Cross-Domain Generalization of Neural Constituency Parsers

Daniel Fried*, Nikita Kitaev*, and Dan Klein

Cross-Domain Transfer

Penn Treebank

But Coleco bounced back with the introduction of the Cabbage Patch dolls, whose sales hit $600 million in 1985.
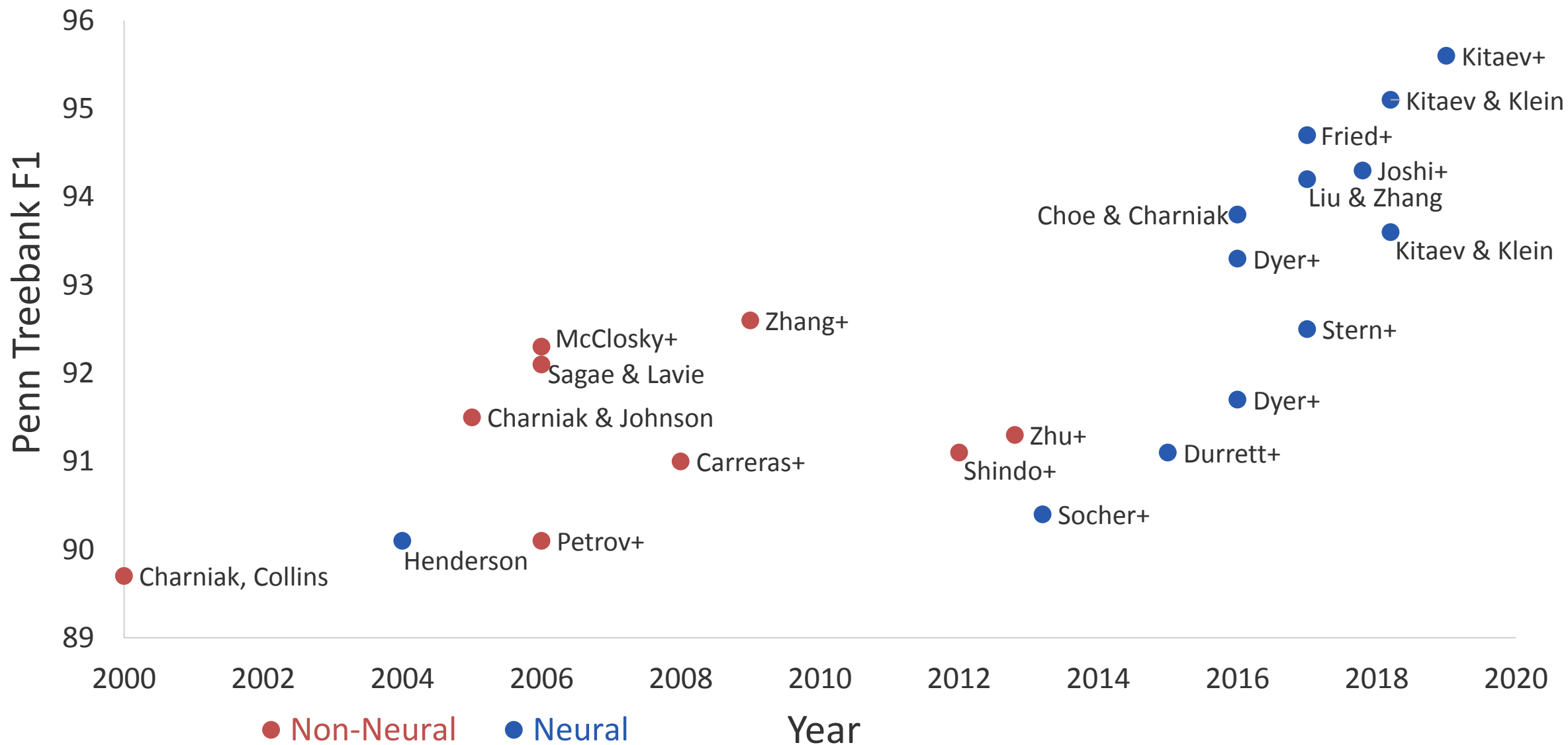

