

# Parse Ranking with Semantic Dependencies and WordNet

Xiaocheng Yin♣ Jungjae Kim♣ Zinaida Pozen◇♣ Francis Bond♣

♣ Nanyang Technological University, Singapore

◇ University of Washington, Seattle

yinx0005@e.ntu.edu.sg, jungjae.kim@ntu.edu.sg,

zpozen@gmail.com, bond@ieee.org

## Abstract

In this paper, we investigate which features are useful for ranking semantic representations of text. We show that two methods of generalization improved results: extended grand-parenting and super-types. The models are tested on a subset of SemCor that has been annotated with both Dependency Minimal Recursion Semantic representations and WordNet senses. Using both types of features gives a significant improvement in whole sentence parse selection accuracy over the baseline model.

## 1 Introduction

In this paper we investigate various features to improve the accuracy of semantic parse ranking. There has been considerable successful work on syntactic parse ranking and reranking (Toutanova et al., 2005; Collins and Koo, 2006; McClosky et al., 2006), but very little that uses pure semantic representations. With recent work on building semantic representations (from deep grammars such as LFG (Butt et al., 1999) and HPSG (Sag et al., 1999), directly through lambda calculus, or as in intermediate step in machine translation) the question of ranking them has become more important.

The closest related work is Fujita et al. (2010) who ranked parses using semantic features from Minimal Recursion Semantics (MRS) and syntactic trees, using a Maximum Entropy Ranker. They experimented with Japanese data, using the Hinoki Treebank (Bond et al., 2008), using primarily elementary dependencies: single arcs between pred-

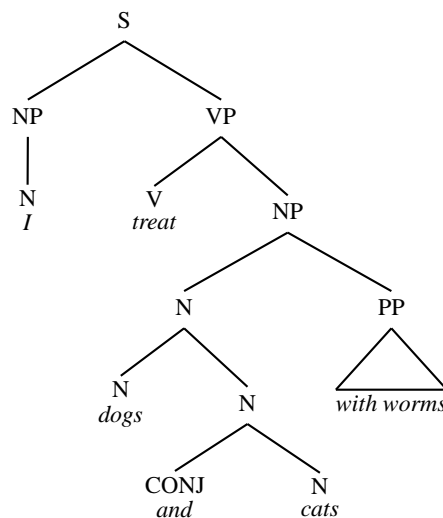


Figure 1: Syntactic view of sentence “*I treat dogs and cats with worms*”.

icates and their arguments. These can miss some important connections between predicates.

An example parse tree for *I treat dogs and cats with worms* is shown in Figure 1.<sup>1</sup>, for the interpretation “I treat both dogs and cats that have worms” (not “I treat, using worms, dogs and cats” or any of the other possibilities)

The semantic representation we use is Dependency Minimal Recursion Semantics (DRMS: Copestake, 2009). The Minimal Recursion Semantics (MRS: Copestake et al., 2005) is a computationally tractable flat semantics that under-specifies quantifier scope. The Dependency MRS is an MRS representation format that keeps all the information from the MRS but is simpler to manipulate. DMRSs differ from syntactic dependency graphs in that the relations are defined between slightly abstract predicates, not between

♣ Currently at PointInside, Inc.

<sup>1</sup>Simplified by omission of non-branching nodes.

surface forms. Some semantically empty surface tokens (such as infinitive *to*) are not included, while some predicates are inserted that are not in the original text (such as the null article).

A simplified MRS representation of our example sentence and its DMRS equivalent are shown in Figure 2.

In the DMRS, the basic links between the nodes are present. However, potentially interesting relations such as that between the verb *treat* and its conjoined arguments *dogs* and *cats* are not linked directly. Similarly, the relation between *dogs and cats* and *worms* is conveyed by the preposition *with*, which links them through its external argument (ARG1: *and*) and internal argument (ARG2: *worms*). There is no direct link. We investigate new features that make these links more direct (Section 3.2).

We also explore the significance of the effectiveness of links between words that are connected arbitrarily far away in the semantic graph (Section 3.2.3).

Finally, we experimented with generalizing over semantic classes. We used WordNet semantic files as supertypes to reduce data sparseness (Section 3.2.4). This will generalize the lexical semantics of the predicates, resulting in a reduction of feature size and ambiguity.

