Cross-lingual Decompositional Semantic Parsing

Sheng Zhang
Johns Hopkins University

Xutai Ma
Johns Hopkins Univeristy

Rachel Rudinger
Johns Hopkins University

Kevin Duh
Johns Hopkins University

Benjamin Van Durme
Johns Hopkins University

Abstract

We introduce the task of cross-lingual decompositional semantic parsing: mapping content provided in a source language into a decompositional semantic analysis based on a target language. We present: (1) a form of decompositional semantic analysis designed to allow systems to target varying levels of structural complexity (shallow to deep analysis), (2) an evaluation metric to measure the similarity between system output and reference semantic analysis, (3) an end-to-end model with a novel annotating mechanism that supports intra-sentential coreference, and (4) an evaluation dataset on which our model outperforms strong baselines by at least 1.75 F_1 score.

1 Introduction

We are concerned here with representing the semantics of multiple natural languages in a single semantic analysis. Renewed interest in semantic analysis has led to a surge of proposed new frameworks, e.g., GMB (Basile et al., 2012), AMR (Banarescu et al., 2013), UCCA (Abend and Rappoport, 2013), and UDS (White et al., 2016), as well as further calls to attend to existing efforts, e.g., Episodic Logic (EL) (Schubert and Hwang, 2000; Schubert, 2000; Hwang and Schubert, 1994; Schubert, 2014), or Discourse Representation Theory (Kamp, 1981; Heim, 1988).

Many of these efforts are limited to the analysis of English, but with a number of exceptions, e.g., recent efforts by Bos et al. (2017), ongoing efforts in Minimal Recursion Semantics (MRS) (Copestake et al., 1995), multilingual FrameNet annotation and parsing (Fung and Chen, 2004; Padó and Lapata, 2005), among others. For many languages, semantic analysis cannot be performed directly, owing to a lack of training data. While there is active work in the community focused on rapid construction of resources for low resource languages (Strassel and Tracey, 2016), it remains an expensive and perhaps infeasible solution to assume in-language annotated resources for developing semantic parsing technologies. In contrast, bitext is easier to get: it occurs often without researcher involvement,[1] and even when not available, it may be easier to find bilingual speakers that can translate a text, than it is to find experts that will create in-language semantic annotations. In addition, we are simply further along in being able to automatically understand English than we are other languages, resulting from the bias in investment in English-rooted resources.

Therefore, we propose the task of cross-lingual decompositional semantic parsing, which aims at transducing a sentence in the source language (e.g., Chinese sentence in Fig. 1b) into a decompositional semantic analysis derived based on English, via bitext. The efforts of decompositional semantics (White et al., 2016) focus on approaches to annotating meaning based on fine-grained scalar judgments which reflect the ambiguity of language, and the underspecification of meaning in context. Our contributions include:

(1) A form of decompositional semantic analysis allowing systems to target varying levels of structural complexity.

(2) An evaluation metric to measure the similarity between system and reference semantic analysis.

(3) An encoder-decoder model for cross-lingual decompositional semantic parsing. With a coreference annotating mechanism, the model solves intra-sentential coreference explicitly.

(4) The first evaluation dataset for cross-lingual decompositional semantic parsing.[2]

Experiments demonstrate our model achieves 38.78% F_1, outperforming strong baselines.

---

[1] For example, owing to a government decree.
[2] http://decomp.io

(a) UDS graph representation.
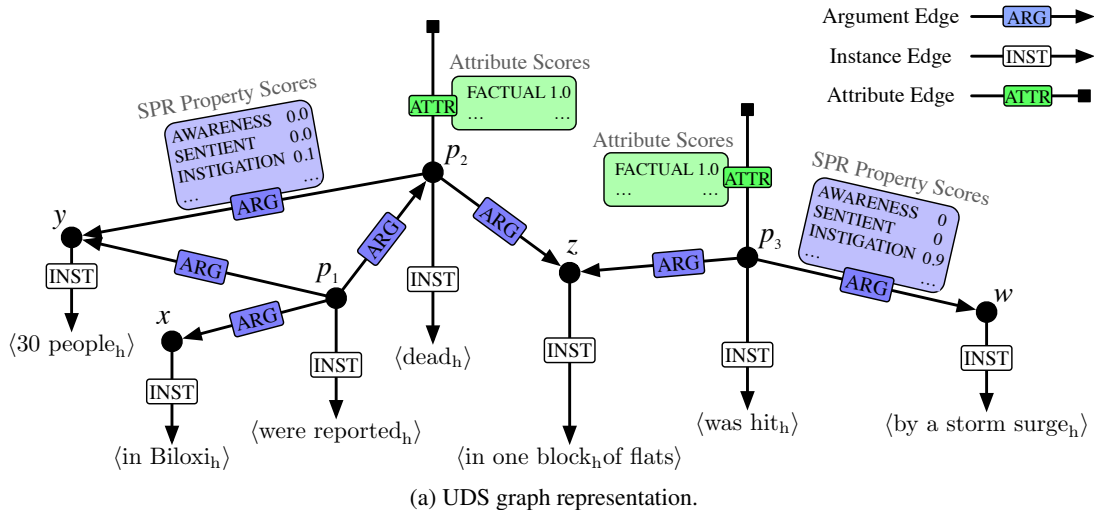
据	报道	，	在	比洛克西	，	三十	人	死	于	一	栋	被	风暴	潮	袭击	的	公寓楼
were	reported	,	in	Biloxi	,	30	people	dead	in	one	block	PTCP	storm	surge	hit	which	flats

"*30 people were reported dead in one block of flats which was hit by a storm surge*."

(b) Chinese sentence with Leipzig gloss.

Figure 1: Input and output of cross-lingual decompositional semantic parsing.

