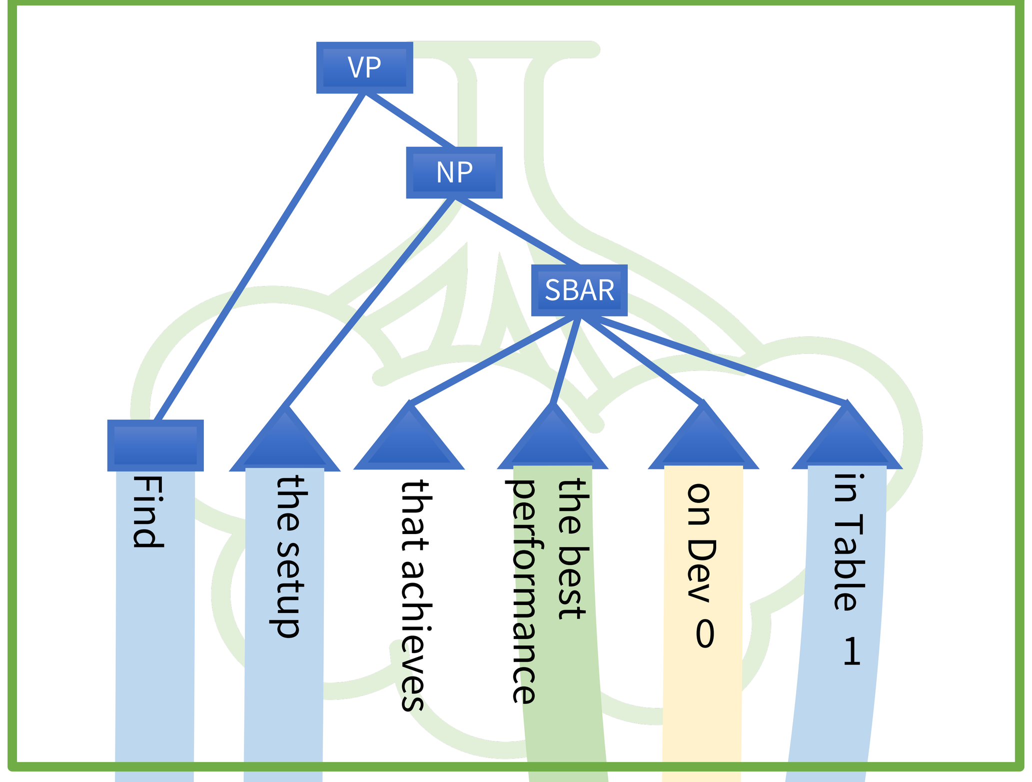
Unsupervised Learning of Neural Semantic Mappings with the Hungarian Algorithm for Compositional Semantics

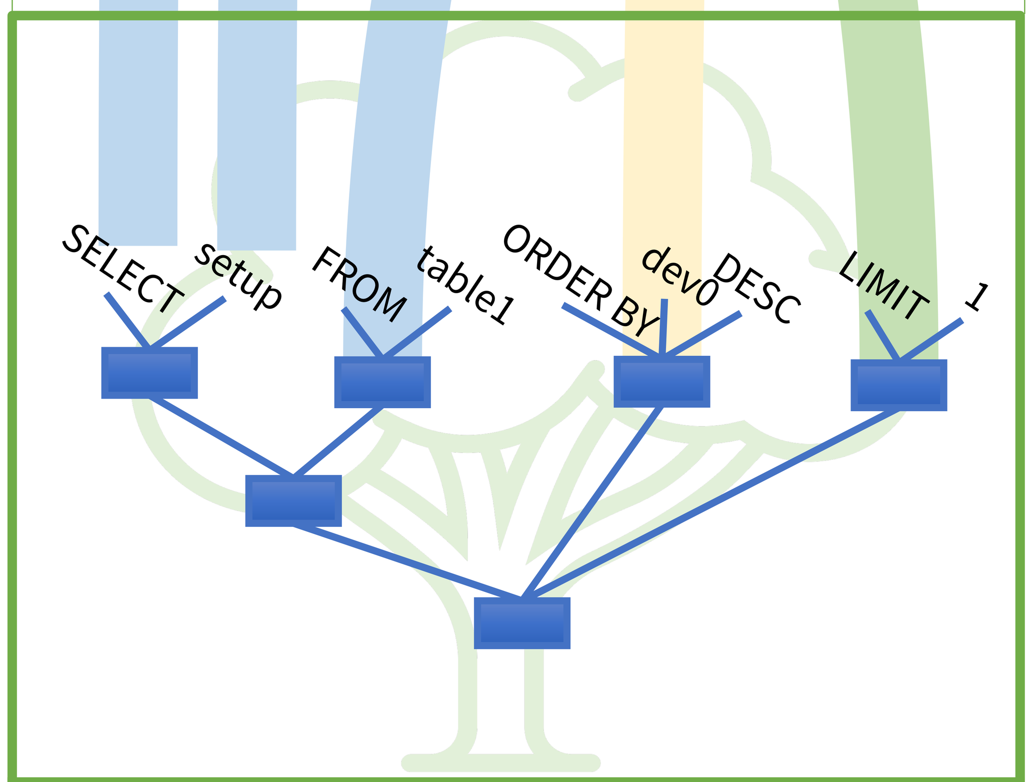
Xiang Zhang[1], Shizhu He[2], Kang Liu[2], Jun Zhao[2]

