Learning Cross-lingual Distributed Logical Representations for Semantic Parsing

Yanyan Zou and **Wei Lu**
Singapore University of Technology and Design
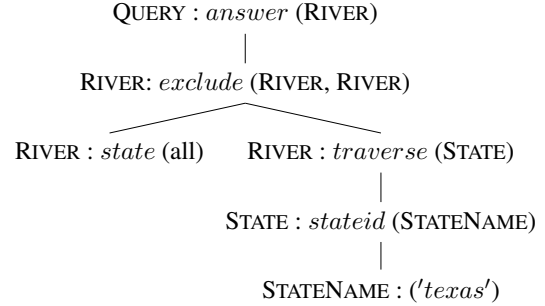8 Somapah Road, Singapore, 487372
`yanyan_zou@mymail.sutd.edu.sg, luwei@sutd.edu.sg`

Abstract

With the development of several multilingual datasets used for semantic parsing, recent research efforts have looked into the problem of learning semantic parsers in a multilingual setup (Duong et al., 2017; Susanto and Lu, 2017a). However, how to improve the performance of a monolingual semantic parser for a specific language by leveraging data annotated in different languages remains a research question that is under-explored. In this work, we present a study to show how learning distributed representations of the logical forms from data annotated in different languages can be used for improving the performance of a monolingual semantic parser. We extend two existing monolingual semantic parsers to incorporate such cross-lingual distributed logical representations as features. Experiments show that our proposed approach is able to yield improved semantic parsing results on the standard multilingual GeoQuery dataset.

1 Introduction

Semantic parsing, one of the classic tasks in natural language processing (NLP), has been extensively studied in the past few years (Zettlemoyer and Collins, 2005; Wong and Mooney, 2006, 2007; Liang et al., 2011; Kwiatkowski et al., 2011; Artzi et al., 2015). With the development of datasets annotated in different languages, learning semantic parsers from such multilingual datasets also attracted attention of researchers (Susanto and Lu, 2017a). However, how to make use of such cross-lingual data to perform cross-lingual semantic parsing – using data annotated for one language to help improve the performance of another lan-

QUERY : *answer* (RIVER)
|
RIVER: *exclude* (RIVER, RIVER)
RIVER : *state* (all)   RIVER : *traverse* (STATE)
|
STATE : *stateid* (STATENAME)
|
STATENAME : ($'texas'$)

English: which rivers do not run through texas ?
German: welche flüsse fliessen nicht durch texas ?

Figure 1: An example of two semantically equivalent sentences (below) and their tree-shaped semantic representation (above).

guage remains a research question that is largely under-explored.

Prior work (Chan et al., 2007) shows that semantically equivalent words coming from different languages may contain shared semantic level information, which will be helpful for certain semantic processing tasks. In this work, we propose a simple method to learn the distributed representations for output structured semantic representations which allow us to capture cross-lingual features. Specifically, following previous work (Wong and Mooney, 2006; Jones et al., 2012; Susanto and Lu, 2017b), we adopt a commonly used tree-shaped form as the underlying meaning representation where each tree node is a semantic unit. Our objective is to learn for each semantic unit a distributed representation useful for semantic parsing, based on multilingual datasets. Figure 1 depicts an instance of such tree-shaped semantic representations, which correspond to the two semantically equivalent sentences in English and German below it.

For such structured semantics, we consider each semantic unit separately. We learn distributed rep-

resentations for individual semantic unit based on multilingual datasets where semantic representations are annotated with different languages. Such distributed representations capture shared information cross different languages. We extend two existing monolingual semantic parsers (Lu, 2015; Susanto and Lu, 2017b) to incorporate such cross-lingual features. To the best of our knowledge, this is the first work that exploits cross-lingual embeddings for logical representations for semantic parsing. Our system is publicly available at http://statnlp.org/research/sp/.

