

# Evaluating Contribution of Deep Syntactic Information to Shallow Semantic Analysis

Sumire Uematsu    Jun'ichi Tsujii

Graduate School of Information Science and Technology

The University of Tokyo

{uematsu,tsujii}@is.s.u-tokyo.ac.jp

## Abstract

This paper presents shallow semantic parsing based only on HPSG parses. An HPSG-FrameNet map was constructed from a semantically annotated corpus, and semantic parsing was performed by mapping HPSG dependencies to FrameNet relations. The semantic parsing was evaluated in a Senseval-3 task; the results suggested that there is a high contribution of syntactic information to semantic analysis.

## 1 Introduction

This paper presents semantic parsing based only on HPSG parses, and examines the contribution of the syntactic information to semantic analysis.

In computational linguistics, many researchers have studied the relationship between syntax and semantics. Its quantitative analysis was formalized as semantic parsing, or semantic role labeling, and has attracted the attention of researchers.

Recently, an improvement in the accuracy and robustness of “deep parsers” has enabled us to directly map deep syntactic dependencies to semantic relations. Deep parsers are based on linguistically expressive grammars; e.g. HPSG, LFG, etc, and less affected by syntactic alternations such as passivization. Their results are therefore expected to closely relate to semantic annotations. For example, the sentences in figure 1 share the same set of semantic roles, and the roles have one-to-one relations to deep syntactic dependencies in the sentences. However, the results of the deep parsers are represented in complex structures, shown in figure 3, and cannot be straightforwardly compared to semantic annotations.

In order to directly map the deep dependencies to semantic relations, we adapted the corpus analysis method of (Frank and Semecký, 2004) for the semantic parsing using HPSG parses. We performed the semantic parsing by mapping paths in

HPSG parses to semantic predicate-argument relations. The analysis of the HPSG paths for the predicate-argument pairs, and the preliminary result of the semantic parsing indicate the contribution of syntactic analysis to semantic parsing.

## 2 Related Work

Besides (Frank and Semecký, 2004)'s work, as mentioned above, there have been several studies on the relationship between deep syntax and semantic parsing. Although the studies did not focus on direct mappings between deep syntax and shallow semantics, they suggested a strong relationship between the two. (Miyao and Tsujii, 2004) evaluated the accuracy of an HPSG parser against PropBank semantic annotations, and showed that the HPSG dependants correlated with semantic arguments of the PropBank, particularly with “core” arguments. In (Gildea and Hockenmaier, 2003) and (Zhang et al., 2008), features from deep parses were used for semantic parsing, together with features from CFG or dependency parses. The deep features were reported to contribute to a performance gain.

## 3 Syntactic and Semantic Parsing

Some semantic relations are easily identified by using syntactic parsing while others are more difficult. This section presents easy and difficult cases in syntax-semantics map construction.

**Trivial when using syntactic analysis:** Syntactic parsing, including CFG analysis, detects semantic similarity of sentences sharing similar phrase structures. For the example sentences a) and b) in figure 1, the parsing provides similar phrase structures, and therefore gives the same syntactic dependency to occurrences of each role.

**Trivial when using deep analysis:** Deep parsing reveals the semantic similarity of sentences

- a) ...,  $I_{\text{Communicator}}$  **praise**  $\text{them}_{\text{Evaluee}}$  **for being 99 percent perfect**,  $\text{Reason}$   
 b) ..., but  $\text{he}_{\text{Communicator}}$  **praised the Irish premier**,  $\text{Evaluee}$  **for making a "sensible" speech**,  $\text{Reason}$   
 c) **The child**,  $\text{Evaluee}$  **is praised for having a dry bed**,  $\text{Reason}$  and ...  
 d) ...,  $\text{She}_{\text{Communicator}}$  was supposed, therefore, to **praise him**,  $\text{Evaluee}$  and then ...

