Cross-Domain Generalization of Neural Constituency Parsers

Daniel Fried*, Nikita Kitaev*, and Dan Klein
But Coleco bounced back with the introduction of the Cabbage Patch dolls, whose sales hit $600 million in 1985.

Several of the heterogeneous clinical manifestations of systemic lupus erythematosus have been associated with specific autoantibodies.

Where can I get morcillas in tampa bay, I will like the Argentinian type, but I will to try another please?
Penn Treebank Parsing by the Numbers

Year
Penn Treebank F1
96 95 94 93 92 91 90 89

No Structure

Choe & Charniak
Dyer+
Stern+
Durrett+
Socher+
Kitaev & Klein
Kitaev+
Fried+
Joshi+
Liu & Zhang
Kitaev & Klein
McClosky+
Sagae & Lavie
Zhang+
Charniak & Johnson
Carreras+
Zhu+
Shindo+
Mcclosky+
Charniak, Collins
Henderson
Petrov+

Non-Neural  Neural
Methodology

**Non-neural:**
- Berkeley [Petrov and Klein 2007]
- BLLIP [Charniak and Johnson 2005]
- ZPar (Chinese) [Zhang and Clark 2011]

**Neural:**
- Self-Attentive Chart [Stern et al. 2017; Kitaev and Klein 2018]
- In-Order Recurrent Neural Network Grammars (RNNG) [Dyer et al. 2016; Kuncoro et al. 2017; Liu and Zhang 2017]
### Methodology

#### Zero-shot generalization setup:

<table>
<thead>
<tr>
<th>Language</th>
<th>Train</th>
<th>Test out-of-domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>English</strong></td>
<td>Newswire (PTB WSJ)</td>
<td>Literature (Brown)</td>
</tr>
<tr>
<td><strong>Chinese</strong></td>
<td>Newswire (CTB v5)</td>
<td>TV News (CTB v8)</td>
</tr>
</tbody>
</table>
Fact or Myth?

Neural parsers transfer less reliably than non-neural?

Pre-trained representations are most useful out-of-domain?

Structure helps in generalization?
Fact or Myth?

Neural parsers transfer less reliably than non-neural?

Pre-trained representations are most useful out-of-domain?

Structure helps in generalization?
Neural vs. Non-Neural Generalization

Chinese Corpora

<table>
<thead>
<tr>
<th>Corpus</th>
<th>In-Domain</th>
<th>Out-of-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTB v5 (mostly newswire)</td>
<td>83.0</td>
<td>77.2</td>
</tr>
<tr>
<td>Broadcast News</td>
<td>77.8</td>
<td>74.3</td>
</tr>
<tr>
<td>Web Forums</td>
<td>75.7</td>
<td>73.9</td>
</tr>
<tr>
<td>Blogs</td>
<td>74.7</td>
<td></td>
</tr>
</tbody>
</table>

F1 scores for ZPar (Non-Neural) and In-Order RNNG (Neural) models.
Fact or Myth?

Neural parsers transfer less reliably than non-neural? **MYTH**

Pre-trained representations are most useful out-of-domain?

Structure helps in generalization?
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-trained representations are most useful out-of-domain?

Structure helps in generalization?
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-trained representations are most useful out-of-domain?

Structure helps in generalization?
Effects of Pre-Trained Representations

<table>
<thead>
<tr>
<th>Corpus</th>
<th>In-Order RNNG</th>
<th>+Word Embeddings</th>
<th>+BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penn Treebank</td>
<td>91.5</td>
<td>92.1</td>
<td>95.7</td>
</tr>
<tr>
<td>Brown Corpus</td>
<td>85.6</td>
<td>86.8</td>
<td>93.5</td>
</tr>
<tr>
<td>Genia</td>
<td>80.3</td>
<td>81.6</td>
<td>87.8</td>
</tr>
<tr>
<td>English Web</td>
<td>79.1</td>
<td>80.5</td>
<td>89.3</td>
</tr>
</tbody>
</table>

% Relative Error Reductions:

- In-Domain: -8%, -7%, -8%, -7%
- Out-of-Domain: -50%, -55%, -38%, -49%
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-trained representations are most useful out-of-domain? **MYTH**

Structure helps in generalization?
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

Structure helps in generalization?
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

Structure helps in generalization?
Structured Decoding?