2 Related Work

Many research efforts on semantic parsing have been made, such as mapping sentences into lambda calculus forms based on CCG (Artzi and Zettlemoyer, 2011; Artzi et al., 2014; Kwiatkowski et al., 2011), modeling dependency-based compositional semantics (Liang et al., 2011; Zhang et al., 2017), or transforming sentences into tree structured semantic representations (Lu, 2015; Susanto and Lu, 2017b). With the development of multilingual datasets, systems for multilingual semantic parsing are also developed. Jie and Lu (2014) employed majority voting to combine outputs from different parsers for certain languages to perform multilingual semantic parsing. Susanto and Lu (2017a) presented an extension of one existing neural parser, SEQ2TREE (Dong and Lapata, 2016), by developing a shared attention mechanism for different languages to conduct multilingual semantic parsing. Such a model allows two types of input signals: single source SL-SINGLE and multi-source SL-MULTI. However, semantic parsing with cross-lingual features has not been explored, while many recent works in various NLP tasks show the effectiveness of shared information cross different languages. Examples include semantic role labeling (Kozhevnikov and Titov, 2013), information extraction (Wang et al., 2013; Pan et al., 2017; Ni et al., 2017), and question answering (Joty et al., 2017), which motivate this work.

Our work involves exploiting distributed output representations for improved structured predictions, which is in line with works of (Srikumar and Manning, 2014; Rocktäschel et al., 2014; Xiao and Guo, 2015). The work of (Rocktäschel et al., 2014) is perhaps the most related to this research. The authors first map first-order logical statements produced by a semantic parser or an information extraction system into expressions in tensor calculus. They then learn low-dimensional embeddings of such statements with the help of a given logical knowledge base consisting of first-order rules so that the learned representations are consistent with these rules. They adopt stochastic gradient descent (SGD) to conduct optimizations. This work learns distributed representations of logical forms from cross-lingual data based on co-occurrence information without relying on external knowledge bases.

3 Approach

3.1 Semantic Parser

In this work, we build our model and conduct experiments on top of the discriminative hybrid tree semantic parser (Lu, 2014, 2015). The parser was designed based on the hybrid tree representation (HT-G) originally introduced in (Lu et al., 2008). The hybrid tree is a joint representation encoding both sentence and semantics that aims to capture the interactions between words and semantic units. A discriminative hybrid tree (HT-D) (Lu, 2014, 2015) learns the optimal latent word-semantics correspondence where every word in the input sentence is associated with a semantic unit. Such a model allows us to incorporate rich features and long-range dependencies. Recently, Susanto and Lu (2017b) extended HT-D by attaching neural architectures, resulting in their neural hybrid tree (HT-D (NN)).

Since the correct correspondence between words and semantics is not explicitly given in the training data, we regard the hybrid tree representation as a latent variable. Formally, for each sentence $\mathbf{n}$ with its semantic representation $\mathbf{m}$ from the training set, we assume the joint representation (a hybrid tree) is $\mathbf{h}$. Now we can define a discriminative log-linear model as follows:

$$P_\Lambda(\mathbf{m}|\mathbf{n}) = \sum_{\mathbf{h} \in \mathcal{H}(\mathbf{n},\mathbf{m})} P_\Lambda(\mathbf{m},\mathbf{h}|\mathbf{n})$$

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

$$F_\Lambda(\mathbf{n},\mathbf{m},\mathbf{h})) = \Lambda \cdot \Phi(\mathbf{n},\mathbf{m},\mathbf{h})) \quad (2)$$

where $\mathcal{H}(\mathbf{n},\mathbf{m})$ returns the set of all possible joint representations that contain both $\mathbf{n}$ and $\mathbf{m}$ exactly, and F is a scoring function that is calculated as a

dot product between a feature function Φ defined over tuple $(\mathbf{m}, \mathbf{n}, \mathbf{h})$ and a weight vector Λ.