Genia

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

English Web Treebank

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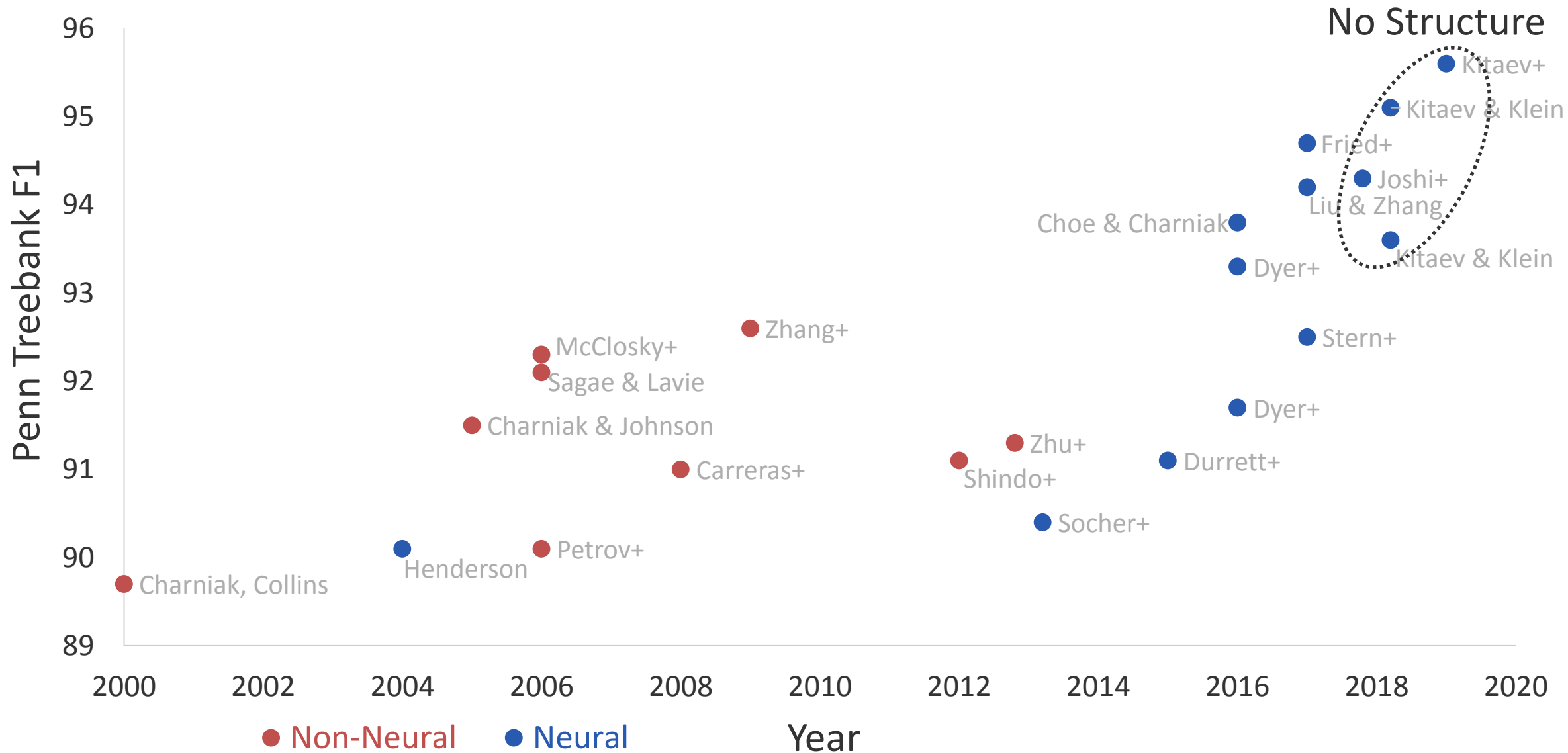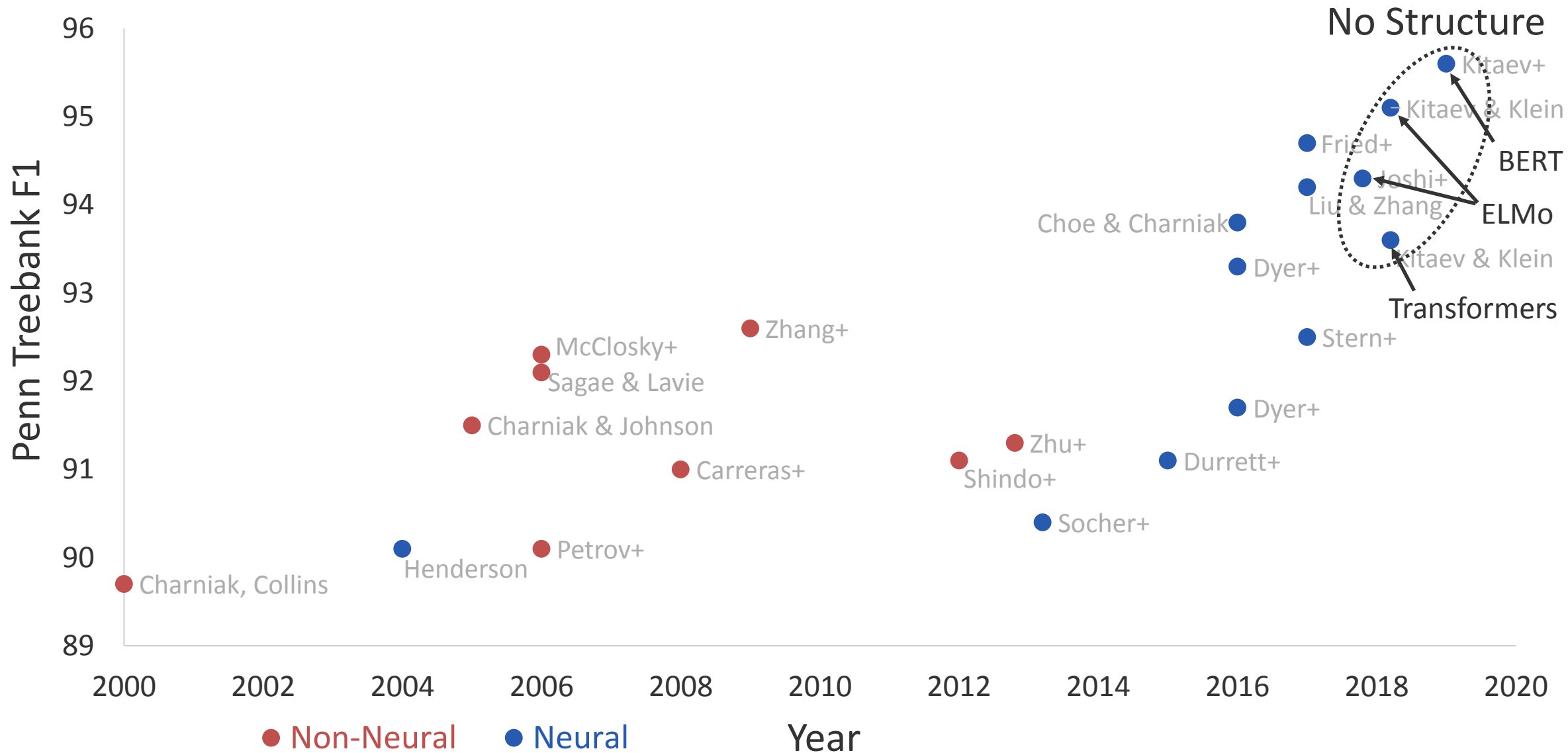
Penn Treebank Parsing by the Numbers

