)

2) (mathematics) in reference to an 'onto' mapping

OVER

1) (above and across) vertically above and moving/extending (at least partially) across (e.g., flew over the plains)

TPP mapping: $\approx 2(1a)_{\subset} \cup 3(1b) \cup 4(2)_{\subset} \cup 13(4b)_{\subset}$

2) (above) vertically above (e.g., the halo over his head)

TPP mapping: $\approx 1(1)_{\subset} \cup 4(2)_{\subset}$

3) expressing passage or trajectory across (but without implied vertical separation); expressing a location that could be reached by travel across (e.g., journey over the desert, the house over the hill)

TPP mapping: $\approx 11(4) \cup 3(1b)_{\subset}$

4) located at the other side of

TPP mapping: $\approx 13(4b)$

5) across so as to form a covering (e.g., spread jam over it)

TPP mapping: $\approx 2(1a)$

6) expressing the means by which something is done (e.g., talk over the phone)

TPP mapping: $\approx 15(6)$

6.1) by means of the resistant surface constituted by (an object), as with striking or hitting (e.g., broke his club over Joe's head)

TPP mapping: $\approx 15(6)\text{-}1$

7) in cases of perceived despite some sort of interference, usually some sort of audible noise (e.g., heard it over the rock music)

TPP mapping: ≈ 9(2e)

8) indicating the subject of (and, sometimes, the reason for) (e.g., fought over the trophy)

TPP mapping: ≈ 16(7)

9) indicating a time period during which something happens or persists (e.g., the food spoiled over the weekend)

TPP mapping: ≈ 14(5)

10) expressing dominance or influence, as of authority, rank, or victory (e.g. power over him)

TPP mapping: ≈ 5(2a)∪6(2b)∪7(2c)∪8(2d)

11) indicating some foodstuff available/consumed during an event/activity (e.g., discuss it over coffee)

No mapping

THROUGH

1) expressing passage, extension, or perception that goes (perhaps to many places) within (and, perhaps, out of) someplace/something; or alternatively, into and then out of (e.g., go through the tunnel)

TPP mapping: ≈ 1(1)∪2(1a)∪3(1b)∪4(1c)∪5(1d)-1∪10(3)-1

1.1) being positioned/located beyond or at the far end of something, as of openings or obstacles (e.g., her house is through the woods)

TPP mapping: ≈ 5(1d)

1.2) expressing the extent of rotation/turning from one orientation to another (e.g., turn through 90 degrees)

TPP mapping: ≈ 6(1e)

2) so as to reach the end of (an experience, situation, or activity, often a tedious or stressful one) (e.g., worked through the process)

TPP mapping: ≈ 7(2)∪8(2a)∪9(2b)

3) indicating a collection, container, inventory, or publication being at least partially searched or inspected (e.g., go through the newspaper)

TPP mapping: ≈ 10(3)

4) indicating a means by which something happens or is done (e.g., got a job through his uncle)

TPP mapping: ≈ 12(5)∪13(5a)

5) up to and including (a particular point in an ordered sequence, often a point in time) (e.g., March 2nd through the 5th)

TPP mapping: ≈ 11(4)

TO

1) expressing the end location of motion or extension (e.g. walked to NYC)

TPP mapping: $\approx$ 1(1)$\cup$2(1a)

1.1) indicating attachment or linkage with something (e.g., attached to the car)

TPP mapping: $\supset$ 13(5)

2) approaching or reaching (a particular state/condition/circumstances/relationship) (e.g., turned to dust)

TPP mapping: $\approx$ 5(2)$\cup$6(2a)

3) expressing an extent or point reached along a particular scale or range (e.g., changed to $5)

TPP mapping: $\approx$ 3(1b)

4.1) indicating the recipient of communication (e.g., talked to Joe)

TPP mapping: $\supset$ 8(3)

4.2) indicating the recipient of possession (e.g., bequeathed the ball to Joe)

TPP mapping: $\approx$ 8(3)

5) indicating the target of focused perception (e.g., listen to Joe)

6) indicating the topic or subject matter (used with a handful of assertive communicative words such as 'testify') (e.g., testify to the truth)

7) indicating a complement of a relational noun, verb, or predicative adjective (e.g., allergic to corn)