## 2 Previous Work

This paper follows up on the work of Fujita et al. (2010) in ranking MRS semantic representations, which was carried out for Japanese. We are conducting a similar investigation for English, and add new features and approaches. Fujita et al. (2010) worked with the Japanese Hinoki Corpus (Bond et al., 2008) data and used hypernym chains from the Goi-Taikai Japanese ontology (Ikehara et al., 1997) for variable-level semantic backoff. This is in contrast to the uniform WordNet semantic file backoff performed here. In addition, this work only focuses on MRS ranking, whereas Fujita et al. (2010) combined MRS features with syntactic features to improve syntactic parse ranking accuracy.

Our use of WordNet Semantic Files (SF) to reduce lexical feature sparseness is inspired by several recent papers. Agirre et al. (2008, 2011) have experimented with replacing open-class words with their SFs. Agirre et al. (2008) have shown an improvement in full parse and PP attachment

scores with statistical constituency parsers using SFs. Agirre et al. (2011) have followed up on those results and re-trained a dependency parser on the data where words were replaced with their SFs. This resulted in a very modest labeled attachment score improvement, but with a significantly reduced feature set. In a recent HPSG work, MacKinlay et al. (2012) attempted to integrate lexical semantic features, including SF backoff, into a discriminative parse ranking model. However, this was not shown to help, presumably because the lexical semantic features were built from syntactic constituents rather than MRS predicates.

The ancestor features found to be helpful here are inspired by the use of grand-parenting in syntactic parse ranking (Toutanova et al., 2005) and chains in dependency parsing ranking (Le Roux et al., 2012).

## 3 Resources and Methodology

In this section we introduce the corpus we work on, and the features we extract from it.

### 3.1 Corpus: SemCor

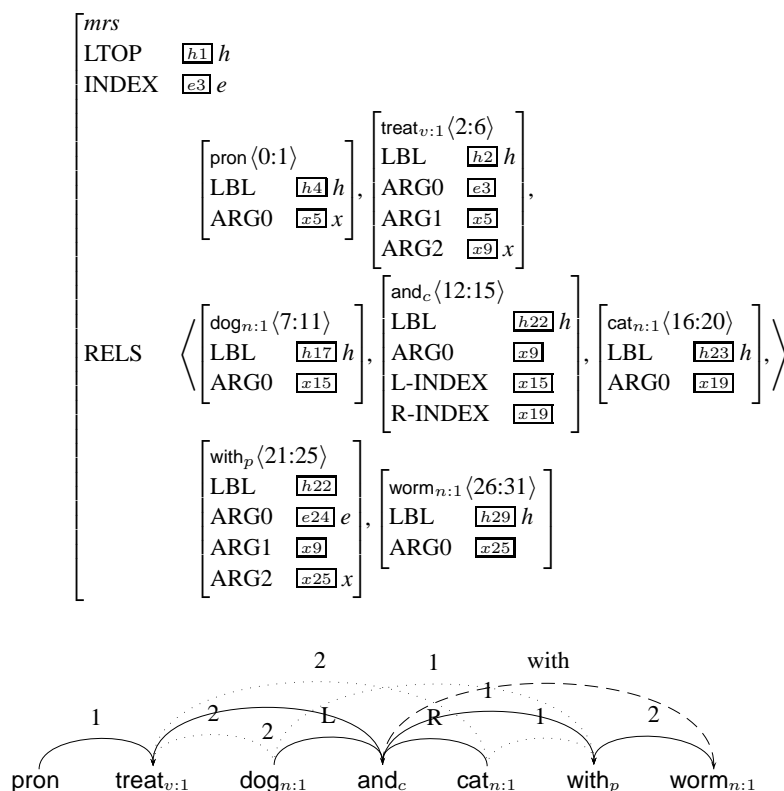
To evaluate our ranking methods, we are using the Redwoods Treebank (Oepen et al., 2004) of manually disambiguated HPSG parses, storing full signs for each analysis and supporting export into a variety of formats, including the Dependency MRS (DMRS) format used in this work.

The HPSG parses in Redwoods are based on the English Resource Grammar (ERG; Flickinger, 2000) – a hand-crafted broad-coverage HPSG grammar of English.

For our experiments, we used a subset of the Redwoods Treebank, consisting of 2,590 sentences drawn from SemCor (Landes et al., 1998). In the SemCor corpus each of the sentences is tagged with WordNet senses created at Princeton University by the WordNet Project research team. The average length of the Redwoods SemCor sentences is 15.4 words, and the average number of parses is 247.

