

# Constrained Semantic Forests for Improved Discriminative Semantic Parsing

Wei Lu

Information Systems Technology and Design  
Singapore University of Technology and Design  
luwei@sutd.edu.sg

## Abstract

In this paper, we present a model for improved discriminative semantic parsing. The model addresses an important limitation associated with our previous state-of-the-art discriminative semantic parsing model – the *relaxed hybrid tree* model by introducing our *constrained semantic forests*. We show that our model is able to yield new state-of-the-art results on standard datasets even with simpler features. Our system is available for download from <http://statnlp.org/research/sp/>.

## 1 Introduction

This paper addresses the problem of parsing natural language sentences into their corresponding semantic representations in the form of formal logical representations. Such a task is also known as *semantic parsing* (Kate and Mooney, 2006; Wong and Mooney, 2007; Lu et al., 2008; Kwiatkowski et al., 2010).

One state-of-the-art model for semantic parsing is our recently introduced *relaxed hybrid tree* model (Lu, 2014), which performs integrated lexicon acquisition and semantic parsing within a single framework utilizing efficient algorithms for training and inference. The model allows natural language phrases to be recursively mapped to semantic units, where certain long-distance dependencies can be captured. It relies on representations called *relaxed hybrid trees* that can jointly represent both the sentences and semantics. The model is essentially discriminative, and allows rich features to be incorporated.

Unfortunately, the relaxed hybrid tree model has an important limitation: it essentially does not allow certain sentence-semantics pairs to be jointly encoded using the proposed relaxed hybrid tree representations. Thus, the model is unable to identify joint representations for certain

sentence-semantics pairs during the training process, and is unable to produce desired outputs for certain inputs during the evaluation process. In this work, we propose a solution addressing the above limitation, which makes our model more robust. Through experiments, we demonstrate that our improved discriminative model for semantic parsing, even when simpler features are used, is able to obtain new state-of-the-art results on standard datasets.

## 2 Related Work

Semantic parsing has recently attracted a significant amount of attention in the community. In this section, we provide a relatively brief discussion of prior work in semantic parsing. The hybrid tree model (Lu et al., 2008) and the Bayesian tree transducer based model (Jones et al., 2012) are generative frameworks, which essentially assume natural language and semantics are jointly generated from an underlying generative process. Such models are efficient, but are limited in their predictive power due to the simple independence assumptions made.

On the other hand, discriminative models are able to exploit arbitrary features and are usually able to give better results. Examples of such models include the WASP system (Wong and Mooney, 2006) which regards the semantic parsing problem as a statistical machine translation problem, the UBL system (Kwiatkowski et al., 2010) which performs CCG-based semantic parsing using a log-linear model, as well as the *relaxed hybrid tree* model (Lu, 2014) which extends the generative hybrid tree model. This extension results in a discriminative model that incorporates rich features and allows long-distance dependencies to be captured. The relaxed hybrid tree model has achieved the state-of-the-art results on standard benchmark datasets across different languages.

Performing semantic parsing under other forms

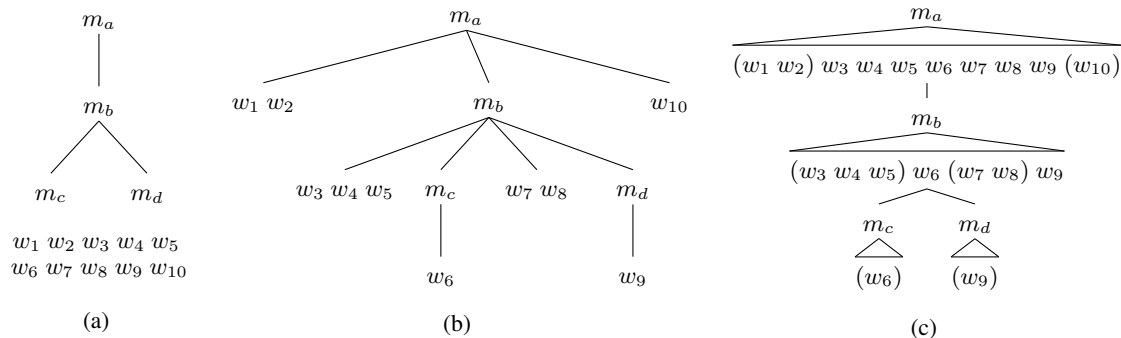


Figure 1: The semantics-sentence pair (a), an example *hybrid tree* (b), and an example *relaxed hybrid tree* (c).