To incorporate neural features, HT-D (NN) defines the following scoring function:

$$F_{\Lambda,\Theta}(\mathbf{n}, \mathbf{m}, \mathbf{h})) = \Lambda \cdot \Phi(\mathbf{n}, \mathbf{m}, \mathbf{h}) + G_{\Theta}(\mathbf{n}, \mathbf{m}, \mathbf{h}) \quad (3)$$

where Θ is the set of parameters of the neural networks and G is the neural scoring function over the $(\mathbf{n},\mathbf{m},\mathbf{h})$ tuple (Susanto and Lu, 2017b). Specifically, the neural features are defined over a fixed-size window surrounding a word in $\mathbf{n}$ paired with its immediately associated semantic unit. Following the work (Susanto and Lu, 2017b), we denote the window size as $J \in \{0, 1, 2\}$.

3.2 Cross-lingual Distributed Semantic Representations

A multilingual dataset used in semantic parsing comes with instances consisting of logical forms annotated with sentences from multiple different languages. In this work, we aim to learn one monolingual semantic parser for each language, while leveraging useful information that can be extracted from other languages. Our setup is as follows. Each time, we train the parser for a target language and regard the other languages as *auxiliary languages*. To learn cross-lingual distributed semantic representations from such data, we first combine all data involving all auxiliary languages to form a large dataset. Next, for each target language, we construct a semantics-word co-occurrence matrix $\mathbf{M} \in R^{m \times n}$ (where m is the number of unique semantic units, n is the number of unique words in the combined dataset). Each entry is the number of co-occurrences for a particular (semantic unit-word) pair. We will use this matrix to learn a low-dimensional cross-lingual representation for each semantic unit. To do so, we first apply singular value decomposition (SVD) to this matrix, leading to:

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^* \quad (4)$$

where $\mathbf{U} \in R^{m \times m}$ and $\mathbf{V} \in R^{n \times m}$ are unitary matrices, $\mathbf{V}^*$ is the conjugate transpose of $\mathbf{V}$, and $\mathbf{\Sigma} \in R^{m \times m}$ is a diagonal matrix. We truncate the diagonal matrix $\mathbf{\Sigma}$ and left multiply it with $\mathbf{U}$:

$$\mathbf{U}\tilde{\mathbf{\Sigma}} \quad (5)$$

where $\tilde{\mathbf{\Sigma}} \in R^{m \times d}$ is a matrix that consists of only the left d columns of $\mathbf{\Sigma}$, containing the d largest

	Rank (d)	F		Rank (d)	F
English	30	88.3	Chinese	10	80.0
Thai	20	85.8	Indonesian	30	88.3
German	30	78.3	Swedish	20	83.3
Greek	10	81.7	Farsi	10	76.7

Table 1: Performance on development set. F: F1-measure (%).

singular values. We leave the rank d as a hyperparameter. Each row in the above matrix is a d-dimensional vector, giving a low-dimensional representation for one semantic unit. Such distributed output representations can be readily used as continuous features in $\Phi(\mathbf{n}, \mathbf{m}, \mathbf{h})$.

3.3 Training and Decoding

During the training process, we optimize the objective function defined over the training set as:

$$\mathcal{L}(\Lambda, \Theta) = \sum_i \log \sum_{\mathbf{h} \in \mathcal{H}(\mathbf{n}_i, \mathbf{m}_i)} e^{F_{\Lambda,\Theta}(\mathbf{n}_i, \mathbf{m}_i, \mathbf{h})}$$
$$- \sum_i \log \sum_{\mathbf{m}', \mathbf{h}' \in \mathcal{H}(\mathbf{n}_i, \mathbf{m}')} e^{F_{\Lambda,\Theta}(\mathbf{n}_i, \mathbf{m}', \mathbf{h}')} \quad (6)$$

We follow the dynamic programming approach used in (Susanto and Lu, 2017b) to perform efficient inference, and follow the same optimization strategy as described there.

In the decoding phase, we are given a new input sentence $\mathbf{n}$, and find the optimal semantic tree $\mathbf{m}^*$:

$$\mathbf{m}^* = \underset{\mathbf{m}, \mathbf{h} \in \mathcal{H}(\mathbf{n}, \mathbf{m})}{\arg \max} F_{\Lambda,\Theta}(\mathbf{n}, \mathbf{m}, \mathbf{h}) \quad (7)$$

Again, the above equation can be efficiently computed by dynamic programming (Susanto and Lu, 2017b).