- e)  $\text{It}_{\text{Evaluee}}$  received high **praise**, ...  
 f)  $\text{Alice}_{\text{Wearer}}$ 's **dress**  
 g) Versace's **dress**

Figure 1: Sentences with a set of semantic roles for the predicate *praise*.

Figure 2: Example phrases for section 3.

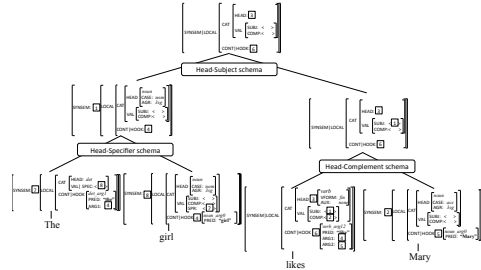


Figure 3: An HPSG parse for *The girl likes Mary*.

containing complex syntactic phenomena, which is not easily detected by CFG analysis. The sentences c) and d) in figure 1 contain passivization and object raising, while deep parsing provides one dependency for each role in the figure.

**Not trivial even when using deep analysis:** Some semantic arguments are not direct syntactic dependants of their predicates - especially of noun predicates. In sentence e) in figure 2, the *Evaluee* phrase depends on the predicate *praise*, through the support verb *receive*. The deep analysis would be advantageous in capturing such dependencies, because it provides *receive* with direct links to the phrases of the role and the predicate.

**Problematic when using only syntactic analysis:** Sometimes, the semantic role of a phrase is strongly dependent on the type of the mentioned entity, rather than on the syntactic dependency. In phrases f) and g) in figure 2, the phrases *Alice* and *Versace*, have the same syntactic relation to the predicate *dress*. However, the *Wearer* role is given only to the former phrase.

#### 4 A Wide-Coverage HPSG Parser

We employed a wide-coverage HPSG parser for semantic parsing, and used deep syntactic dependencies encoded in a Predicate Argument Structure (PAS) in each parse node.

In our experiments, the parser results were considered as graphs, as illustrated by figures 3 and 4, to extract HPSG dependencies conveniently. The

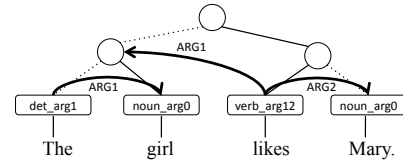


Figure 4: A simplified representation of figure 3.

graph is obtained by ignoring most of the linguistic information in the original parse nodes, and by adding edges directing to the PAS dependants. The PAS information is represented in the graph, by the terminal nodes' *PAS types*, e.g. *verb\_arg12*, etc., and by the added edges. Note that the interpretation of the edge labels depends on the PAS type. If the PAS type is *verb\_arg12*, the *ARG2* dependant is the object of the transitive verb or its equivalence (the subject of the passive, etc.). If the PAS type is *prep\_arg12*, then the dependant is the NP governed by the preposition node.

#### 5 Semantic Parsing Based on FrameNet

We employed FrameNet (FN) as a semantic corpus. Furthermore, we evaluated our semantic parsing on the SRL task data of Senseval-3 (Litkowski, 2004), which consists of FN annotations.

In FN, semantic *frames* are defined, and each frame is associated with predicates that evoke the frame. For instance, the verb and noun *praise* are predicates of the *Judgment\_communication* frame, and they share the same set of semantic roles.

The Senseval-3 data is a standard for evaluation of semantic parsing. The task is defined as identifying phrases and their semantic roles for a given sentence, predicate, and frame. The data includes *null instantiations* of roles<sup>1</sup>, which are "conceptually salient", but do not appear in the text.

#### 6 Methods

The semantic parsing using an HPSG-FN map consisted of the processes shown in figure 5.

<sup>1</sup>An example of a null instantiation is the *Communicator* role in the sentence, "All in all the conference was **acclaimed** as a considerable success."