2 Semantic Analysis

The goal of cross-lingual decompositional semantic parsing is to provide a semantic analysis which can be used for various types of deep and shallow processing on the target language side. Many forms of semantic analysis are potentially suitable for this goal, e.g., AMR (Banarescu et al., 2013), UCCA (Abend and Rappoport, 2013), and Universal Decompositional Semantics (White et al., 2016). Here we choose Universal Decompositional Semantics (UDS), but note that our approach is applicable to other potential graph semantic formalisms.

The reasons for choosing UDS are three-fold: (1) **Compatibility**: UDS relates to Robust Minimal Recursion Semantics (RMRS) (Copestake, 2007), aiming for a maximal degree of semantic compatibility. With UDS, shallow analysis, such as predicate-argument extraction (Zhang et al., 2017a), can be regarded as producing a semantics which is underspecified and reusable with respect to deeper analysis, such as lexical semantics and inference (White et al., 2016). (2) **Robustness and Speed**: There exists a robust framework, PredPatt (White et al., 2016), for automatically creating UDS from raw sentences and their Universal Dependencies. PredPatt has been shown to be fast and accurate enough to process large volumes of text (Zhang et al., 2017c). (3) **Cross-lingual validity**: PredPatt is based purely on non-lexical and linguistically well-founded patterns from Universal Dependencies, which is designed to be cross-linguistically consistent.

There are three forms to represent UDS: flat, graph, or linearized representations. They are created for different purposes, and are inter-convertible. Flat representation relates to RMRS (Copestake, 2007), and we defer its description to Appendix A.

2.1 Graph Representation

The graph representation as shown in Fig. 1a is developed to improve ease of readability, parser evaluation, and integration with lexical semantics. The structure of the graph representation is a tuple $\mathcal{G} = \langle V, E \rangle$: a set of variables V (e.g., p_1 and x), and a set of edges E. There are 3 types of edges: (1) Argument edges describe argument relations between variable pairs. Deeper analysis such as Semantic Proto-Role (SPR) properties (Reisinger et al., 2015) can be attached to argument edges. SPR analysis can be considered as a scalar regression problem (White et al., 2016), where each predicate-argument pair is annotated with scalar values for different SPR properties. (2) Instance
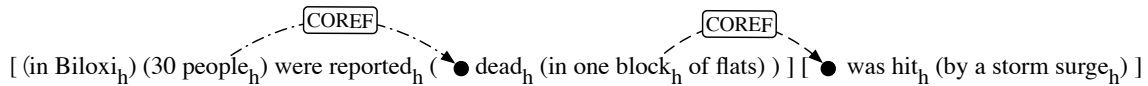
1665

Figure 2: UDS linearized representation. Deeper analysis such as SPR and factuality is not shown.

edges describe instances of variables in the target language. The subscript "h" indicates the syntactic head of an instance. (3) Attribute edges are unary, which describe various attributes of variables, such as event factuality (Saurí and Pustejovsky, 2009) and word senses (Miller, 1995). The graph representation can be viewed as an underspecified version of Dependency Minimal Recursion Semantics (DMRS) (Copestake, 2009) due to the underspecification of scope. Different from DMRS, the graph representation is linked cleanly to Universal Dependency syntax via PredPatt.

2.2 Linearized Representation

The linearized representation aims to facilitate learning of semantic parsers. Recently parsers based on RNN that make use of linearized representation have achieved state-of-the-art performance in constituency parsing (Choe and Charniak, 2016), logical form prediction (Dong and Lapata, 2016; Jia and Liang, 2016), and AMR parsing (Barzdins and Gosko, 2016; Peng et al., 2017). There was also work on predicting linearized semantic representations before RNN based approaches (Wong and Mooney, 2006).

Fig. 2 shows an example of UDS linearized representation. Intra-sentential coreference occurs when an instance refers to an antecedent, where we replace the instance with a special symbol "●" and add a COREF link between "●" and its antecedent. The linearized representation can be viewed as a sequence of tokens with a list of COREF links. Brackets, parentheses, and the special symbol "●" are all considered as tokens in this representation. The COREF links are drawn as a visual convenience, and the actual linearized representation achieves this via co-indexing, and is thus fully linear. We describe the procedure of converting graph representation to linearized representation in Appendix B.

3 Related Work

Our work synthesizes two strands of research, semantic analysis and cross-lingual learning.

The semantic analysis targeted in this work is akin to that of Hobbs (2003), but our eventual goal is to transduce texts from arbitrary human languages into a *"...broad, language-like, inference-enabling [semantic representation] in the spirit of Montague..."* (Schubert, 2015). Unlike efforts such as by Schubert and colleagues that directly target such an analysis, we are pursuing a strategy that incrementally increases the complexity of the target analysis in accordance with our ability to fashion models capable of producing it.[3] Embracing underspecification in the name of tractability is exemplified by MRS (Copestake et al., 2005; Copestake, 2009), the so-called *slacker semantics*, and we draw inspiration from that work. Analyses such as AMR (Banarescu et al., 2013) also make use of underspecification, but usually this is only implicit: certain aspects of meaning are simply not annotated. Unlike AMR, but akin to decisions made in PropBank (Palmer et al., 2005) (which forms the majority of the AMR ontological backbone), we target an analysis with a close correspondence to natural language syntax. Unlike interlingua (Mitamura et al., 1991; Dorr and Habash, 2002) that maps the source language into an intermediate analysis, and then maps it into the target language, we are not concerned with generating text from the semantic analysis. Substantial prior work on semantic analyses exists, including HPSG-based analyses (Copestake et al., 2005), CCG-based analyses (Steedman, 2000; Baldridge and Kruijff, 2002; Bos et al., 2004), and Universal Dependencies based analyses (White et al., 2016; Reddy et al., 2017). See (Schubert, 2015; Abend and Rappoport, 2017) for further discussion.