4 Experiments and Results

4.1 Datasets and Settings

We evaluate our approach on the standard GeoQuery dataset annotated in eight languages (Wong and Mooney, 2006; Jones et al., 2012; Susanto and Lu, 2017b). We follow a standard practice for evaluations which has been adopted in the literature (Lu, 2014, 2015; Susanto and Lu, 2017b). Specifically, to evaluate the proposed model, predicted outputs are transformed into Prolog queries. An output is considered as correct if answers that queries retrieve from GeoQuery database are the same as the gold ones . The dataset consists of 880 instances. In all experiments, we follow the experimental settings and procedures in (Lu, 2014,

		English		Thai		German		Greek		Chinese		Indonesian		Swedish		Farsi	
		Acc.	*F*	*Acc.*	*F*	*Acc.*	*F*	*Acc.*	*F*	*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	48.2	51.6	74.6	**79.8**	63.9	71.5	46.8	54.1
HT-G		76.8	81.0	73.6	76.7	62.1	68.5	69.3	74.6	56.1	58.4	66.4	72.8	61.4	70.5	51.8	58.6
UBL-S		82.1	82.1	66.4	66.4	75.0	75.0	73.6	73.7	63.8	63.8	73.8	73.8	78.1	78.1	64.4	64.4
TREETRANS		79.3	79.3	78.2	78.2	74.6	74.6	75.4	75.4	-	-	-	-	-	-	-	-
SEQ2TREE†		84.5	-	71.9	-	70.3	-	73.1	-	73.3	-	**80.7**	-	**80.8**	-	70.5	-
SL-SINGLE †		83.5	-	72.1	-	69.3	-	74.2	-	74.9	-	79.8	-	77.5	-	72.2	-
HT-D		**86.8**	**86.8**	80.7	80.7	**75.7**	**75.7**	79.3	79.3	76.1	76.1	75.0	75.0	79.3	**79.3**	73.9	73.9
HT-D (+O)		86.1	86.1	**81.1**	**81.1**	73.6	73.6	**81.4**	**81.4**	**77.9**	**77.9**	79.6	79.6	79.3	**79.3**	**75.7**	**75.7**
	J=0	87.9	87.9	82.1	82.1	75.7	75.7	81.1	81.1	76.8	76.8	76.1	76.1	81.1	81.1	75.0	75.0
HT-D (NN)	*J=1*	88.6	88.6	84.6	84.6	**76.8**	**76.8**	79.6	79.6	75.4	75.4	78.6	78.6	82.9	82.9	76.1	76.1
	J=2	**90.0**	**90.0**	82.1	82.1	73.9	73.9	80.7	80.7	81.1	81.1	81.8	81.8	**83.9**	**83.9**	74.6	74.6
	J=0	86.1	86.1	83.6	83.6	73.9	73.9	82.1	82.1	77.9	77.9	81.1	81.1	82.1	82.1	74.6	74.6
HT-D (NN+O)	*J=1*	86.1	86.1	**86.1**	**86.1**	72.5	72.5	80.4	80.4	81.4	81.4	82.5	82.5	82.5	82.5	75.7	75.7
	J=2	89.6	89.6	84.6	84.6	72.1	72.1	**83.2**	**83.2**	**82.1**	**82.1**	**83.9**	**83.9**	83.6	83.6	**76.8**	**76.8**

Table 2: Performance on multilingual datasets. *Acc.*: accuracy (%), *F*: F1-measure (%). +O: including distributed representations for semantic units as features. († indicates systems that make use of lambda calculus expressions as meaning representations.)

2015; Susanto and Lu, 2017b). In particular, we use 600 instances for training and 280 for test and set the maximum optimization iteration to 150. In order to tune the rank d, we randomly select 80% of the training instances for learning and use the rest 20% for development. We report the value of d for each language in Table 1 and the F1 scores on the development set.