Penn Treebank F1 (y-axis, 89 to 96)
Year (x-axis, 2000 to 2020)

Data points:
- Charniak, Collins (Non-Neural)
- Henderson (Neural)
- Petrov+ (Non-Neural)
- Charniak & Johnson (Non-Neural)
- Sagae & Lavie (Non-Neural)
- McClosky+ (Non-Neural)
- Carreras+ (Non-Neural)
- Zhang+ (Non-Neural)
- Shindo+ (Non-Neural)
- Zhu+ (Non-Neural)
- Socher+ (Neural)
- Durrett+ (Neural)
- Dyer+ (Neural)
- Stern+ (Neural)
- Dyer+ (Neural)
- Choe & Charniak (Neural)
- Liu & Zhang (Neural)
- Fried+ (Neural)
- Joshi+ (Neural)
- Kitaev & Klein (Neural)
- Kitaev & Klein (Neural)
- Kitaev+ (Neural)

Legend: ● Non-Neural ● Neural

Penn Treebank Parsing by the Numbers

Penn Treebank Parsing by the Numbers

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:

	Train	Test out-of-domain		
English	Newswire (PTB WSJ)	Literature (Brown)	Biomedical (Genia)	Web newsgroups, reviews, questions (EWT)
Chinese	Newswire (CTB v5)	TV News (CTB v8)	Web forums (CTB v8)	Blogs (CTB v8)

Fact or Myth?

Neural parsers transfer less reliably than non-neural?

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

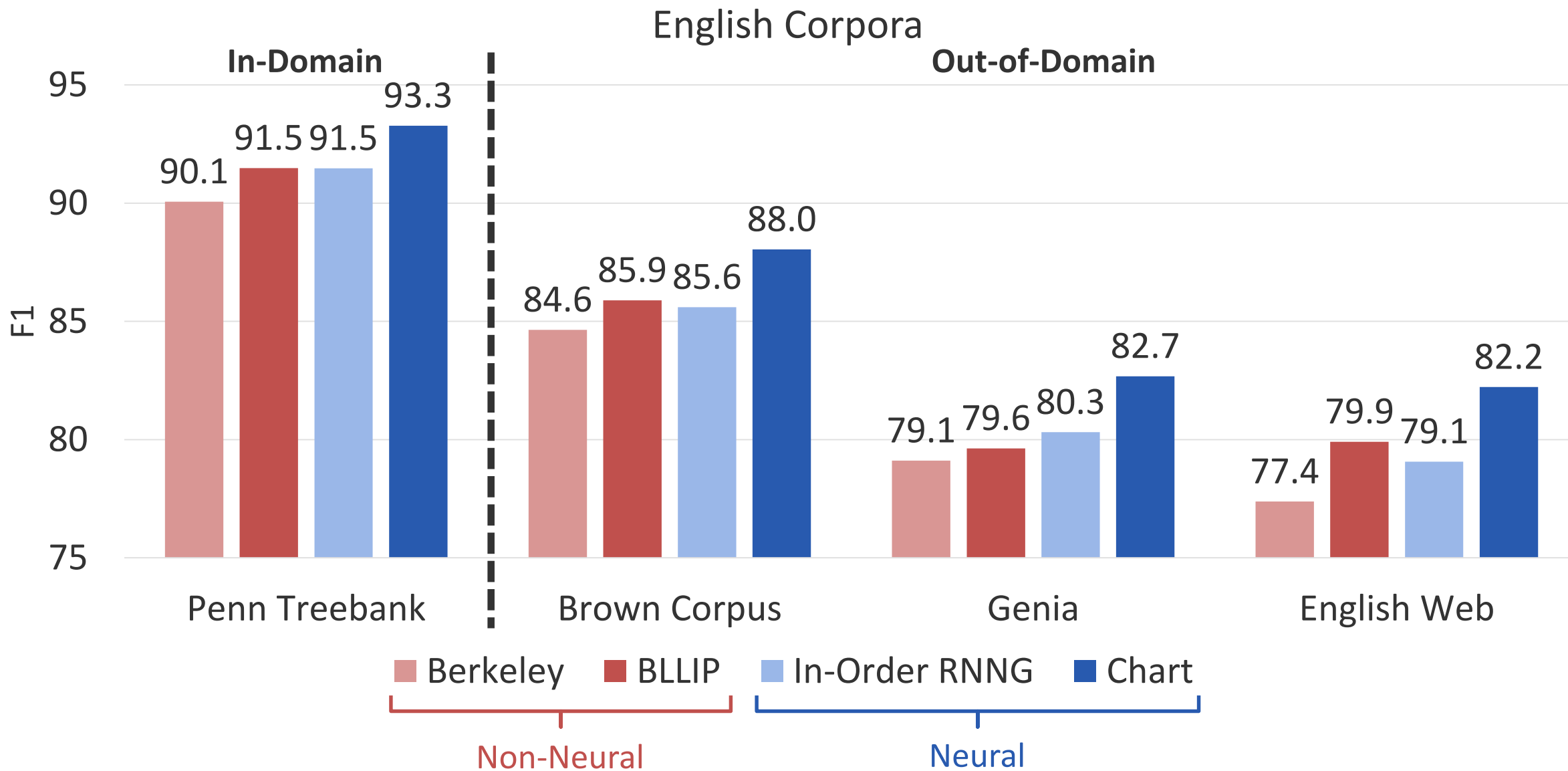Structure helps in generalization?

Fact or Myth?

▶ Neural parsers transfer less reliably than non-neural?