From the treebank we can export the DMRS. The choice of which words become predicates is slightly different in the SemCor/WordNet and the ERG. The ERG lexicon groups together all senses that have the same syntactic properties, making them underspecified for many sense differences. Thus elementary predicate  $cat_{n:1}$  could be any of the WordNet senses  $cat_{n:1}$  “feline mammal usu-

*I treat dogs and cats with worms.*



Simplified by omission of quantifiers  
 Dashed lines show Preposition (P) features  
 Dotted lines show Conjunction (LR) features  
 Arc labels show the roles: 1 is ARG1, 2 is ARG2, ...

Figure 2: MRS and DMRS for *I treat cats and dogs with worms.*

<b>Tops</b> <sub>n</sub>	<b>act</b> <sub>n</sub>
<b>animal</b> <sub>n</sub>	<b>artifact</b> <sub>n</sub>
<b>attribute</b> <sub>n</sub>	<b>body</b> <sub>n</sub>
<b>cognition</b> <sub>n</sub>	<b>communication</b> <sub>n</sub>
<b>event</b> <sub>n</sub>	<b>feeling</b> <sub>n</sub>
<b>food</b> <sub>n</sub>	<b>group</b> <sub>n</sub>
<b>location</b> <sub>n</sub>	<b>motive</b> <sub>n</sub>
<b>object</b> <sub>n</sub>	<b>person</b> <sub>n</sub>
<b>phenomenon</b> <sub>n</sub>	<b>plant</b> <sub>n</sub>
<b>possession</b> <sub>n</sub>	<b>process</b> <sub>n</sub>
<b>quantity</b> <sub>n</sub>	<b>relation</b> <sub>n</sub>
<b>shape</b> <sub>n</sub>	<b>state</b> <sub>n</sub>
<b>substance</b> <sub>n</sub>	<b>time</b> <sub>n</sub>

Table 1: WordNet Noun Semantic Files.

ally having thick soft fur and no ability to roar”, *cat*<sub>n:2</sub> “an informal term for a youth or man” and six more.<sup>2</sup> In some cases, DMRS decomposes a single predicate into multiple predicates (e.g. *here* into *in<sub>p</sub> this<sub>q</sub> place<sub>n</sub>*). The ERG and WordNet also often make different decisions about what constitutes a multiword expression. For these reasons the mapping between the two annotations is not always straightforward. In this paper we use the mapping between the DRMS and WordNet annotations produced by Pozen (2013).

Using the mapping, we exploited the sense tagging of the SemCor in several ways. We experimented both with replacing elementary predicates with their synsets, their hypernyms at various levels and with their semantic files (Landes et al., 1998), which generalize the meanings of words that belong to the same broad semantic categories.<sup>3</sup> These dozens of generalized semantic tags help to address the issue of feature sparseness, compared to thousands of synsets. We show the semantic files for nouns and verbs in Tables 1 and 2. In this paper, we only report on the parse selection accuracy using semantic files to reduce ambiguity, as it gave the best results.

### 3.2 Semantic Dependency Features

In this section we introduce the baseline features for parse ranking.

Table 3 shows example features extracted from the DMRS depicted in Figure 2. Features 1–16 are

<sup>2</sup>Elementary predicates are shown in sans-serif font, *WordNet senses* in bold italic, **WordNet semantic files** are shown in bold typewriter.

<sup>3</sup>Semantic Files are also sometimes referred to as Semantic Fields, Lexical Fields or Supersenses.

<b>body</b> <sub>v</sub>	<b>change</b> <sub>v</sub>
<b>cognition</b> <sub>v</sub>	<b>communication</b> <sub>v</sub>
<b>competition</b> <sub>v</sub>	<b>consumption</b> <sub>v</sub>
<b>contact</b> <sub>v</sub>	<b>creation</b> <sub>v</sub>
<b>emotion</b> <sub>v</sub>	<b>motion</b> <sub>v</sub>
<b>perception</b> <sub>v</sub>	<b>possession</b> <sub>v</sub>
<b>social</b> <sub>v</sub>	<b>stative</b> <sub>v</sub>
<b>weather</b> <sub>v</sub>	

Table 2: WordNet Verb Semantic Files.

the semantic dependency features (Baseline). 17–18 are the conjunctive features (LR). 19–22 are the preposition role features (PR).