Cross-lingual learning has previously been applied to various NLP tasks. Yarowsky et al. (2001); Padó and Lapata (2009); Evang and Bos (2016); Faruqui and Kumar (2015) focused on projecting existing annotations on source-language text to the target language. Zeman and Resnik (2008); Ganchev et al. (2009); McDonald et al. (2011); Naseem et al. (2012); Wang and Manning (2014) enabled model transfer by shar-

---

[3]E.g., in Fig. 1a we recognize *"by a storm surge"* as an initial structural unit, with multiple potential analysis, which may be further refined based on the capabilities of a given cross-lingual semantic parser.

ing features or model parameters for different languages. Sudo et al. (2004); Zhang et al. (2017a,b); Mei et al. (2018) worked on cross-lingual information extraction and demonstrated the advantages of end-to-end learning. In this work, we explore end-to-end cross-lingual learning.

4   Evaluation Metric S

UDS can be represented in three forms. Evaluating such forms is crucial to the development of parsing algorithms. However, there is no method directly available for evaluation. Related methods come from semantic parsing, whose results are mainly evaluated in three ways: (1) task correctness (Tang and Mooney, 2001), which evaluates on a specific NLP task that uses the parsing results; (2) whole-parse correctness (Zettlemoyer and Collins, 2005), which counts the number of parsing results that are completely correct; and (3) Smatch (Cai and Knight, 2013), which computes the number of exactly matched edges between two semantic structures.

Nevertheless, our task needs an evaluation metric that can be used regardless of specific tasks or domains, and is able to differentiate two UDS graph representations with similar instances, SPR analysis, or attributes. We design an evaluation metric S that computes the similarity between two graph representations.

As described in Section 2.1, the graph representation is a tuple $\mathcal{G} = (V, E)$. For two graphs $\mathcal{G}_1 = (V_1, E_1)$ and $\mathcal{G}_2 = (V_2, E_2)$, we define the score S as the maximum soft edge matching score between $\mathcal{G}_1$ and $\mathcal{G}_2$:

$$S(\mathcal{G}_1, \mathcal{G}_2) = \max_{m \in \mathcal{M}} \Big[\sum_{(e_1^{(i)}, e_2^{(j)}) \in \mathcal{P}} f_T(e_1^{(i)}, e_2^{(j)}) \Big]$$

where m is a mapping from variables in V_1 to variables in V_2. Given a mapping m, $\mathcal{P}$ is a set of edge pairs: for each pair $(e_1^{(i)}, e_2^{(j)})$, variables(s) in $e_1^{(i)}$ are mapped to variables(s) in $e_2^{(j)}$ via m. f_T computes the matching score for a pair of edges belonging to type $T \in \{\text{ARG, INST, ATTR}\}$. The matching score is normalized to $[0, 1]$.

The precision and recall are computed by $S(\mathcal{G}_1, \mathcal{G}_2)/|E_1|$, and $S(\mathcal{G}_1, \mathcal{G}_2)/|E_2|$ respectively.

In this work, $f_{\text{ARG}} = f_{\text{ATTR}} = e^{-\text{MAE}}$, where MAE computes the mean absolute error between two set of scores s_1 and s_2: $\sum_i^n |s_1^{(i)} - s_2^{(i)}|/n$.

$f_{\text{INST}} = \text{BLEU}$ (Papineni et al., 2002) which compute the BLEU score of an instance pair.[4]

Finding an optimal variable mapping m that yields the highest S is NP-complete. We instead adopt a strategy used in Smatch (Cai and Knight, 2013) that does a hill-climbing search with smart initialization plus 4 random restarts, and has been shown to give the best trade-off between accuracy and speed. Smatch for evaluating semantic structures can be considered as a special case of S, where $f_T = \delta$, the Kronecker delta.

5   Model

We formulate the task of cross-lingual decompositional semantic parsing as a joint problem of sequence-to-sequence learning, coreference resolution and decompositional semantic analysis. The input is a sentence X in the source language, e.g., the Chinese sentence in Fig. 1b. The output is a UDS linearized representation (Y, C, D) based on the target language: Y is a sequence of tokens; C is a set of COREF links; and D is a set of scores for decompositional analysis, such as SPR and factuality.

The goal is to learn a conditional probability distribution $P(Y, C, D|X)$ whose most likely configuration, given the input sentence, outputs the true UDS linearized representation with decompositional analysis. While the standard encoder-decoder framework shows the state-of-the-art performance in sequence-to-sequence learning (Choe and Charniak, 2016; Jia and Liang, 2016; Barzdins and Gosko, 2016), it cannot directly solve intra-sentential conference and decompositional semantic analyses in our task. To achieve this goal, we propose an encoder-decoder architecture incorporated with a *coreference annotating* mechanism[5] and *decompositional analysis*. As illustrated in Fig. 3, **Encoder** transforms the input sequence into hidden states; **Decoder** reads the hidden states, and then at each time step generates a token and creates its COREF link; **Decompositional Analysis**, based on the decoder output, performs SPR analysis for predicate-argument pairs, and factuality analysis for predicates.

---

[4]Future work could consider, e.g., a modified BLEU that considers Levenshtein distance between tokens for a more robust partial-scoring in the face of transliteration errors.