Compositional Semantics



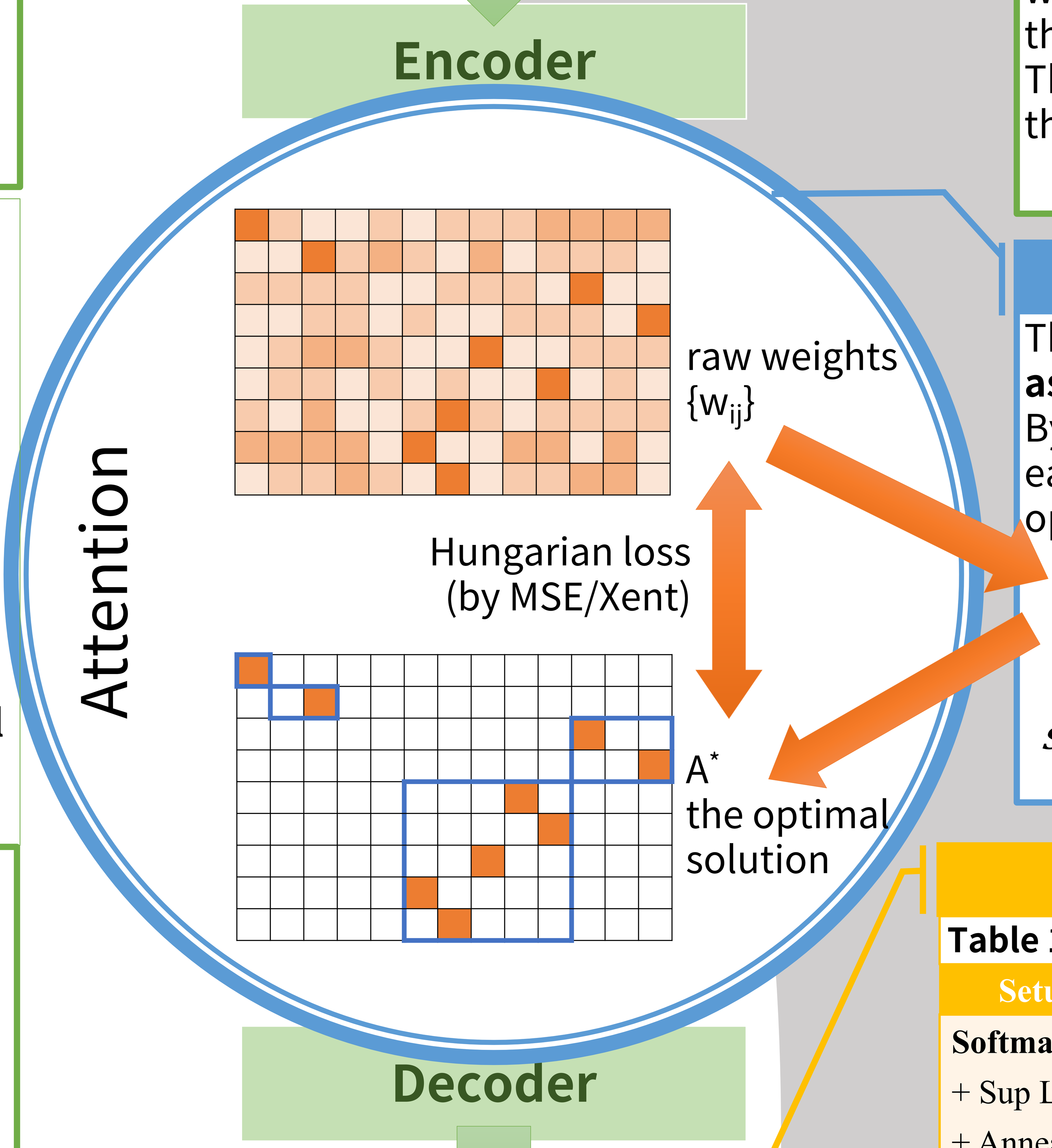
The "Compositional Semantics" usually denotes that the underlying semantic theory is compositional, in which a large expression has the meaning composed by those meanings of smaller expressions.