#	Sample Features
0	⟨0 treat <sub>v:1</sub> ARG1 pron ARG2 and <sub>c</sub> ⟩
1	⟨0 and <sub>c</sub> L-IND dog <sub>n:1</sub> R-IND cat <sub>n:1</sub> ⟩
2	⟨0 with <sub>p</sub> ARG1 and <sub>c</sub> ARG2 worm <sub>n:1</sub> ⟩
3	⟨1 treat <sub>v:1</sub> ARG1 pron⟩
4	⟨1 treat <sub>v:1</sub> ARG2 and <sub>c</sub> ⟩
5	⟨1 and <sub>c</sub> L-IND dog <sub>n:1</sub> ⟩
6	⟨1 and <sub>c</sub> R-IND cat <sub>n:1</sub> ⟩
7	⟨1 with <sub>p</sub> ARG1 and <sub>c</sub> ⟩
8	⟨1 with <sub>p</sub> ARG2 worm <sub>n:1</sub> ⟩
9	⟨2 treat <sub>v:1</sub> pron and <sub>c</sub> ⟩
10	⟨2 with <sub>p</sub> and <sub>c</sub> worm <sub>n:1</sub> ⟩
11	⟨3 treat <sub>v:1</sub> pron⟩
12	⟨3 treat <sub>v:1</sub> and <sub>c</sub> ⟩
13	⟨3 and <sub>c</sub> dog <sub>n:1</sub> ⟩
14	⟨3 and <sub>c</sub> cat <sub>n:1</sub> ⟩
15	⟨3 with <sub>p</sub> and <sub>c</sub> ⟩
16	⟨3 with <sub>p</sub> worm <sub>n:1</sub> ⟩
17	⟨1 treat <sub>v:1</sub> ARG2 dog <sub>n:1</sub> ⟩
18	⟨1 treat <sub>v:1</sub> ARG2 cat <sub>n:1</sub> ⟩
19	⟨0 and <sub>c</sub> L-IND dog <sub>n:1</sub> R-IND cat <sub>n:1</sub> with <sub>p</sub> worm <sub>n:1</sub> ⟩
20	⟨1 and <sub>c</sub> with <sub>p</sub> worm <sub>n:1</sub> ⟩
21	⟨2 and <sub>c</sub> worm <sub>n:1</sub> ⟩
22	⟨3 and <sub>c</sub> worm <sub>n:1</sub> ⟩

Table 3: Features for the DMRS in Fig 2.

Baseline features are those that directly reflect the dependencies of the DMRS. In Table 3, feature type ⟨0⟩ (0–2) shows predicates with all their arguments. Feature type ⟨1⟩ (3–8) shows each argument individually. Feature type ⟨2⟩ shows all arguments without the argument types. Feature type ⟨3⟩ is the least specified, showing individual arguments without the labels. These types are the same as the MRS features of Toutanova et al. (2005) and

the SEM-DEP features of Fujita et al. (2010).

### 3.2.1 Conjunctive Features

We further create two more features, called Left/Right Handle Features (LR), to link directly the two arguments of conjunctive relations with their parent, independently from the other argument. In Table 1, for example, the feature  $\langle \text{treat}_{v:1} \text{ ARG2 and}_c \rangle$ , although valid, does not convey the meaning of the sentence. Instead, we add the two LR features  $\langle \text{treat}_{v:1} \text{ ARG2 dog}_{n:1} \rangle$  (feature 17) and  $\langle \text{treat}_{v:1} \text{ ARG2 cat}_{n:1} \rangle$  (feature 18), which better model the conjunction relation.

### 3.2.2 Preposition Role Features

As shown in Figure 2, the node with<sub>p</sub> has two links: to and<sub>c</sub> (ARG1) and to worm<sub>n:1</sub> (ARG2). The two relations together indicate a noun-preposition-noun relationship. Instead of breaking the relationship into the two separate features, we introduce it, as a whole, as a new type of feature, where the two arguments of the preposition (e.g. and<sub>c</sub>, worm<sub>n:1</sub>) will have a direct relation via the preposition (e.g. with<sub>p</sub>). We name these Preposition Role features (PR), as they are similar in spirit to semantic roles. Some sample PR features are given in Table 3, features 19–22.

The new features explicitly convey, for example, noun-preposition-noun relations. Parses containing features like *something at somewhere* can be further distinguished from parses containing *at somewhere* and *something at* separately. When the features become more representative, active parses are more likely to be selected, though with the cost of a larger feature set size.