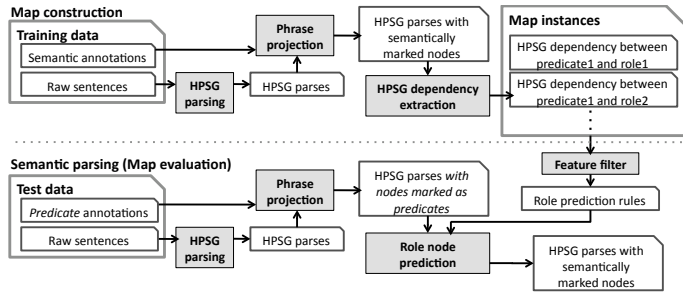


Figure 5: Processes in the map construction and evaluation.

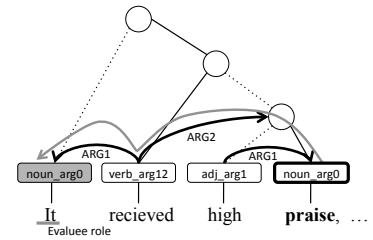


Figure 6: an HPSG path for a semantic relation.

<b>Predicate base:</b> The base form of the semantic predicate word. ( <i>praise</i> in the case of figure 6).
<b>Predicate type:</b> The <i>PAS type</i> of the HPSG terminal node for the predicate - see section 4. ( <i>noun_arg0</i> in figure 6).
<b>Intermediate word base:</b> The base form of the <i>intermediate word</i> , corresponding to a terminal passed by the path, and satisfying pre-defined conditions. The word may be a support verb. - see figure 6. ( <i>receive</i> in figure 6).
<b>Intermediate word type:</b> The <i>PAS type</i> of the <i>intermediate word</i> . ( <i>verb_arg12</i> in figure 6).
<b>Dependency label sequence:</b> The labels of the path's edges. We omitted labels presenting head-child relations, for identifying a phrase with another phrase sharing the same head word. ( <i>Reverse of ARG2, ARG1</i> in figure 6).

Table 1: Features used to represent a HPSG path.

Filter	Pred. base type	Inter. base type	Dep. label
Same	✓	✓	✓
AllInter	✓	✓	✓
AllPred		✓	✓
AllPred-AllInter	✓	✓	✓

Table 2: Syntactic features for role prediction.

**Phrase projection:** Because we used FN annotations, which are independent of any syntactic framework, role phrases needed to be projected to appropriate HPSG nodes. We projected the phrases based on *maximal projection*, which was generally employed, with heads defined in the HPSG.

**HPSG dependency extraction:** As an HPSG dependency for a predicate-argument pair, we used the shortest path between the predicate node and the argument node in the HPSG parse. The

path was then represented by pre-defined features, listed in table 1. The search for the shortest path was done in the simplified graph of the HPSG parse (see figure 4), with the edges denoting deep dependencies, and head-child relations. An instance of the HPSG-FN map consisted of the path's features, the *FN frame*, and the *role label*.

**Role node prediction:** The role prediction was based on simple rules with scores. The rules were obtained by filtering features of the map instances. Table 2 shows the feature filters. The score of a rule was the number of map instances matching the rule's features. In the test, for each node of a HPSG parse, the role label with the highest score was selected as the result, where the score of a label was that of the rule providing the label.

## 7 Experiments

For the experiments, we employed a wide coverage HPSG parser, Enju version 2.3.1<sup>2</sup>, and the data for the Semantic Role Labeling task of Senseval-3.

### 7.1 Analysis of Map Instances

We extracted 41,193 HPSG-FN map instances from the training set, the training data apart from the development set. The instances amounted to 97.7 % (41,193 / 42,163) of all the non-null instantiated roles in the set, and HPSG paths were short for many instances. Paths to syntactic arguments were almost directly mapped to semantic roles, while roles for other phrases were more ambiguous.