Pre-trained representations are most useful out-of-domain?
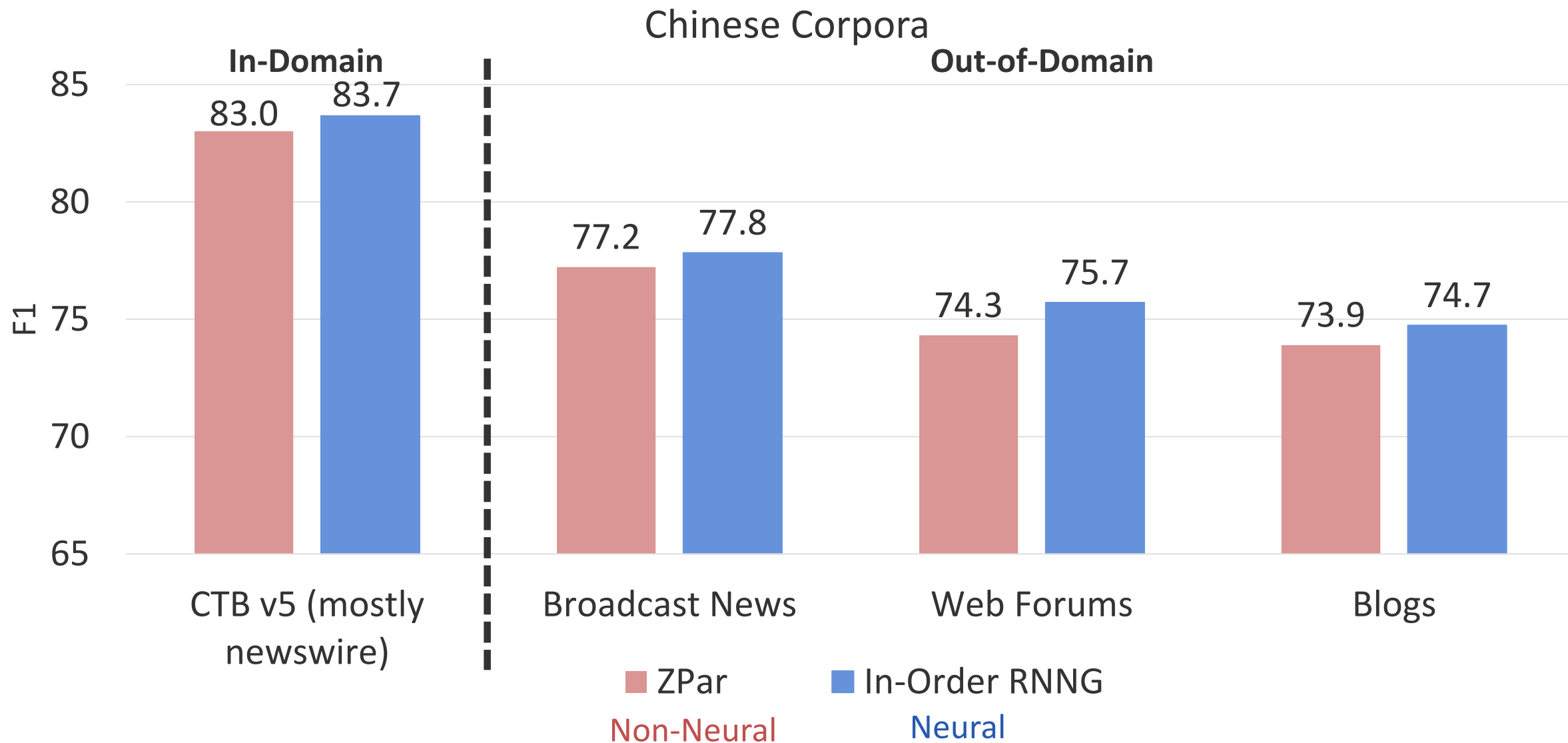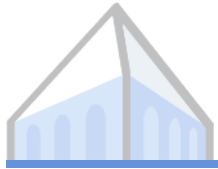
Structure helps in generalization?

Neural vs. Non-Neural Generalization

English Corpora

In-Domain | **Out-of-Domain**

F1

Penn Treebank — In-Domain:
- Berkeley: 90.1
- BLLIP: 91.5
- In-Order RNNG: 91.5
- Chart: 93.3

Brown Corpus:
- Berkeley: 84.6
- BLLIP: 85.9
- In-Order RNNG: 85.6
- Chart: 88.0

Genia:
- Berkeley: 79.1
- BLLIP: 79.6
- In-Order RNNG: 80.3
- Chart: 82.7

English Web:
- Berkeley: 77.4
- BLLIP: 79.9
- In-Order RNNG: 79.1
- Chart: 82.2

Legend: Berkeley, BLLIP, In-Order RNNG, Chart

Non-Neural | Neural

Neural vs. Non-Neural Generalization

Chinese Corpora

Fact or Myth?