As 4 types of features can be developed based on one relationship, a Preposition Role link would have 4 separate features. While the Conjunctive features mentioned in previous section give 2 to 4 additional features, Baseline-PR features normally give 4 more. Thus, the feature size of Baseline-PR model is larger than that of the Baseline-LR model.

### 3.2.3 Ancestor Features

While the semantic dependency features correspond to direct dependencies, we introduce a new type of features that represent indirect dependencies between ancestors and their descendants in the DMRS. For each predicate, we collect all its descendants linked through more than one dependency and create features to represent the indirect

#	Sample Features
0	$\langle 0 \text{ treat}_{v:1} \text{ ARG1 pron ARG2 and}_c \rangle$
1	$\langle 0 \text{ and}_c \text{ L-IND dog}_{n:1} \text{ R-IND cat}_{n:1} \rangle$
2	$\langle 0 \text{ treat}_{v:1} \text{ ARG2 dog}_{n:1} \text{ ARG2 cat}_{n:1} \rangle$
3	$\langle 1 \text{ treat}_{v:1} \text{ ARG1 pron} \rangle$
4	$\langle 1 \text{ treat}_{v:1} \text{ ARG2 and}_c \rangle$
5	$\langle 1 \text{ treat}_{v:1} \text{ ARG2 dog}_{n:1} \rangle$
6	$\langle 1 \text{ treat}_{v:1} \text{ ARG2 cat}_{n:1} \rangle$
7	$\langle 1 \text{ and}_c \text{ L-IND dog}_{n:1} \rangle$
8	$\langle 1 \text{ and}_c \text{ R-IND cat}_{n:1} \rangle$
9	$\langle 2 \text{ treat}_{v:1} \text{ pron and}_c \rangle$
10	$\langle 2 \text{ treat}_{v:1} \text{ dog}_{n:1} \text{ cat}_{n:1} \rangle$
11	$\langle 3 \text{ treat}_{v:1} \text{ pron} \rangle$
12	$\langle 3 \text{ treat}_{v:1} \text{ and}_c \rangle$
13	$\langle 3 \text{ treat}_{v:1} \text{ dog}_{n:1} \rangle$
14	$\langle 3 \text{ treat}_{v:1} \text{ cat}_{n:1} \rangle$

Table 4: Ancestor Features (AF).

dependencies between the predicate and the descendants. We name these features Ancestor Features (AF).

Table 4 has some sample AF features such as that linking from  $\text{treat}_{v:1}$  to  $\text{dog}_{n:1}$  and  $\text{cat}_{n:1}$  (i.e. feature 2). This is a one-level ancestor, involving two predicates, while multi-level ancestors deal with more than two predicates linked in a sequence. Note that these are different from the LR features (features 15, 16 in Table 1), in that AF features include both arguments of a conjunction, for example, connecting the predicate  $\text{treat}_{v:1}$  to its grandchildren  $\text{dog}_{n:1}$  and  $\text{cat}_{n:1}$  via the argument role of and<sub>c</sub> in the predicate (feature 2 in Table 4).

When a sentence has  $n$  dependencies, our method generates  $O(\frac{n(n-1)}{2}) = O(n^2)$  AF features. In the corpus we use, the dependency structure of a sentence typically has 4 levels. In practice the number of AF features is roughly triple the number of Baseline features. In the evaluation experiments, we investigated all the eight combinations of the three types of LR, PR, and AF features, where each combination is combined with the baseline features.

### 3.2.4 Semantic File Features

In the features up until now, words have been represented as elementary predicate semantic dependencies (SD). Because SemCor also has WordNet senses, we experiment with replacing open class words with their supertypes, in this case using the WordNet semantic files (SF). If a word is not matched to a WordNet synset we continue to use