**The length distribution of HPSG paths:** 64 % (26410 / 41193) of the obtained HPSG paths were length-one, and 8 % (3390 / 41193) were length-two, due to the effect of direct links provided by HPSG parsing. The length of a path was defined

<sup>2</sup><http://www-tsujii.is.s.u-tokyo.ac.jp/enju/>

Pred.	Freq.	Feature representation	Interpretation
Verb	3792	verb_arg12/--/ARG2	The object of the transitive predicate
	3191	verb_arg12/--/ARG1	The subject of the transitive predicate
Noun	7468	noun_arg0/--/--	NP headed by the predicate
	1161	noun_arg0/of/prep_arg12/Rev-ARG1	The PP headed by “of”, attaching to the predicate
Adj	1595	adj_arg1/--/ARG1	The modifiee of the predicate
	274	verb_arg12/--/ARG2	The modifiee of the predicate treated as a verb

Table 3: Most frequent syntactic paths extracted for predicates of each POS.

as the number of the labels in the **Dep. label seq.** of the path. Most of the one-length paths were paths directing to syntactic arguments, and to PPs attaching to the predicates. The two-length paths included paths using support verbs (see figure 6).

**Most frequent HPSG dependencies:** The most frequent paths are shown in table 3; syntactic dependencies are presented and counted as tuples of **Pred. type, Inter. base, Inter. type, and Dep. label seq.** The interpretation column describes the syntactic dependencies for the tuples. Note that the column denotes normalized dependencies, in which *object* indicates objects of active voice verbs, subjects of passive-voiced verbs, etc.

## 7.2 Performance of Semantic Parsing

Finally, semantic parsing was evaluated on the test data. Table 4 shows the overall performance. The scores were measured by the Senseval-3 official script, in the *restrictive* setting, and can be directly compared to other systems’ scores. Since our preliminary system of semantic parsing ignored null instantiations of roles, it lost around 0.10 point of the recalls. We believe that such instantiations may be separately treated. Although the system was based on only the syntactic information, and was very naïve, the system’s performance was promising, and showed the high contribution of syntactic dependencies for semantic parsing.

## 8 Conclusion

This paper presents semantic parsing based on only HPSG parses, and investigates the contribution of syntactic information to semantic parsing.

We constructed an HPSG-FN map by finding the HPSG paths that corresponded to semantic relations, and used it as role prediction rules in semantic parsing. The semantic parsing was evaluated on the SRL task data of Senseval-3. Although the preliminary system used only the syntactic information, the performance was promising, and

Rule set	Prec.	Overlap	Recall
Same	0.799	0.783	0.518
AllInter	0.599	0.586	0.589
AllPred	0.472	0.462	0.709
AllPred-AllInter	0.344	0.335	0.712
Senseval-3 best	0.899	0.882	0.772
Senseval-3 4th best	0.802	0.784	0.654

Table 4: Semantic parsing result on the test data.

indicated that syntactic dependencies may make significant contribution to semantic analysis.

This paper also suggests a limit of the semantic analysis based purely on syntax. A next step for accurate HPSG-FN mapping could be analysis of the interaction between the HPSG-FN map and other information, such as named entity types which were shown to be effective in many studies.

## Acknowledgments

This work was partially supported by Grant-in-Aid for Specially Promoted Research (MEXT, Japan) and Special Coordination Funds for Promoting Science and Technology (MEXT, Japan).

## References

- Anette Frank and Jiří Semecký. 2004. Corpus-based induction of an LFG syntax-semantics interface for frame semantic processing. In *Proc. of International Workshop on Linguistically Interpreted Corpora*.
- Daniel Gildea and Julia Hockenmaier. 2003. Identifying semantic roles using combinatory categorial grammar. In *Proc. of EMNLP*.
- Ken Litkowski. 2004. Senseval-3 task: Automatic labeling of semantic roles. In *Proc. of Senseval-3*.
- Yusuke Miyao and Jun’ichi Tsujii. 2004. Deep linguistic analysis for the accurate identification of predicate-argument relations. In *Proc. of Coling*.
- Yi Zhang, Rui Wang, and Hans Uszkoreit. 2008. Hybrid learning of dependency structures from heterogeneous linguistic resources. In *Proc. of CoNLL*.