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

Pre-trained representations are most useful out-of-domain?
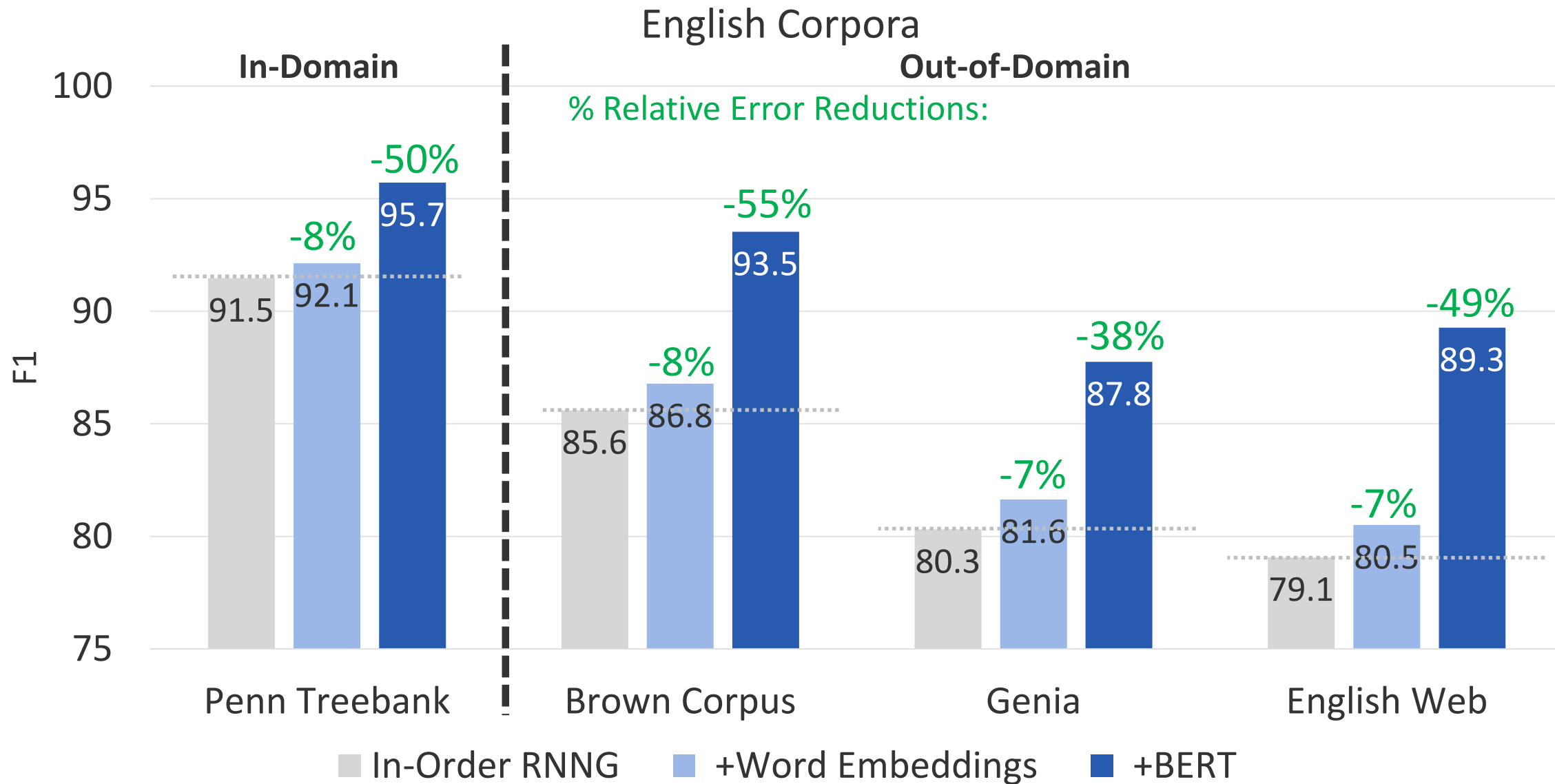
Structure helps in generalization?

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

English Corpora

In-Domain

Out-of-Domain

% Relative Error Reductions:

	In-Order RNNG	+Word Embeddings	+BERT
Penn Treebank	91.5	92.1 (-8%)	95.7 (-50%)
Brown Corpus	85.6	86.8 (-8%)	93.5 (-55%)
Genia	80.3	81.6 (-7%)	87.8 (-38%)
English Web	79.1	80.5 (-7%)	89.3 (-49%)

F1

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.

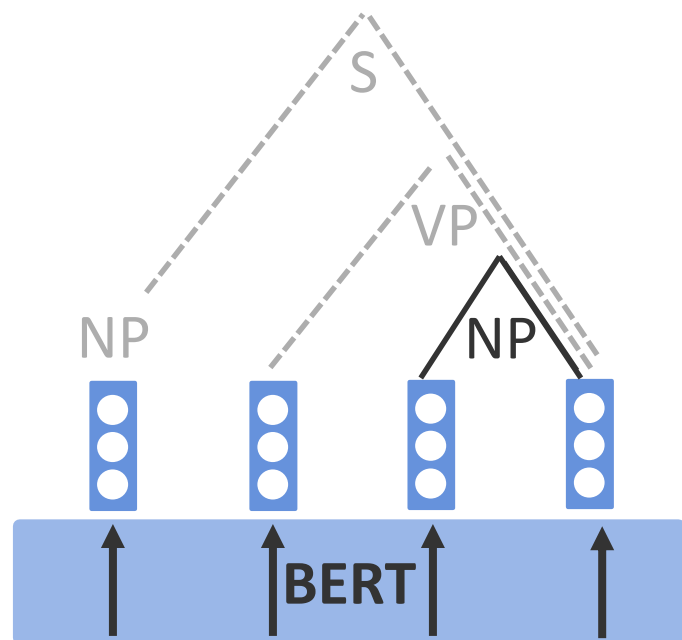▶ Pre-training helps across domains.

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

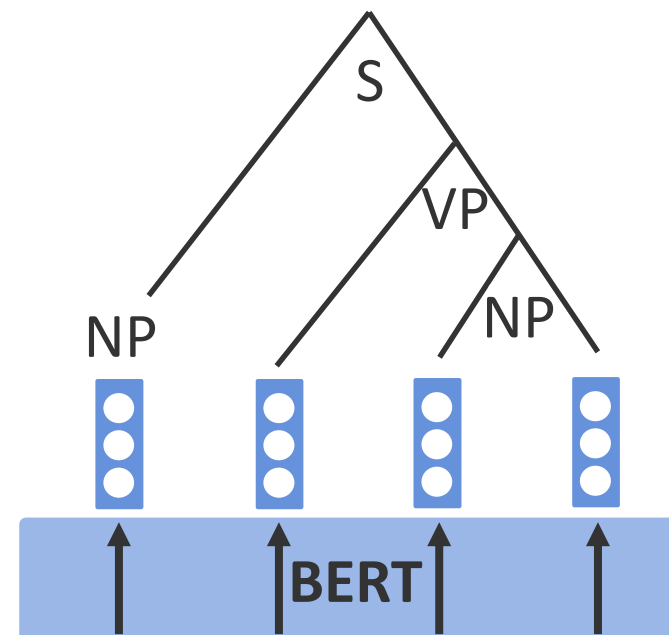Self-Attentive Chart Parser

[Stern et al. 2017, Kitaev and Klein 2018]