#	Sample Features
0	$\langle 0 \text{ body}_v \text{ ARG1 pron ARG2 and}_c \rangle$
1	$\langle 0 \text{ and}_c \text{ L-IND animal}_n \text{ R-IND animal}_n \rangle$
2	$\langle 0 \text{ with}_p \text{ ARG1 and}_c \text{ ARG2 animal}_n \rangle$
3	$\langle 1 \text{ body}_v \text{ ARG1 pron} \rangle$
4	$\langle 1 \text{ body}_v \text{ ARG2 and}_c \rangle$
5	$\langle 1 \text{ and}_c \text{ L-IND animal}_n \rangle$
6	$\langle 1 \text{ and}_c \text{ R-IND animal}_n \rangle$
7	$\langle 1 \text{ with}_c \text{ ARG1 animal}_n \rangle$
8	$\langle 1 \text{ with}_c \text{ ARG2 animal}_n \rangle$
9	$\langle 2 \text{ body}_v \text{ pron and}_c \rangle$
10	$\langle 2 \text{ with}_p \text{ and}_c \text{ animal}_n \rangle$
11	$\langle 3 \text{ body}_v \text{ pron} \rangle$
12	$\langle 3 \text{ body}_v \text{ and}_c \rangle$
13	$\langle 3 \text{ and}_c \text{ animal}_n \rangle$
14	$\langle 3 \text{ and}_c \text{ animal}_n \rangle$
15	$\langle 3 \text{ with}_c \text{ and}_c \rangle$
16	$\langle 3 \text{ with}_c \text{ animal}_n \rangle$

Table 5: Baseline features with Semantic Files (SF).

the elementary predicate. This SF representation is also applied to the eight combinations of feature types. A sample of the features in the SF representations are given in Table 5.

Sometimes two features, such as 13 and 14 in Table 3, are replaced with the same feature, like 9 in Table 5, because  $\text{dog}_{n:1}$  and  $\text{cat}_{n:1}$  are both replaced with  $\text{animal}_n$ . There are about half as many Semantic File features as there are SD features.

## 4 Results

We set up the evaluation task as reranking of the top 500 Redwoods analyses, previously selected by the syntactic MaxEnt ranker. The subset of SemCor introduced in Section 3.1 is trained and tested with the features introduced in Section 3.2. We grouped the feature sets into two according to the two word representation of basic Semantic Dependencies (SD) and generalized Semantic Files (SF). Sometime two or more different parses of a sentence have the same set of features. That is, the features failed to distinguish between two parses: often because of spurious syntactic ambiguity that had no effect on the semantics. In this case we merged duplicate feature sets to reduce the ambiguity in machine learning. If an inactive parse has the same set of features as that of the active one, the resulting merged parse was treated as active.

Features	Accuracy (%)	Features ( $\times 1,000$ )
SD-Baseline	25.4	454
SD+LR	25.3	469
SD+PR	25.8	563
SD+LR+PR	25.6	582
SD+AF	24.8	1,430
SD+AF+LR	<b>27.1</b>	1,497
SD+AF+PR	25.8	1,761
SD+AF+LR+PR	26.3	1,842

Table 6: Parse selection results with SD.

Features	Accuracy (%)	Features ( $\times 1,000$ )
SF-Baseline	25.0	223
SF+LR	25.1	235
SF+PR	26.3	306
SF+LR+PR	26.3	321
SF+AF	<b>28.2</b>	1,051
SF+AF+LR	28.0	1,101
SF+AF+PR	28.1	1,310
SF+AF+LR+PR	27.7	1,375

Table 7: Parse selection results with SF.

We used TADM (Toolkit for Advanced Discriminative Modeling; Malouf, 2002) for the training and testing of our machine learning model, following Fujita et al. (2010). We carried out 10-fold cross-validation for evaluation. We measured the parse selection accuracy at the sentence level. A parse was considered correct only when all the dependencies of the parse are correct.

The results of parse selection based on SD and SF representations are shown in Tables 6 and 7. The addition of the ancestor features (AF) gives the most increase in the parse selection accuracy. This result indicates that indirect dependencies as well as direct dependencies in a successful parse frequently appear in other active parses. Second, the SF representation shows better results than the SD representation in most cases. The semantic abstraction of the semantic files reduces the problem of feature sparseness and is enough to effectively rerank parses, whose syntactic properties are already to some extent validated during parsing.

Third, the addition of the PR features also usually increases the parse selection accuracy. We plan to (semi-)automatically find more such multi-dependency structures whose combination shows better performance than the individual dependen-

cies. Fourth, the LR features do not improve the accuracy significantly in most cases, though the SD+AF+LR combination shows the best results among the feature sets of the SD representation. This is understandable since the number of the LR features in our corpus is much smaller than those of the other features of SD, PR and AF. We need to test it with a bigger corpus.