of supervision is also possible. Clarke et al. (2010) proposed a model that learns a semantic parser for answering questions without relying on semantic annotations. Goldwasser et al. (2011) presented a confidence-driven approach to semantic parsing based on self-training. Liang et al. (2013) introduced semantic parsers based on dependency based semantics (DCS) that map sentences into their denotations. In this work, we focus on parsing sentences into their formal semantic representations.

### 3 Relaxed Hybrid Trees

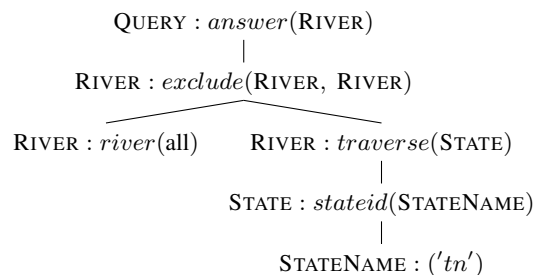
We briefly discuss our previously proposed *relaxed hybrid tree* model (Lu, 2014) in this section. The model is a discriminative semantic parsing model which extends the generative *hybrid tree* model (Lu et al., 2008). Both systems are publicly available<sup>1</sup>.

Let us use  $\mathbf{m}$  to denote a complete semantic representation,  $\mathbf{n}$  to denote a complete natural language sentence, and  $\mathbf{h}$  to denote a complete latent structure that jointly represents both  $\mathbf{m}$  and  $\mathbf{n}$ . The model defines the conditional probability for observing a  $(\mathbf{m}, \mathbf{h})$  pair for a given natural language sentence  $\mathbf{n}$  using a log-linear approach:

$$P_{\Lambda}(\mathbf{m}, \mathbf{h}|\mathbf{n}) = \frac{e^{\Lambda \cdot \Phi(\mathbf{n}, \mathbf{m}, \mathbf{h})}}{\sum_{\mathbf{m}', \mathbf{h}' \in \mathcal{H}(\mathbf{n}, \mathbf{m}')} e^{\Lambda \cdot \Phi(\mathbf{n}, \mathbf{m}', \mathbf{h}')}} \quad (1)$$

where  $\Lambda$  is the set of parameters (weights of features) used by the model. Figure 1 (a) gives an example sentence-semantics pair. A real example taken from the GeoQuery dataset is shown in Figure 2.

Note that  $\mathbf{h}$  is a complete latent structure that jointly represents a natural language sentence and



What rivers do not run through Tennessee ?

Figure 2: An example tree-structured semantic representation (above) and its corresponding natural language sentence (below).

its corresponding semantic representation. Typically, to limit the space of latent structures, certain assumptions have to be made to  $\mathbf{h}$ . In our work, we assume that  $\mathbf{h}$  must be from a space consisting of *relaxed hybrid tree* structures (Lu, 2014).

The relaxed hybrid trees are analogous to the hybrid trees, which was earlier introduced as a generative framework. One major distinction between these two types of representations is that the relaxed hybrid tree representations are able to capture unbounded long-distance dependencies in a principled way. Such dependencies were unable to be captured by hybrid tree representations largely due to their generative settings. Figure 1 gives an example of a hybrid tree and a relaxed hybrid tree representation encoding the sentence  $w_1 w_2 w_3 w_4 w_5 w_6 w_7 w_8 w_9 w_{10}$  and the semantics  $m_a(m_b(m_c, m_d))$ .