**Unstructured**
Self-Attentive Chart Parser
[Stern et al. 2017, Kitaev and Klein 2018]

Condition on sentence only

---

**Structured**
In-Order RNNG
[Dyer et al. 2016, Liu and Zhang 2017]

Also condition on predicted structure

---

We wanted more structure
Structure Helps More Out-of-Domain

% Relative Error Reductions:

<table>
<thead>
<tr>
<th>Dataset</th>
<th>In-Domain</th>
<th>Out-of-Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penn Treebank</td>
<td>95.6</td>
<td>95.7</td>
</tr>
<tr>
<td>Brown Corpus</td>
<td>93.1</td>
<td>93.5</td>
</tr>
<tr>
<td>Genia</td>
<td>88.7</td>
<td>89.3</td>
</tr>
<tr>
<td>English Web</td>
<td>87.5</td>
<td>87.8</td>
</tr>
</tbody>
</table>

Relative Error Reductions:
- Penn Treebank: -2%
- Brown Corpus: -5%
- Genia: -2%
- English Web: -2%
Structure Helps More Out-of-Domain

<table>
<thead>
<tr>
<th></th>
<th>In-Domain</th>
<th>Labeled Bracket F1 (English)</th>
<th>Out-of-Domain</th>
<th>% Relative Error Reductions:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Penn Treebank</td>
<td>95.6</td>
<td>95.7</td>
<td>-2%</td>
</tr>
<tr>
<td></td>
<td>Brown Corpus</td>
<td>93.1</td>
<td>93.5</td>
<td>-6%</td>
</tr>
<tr>
<td></td>
<td>Genia</td>
<td>88.7</td>
<td>89.3</td>
<td>-5%</td>
</tr>
<tr>
<td></td>
<td>English Web</td>
<td>87.5</td>
<td>87.8</td>
<td>-2%</td>
</tr>
</tbody>
</table>

Unstructured (Chart + BERT) | Structured (In-Order + BERT)
Structure Helps with Larger Spans

F1 by minimum span length, on English Web
Structure Improves Exact Match

Exact Parse Match Accuracies (English)

% Relative Error Reductions:

-4%  -6%  -4%  -1%

Exact Match %

Penn Treebank 55.1 57.1
Brown Corpus 49.2 52.0
Genia 41.8 44.0
English Web 17.5 18.0

Unstructured (Chart + BERT)  Structured (In-Order + BERT)
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

Structure helps in generalization? **FACT**
Fact or Myth?

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

Structure helps in domain transfer, longer spans, and whole parses.
Conclusions

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

Structure helps in domain transfer, longer spans, and whole parses.
Thank you!

Code and models:

Chart + BERT: parser.kitaev.io
In-Order RNNG + BERT: github.com/dpfried/rnng-bert
Syntactic parses can improve system performance, even for neural models

[Roth and Lapata 2016; Andreas et al. 2016; Aharoni and Goldberg 2017; Strubell et al. 2018; Swayamdipta et al. 2018; Hale et al. 2018; Kuncoro et al. 2018; Kim et al. 2019; He et al. 2019]
Structure Helps with Larger Spans

F1 by minimum span length, on Genia corpus
Effects of Pre-Trained Representations

Chinese Corpora

% Relative Error Reductions:

CTB v5 (newswire)
- In-Domain: 83.7, -50%
- Out-of-Domain: 85.7, -12%

Broadcast News
- In-Domain: 77.8, -17%
- Out-of-Domain: 81.6, -48%

Web Forums
- In-Domain: 75.7, -15%
- Out-of-Domain: 79.4, -47%

Blogs
- In-Domain: 74.7, -14%
- Out-of-Domain: 78.2, -38%