TPP mapping: $\approx$ 8(3)-1$\cup$9(4)$\cup$10(4a)$\cup$14(6)

8) (telling the time) until/before (the time/hour specified) (e.g., 5 minutes to noon)

TPP mapping: ≈ 4(1c)

9) indicating a rate of return on something, for example the distance travelled in exchange for fuel used (e.g., it gets 40 miles to the gallon)

TPP mapping: ≈ 11(4b)

10) (mathematics) indicating the power (exponent) to which a number is raised

TPP mapping: ≈ 12(4c)

11) in comparison with (e.g., the club's nothing to what it once was)

TPP mapping: ≈ 15(7)

12) governing a phrase expressing someone's reaction to something (e.g., to her astonishment, he smiled)

TPP mapping: ≈ 7(2b)

TOWARD

1) indicating a location something moves, extends, or is oriented in the direction of (e.g., move towards him)

TPP mapping: ≈ 1(1)

1.1) indicating a point in TIME that is near and/or being approached (e.g., towards the end of November)

TPP mapping: ≈ 3(1b)

1.2) getting closer to achieving a goal; indicating a state/form/situation/action being approached

TPP mapping: $\approx$ 2(1a)

2) indicating the person or thing that a manner of behavior is directed at (e.g., be polite towards him)

TPP mapping: $\approx$ 4(2)

3) indicating the complement of a cognitive/emotional word (e.g., resentment towards him)

TPP mapping: $\supset$ 4(2)-1

4) indicating the complement of a word indicating a tendency or bias (e.g., bias towards Harvard graduates)

TPP mapping: $\supset$ 4(2)-1

5) indicating a cost being payed for (e.g., $5 towards the cost of lunch)

TPP mapping: $\approx$ 5(3)

WITH

1) indicating a participating entity/thing in an action/activity (e.g., talk with Joe)

TPP mapping: $\approx 6(4) \cup 1(1)_{\subset}$

2) indicating spatial accompaniment / proximity (e.g., he is with the bike)

TPP mapping: $\supset$ 1(1)

3) indicating the group in a member-of-group relationship (e.g., Joe is with the marines, grouped with alternative medicine)

TPP mapping: $\approx$ 13(8)

4) indicating the means or material used to perform an action or acting as the complement of similar participle adjectives (e.g., crammed with, coated with, covered with)

TPP mapping: $\approx$ 4(3)$\cup$5(3a)

5) because of / due to (the physical/mental presence of) (e.g., boiling with anger, shining with dew)

TPP mapping: $\approx$ 11(7b)

6) indicating the manner or circumstances (but not cause or motivation) of something (e.g., fix with precision)

TPP mapping: $\approx$ 7(5)

7) expressing a possessive relationship (e.g., the man with the hat)

TPP mapping: $\approx$ 2(2)$_{\subset}\cup$3(2a)$\cup$8(6)$\cup$10(7a)

7.1) indicating a part in a part-of-whole relationship (e.g., man with the red nose)

TPP mapping: $\supset$ 2(2)

7.2) connecting an experiencer with some sort of experience, primarly ailments (e.g., person with schizophrenia)

TPP mapping: $\approx$ 2(2)$_{\subset}\cup$3(2a)$\cup$8(6)$\cup$10(7a)

8) indicating the information content brought to a communication/discussion (e.g., phoned him with the question)

TPP mapping: $\approx$ 4(3)-1

9) expressing correlation (e.g., increases with age)

TPP mapping: $\approx$ 12(7c)

10) *miscellaneous complements*

10.1) indicating the complement of a cognitive/emotional word (whatever it is oriented toward) (e.g., angry with her)

TPP mapping: $\supset$ 9(7)

10.2) indicating the complement of an adherence/conformity word (e.g., adheres with, conforms with)

TPP mapping: $\approx$ 9(7)$_{\subset}\cup$15(9)-1

10.3) in respect to; in relation to; indicating something(s)/someone(s) towards which some manner of behavior or being exists or is expressed (and doesn't fit elsewhere) (e.g., be polite with her)

TPP mapping: $\approx$ 9(7)$_{\subset}\cup$16(10)

11) in the same direction as (e.g., shave with the grain)

TPP mapping: $\supset$ 15(9)