UCCA supports reentrancy, discontinuity and non-terminal nodes, which are essential for representing natural language semantics. We present the first parser for UCCA and the first to support these properties.

### Experimental Setup

**Corpora.** English Wikipedia (in-domain), English part of Twenty Thousand Leagues Under the Sea (English-French parallel corpus) out-of-domain.

**Evaluation.** Labeled precision, recall and F-score on graph edges, represented by their terminal yields. Primary and remote evaluated separately.

**Baselines.** Parsers trained on bilexical graphs and trees converted from UCCA training set, and evaluated by converting test set output to UCCA.

### Results

{TUPA}_{LSTM} obtains the highest F-scores in all metrics:

<table>
<thead>
<tr>
<th>Wiki (in-domain)</th>
<th>20K Leagues (out-of-domain)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20K</td>
</tr>
<tr>
<td></td>
<td>Gold</td>
</tr>
<tr>
<td></td>
<td>Wiki</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

|                  |       |      |       |      |
|                  |       |      |       |      |

As no direct comparison with existing parsers is possible, we compare TUPA to bilexical dependency graph parsers, which support reentrancy and discontinuity but not terminal nodes. We also convert UCCA to bilexical trees and evaluate constituency and dependency tree parsers on them, by simply removing remote edges from the graph.

### Conclusions

We present TUPA, the first parser for UCCA, and show that with a NN classifier and BLSTM feature extractor, it accurately predicts UCCA graphs from text, outperforming various strong baselines. Future work will explore different target representations, and apply the parser to more languages, demonstrating the importance of broad-coverage parsers. A parser for UCCA will enable the framework for new tasks.

### References