In the hybrid tree structure, each word is strictly associated with a semantic unit. For example the word  $w_3$  is associated with the semantic unit  $m_b$ . In the relaxed hybrid tree, however, each word is not only directly associated with exactly one semantic unit  $m$ , but also indirectly associated with

<sup>1</sup><http://statnlp.org/research/sp/>

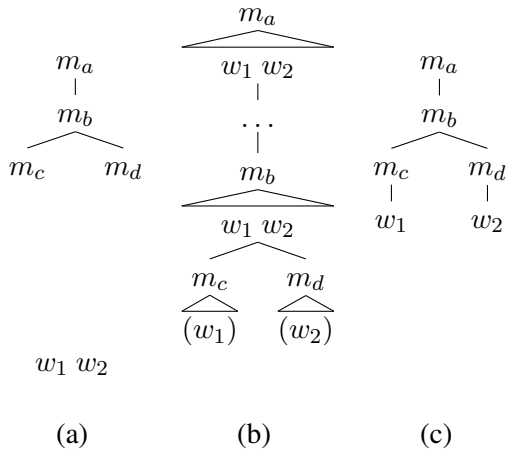


Figure 3: (a) Example semantics-sentence pair that cannot be jointly represented with *relaxed hybrid trees* if pattern **X** is disallowed. (b) Example *relaxed hybrid tree* that consists of an infinite number of nodes when pattern **X** is allowed. (c) Example *hybrid tree* jointly representing both the semantics and the sentence (where pattern **X** is allowed).

all other semantic units that are predecessors of  $m$ . For example, the word  $w_3$  now is directly associated with  $m_b$ , but is also indirectly associated with  $m_a$ . These indirect associations allow the long-distance dependencies to be captured.

Both the hybrid tree and relaxed hybrid tree models define *patterns* at each level of their latent structure which specify how the words and child semantic units are organized at each level. For example, within the semantic unit  $m_a$ , we have a pattern  $\mathbf{wXw}$  which states that we first have words that are directly associated with  $m_a$ , followed by some words covered by its first child semantic unit, then another sequence of words directly associated with  $m_a$ .

### 3.1 Limitations

One important difference between the hybrid tree representations and the relaxed hybrid tree representations is the exclusion of the pattern **X** in the latter. This ensured relaxed hybrid trees with an infinite number of nodes were not considered (Lu, 2014) when computing the denominator term of Equation 1. In relaxed hybrid tree,  $\mathcal{H}(\mathbf{n}, \mathbf{m})$  was implemented as a packed forest representation for exponentially many possible relaxed hybrid trees where pattern **X** was excluded.

By allowing pattern **X**, we allow certain semantic units with no natural language word counter-

#Args	Patterns
0	$\mathbf{w}$
1	$[\mathbf{w}]\mathbf{X}[\mathbf{w}]$
2	$[\mathbf{w}]\mathbf{X}[\mathbf{w}]\mathbf{Y}[\mathbf{w}], [\mathbf{w}]\mathbf{Y}[\mathbf{w}]\mathbf{X}[\mathbf{w}]$

Table 1: The patterns allowed for our model.  $[\mathbf{w}]$  denotes an optional sequence of natural language words. *E.g.*,  $[\mathbf{w}]\mathbf{X}[\mathbf{w}]$  refers to the following 4 patterns:  $\mathbf{wX}$ ,  $\mathbf{Xw}$ ,  $\mathbf{wXw}$ , and  $\mathbf{X}$  (the pattern excluded by the relaxed hybrid tree model).

part to exist in the joint relaxed hybrid tree representation. This may lead to possible relaxed hybrid tree representations consisting of an infinite number of internal nodes (semantic units), as seen in Figure 3 (b). When pattern **X** is allowed, both  $m_a$  and  $m_b$  are not directly associated with any natural language word, so we are able to further insert arbitrarily many (compatible) semantic units between the two units  $m_a$  and  $m_b$  while the resulting relaxed hybrid tree remains valid. Therefore we can construct a relaxed hybrid tree representation that contains the given natural language sentence  $w_1 w_2$  with an infinite number of nodes. This issue essentially prevents us from computing the denominator term of Equation 1 since it involves an infinite number of possible  $\mathbf{m}'$  and  $\mathbf{h}'$ .