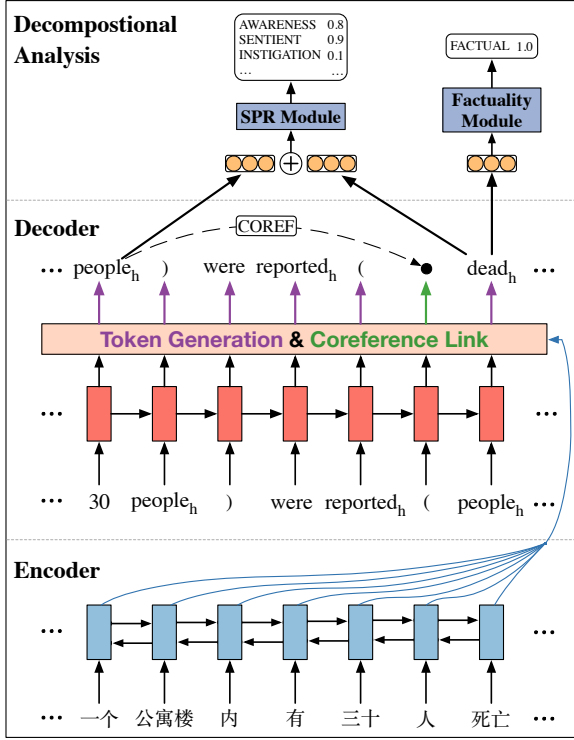[5]Similar coreference mechanism has been proposed by Ji et al. (2017).

Figure 3: Illustration of the model architecture.

5.1 Encoder

The encoder employs a bidirectional recurrent neural network (Schuster and Paliwal, 1997) with LSTM units (Hochreiter and Schmidhuber, 1997). It encodes the input $X = x_1, \ldots, x_N$[6] into a sequence of hidden states $\boldsymbol{h} = h_1, \ldots, h_N$. Each hidden state h_i is a concatenation of a left-to-right hidden state $\overrightarrow{h_i}$ and a right-to-left hidden state $\overleftarrow{h_i}$,

5.2 Decoder

Given the encoder hidden states, the decoder predicts the linearized representation (as shown in Fig. 2) according to the conditional probability $P(Y, C \mid X)$ which is decomposed as a product of the decoding probabilities at each time step t:

$$P(Y, C \mid X) = \prod_{t=1}^{M} P(y_t, c_t \mid y_{<t}, c_{<t}, X) \quad (1)$$

where y_t is the decoded token at time step t, and c_t is the source of the COREF link for y_t, i.e., the antecedent of y_t. The set of possible antecedents of y_t is $\mathcal{A}(t) = \{\epsilon, y_1, \ldots, y_{t-1}\}$: a dummy antecedent ϵ and all preceding tokens. ϵ represents a scenario, where the token is not a special symbol "•", and it refers to none of the preceding tokens.

---

[6] For simplicity, we use X (and Y) to represent both tokens as well as their word embeddings.

$y_{<t}$ and $c_{<t}$ are the preceding tokens and their antecedents. We omit $y_{<t}$ and $c_{<t}$ from the notation when the context is unambiguous.

The decoding probability at each time step t is decomposed as

$$P(y_t, c_t) = P(y_t)P(c_t|y_t) \quad (2)$$

where $P(y_t)$ is the token generation probability, and $P(c_t|y_t)$ is the antecedent probability.

Token Generation: The probability distribution of the generated token y_t is defined as

$$P(y_t) = \text{softmax}(\text{FFNN}_g(s_t, a_t)) \quad (3)$$

where FFNN_g is a two-layer feed-forward neural network over the decoder hidden state s_t and the attention-weighted vector a_t. s_t is computed by

$$s_t = \text{RNN}(y_{t-1}, s_{t-1}), \quad (4)$$

where RNN is a recurrent neural network using LSTM. a_t is computed by the attention mechanism (Bahdanau et al., 2014; Luong et al., 2015),

$$a_t = \sum_{i}^{N} \alpha_{t,i} h_i, \quad (5)$$

$$\alpha_{t,i} = \frac{\exp\left(s_t^\top (W_a h_i + b_a)\right)}{\sum_{j=1}^{N} \exp\left(s_t^\top (W_a h_j + b_a)\right)}, \quad (6)$$

where W_a is a transform matrix and b_a is a bias.

Coref Link: The probability of y_t referring to the preceding token y_k, i.e., $c_t = y_k$, is defined as

$$P(c_t = y_k|y_t) = \frac{\exp\left(\text{SCORE}(y_t, y_k)\right)}{\sum_{y_k' \in \mathcal{A}(t)} \exp\left(\text{SCORE}(y_t, y_{k'})\right)}, \quad (7)$$

$\text{SCORE}(y_t, y_k)$ is a pairwise score for a COREF link from y_k to y_t, defined as:

$$\text{SCORE}(y_t, y_k) = s_c(y_t) + s_p(y_k) + s_a(y_t, y_k) \quad (8)$$

There are three factors in this pairwise score, which is akin to Lee et al. (2017): (1) $s_c(y_t)$, whether y_t should refer to a preceding instance; (2) $s_p(y_k)$, whether y_k shoud be a candidate source of such a coreference; and (3) $s_a(y_t, y_k)$, whether y_k is an antecedent of y_t.

Fig. 4 shows the details of the scoring architecture. At the core of the three factors are vector representations $\gamma(y_t)$ for each token y_t, which is described in detail in the following section. Given
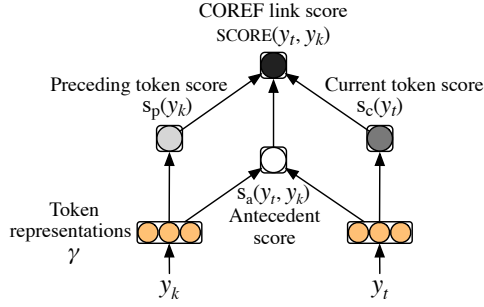
Figure 4: Scoring architecture in the copy mechanism between a preceding token y_k and the currently considered token y_t.

the currently considered token y_t and a preceding token y_k, the scoring functions above are computed via standard feed-foward neural networks:

$$s_c(y_t) = w_c \cdot \text{FFNN}_c(\gamma(y_t)) \quad (9)$$

$$s_p(y_k) = w_p \cdot \text{FFNN}_p(\gamma(y_k)) \quad (10)$$

$$s_a(y_t, y_k) = w_a \cdot \text{FFNN}_a\big([\gamma(y_t), \gamma(y_k),$$
$$\gamma(y_t) \circ \gamma(y_k)]\big) \quad (11)$$

where $\cdot$ denotes dot product, $\circ$ denotes element-wise multiplication, and FFNN denotes a two-layer feed-foward neural network over the input. The input of FFNN_a is a concatenation of vector representations $\gamma(y_t)$ and $\gamma(y_k)$, and their explicit element-wise similarity $\gamma(y_t) \circ \gamma(y_k)$.