## 5 Discussion

These results show the validity of our assumption that long distance features and supertypes are both useful for selecting the correct interpretation of a sentence. Currently the SD+AF+LR model is the best for using the elementary predicates. However the best overall results come from the SF+AF model when we generalize to the semantic files. In future work we will investigate on larger-sized and more richly annotated corpora so that we can discover more about the relation between feature size and parse selection accuracy. In addition, we expect that increasing the corpus size will lead directly to higher accuracy. Other avenues we would like to explore is backing off not to the semantic files, but rather to WordNet hypernyms at various levels.

These results show that generalizing to semantic supertypes allows us to build semantic ranking models that are not only smaller, but more accurate. In general, learning time was roughly proportional to the number of features, so a smaller model can be learned faster. We hypothesize that it is the combination of dependencies and supertypes that makes the difference: approaches that used semantic features on phrase structure trees (such as Bikel (2000) and MacKinlay et al. (2012)) have in general failed to get much improvement.

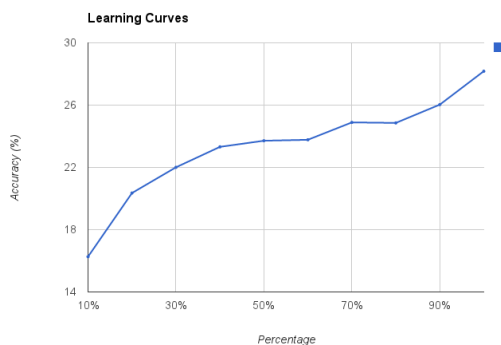


Figure 3: Learning curve for SF+AF.

The overall accuracy is still quite low, due principally to the lack of training data. We show the learning curves for the SF+AF configuration in Figure 3 (the other configurations are similar). The curve is still clearly rising: the accuracy of parse selection on our corpus is far from saturated. This observation gives us confidence that with a larger corpus the accuracy of parse selection will improve considerably. The learning curve in Fujita et al. (2010) showed similar results for the same amount of data, and increased rapidly with more (they had a larger corpus for Japanese).

As there are so far still very few corpora with both structural and lexical semantic annotation, we are currently investigating the use of automatic word sense disambiguation to create the features, in a similar way to Agirre et al. (2008). Finally, we would like to investigate even more features, such as the dependency chains of Le Roux et al. (2012).

One exciting possibility is projecting ranking features across languages: wordnet semantic files are the supertypes for all wordnets linked to the Princeton Wordnet, of which there are many (Bond and Foster, 2013). The predicates that are not in the wordnets are generally either named entities or from smallish closed sets of function words such as conjunctions, prepositions and pronouns. We are currently investigating mapping these between Japanese and English using transfer rules from an existing machine translation system (Bond et al., 2011). In principal, a small set of mappings for closed class words could allow us to quickly boot-strap a semantic ranking model for any language with a wordnet.

## 6 Conclusion

In summary, we showed some features that help parse selection. In the SD group, LR features together with AF features achieved a 1.75% improvement in accuracy over the basic Baseline model (25.36%  $\rightarrow$  27.12%). However, LR feature alone and AF feature alone both decrease the accuracy (25.36%  $\rightarrow$  25.28% and 25.36%  $\rightarrow$  24.84%). PR features and combination of PR and AF features both achieved small improvements (0.416% Baseline  $\rightarrow$  Baseline+PR, 0.410% Baseline  $\rightarrow$  Baseline-PR+AF). LR combined with PR features did not improve the accuracy.

When features get generalized to supertypes, as shown in the SF group, models with more features achieved higher accuracies with the best be-

ing the model with ancestor features (AF) added. This (SF+AF) achieved an improvement of 3.21% absolute over the baseline model (24.97% → 28.18%). Adding more features to AF only decreases the accuracy. Generalizing to semantic supertypes allows us to build dependency ranking models that are not only smaller, but more accurate.

## Acknowledgments

The authors are grateful to Mathieu Morey and other members of the Deep Linguistic Processing with HPSG Initiative along with other members of their research groups for many extremely helpful discussions.