To eliminate relaxed hybrid trees consisting of an infinite number of nodes, pattern **X** is disallowed in the relaxed hybrid trees model (Lu, 2014). However, disallowing pattern **X** has led to other issues. Specifically, for certain semantics-sentence pairs, it is not possible to find relaxed hybrid trees that jointly represent them. In the example semantics-sentence pair given in Figure 3 (a), it is not possible to find any relaxed hybrid tree that contains both the sentence and the semantics since each semantic unit which takes one argument must be associated with at least one word. On the other hand, it is still possible to find a hybrid tree representation for both the sentence and the semantics where pattern **X** is allowed (see Figure 3 (c)).

In practice, we can alleviate this issue by extending the lengths of the sentences. For example, we can append the special beginning-of-sentence symbol  $\langle s \rangle$  and end-of-sentence symbol  $\langle /s \rangle$  to all sentences to increase their lengths, allowing the relaxed hybrid trees to be constructed for certain sentence-semantics pairs with short sentences. However, such an approach does not resolve the theoretical limitation of the model.

## 4 Constrained Semantic Forests

To address this limitation, we allow pattern  $\mathbf{X}$  to be included when building our new discriminative semantic parsing model. However, as mentioned above, doing so will lead to latent structures (relaxed hybrid tree representations) of infinite heights. To resolve such an issue, we instead add an additional constraint – limiting the height of a semantic representation to a fixed constant  $c$ , where  $c$  is larger than the maximum height of all the trees appearing in the training set.

Table 1 summarizes the list of patterns that our model considers. This is essentially the same as those considered by the hybrid tree model.

Our new objective function is as follows:

$$P_{\Lambda}(\mathbf{m}, \mathbf{h}|\mathbf{n}) = \frac{e^{\Lambda \cdot \Phi(\mathbf{n}, \mathbf{m}, \mathbf{h})}}{\sum_{\mathbf{m}' \in \mathcal{M}, \mathbf{h}' \in \mathcal{H}'(\mathbf{n}, \mathbf{m}')} e^{\Lambda \cdot \Phi(\mathbf{n}, \mathbf{m}', \mathbf{h}')}} \quad (2)$$

where  $\mathcal{M}$  refers to the set of all possible semantic trees whose heights are less than or equal to  $c$ , and  $\mathcal{H}'(\mathbf{n}, \mathbf{m}')$  refers to the set of possible relaxed hybrid tree representations where the pattern  $\mathbf{X}$  is allowed.

The main challenge now becomes the computation of the denominator term in Equation 2, as the set  $\mathcal{M}$  is still very large. To properly handle all such semantic trees in an efficient way, we introduce a *constrained semantic forest* (CSF) representation of  $\mathcal{M}$  here. Such a constrained semantic forest is a packed forest representation of exponentially many possible unique semantic trees, where we set the height of the forest to  $c$ . By contrast, it was not possible in our previous relaxed hybrid tree model to introduce such a compact representation over all possible semantic trees. In our previous model’s implementation, we directly constructed for each sentence  $\mathbf{n}$  a different compact representation over all possible *relaxed hybrid trees* containing  $\mathbf{n}$ .

Setting the maximum height to  $c$  effectively guarantees that all semantic trees contained in the constrained semantic forest have a height no greater than  $c$ . We then constructed the (exponentially many) relaxed hybrid tree representations based on the constrained semantic forest  $\mathcal{M}$  and each input sentence  $\mathbf{n}$ . We used a single packed forest representation to represent all such relaxed hybrid tree representations. This allows the computation of the denominator to be performed ef-

ficiently using similar dynamic programming algorithms described in (Lu, 2014). Optimization of the model parameters were done by using L-BFGS (Liu and Nocedal, 1989), where the gradients were computed efficiently using an analogous dynamic programming algorithm.