Token representations: To accurately predict COREF link scores as well as decompositional analysis (which is described in the following section), we consider three types of information in each token representation $\gamma(y_t)$: (1) the token itself y_t, (2) on the decoder side, the preceding context $y_{<t}$, and (3) on the encoder side, the input sequence $X = x_1, \ldots, x_N$.

The lexical information of the token itself y_t is represented by its word embedding e_t. The preceding context $y_{<t}$ is encoded by the decoder RNN in Equation (4). We use the decoder hidden state s_t to represent the preceding context information. The encoder-side context is represented by an attention-weighted weight a_t defined in Equation (6). All the above information is concatenated to produce the final token representation $\gamma(y_t)$:

$$\gamma(y_t) = [e_t, s_t, a_t] \quad (12)$$

5.3 Decompositional Analyses

The decompositional analyses D contains scores for Semantic Proto-Role (SPR) properties D_{SPR}, and scores for event factuality D_{FACT}.

SPR: Given a predicate-argument pair (y_i, y_j), we denote the score for SPR property p as $D_{\text{SPR}_p}^{(y_i, y_j)}$. As shown in Fig. 3, we concatenate the token representations of predicate and argument head tokens $\gamma(y_i)$ and $\gamma(y_j)$ as the input to a SPR module. We employ the state-of-the-art SPR module in Rudinger et al. (2018a), defined as:

$$\hat{D}_{\text{SPR}_p}^{(y_i, y_j)} = W_{\text{SPR}_p}\text{ReLU}(W_{\text{shared}}[\gamma(y_i), \gamma(y_j)]) \quad (13)$$

where $\mathbf{W}_{\text{shared}}$ is the weight matrix shared across all properties. $\mathbf{W}_{\text{SPR}_p}$ is the weight matrix for SPR property p. Then, the log-likelihood of the score of SPR property p is defined as the negative L_2 loss, i.e., $-|\hat{D}_{\text{SPR}_p}^{(y_i, y_j)} - D_{\text{SPR}_p}^{(y_i, y_j)}|^2$.

Factuality: We consider predicting event factuality as a scalar regression problem (White et al., 2016), and denote the factuality score of predicate y_k as $D_{\text{FACT}}^{(y_k)}$. As shown in Fig. 3, we take the token representation of predicate head token $\gamma(y_k)$ as the input to the state-of-the-art factuality module (Rudinger et al., 2018c):

$$\hat{D}_{\text{FACT}}^{(y_k)} = V_2\text{ReLU}\left(V_1\gamma(y_k) + b_1\right) + b_2, \quad (14)$$

where V_1 and V_2 are weight matrices, and b_1 and b_2 are biases. The log-likelihood of factuality score is defined as negative of the Huber loss (Huber, 1964) with $\delta = 1$.

We assume conditional independence among decompositional analysis:

$$P(D|X, Y, C) = \prod_{(y_i, y_j)} \prod_p (D_{\text{SPR}_p}^{(y_i, y_j)}|X, Y, C)$$
$$\prod_{y_k} P(D_{\text{FACT}}^{(y_k)}|X, Y, C) \quad (15)$$

5.4 Learning

Given the input sentence X, the output sequence of tokens Y, and the COREF links C, and the decompositional analysis D, the objective is to minimize the below negative log-likelihood:

$$\mathcal{L} = -\log P(Y, C, D|X)$$
$$= -\sum_{t=1}^{M}[\mu_1 \log P(y_t) + \mu_2 \log P(c_t|y_t)] -$$
$$\mu_3 \log P(D|X, Y, C)$$

To increase the convergence rate, we pretrain the model by setting the weights $\mu_1 = 1$ and $\mu_2 = \mu_3 = 0$ to only optimize the token generation accuracy. After the model converges, we set $\mu_2 = \mu_3 = 1$ and lower $\mu_1 = 0.1$.

	S metric			$\text{BLEU}_{\text{INST}}$	MAE_{SPR}	MAE_{FACT}
	Precision	Recall	F1			
Pipeline	35.08	30.10	32.39	15.03	N/A	N/A
Variant (a)	39.31	32.93	35.84	16.74	0.75	1.11
Variant (b)	42.76	33.20	37.38	17.71	0.74	1.14
Variant (c)	41.74	33.28	37.03	18.01	0.80	1.14
Our model	**45.33**	**33.88**	**38.78**	**19.61**	**0.71**	**1.06**

Table 1: Evaluation of results on the **test** set.
(**In-domain test** results are shown in Appendix D.)

6 Experiments

We now describe the evaluation data, baselines, and experimental results. Hyperparameter settings are reported in Appendix C.