Structured

In-Order RNNG

[Dyer et al. 2016, Liu and Zhang 2017]

Condition on sentence only

Also condition on predicted structure

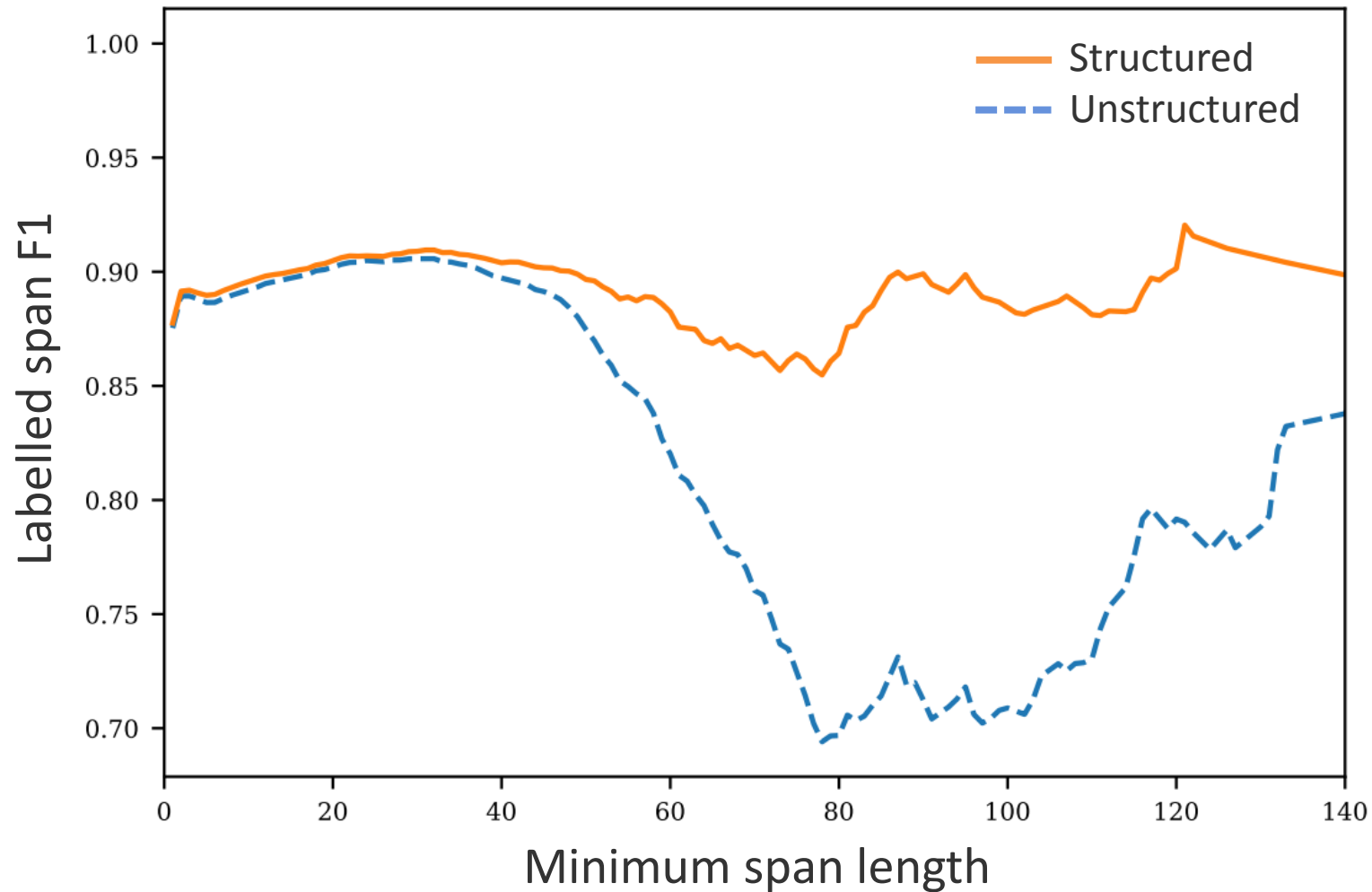# Structure Helps More Out-of-Domain
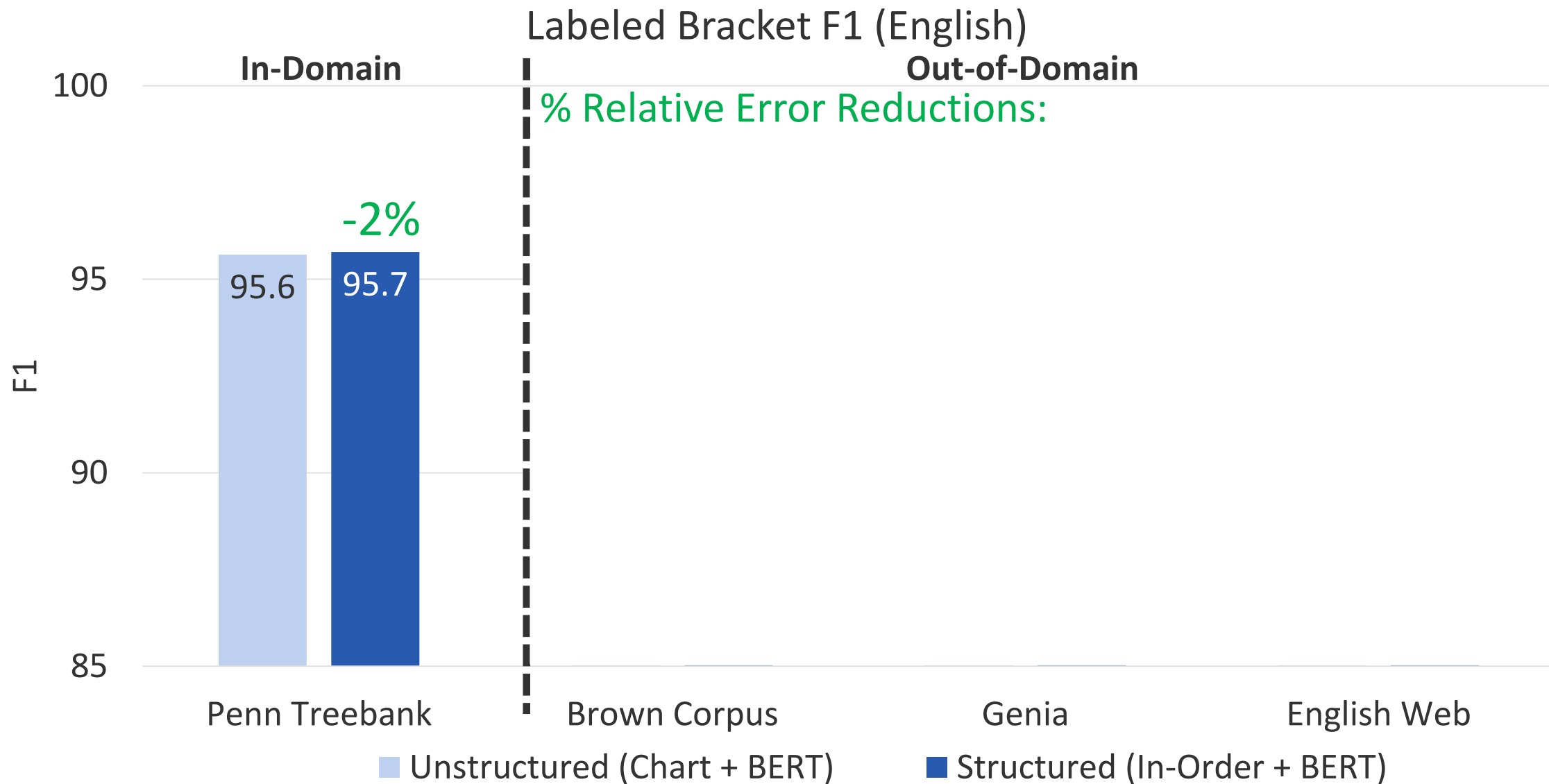
Labeled Bracket F1 (English)