The compositionality is claimed to be a crucial feature of the natural languages but is only strictly defined in the formal languages. Thus, the latter usually serve as the semantical notations/representations for natural languages.



Compositional Semantics

Find the setup that achieves the best performance on Dev0 in Table 1



SELECT setup FROM table1
ORDER BY dev0 DESC LIMIT 1

Goal

INPUT: natural language (NL)
OUTPUT: formal language (FL)
such that: they have equivalent semantics.

According to the Compositional Semantics, we assume there are mappings between the surface forms of the input and output. The Hungarian algorithm is brought to find the mappings without human labels.

Method

The Hungarian algorithm solves the **assignment problem** in multinomial time. By casting the mappings into assignments, each attention weight matrix defines an optimization problem as:

$$\max_A R(A) = \sum_{a_{ij} \in A} a_{ij} w_{ij}$$

$$s.t.\ a_{ij} \in \{0,1\}, \sum_j a_{ij} \le 1, \sum_i a_{ij} \le 1$$

Results (Performance)

Table 1. Acc (%) on various Dev splits of SQuALL data.

Setup	Dev0	Dev1	Dev2	Dev3	Dev4	Mean
Softmax	40.6	44.8	43.6	**46.9**	45.2	44.2
+ Sup Loss	43.4	**47.6**	**43.7**	45.7	**46.4**	**45.4**
+ Annealing	38.6	41.6	39.5	40.6	44.7	41.0
+ **Hungarian**	**43.5**	47.2	42.7	44.3	**46.4**	44.8
SparseMAX	36.6	40.1	35.8	35.0	39.2	37.3
+ Sup Loss	**42.7**	**46.3**	**43.6**	**44.4**	45.0	**44.4**
+ **Hungarian**	42.3	44.6	40.8	43.4	**45.5**	43.3
Oracle	59.5	64.0	59.5	58.7	61.0	60.5
+Sup Loss	**61.8**	**65.4**	**61.4**	**60.8**	**62.5**	**62.4**

1. The best results are obtained by **Oracle**, where only mapping arcs with supervised labels will receive non-zero attention weights.
2. Our **Hungarian tweaks** surpass sparse baselines, including **SparseMAX** and **Annealing.** They are also very close to the supervised training.

Reverse Hungarian

Table 3. Results (Acc, Gini) on the merged dataset with ATIS, Geo, Scholar, Advising

Setup	Dev	Test	Gini
Seq2Seq	71.35±1.13	69.87±0.81	0.8199
+ Hungarian	65.81±1.73	64.52±0.95	**0.8757**
+ Reverse Hungarian	**71.50**±0.80	**70.42**±1.26	0.8535

1. The **reverse Hungarian** tweak does not find the optimal solution to build an attention loss but finds the worst solution to remove the least possible mappings as noises from the attention.
2. The simple tweak is applicable to datasets which cannot be reduced to 1-to-1 assignment problem.
3. The reverse tweak encourages the sparsity moderately (to 0.8535), but also improves the results when manual labels are not available.