6.1 Data

We choose Chinese as the source language and English as the target language. For **test**, we selected 270 sentences from the Universal Dependencies (UD) English Treebank (Silveira et al., 2014) test set, which have human-annotated SPR (White et al., 2016) and factuality (Rudinger et al., 2018c) analyses. We then created linearized representations for these sentences using PredPatt based on their gold UD syntax. Meanwhile, the Chinese translations of these sentences were created by crowdworkers on Amazon Mechanical Turk. The test dataset will be released upon publication. For **training**, we first collected about 1.8M Chinese-English sentence bitexts from the GALE project (Cohen, 2007), then tokenized Chinese sentences with Stanford Word Segmenter (Chang et al., 2008). We created linearized representations for English sentences using PredPatt based on automatic UD syntax generated by SyntaxNet Parser (Andor et al., 2016), and added SPR and factuality annotations using the state-of-the-art models (Rudinger et al., 2018b,c) trained on SPR v2.x and It-happened v2.0 respectively.[7] We hold out 20K training sentences for **validation** and **in-domain test**. Table 2 reports the dataset statistics.

6.2 Variants

We evaluate our model described in Section 5 and three variants: **(a)** We replace the coreference annotating mechanism by randomly choosing an an-

---
[7] Both datasets are available at `http://decomp.net`

	No. sents	Source
Train	1,879,172	GALE
Validation	10,000	GALE
In-domain Test	10,000	GALE
Test	270	UD Treebank

Table 2: Statistics of the evaluation data.

tecedent from all preceding instances. **(b)** We preprocess the data by replacing the special symbol "•" with the syntactic head of its antecedent. During training and testing, we replace the coreference annotating mechanism with a heuristic that solves coreference by randomly choosing an antecedent among preceding instances which have the same syntactic head. **(c)** We remove the decoder-side information in the token representation $\gamma(y_t)$ defined in Equation (12) and only keep the encoder-side information a_t. We also include a **Pipeline** approach where Chinese sentences are first translated into English by a neural machine translation system (Klein et al., 2017) and are then annotated by a UD parser (Andor et al., 2016). The UDS linearized representation of **Pipeline** are created by PredPatt based the automatic UD parses.

6.3 Results

Table 1 reports the experimental results on the **test** set. Results on the **in-domain test** set are similar and shown in Appendix D. In Table 1, S **metric** (defined in Section 4) measures the similarity between predicted and reference graph representations. Based on the optimal variable mapping provided by the S **metric**, we are able to evaluate our model and the variants in different aspects: $\text{BLEU}_{\text{INST}}$ measures the BLEU score of all matched instance edges; MAE_{SPR} measures the

mean absolute error of SPR property scores of all matched argument edges; and **MAE**$_{FACT}$ measures the mean absolute error of factuality scores of all matched attribute edges.

Overall, our proposed model outperforms the variants in every aspect. Variants (a) and (b) use simple heuristics to solve coreference, and achieve reasonable results: they both employ sequence-to-sequence models to predict graph representations, which can be considered a replica of state-of-the-art approaches for structured prediction (Choe and Charniak, 2016; Barzdins and Gosko, 2016; Peng et al., 2017). Compared to our model which employs the coreference annotating mechanism, these two variants suffer notable loss in the precision of S metric. As a result, their performance drops on the other metrics. Variant (c) only uses the encoder-side information for token representation, resulting in significant loss in MAE$_{SPR}$ and MAE$_{FACT}$. In the pipeline approach, each component is trained independently. During test, residual errors from each component are propagated through the pipeline. As expected, it shows a significant performance drop.

	Precision	Recall	F1
Variant (a)	10.38	31.23	15.58
Variant (b)	88.42	50.59	64.36
Variant (c)	84.12	35.99	50.41
Our model	**96.63**	**97.62**	**97.12**

Table 3: Coreference evaluation (MUC) based on forced decoding.

Coreference occurs 589 times in the test set. To evaluate the coreference accuracy of our model, we force the decoder to generate the reference target sequence, and only predict coreference via the copy mechanism, or its variants. In Table 3, we report the precision, recall, and F_1 for the standard MUC using the official coreference scorer of the CoNLL-2011/2012 shared tasks (Pradhan et al., 2014). Since coreference in our setup occurs at the sentence level, our model achieves high performance. Variant (a) randomly choosing antecedents performs poorly, whereas variant (b), which solves coreference only based on syntactic heads, achieves a relatively high score. Variant (c) demonstrates that only using encoder-side information in the coreference annotating mechanism leads a significant performance drop.

	Our Model	Monolingual SOTA
awareness	0.852	0.879
change location	0.491	0.492
change possession	0.448	0.488
changed	0.307	0.352
change state	0.362	0.373
existed after	0.426	0.478
existed before	0.602	0.618
existed during	0.336	0.358
instigation	0.597	0.59
partitive	0.317	0.359
sentient	0.849	0.88
volition	0.818	0.837
was for benefit	0.566	0.578
was used	0.268	0.203

Table 4: Pearson coefficient of each SPR property.

Since our model and the state-of-the-art monolingual SPR model (Rudinger et al., 2018c) use the same test set, we are able to compare the performance of our model against the monolingual model by forcing the decoder and the coreference mechanism to create the reference graph representation and only predicting the SPR property scores. Table 4 shows the Pearson coefficient of each SPR property. While our model only has the access to the sentence in the source language during the encoding stage,[8] the performance is comparable to the state-of-the-art monolingual model.

7 Conclusions

We introduce the task of cross-lingual decompositional semantic parsing, which maps content provided in a source language into decompositional analysis based on a target language. We present: UDS graph/linearized representations as the target semantic interface, the S metric for evaluation, and the Chinese-English decompositional semantic parsing dataset. We propose an end-to-end learning approach with a coreference annotating mechanism which outperforms three strong baselines. We separately evaluate the coreference mechanism and SPR prediction, showing promising results. The representations for cross-lingual decompositional semantics, the evaluation metric, and the evaluation dataset provided in this work

---

[8] The state-of-the-art monolingual SPR model directly encodes the sentence in the target language.

will be beneficial to the increasing interests in semantic analysis and cross-lingual applications.