## 5 Experiments

Our experiments were conducted on the publicly available multilingual GeoQuery dataset. Various previous works on semantic parsing used this dataset for evaluations (Wong and Mooney, 2006; Kate and Mooney, 2006; Lu et al., 2008; Jones et al., 2012). The dataset consists of 880 natural language sentences where each sentence is coupled with a formal tree-structured semantic representation. The early version of this dataset was annotated with English only (Wong and Mooney, 2006; Kate and Mooney, 2006), and Jones et al. (2012) released a version that is annotated with three additional languages: German, Greek and Thai. To make our system directly comparable to previous works, we used the same train/test split used in those works (Jones et al., 2012; Lu, 2014) for evaluation. We also followed the standard approach for evaluating the correctness of an output semantic representation from our system. Specifically, we used a standard script to construct Prolog queries based on the outputs, and used the queries to retrieve answers from the GeoQuery database. Following previous works, we regarded an output semantic representation as correct if and only if it returned the same answers as the gold standard (Jones et al., 2012; Lu, 2014).

The results of our system as well as those of several previous systems are given in Table 2. We compared our system’s performance against those of several previous works. The WASP system (Wong and Mooney, 2006) is based on statistical machine translation technique while the HYBRIDTREE+ system (Lu et al., 2008) is based on the generative hybrid tree model augmented with a discriminative re-ranking stage where certain global features are used. UBL-S (Kwiatkowski et al., 2010) is a CCG-based semantic parsing system. TREETRANS (Jones et al., 2012) is the system based on tree transducers. RHT (Lu, 2014) is the discriminative semantic parsing system based on relaxed hybrid trees.

In practice, we set  $c$  (the maximum height of a semantic representation) to 20 in our experi-

System	English		Thai		German		Greek	
	Acc.	F	Acc.	F	Acc.	F	Acc.	F
WASP	71.1	77.7	71.4	75.0	65.7	74.9	70.7	78.6
HYBRIDTREE+	76.8	81.0	73.6	76.7	62.1	68.5	69.3	74.6
UBL-S	82.1	82.1	66.4	66.4	75.0	75.0	73.6	73.7
TREETRANS	79.3	79.3	78.2	78.2	74.6	74.6	75.4	75.4
RHT ( <i>all features</i> )	83.6	83.6	79.3	79.3	74.3	74.3	78.2	78.2
This work	<b>86.8</b>	<b>86.8</b>	<b>80.7</b>	<b>80.7</b>	<b>75.7</b>	<b>75.7</b>	<b>79.3</b>	<b>79.3</b>

Table 2: Performance of various works across four different languages. Acc.: accuracy percentage, F:  $F_1$ -measure percentage.

ments, which we determined based on the heights of the semantic trees that appear in the training data. Results showed that our system consistently yielded higher results than all the previous systems, including our state-of-the-art relaxed hybrid tree system (the full model, when all the features are used), in terms of both accuracy score and  $F_1$ -measure. We would like to highlight two potential advantages of our new model over the old RHT model. First, our model is able to handle certain sentence-semantics pairs which could not be handled by RHT during both training and evaluation as discussed in Section 3.1. Second, our model considers the additional pattern  $\mathbf{X}$  and therefore has the capability to capture more accurate dependencies between the words and semantic units.