## References

- Eneko Agirre, Timothy Baldwin, and David Martinez. 2008. Improving parsing and PP attachment performance with sense information. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL HLT 2008)*, pages 317–325. Columbus, USA.
- Eneko Agirre, Kepa Bengoetxea, Koldo Gojenola, and Joakim Nivre. 2011. Improving dependency parsing with semantic classes. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 699–703.
- Daniel M. Bikel. 2000. A statistical model for parsing and word-sense disambiguation. In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 155–163. Hong Kong.
- Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In *51st Annual Meeting of the Association for Computational Linguistics: ACL-2013*, pages 1352–1362. Sofia. URL <http://aclweb.org/anthology/P13-1133>.
- Francis Bond, Sanae Fujita, and Takaaki Tanaka. 2008. The Hinoki syntactic and semantic treebank of Japanese. *Language Resources and Evaluation*, 42(2):243–251. URL <http://dx.doi.org/10.1007/s10579-008-9062-z>, (Re-issue of DOI 10.1007/s10579-007-9036-6 as Springer lost the Japanese text).
- Francis Bond, Stephan Oepen, Eric Nichols, Dan Flickinger, Erik Velldal, and Petter Haugereid. 2011. Deep open source machine translation. *Machine Translation*, 25(2):87–105. URL <http://dx.doi.org/10.1007/s10590-011-9099-4>, (Special Issue on Open source Machine Translation).
- Miriam Butt, Tracy Holloway King, María-Eugenia Niño, and Frédérique Segond. 1999. *A Grammar Writer’s Cookbook*. CSLI publications.
- Michael Collins and Terry Koo. 2006. Discriminative reranking for natural language parsing. *Computational Linguistics*, 31(1).
- Ann Copestake. 2009. Slacker semantics: Why superficiality, dependency and avoidance of commitment can be the right way to go. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 1–9. Athens.
- Ann Copestake, Dan Flickinger, Ivan A. Sag, and Carl Pollard. 2005. Minimal Recursion Semantics. An introduction. *Research on Language and Computation*, 3(4):281–332.
- Dan Flickinger. 2000. On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1):15–28.
- Sanae Fujita, Francis Bond, Takaaki Tanaka, and Stephan Oepen. 2010. Exploiting semantic information for HPSG parse selection. *Research on Language and Computation*, 8(1):1–22. URL <http://dx.doi.org/10.1007/s11168-010-9069-7>.
- Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, Akio Yokoo, Hiromi Nakaiwa, Kentaro Ogura, Yoshifumi Ooyama, and Yoshihiko Hayashi. 1997. *Goi-Taikei — A Japanese Lexicon*. Iwanami Shoten, Tokyo. 5 volumes/CDROM.
- Shari Landes, Claudia Leacock, and Christiane Fellbaum. 1998. Building semantic concordances. In Christine Fellbaum, editor, *WordNet: An Electronic Lexical Database*, chapter 8, pages 199–216. MIT Press.
- Joseph Le Roux, Benoit Favre, Alexis Nasr, and Seyed Abolghasem Mirroshandel. 2012. Generative constituent parsing and discriminative dependency reranking: Experiments on english and french. In *Proceedings of the ACL 2012 Joint Workshop on Statistical Parsing and Semantic Processing of Morphologically Rich Languages*, pages 89–99. Association for Computational Linguistics, Jeju, Republic of Korea. URL <http://www.aclweb.org/anthology/W12-3412>.
- Andrew MacKinlay, Rebecca Dridan, Diana McCarthy, and Timothy Baldwin. 2012. The effects of semantic annotations on precision parse ranking. In *SEM 2012: The First Joint Conference on Lexical and Computational Semantics-*, volume 2, pages 228–236. Association for Computational Linguistics.
- Robert Malouf. 2002. A comparison of algorithms for maximum entropy parameter estimation. In *CONLL-2002*, pages 49–55. Taipei, Taiwan.
- David McClosky, Eugene Charniak, and Mark Johnson. 2006. Reranking and self-training for parser adaptation. In *44th Annual Meeting of the Association for Computational Linguistics and 21st International Conference on Computational Linguistics: COLING/ACL-2006*, pages 337–344.
- Stephan Oepen, Dan Flickinger, Kristina Toutanova, and Christopher D. Manning. 2004. LinGO Redwoods: A rich and dynamic treebank for HPSG. *Research on Language and Computation*, 2(4):575–596.
- Zinaida Pozen. 2013. *Using Lexical and Compositional Semantics to Improve HPSG Parse Selection*. Master’s thesis, University of Washington. URL <https://digital.lib.washington.edu/researchworks/handle/1773/%23469>.
- Ivan A. Sag, Tom Wasow, and Emily Bender. 1999. *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, second edition.
- Kristina Toutanova, Christopher D. Manning, Dan Flickinger, and Stephan Oepen. 2005. Stochastic HPSG parse disambiguation using the Redwoods corpus. *Research on Language and Computation*, 3(1):83–105.