Acknowledgments

References

Omri Abend and Ari Rappoport. 2013. Universal conceptual cognitive annotation (ucca). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 228–238, Sofia, Bulgaria. Association for Computational Linguistics.

Omri Abend and Ari Rappoport. 2017. The state of the art in semantic representation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 77–89, Vancouver, Canada. Association for Computational Linguistics.

Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov, and Michael Collins. 2016. Globally normalized transition-based neural networks. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2442–2452, Berlin, Germany. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Jason Baldridge and Geert-Jan Kruijff. 2002. Coupling ccg and hybrid logic dependency semantics. In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*, pages 319–326, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Laura Banarescu, Claire Bonial, Shu Cai, Madalina Georgescu, Kira Griffitt, Ulf Hermjakob, Kevin Knight, Philipp Koehn, Martha Palmer, and Nathan Schneider. 2013. Abstract meaning representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186.

Guntis Barzdins and Didzis Gosko. 2016. Riga at semeval-2016 task 8: Impact of smatch extensions and character-level neural translation on amr parsing accuracy. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 1143–1147, San Diego, California. Association for Computational Linguistics.

Valerio Basile, Johan Bos, Kilian Evang, and Noortje Venhuizen. 2012. Developing a large semantically annotated corpus. In *LREC 2012, Eighth International Conference on Language Resources and Evaluation*.

Johan Bos, Stephen Clark, Mark Steedman, James R. Curran, and Julia Hockenmaier. 2004. Wide-coverage semantic representations from a ccg parser. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA. Association for Computational Linguistics.

Johan Bos, Kilian Evang, Johannes Bjerva, Lasha Abzianidze, Hessel Haagsma, Rik van Noord, Pierre Ludmann, and Duc-Duy Nguyen. 2017. The parallel meaning bank: Towards a multilingual corpus of translations annotated with compositional meaning representations. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 2: Short Papers*, pages 242–247.

Shu Cai and Kevin Knight. 2013. Smatch: an evaluation metric for semantic feature structures. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 748–752, Sofia, Bulgaria. Association for Computational Linguistics.

Pi-Chuan Chang, Michel Galley, and Christopher D. Manning. 2008. Optimizing Chinese word segmentation for machine translation performance. In *Proceedings of the Third Workshop on Statistical Machine Translation*, pages 224–232, Columbus, Ohio. Association for Computational Linguistics.

Do Kook Choe and Eugene Charniak. 2016. Parsing as language modeling. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2331–2336, Austin, Texas. Association for Computational Linguistics.

Jordan Cohen. 2007. The gale project: A description and an update. In *Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on*, pages 237–237. IEEE.

Ann Copestake. 2007. Semantic composition with (robust) minimal recursion semantics. In *Proceedings of the Workshop on Deep Linguistic Processing*, pages 73–80. Association for Computational Linguistics.

Ann Copestake. 2009. *Invited Talk:* 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, Greece. Association for Computational Linguistics.

Ann Copestake, Dan Flickinger, Rob Malouf, Susanne Riehemann, and Ivan Sag. 1995. Translation using minimal recursion semantics. In *In Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation*.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language and Computation*, 3(2-3):281–332.

Li Dong and Mirella Lapata. 2016. Language to logical form with neural attention. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 33–43, Berlin, Germany. Association for Computational Linguistics.

Bonnie Dorr and Nizar Habash. 2002. Interlingua approximation: A generation-heavy approach. In *In Proceedings of AMTA-2002*. University of Chicago Press.

Kilian Evang and Johan Bos. 2016. Cross-lingual learning of an open-domain semantic parser. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 579–588. The COLING 2016 Organizing Committee.

Manaal Faruqui and Shankar Kumar. 2015. Multilingual open relation extraction using cross-lingual projection. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1351–1356, Denver, Colorado. Association for Computational Linguistics.

Pascale Fung and Benfeng Chen. 2004. Biframenet: Bilingual frame semantics resources construction by cross-lingual induction. In *In Proceedings of the 20th International Conference on Computational Linguistics*, pages 931–935.

Kuzman Ganchev, Jennifer Gillenwater, and Ben Taskar. 2009. Dependency grammar induction via bitext projection constraints. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, pages 369–377, Suntec, Singapore. Association for Computational Linguistics.

Irene Heim. 1988. *The Semantics of Definite and Indefinite Noun Phrases*. Ph. d. dissertation, New York and London.

Jerry R. Hobbs. 2003. Discourse and Inference.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Peter J Huber. 1964. Robust estimation of a location parameter. *The annals of mathematical statistics*, pages 73–101.

Chung Hee Hwang and Lenhart K. Schubert. 1994. Interpreting tense, aspect and time adverbials: A compositional, unified approach. In *Temporal Logic*, pages 238–264, Berlin, Heidelberg. Springer Berlin Heidelberg.

Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, and Noah A. Smith. 2017. Dynamic entity representations in neural language models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1830–1839, Copenhagen, Denmark. Association for Computational Linguistics.

Robin Jia and Percy Liang. 2016. Data recombination for neural semantic parsing. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12–22, Berlin, Germany. Association for Computational Linguistics.

H Kamp. 1981. A theory of truth and semantic representation, 277-322, jag groenendijk, tmv janssen and mbj stokhof, eds.

G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. 2017. OpenNMT: Open-Source Toolkit for Neural Machine Translation. *ArXiv e-prints*.

Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 188–197, Copenhagen, Denmark. Association for Computational Linguistics.

Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

Ryan McDonald, Slav Petrov, and Keith Hall. 2011. Multi-source transfer of delexicalized dependency parsers. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 62–72, Edinburgh, Scotland, UK. Association for Computational Linguistics.