In-Domain

Out-of-Domain

% Relative Error Reductions:

-2%

95.6 95.7

F1

100

95

90

85

Penn Treebank   Brown Corpus   Genia   English Web

■ Unstructured (Chart + BERT)   ■ Structured (In-Order + BERT)

Structure Helps More Out-of-Domain

Labeled Bracket F1 (English)

In-Domain

Out-of-Domain

% Relative Error Reductions:

F1

100

95

90

85

-2%

95.6 | 95.7

-6%

93.1 | 93.5

-5%

88.7 | 89.3

-2%

87.5 | 87.8

Penn Treebank | Brown Corpus | Genia | English Web
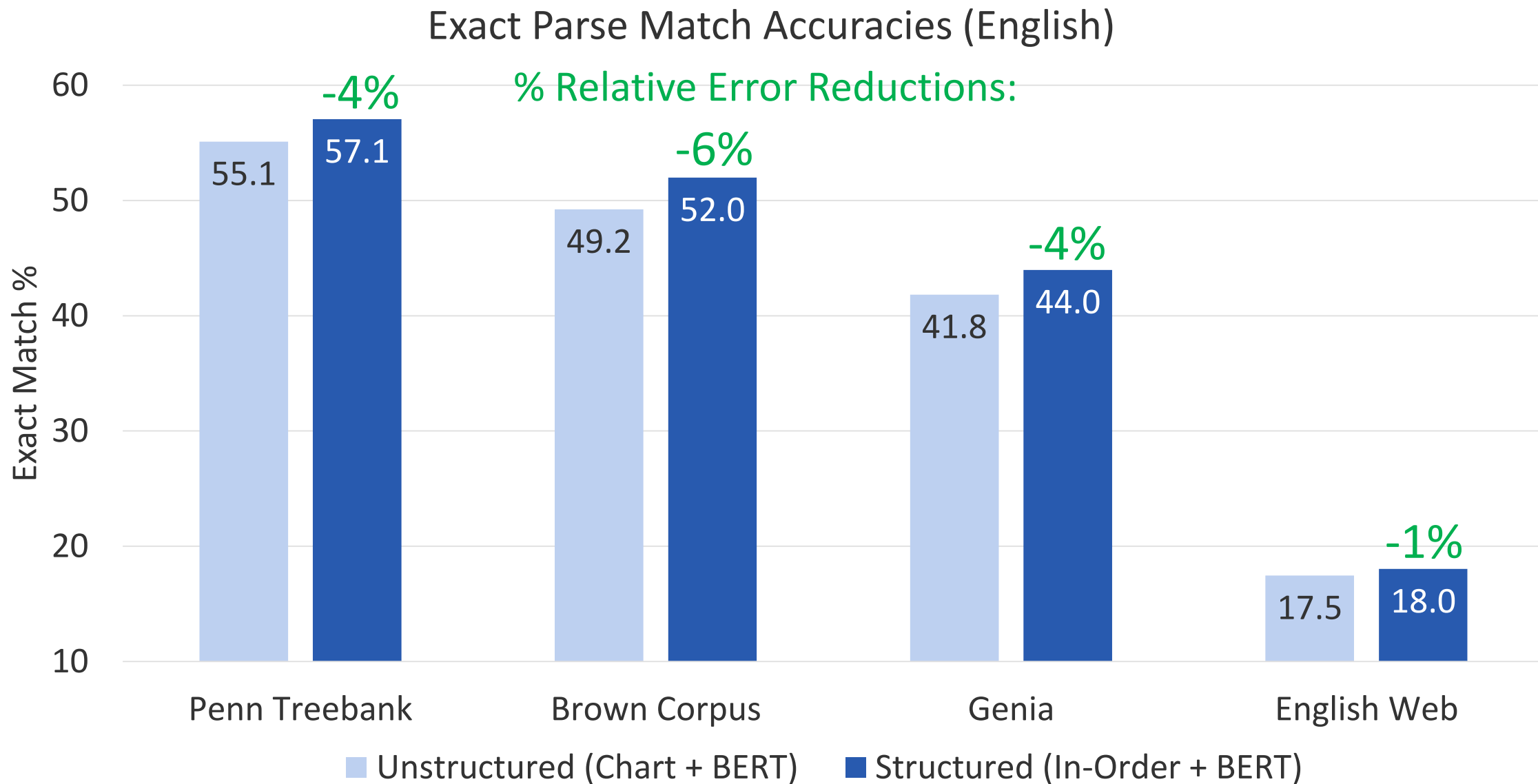