4.2 Results

We compare our models against different existing systems, especially the two baselines HT-D (Lu, 2015) and HT-D (NN) (Susanto and Lu, 2017b) with different word window sizes $J \in \{0, 1, 2\}$.

WASP (Wong and Mooney, 2006) is a semantic parser based on statistical phrase-based machine translation. UBL-S (Kwiatkowski et al., 2010) induced probabilistic CCG grammars with higher-order unification that allowed to construct general logical forms for input sentences. TREETRANS (Jones et al., 2012) is built based on a Bayesian inference framework. We run WASP, UBL-S, HT-G, UBL-S, SEQ2TREE and SL-SINGLE [1] for comparisons. Note that there exist multiple versions of logical representations used in the GEO-QUERY dataset. Specifically, one version is based on lambda calculus expression, and the other is based on the variable free tree-shaped represen-

tation. We use the latter representation in this work, while the SEQ2TREE and SL-SINGLE employ the lambda calculus expression. It was noted in Kwiatkowski et al. (2010); Lu (2014) that evaluations based on these two versions are not directly comparable – the version that uses tree-shaped representations appears to be more challenging. We do not compare against (Jie and Lu, 2014) due to their different setup from ours.[2]

Table 2 shows results that we have conducted on eight different languages. The highest scores are highlighted. We can observe that when distributed logical representations are included, both HT-D and HT-D (NN) can lead to competitive results. Specifically, when such features are included, evaluation results for 5 out of 8 languages get improved.

We found that the shared information cross different languages could guide the model so that it can make more accurate predictions, eliminating certain semantic level ambiguities associated with the semantic units. This is exemplified by a real instance from the English portion of the dataset:

Input:	Which states have a river?
Gold:	$answer(state(loc(river(all))))$
Output:	$answer(state(traverse(river(all))))$
Output (+O):	$answer(state(loc(river(all))))$

[1] Note that in Dong and Lapata (2016), they adopted a different split – using 680 instances as training examples and the rest 200 for evaluation. We ran the released source code for SEQ2TREE (Dong and Lapata, 2016) over eight different languages. For each language, we repeated the experiments 3 times with different random seed values, and reported the average results as shown in Table 2 to make comparisons. Likewise, we ran SL-SINGLE following the same procedure.

[2] They trained monolingual semantic parsers. In the evaluation phase, they fed parallel text from different languages to each individual semantic parser, then employed a majority voting based ensemble method to combine predictions. Differently, we train monolingual semantic parsers augmented with cross-lingual distributed semantic information. In the evaluation phase, we only have one monolingual semantic parser. Hence, these two efforts are not directly comparable.
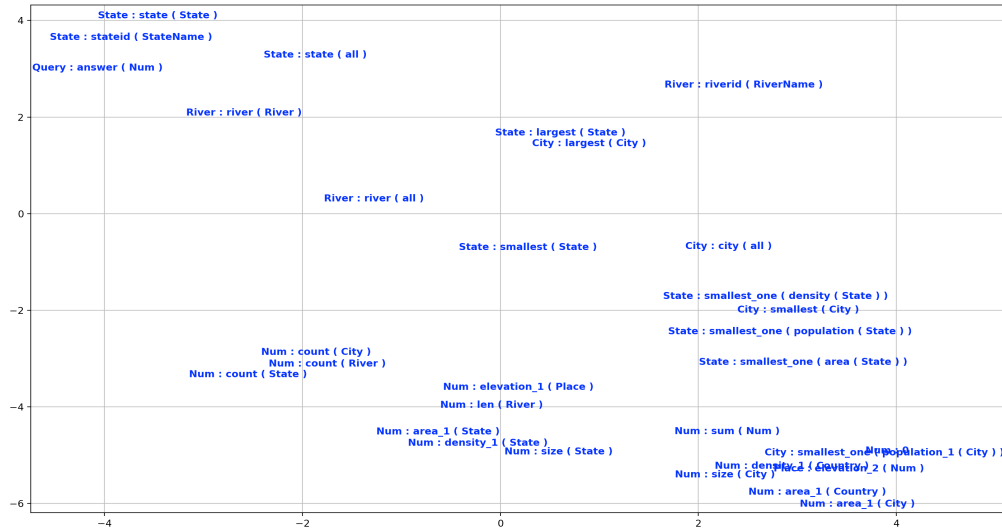
Figure 2: 2-D projection of learned distributed representations for semantics.