Results (Decoded Mapping)

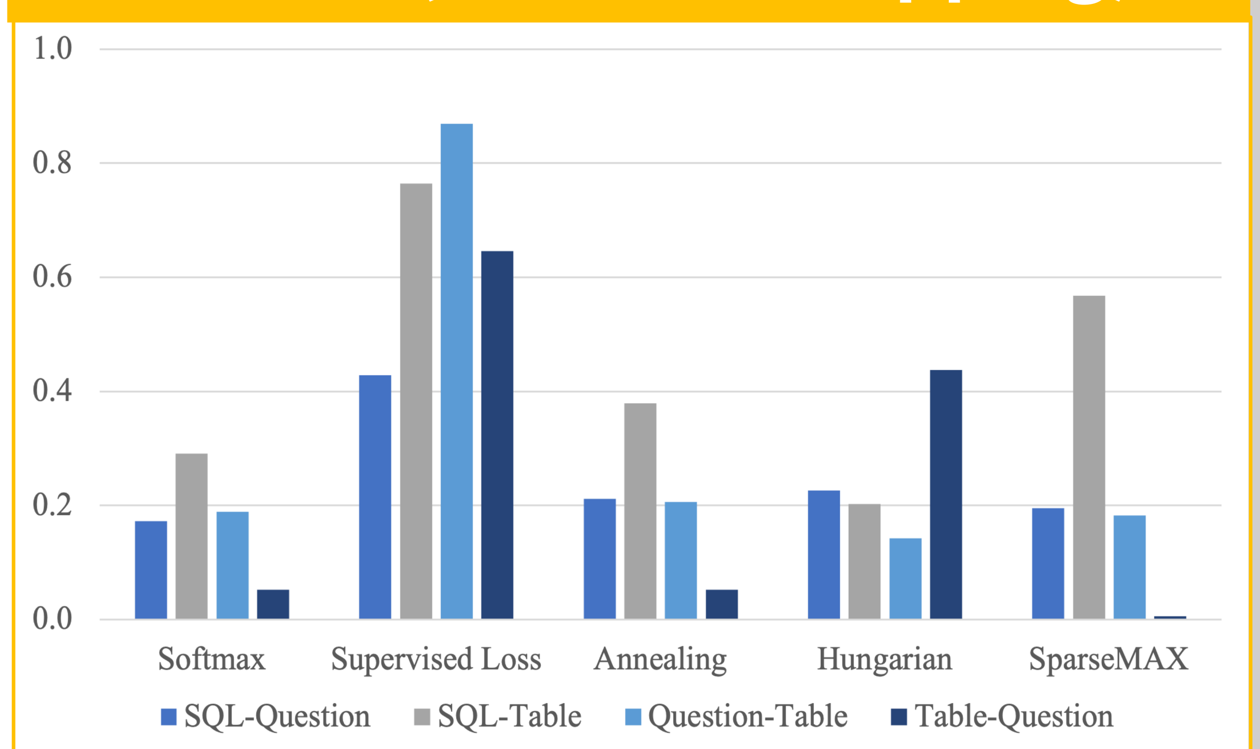


Figure 1. Gini indices on Dev0 for various setups.

The Hungarian tweaks significantly improve the recall for Table-Question attentions, i.e., more then 40% mappings of table-question are predicted correctly, only left behind the Supervised training.

Results (Sparsity)

Table 2. Gini indices on Dev0 splits for various setups.

Setup	S-Q	S-T	Q-T	T-Q
Softmax	**0.7685**	**0.8272**	**0.9141**	**0.9013**
+ Sup Loss	0.4885	0.2490	0.2351	0.3499
+ annealing	0.8949	0.9091	0.9648	0.9981
+ **Hungarian**	**0.4455**	**0.2368**	**0.1327**	**0.5577**
SparseMAX	0.9004	0.8280	0.8295	0.9010
+ Sup Loss	0.5434	0.2497	0.2506	0.4166
+ **Hungarian**	0.5892	0.3023	0.2700	0.6970
Oracle	**0.0173**	**0.7409**	**0.7157**	**0.8426**
+ Sup Loss	0.0178	0.7467	0.7240	0.8235

The **supervised** training and the **Oracle** indicate the attentions are **not that significantly sparse**, especially on the SQL-Question attention (only a few are labeled as real mappings, most of them are not related).
The **Hungarian tweaks** can reduce the unwanted S-Q sparsity and encourage the T-Q attentions.

[1]School of Artificial Intelligence, University of Chinese Academy of Sciences
[2]The Laboratory of Cognition and Decision Intelligence for Complex Systems, Institute of Automation, Chinese Academy of Sciences