Background

- Do we need linguistics in NLP?
- Empirical success of stochastic NLP
  - Black box problem of neural NLP
- Symbolic NLP and Grammar Engineering
  - Show high precision but narrow coverage (Bender and Emerson, 2021)
- Hajdik et al. (2019)
  - HPSG-based engineered grammar (ERG) with neural machine translation (BiLSTM + Attention) for NLG

Suggestion


Method

- Data
  - Gold Dataset: Redwoods
  - Manually inspected MRS representations
  - Silver Dataset: MRS with ERG with no manual inspection from Gigawords Corpus
  - Total data: 984,679 MRS-sentence pairs
  - Anonymized according to ERG's NER to reduce data sparsity
  - Data and processor by Hajdik et al. (2019)
- Linearization:
  - MRS – DMRS – PENMAN – Single line string

- Model
  - Transformer
  - OpenNMT-py
- Hyperparameters
  - From the manual (mimics the original Transformer)
  - Validation on every 5,000 steps to save model frequently
- Evaluation: SacreBLEU

RESULTS

- BLEU score
  - BLEU measured for every 5,000 steps
  - Score peaked at 30,000 steps with 64.2 BLEU
  - 12.97 lower than Hajdik et al. (2019)
  - Score decreased after 30,000 steps with overfitting

Translation Samples

1. a. prediction: If I am correct, they will help you understand exactly what it is saying the Linux community of good software—and perhaps they will help you become more productive yourself.
   b. answer: If I’m correct, they’ll help you understand exactly what it is that makes the Linux community such a fountain of good software—and, perhaps, they will help you become more productive yourself.

2. a. prediction: The myth and the sword.
   b. answer: The Cathedral and the Bazaar

Error Analysis

<table>
<thead>
<tr>
<th>Error</th>
<th>Number</th>
<th>Sample Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Error</td>
<td>57</td>
<td>Okay, we have some small issues.</td>
</tr>
<tr>
<td>Formulation</td>
<td>32</td>
<td>I assume there is a sale on the shipping costs.</td>
</tr>
<tr>
<td>Lexical &amp; Grammatical</td>
<td>6</td>
<td>“cereal”</td>
</tr>
<tr>
<td>Missing Argument</td>
<td>6</td>
<td>This is light — the high page cost is generally a major UNK contributor for UNK-advertisers who want to appear regularly in a publication or not at all.</td>
</tr>
<tr>
<td>Punctuation</td>
<td>5</td>
<td>I’ll be happy to chat with you</td>
</tr>
<tr>
<td>Missing Argument</td>
<td>3</td>
<td>Requesting immediately.</td>
</tr>
<tr>
<td>Syntactic</td>
<td>2</td>
<td>A part of speech</td>
</tr>
</tbody>
</table>

- Attention Weight Distribution

DISCUSSION

- Summary
  - Reproduced Hajdik et al. (2019) with Transformer
  - Great syntactic reconstruction, failed at lexical choices
  - Attention-based approach is suboptimal for MRS

- Significance of the Research
  - ERG is already a robust system, but with narrow coverage
  - Broadening ERG with neural networks

References