Here the input sentence in English is "Which states have a river?", and the correct output is shown below the sentence. Output is the prediction from HT-D (NN) and Output (+O) is the parsing result given by HT-D (NN+O) where the learned cross-lingual representations of semantics are included. We observe that, by introducing our learned cross-lingual semantic information, the system is able to distinguish the two semantically related concepts, loc (located in) and $traverse$ (traverse), and further make more promising predictions.

Interestingly, for German, the results become much lower when such features are included, indicating such features are not helpful in the learning process when such a language is considered. Reasons for this need further investigations. We note, however, previously it was also reported in the literature that the behavior of the performance associated with this language is different than other languages in the presence of additional features (Lu, 2014).

4.3 Visualizing Output Representations

To qualitatively understand how good the learned distributed representations are, we also visualize the learned distributed representations for semantic units. In the Figure 2, we plot the embeddings of a small set of semantic units which are learned from all languages other than English. Each representation is a 30-dimensional vector and is projected into a 2-dimensional space using Barnes-Hut-SNE (Maaten, 2013) for visualization. In general, we found that semantic units expressing similar meanings tend to appear together. For example, the two semantic units STATE : $smallest\_one$ ($density$ (STATE)) and STATE : $smallest\_one$ ($population$ (STATE)) share similar representations. However, we also found that occasionally semantic units conveying opposite meanings are also grouped together. This reveals the limitations associated with such a simple co-occurrence based approach for learning distributed representations for logical expressions.

5 Conclusions

In this paper, we empirically show that the distributed representations of logical expressions learned from multilingual datasets for semantic parsing can be exploited to improve the performance of a monolingual semantic parser. Our approach is simple, relying on an SVD over semantics-word co-occurrence matrix for finding such distributed representations for semantic units. Future directions include investigating better ways of learning such distributed representations as well as learning such distributed representations and semantic parsers in a joint manner.

References

Yoav Artzi, Dipanjan Das, and Slav Petrov. 2014. Learning compact lexicons for ccg semantic parsing. In *Proc. of EMNLP*. https://doi.org/10.3115/v1/D14-1134.

Yoav Artzi, Kenton Lee, and Luke S. Zettlemoyer. 2015. Broad-coverage ccg semantic parsing with amr. In *Proc. of EMNLP*. https://doi.org/10.18653/v1/D15-1198.

Yoav Artzi and Luke S. Zettlemoyer. 2011. Bootstrapping semantic parsers from conversations. In *Proc. of EMNLP*. http://www.aclweb.org/anthology/D11-1039.

Yee Seng Chan, Hwee Tou Ng, and David Chiang. 2007. Word sense disambiguation improves statistical machine translation. In *Proc. of ACL*. http://www.aclweb.org/anthology/P07-1005.

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proc. of ACL*. https://doi.org/10.18653/v1/P16-1004.

Long Duong, Hadi Afshar, Dominique Estival, Glen Pink, Philip Cohen, and Mark Johnson. 2017. Multilingual semantic parsing and code-switching. In *Proc. of CoNLL*. https://doi.org/10.18653/v1/K17-1038.

Zhanming Jie and Wei Lu. 2014. Multilingual semantic parsing : Parsing multiple languages into semantic representations. In *Proc. of COLING*. http://www.aclweb.org/anthology/C14-1122.

Bevan Jones, Mark Johnson, and Sharon Goldwater. 2012. Semantic parsing with bayesian tree transducers. In *Proc. of ACL*. http://www.aclweb.org/anthology/P12-1051.