Hongyuan Mei, Sheng Zhang, Kevin Duh, and Benjamin Van Durme. 2018. Halo: Learning Semantics-Aware Representations for Cross-Lingual Information Extraction. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*, pages 142–147. Association for Computational Linguistics.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Teruko Mitamura, Eric H. Nyberg, and Jaime G. Carbonell. 1991. An efficient interlingua translation system for multi-lingual document production. In *Proceedings of Machine Translation Summit III*, pages 2–4.

Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 629–637, Jeju Island, Korea. Association for Computational Linguistics.

Sebastian Padó and Mirella Lapata. 2005. Cross-linguistic projection of role-semantic information. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 859–866, Stroudsburg, PA, USA. Association for Computational Linguistics.

Sebastian Padó and Mirella Lapata. 2009. Cross-lingual annotation projection of semantic roles. *J. Artif. Int. Res.*, 36(1):307–340.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1):71–106.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 311–318. Association for Computational Linguistics.

Xiaochang Peng, Chuan Wang, Daniel Gildea, and Nianwen Xue. 2017. Addressing the data sparsity issue in neural amr parsing. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 366–375, Valencia, Spain. Association for Computational Linguistics.

Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. 2014. Scoring coreference partitions of predicted mentions: A reference implementation. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 30–35, Baltimore, Maryland. Association for Computational Linguistics.

Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata. 2017. Universal semantic parsing. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 89–101, Copenhagen, Denmark. Association for Computational Linguistics.

Drew Reisinger, Rachel Rudinger, Francis Ferraro, Craig Harman, Kyle Rawlins, and Benjamin Van Durme. 2015. Semantic proto-roles. *Transactions of the Association for Computational Linguistics*, 3:475–488.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018a. Neural davidsonian semantic proto-role labeling. *arXiv preprint arXiv:1804.07976*.

Rachel Rudinger, Adam Teichert, Ryan Culkin, Sheng Zhang, and Benjamin Van Durme. 2018b. Neural davidsonian semantic proto-role labeling. *arXiv preprint arXiv:1804.07976*.

Rachel Rudinger, Aaron Steven White, and Benjamin Van Durme. 2018c. Neural models of factuality. In *Proceedings of the Annual Meeting of the North American Association of Computational Linguistics (NAACL)*.

Roser Saurí and James Pustejovsky. 2009. Factbank: a corpus annotated with event factuality. *Language Resources and Evaluation*, 43(3):227.

Lenhart Schubert. 2014. From treebank parses to episodic logic and commonsense inference. In *Proceedings of the ACL 2014 Workshop on Semantic Parsing*, pages 55–60, Baltimore, MD. Association for Computational Linguistics.

Lenhart Schubert. 2015. Semantic representation. In *Proceedings of AAAI Conference on Artificial Intelligence*.

Lenhart K Schubert. 2000. The situations we talk about. In *Logic-based artificial intelligence*, pages 407–439. Springer.

Lenhart K. Schubert and Chung Hee Hwang. 2000. Natural language processing and knowledge representation. chapter Episodic Logic Meets Little Red Riding Hood: A Comprehensive Natural Representation for Language Understanding, pages 111–174. MIT Press, Cambridge, MA, USA.

Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681.

Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel Bowman, Miriam Connor, John Bauer, and Christopher D. Manning. 2014. A gold standard dependency corpus for English. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.

Mark Steedman. 2000. *The Syntactic Process*. MIT Press, Cambridge, MA, USA.

Stephanie Strassel and Jennifer Tracey. 2016. Lorelei language packs: Data, tools, and resources for technology development in low resource languages.

In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Paris, France. European Language Resources Association (ELRA).

Kiyoshi Sudo, Satoshi Sekine, and Ralph Grishman. 2004. Cross-lingual information extraction system evaluation. In *Proceedings of the 20th international Conference on Computational Linguistics*, page 882. Association for Computational Linguistics.

Lappoon R Tang and Raymond J Mooney. 2001. Using multiple clause constructors in inductive logic programming for semantic parsing. In *European Conference on Machine Learning*, pages 466–477. Springer.

Mengqiu Wang and Christopher D. Manning. 2014. Cross-lingual projected expectation regularization for weakly supervised learning. *Transactions of the Association of Computational Linguistics*, 2:55–66.

Aaron Steven White, Drew Reisinger, Keisuke Sakaguchi, Tim Vieira, Sheng Zhang, Rachel Rudinger, Kyle Rawlins, and Benjamin Van Durme. 2016. Universal decompositional semantics on universal dependencies. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1713–1723, Austin, Texas. Association for Computational Linguistics.

Yuk Wah Wong and Raymond Mooney. 2006. Learning for semantic parsing with statistical machine translation. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 439–446, New York City, USA. Association for Computational Linguistics.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*, HLT '01, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Daniel Zeman and Philip Resnik. 2008. Cross-language parser adaptation between related languages. In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*.

Luke S. Zettlemoyer and Michael Collins. 2005. Learning to map sentences to logical form: Structured classification with probabilistic categorial grammars. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, UAI'05, pages 658–666, Arlington, Virginia, United States. AUAI Press.

Sheng Zhang, Kevin Duh, and Benjamin Van Durme. 2017a. MT/IE: Cross-lingual open information extraction with neural sequence-to-sequence models. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 64–70, Valencia, Spain. Association for Computational Linguistics.

Sheng Zhang, Kevin Duh, and Benjamin Van Durme. 2017b. Selective Decoding for Cross-lingual Open Information Extraction. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 832–842, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Sheng Zhang, Rachel Rudinger, and Benjamin Van Durme. 2017c. An Evaluation of PredPatt and Open IE via Stage 1 Semantic Role Labeling. In *IWCS 2017 — 12th International Conference on Computational Semantics — Short papers*.