# Cross-lingual Decompositional Semantic Parsing Supplemental Material

**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

## A Flat Representation

Predicates:
$\langle \text{were reported}_\text{h} \rangle(p_1)$, $\langle \text{dead}_\text{h} \rangle(p_2)$,
$\langle \text{was hit}_\text{h} \rangle(p_3)$,
$\langle \text{in Biloxi}_\text{h} \rangle(x)$, $\langle \text{30 people}_\text{h} \rangle(y)$,
$\langle \text{in one block}_\text{h} \text{ of flats} \rangle(z)$, $\langle \text{by a storm surge}_\text{h} \rangle(w)$

Argument Relations:
$\text{ARG}(p_1, x)$, $\text{ARG}(p_1, y)$, $\text{ARG}(p_1, p_2)$,
$\text{ARG}(p_2, y)$, $\text{ARG}(p_2, z)$,
$\text{ARG}(p_3, z)$, $\text{ARG}(p_3, w)$,

Figure 1: UDS "flat" representation. Deeper analysis such as SPR and factuality is not shown.

The non-recursive or "flat" representation can be viewed as a Parson-style (Parsons, 1990) and underspecified version of neo-Davidsonianized RMRS (Copestake, 2007). As shown in Figure 1, the flat representation is a tuple $\mathcal{F} = \langle P, A \rangle$ where $P$ is a bag of predicates that are all maximally unary, and $A$ is a bag of arguments represented by separate binary relations.

**Predicate**: Predicates in PredPatt representation are referred as *complex predicates*: they are open-class predicates represented in the target language. Scope and lexical information in the predicates are left unresolved, yet can be recovered incrementally in deep semantic parsing. From the perspective of RMRS, complex predicates are conjunctions of underspecified *elementary predications* (Copestake et al., 2005) where handles are ignored, but syntax properties from Universal Dependencies are retained. For instance, in Figure 1, the subscript "h" in the predicate "$\langle$were reported$_\text{h}\rangle$" indicates that "reported" is a syntactic head in the predicate.

**Argument Relation**: The Parson-style flat representation makes arguments first-class predications $\text{ARG}(\cdot, \cdot)$. Using this style allows incremental addition of arguments, which is useful in shallow semantics where the arity of open-class predicate and the argument indexation are underspecified. They can be recovered when lexicon is available in deep analysis (Dowty, 1989; Copestake, 2007).

## B Linearizing Graph Representation

Figure 2 and figure 3 shows the UDS graph and linearized representations without deeper analysis such as SPR and factuality. The procedure of converting figure 2 to Figure 3 is following: Starting at the root node of the dependency tree (i.e., "reported$_\text{h}$"), we take an in-order traversal of its spanning tree. As the tree is expanded, brackets are inserted to denote the beginning or end of a predicate span, and parentheses are inserted to denote the beginning or end of an argument span. The subscript "h" indicates the syntactic head of each span. Intra-sentential coreference occurs when an instance refers to one of its preceding nodes, where we replace the instance with a special symbol "•" and add a coreference link between "•" and its antecedent.

## C Hyperparameters

**Encoder**: Word embeddings are randomly initialized 300d vectors sampled from $\mathcal{U}(-0.1, 0.1)$. The encoder RNN uses 2-layer bidirectional LSTMs with hidden state size of 500 and dropout rate at 0.3. Hidden states are zero initialized. All other parameters are sampled from $\mathcal{U}(-0.1, 0.1)$.

**Decoder**: Word embeddings are initialized by open-source GloVe vectors (Pennington et al., 2014) trained on Common Crawl 840B with 300 dimensions. The decoder RNN uses 2-layer LSTMs with hidden state size of 500 and dropout rate at 0.3. Hidden states are initialized by the last left-to-right hidden states of encoder. All other pa-

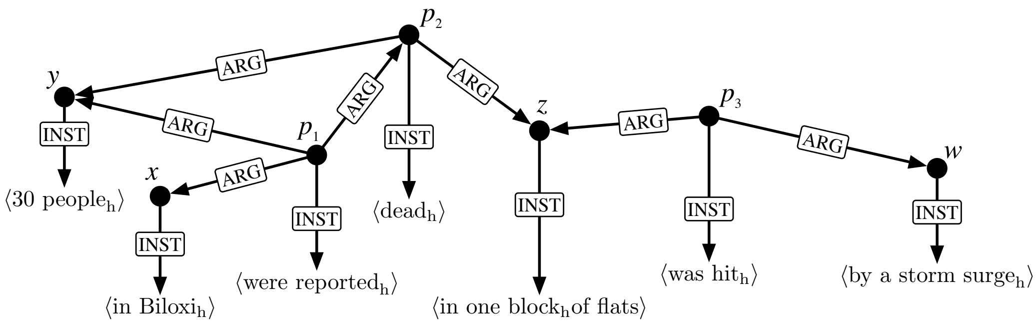


Figure 2: UDS graph representation.