Shafiq Joty, Preslav Nakov, Lluís Màrquez, and Israa Jaradat. 2017. Cross-language learning with adversarial neural networks. In *Proc. of CoNLL*. https://doi.org/10.18653/v1/K17-1024.

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proc. of ACL*. http://www.aclweb.org/anthology/P13-1117.

Tom Kwiatkowski, Luke Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2011. Lexical generalization in ccg grammar induction for semantic parsing. In *Proc. of EMNLP*. http://www.aclweb.org/anthology/D11-1140.

Tom Kwiatkowski, Luke S. Zettlemoyer, Sharon Goldwater, and Mark Steedman. 2010. Inducing probabilistic ccg grammars from logical form with higher-order unification. In *Proc. of EMNLP*. http://www.aclweb.org/anthology/D10-1119.

Percy Liang, Michael Jordan, and Dan Klein. 2011. Learning dependency-based compositional semantics. In *Proc. of ACL*. http://www.aclweb.org/anthology/P11-1060.

Wei Lu. 2014. Semantic parsing with relaxed hybrid trees. In *Proc. of EMNLP*. https://doi.org/10.3115/v1/D14-1137.

Wei Lu. 2015. Constrained semantic forests for improved discriminative semantic parsing. In *Proc. of ACL*. https://doi.org/10.3115/v1/P15-2121.

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*. https://doi.org/10.3115/1613715.1613815.

Laurens van der Maaten. 2013. Barnes-hut-sne. In *Proc. of ICLR*. https://arxiv.org/pdf/1301.3342.pdf.

Jian Ni, Georgiana Dinu, and Radu Florian. 2017. Weakly supervised cross-lingual named entity recognition via effective annotation and representation projection. In *Proc. of ACL*. https://doi.org/10.18653/v1/P17-1135.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proc. of ACL*. https://doi.org/10.18653/v1/P17-1178.

Tim Rocktäschel, Matko Bošnjak, Sameer Singh, and Sebastian Riedel. 2014. Low-dimensional embeddings of logic. In *Proc. of the ACL 2014 Workshop on Semantic Parsing*. https://doi.org/10.3115/v1/W14-2409.

Vivek Srikumar and Christopher D Manning. 2014. Learning distributed representations for structured output prediction. In *Proc. of NIPS*. http://papers.nips.cc/paper/5323-learning-distributed-representations-for-structured-output-prediction.pdf.

Raymond Hendy Susanto and Wei Lu. 2017a. Neural architectures for multilingual semantic parsing. In *Proc. of ACL*. https://doi.org/10.18653/v1/P17-2007.

Raymond Hendy Susanto and Wei Lu. 2017b. Semantic parsing with neural hybrid trees. In *Proc. of AAAI*. https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14843.

Zhigang Wang, Zhixing Li, Juanzi Li, Jie Tang, and Jeff Z. Pan. 2013. Transfer learning based cross-lingual knowledge extraction for wikipedia. In *Proc. of ACL*. http://www.aclweb.org/anthology/P13-1063.

Yuk Wah Wong and Raymond Mooney. 2006. Learning for semantic parsing with statistical machine translation. In *Proc. of NAACL*. https://doi.org/10.3115/1220835.1220891.

Yuk Wah Wong and Raymond Mooney. 2007. Learning synchronous grammars for semantic parsing with lambda calculus. In *Proc. of ACL*. http://www.aclweb.org/anthology/P07-1121.

Min Xiao and Yuhong Guo. 2015. Learning hidden markov models with distributed state representations for domain adaptation. In *Proc. of ACL-IJCNLP*. https://doi.org/10.3115/v1/P15-2086.

Luke S Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proc. of UAI*. http://dl.acm.org/citation.cfm?id=3020336.3020416.

Yuchen Zhang, Panupong Pasupat, and Percy Liang. 2017. Macro grammars and holistic triggering for efficient semantic parsing. In *Proc. of EMNLP*. https://doi.org/10.18653/v1/d17-1125.