We note that in our experiments we used a small subset of the features used by our relaxed hybrid tree work. Specifically, we did not use any long-distance features, and also did not use any character-level features. As we have mentioned in (Lu, 2014), although the RHT model is able to capture unbounded long-distance dependencies, for certain languages such as German such long-distance features appeared to be detrimental to the performance of the system (Lu, 2014, Table 4). Here in this work, we only used simple unigram features (concatenation of a semantic unit and an individual word that appears directly below that unit in the joint representation), pattern features (concatenation of a semantic unit and the pattern below that unit) as well as transition features (concatenation of two semantic units that form a parent-child relationship) described in (Lu, 2014). While additional features could potentially lead to better results, using simpler features would make our model more compact and more interpretable. We summarized in Table 3 the number of features used in both the previous RHT system and our system across four different languages. It can be seen that our system only required about 2-3% of the

System	English	Thai	German	Greek
RHT	$2.1 \times 10^6$	$2.3 \times 10^6$	$2.7 \times 10^6$	$2.6 \times 10^6$
This work	$5.4 \times 10^4$	$5.2 \times 10^4$	$7.5 \times 10^4$	$6.9 \times 10^4$

Table 3: Number of features involved for both the RHT system and our new system using constrained semantic forests, across four different languages.

features used in the previous system.

We also note that the training time for our model is longer than that of the relaxed hybrid tree model since the space for  $\mathcal{H}'(\mathbf{n}, \mathbf{m}')$  is now much larger than the space for  $\mathcal{H}(\mathbf{n}, \mathbf{m}')$ . In practice, to make the overall training process faster, we implemented a parallel version of the original RHT algorithm.

## 6 Conclusion

In this work, we presented an improved discriminative approach to semantic parsing. Our approach does not have the theoretical limitation associated with our previous state-of-the-art approach. We demonstrated through experiments that our new model was able to yield new state-of-the-art results on a standard dataset across four different languages, even though simpler features were used. Since our new model involves simpler features, including unigram features defined over individual semantic unit – word pairs, we believe our new model would aid the joint modeling of both distributional and logical semantics (Lewis and Steedman, 2013) within a single framework. We plan to explore this avenue in the future.

## Acknowledgments

The author would like to thank the anonymous reviewers for their helpful comments. This work was supported by SUTD grant SRG ISTD 2013 064 and was partially supported by project 61472191 under the National Natural Science Foundation of China.

## References

- James Clarke, Dan Goldwasser, Ming-Wei Chang, and Dan Roth. 2010. Driving semantic parsing from the world's response. In *Proc. of CONLL '10*, pages 18–27.
- Dan Goldwasser, Roi Reichart, James Clarke, and Dan Roth. 2011. Confidence driven unsupervised semantic parsing. In *Proc. of ACL '11*, pages 1486–1495.
- Bevan Keeley Jones, Mark Johnson, and Sharon Goldwater. 2012. Semantic parsing with bayesian tree transducers. In *Proc. of ACL '12*, pages 488–496.
- Rohit J. Kate and Raymond J. Mooney. 2006. Using string-kernels for learning semantic parsers. In *Proc. of COLING/ACL*, pages 913–920.
- Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing probabilistic ccg grammars from logical form with higher-order unification. In *Proc. EMNLP'10*, pages 1223–1233.
- Mike Lewis and Mark Steedman. 2013. Combining distributional and logical semantics. *Transactions of the Association for Computational Linguistics*, 1:179–192.
- Percy Liang, Michael I Jordan, and Dan Klein. 2013. Learning dependency-based compositional semantics. *Computational Linguistics*, 39(2):389–446.
- D. C. Liu and J. Nocedal. 1989. On the limited memory bfgs method for large scale optimization. *Math. Program.*, 45(3):503–528, December.
- Wei Lu, Hwee Tou Ng, Wee Sun Lee, and Luke S. Zettlemoyer. 2008. A generative model for parsing natural language to meaning representations. In *Proc. of EMNLP '08*, pages 783–792.
- Wei Lu. 2014. Semantic parsing with relaxed hybrid trees. In *Proc. of EMNLP '14*.
- Yuk Wah Wong and Raymond J. Mooney. 2006. Learning for semantic parsing with statistical machine translation. In *Proc. of HLT/NAACL '06*, pages 439–446.
- Yuk Wah Wong and Raymond J Mooney. 2007. Learning synchronous grammars for semantic parsing with lambda calculus. In *Proc. of ACL '07*.