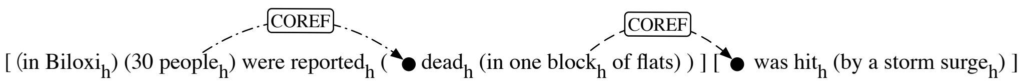


Figure 3: UDS Linearized representation.

rameters are sampled from $\mathcal{U}(-0.1, 0.1)$.
**Token Generation**: The feed-forward neural network is defined as

$$\text{FFNN}_\text{g}(s_t, c_t) = \tanh(W_\text{g} \begin{bmatrix} s_t \\ c_t \end{bmatrix} + b_\text{g}) \quad (1)$$

All transform matrices and bias used in generation are all sampled from $\mathcal{U}(-0.1, 0.1)$.
**Coref Link**: All feed-forward neural networks in the coreference annotating mechanism are defined as

$$\begin{aligned} \text{FFNN}(\boldsymbol{x}) =& W_3 \text{ReLU}(W_2 \text{ReLU}(W_1 \boldsymbol{x} + b_1) + b_2) \\ &+ b_3 \end{aligned}$$

where the sizes of $W_1$, $W_2$ and $W_3$ are $1000 \times 500$, $500 \times 500$ and $500 \times 1$ respectively. Dropout at rate of 0.3 is applied to the output of each layer. All transform matrices and bias used in the copying mechanism are all sampled from $\mathcal{U}(-0.1, 0.1)$.
**SPR module**: The SPR model is a two-layer perceptron:

$$\hat{D}^{(y_i, y_j)}_{\text{SPR}_p} = W_{\text{SPR}_p} \text{ReLU}(W_\text{shared}[\gamma(y_i), \gamma(y_j)]) \quad (2)$$

where size of $W_\text{shared}$ is $2648 \times 2648$ and sizes of all $W_{\text{SPR}_p}$ are $2648 \times 1$. All transform matrices used in the SPR model are all sampled from $\mathcal{U}(-0.1, 0.1)$.
**Factuality module**: The Factuality model is a two-layer perceptron:

$$\hat{D}^{(y_k)}_\text{FACT} = V_2 \text{ReLU}\left(V_1 \gamma(y_k) + b_1\right) + b_2, \quad (3)$$

where size of $V1$ and $V2$ are $1324 \times 1324$ and $1324 \times 1$. All transform matrices used in the factuality model are all sampled from $\mathcal{U}(-0.1, 0.1)$.

**Learning**: Adam optimizer (Kingma and Ba, 2014) with mini-batch gradient is used for optimization. The batch size is 64.

## D In-domain Test

Table 1 reports the experimental results on the **test** set. Results on the **in-domain test** set are similar and shown in the appendix. In Table 1, ***S* metric** 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: **BLEU**$_\text{INST}$ measures the BLEU score of all matched instance edges; **MAE**$_\text{SPR}$ measures the mean absolute error of SPR property scores of all matched argument edges; and **MAE**$_\text{FACT}$ measures the mean absolute error of factuality scores of all matched attribute edges.

| | $S$ **metric** | | | **BLEU**$_{\text{INST}}$ | **MAE**$_{\text{SPR}}$ | **MAE**$_{\text{FACT}}$ |
|---|---|---|---|---|---|---|
| | Prec. | Rec. | F1 | | | |
| Pipeline | 35.42 | 23.53 | 28.27 | 14.80 | N/A | N/A |
| Variant (a) | 37.46 | 25.91 | 30.63 | 15.67 | 0.45 | 0.77 |
| Variant (b) | 41.27 | 26.41 | 32.21 | 16.89 | 0.44 | 0.79 |
| Variant (c) | 40.27 | 26.40 | 31.89 | 16.60 | 0.43 | **0.74** |
| Our model | **44.17** | **27.04** | **33.55** | **17.90** | **0.43** | 0.78 |

Table 1: Evaluation of results on the in-domain test set.

## References


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

David R Dowty. 1989. On the semantic content of the notion of thematic role. In *Properties, types and meaning*, pages 69–129. Springer.

Diederik Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Terence Parsons. 1990. *Events in the Semantics of English*, volume 5. Cambridge, Ma: MIT Press.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.