■ Unstructured (Chart + BERT)   ■ Structured (In-Order + BERT)
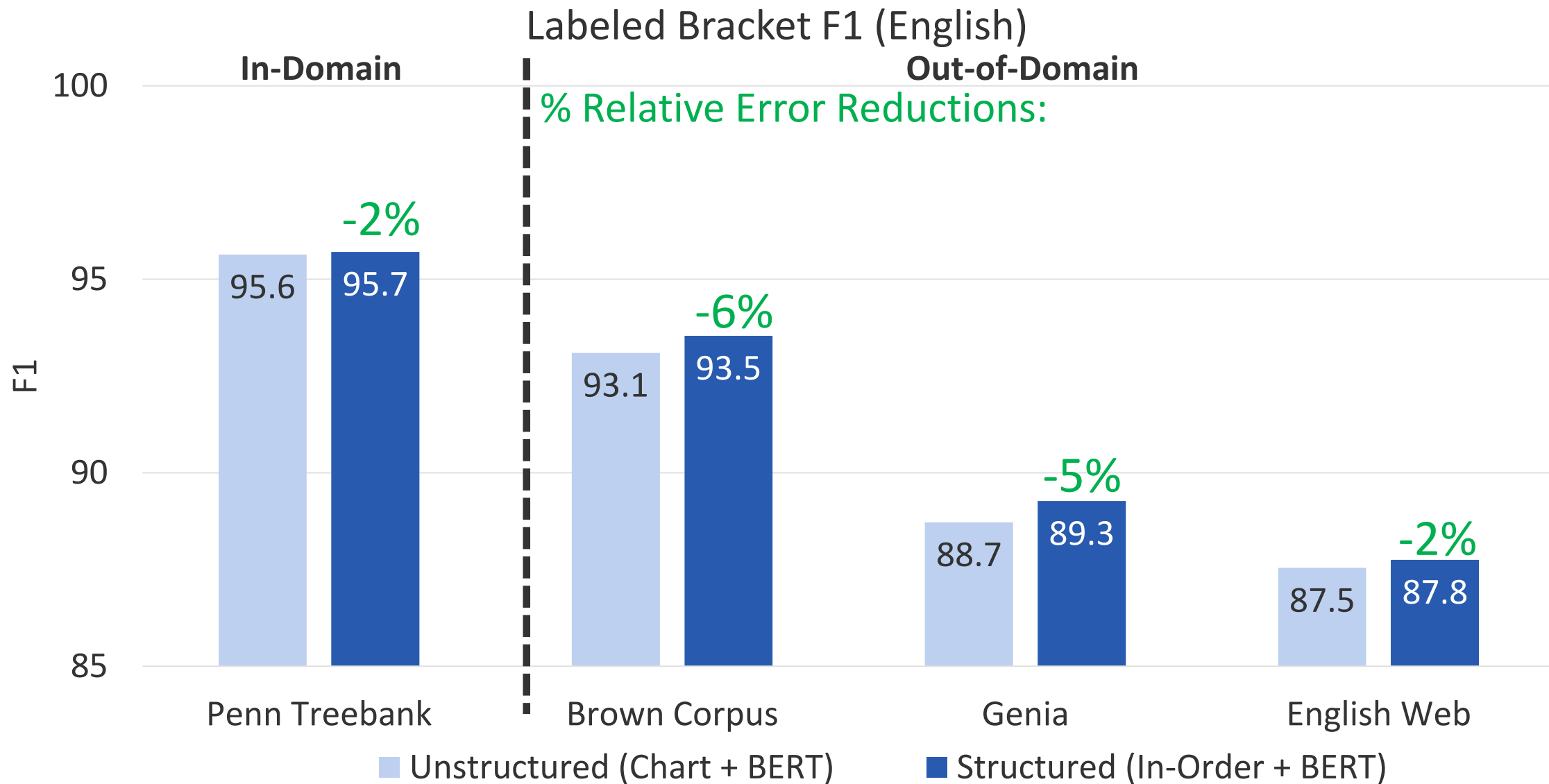
Structure Helps with Larger Spans

F1 by minimum span length, on English Web

Structure Improves Exact Match

Exact Parse Match Accuracies (English)

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.

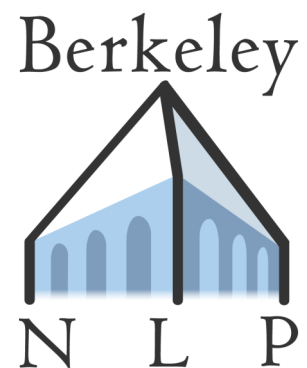Structure helps in domain transfer, longer spans, and whole parses.

Conclusions

Neural and non-neural parsers transfer similarly.

Pre-training helps across domains.

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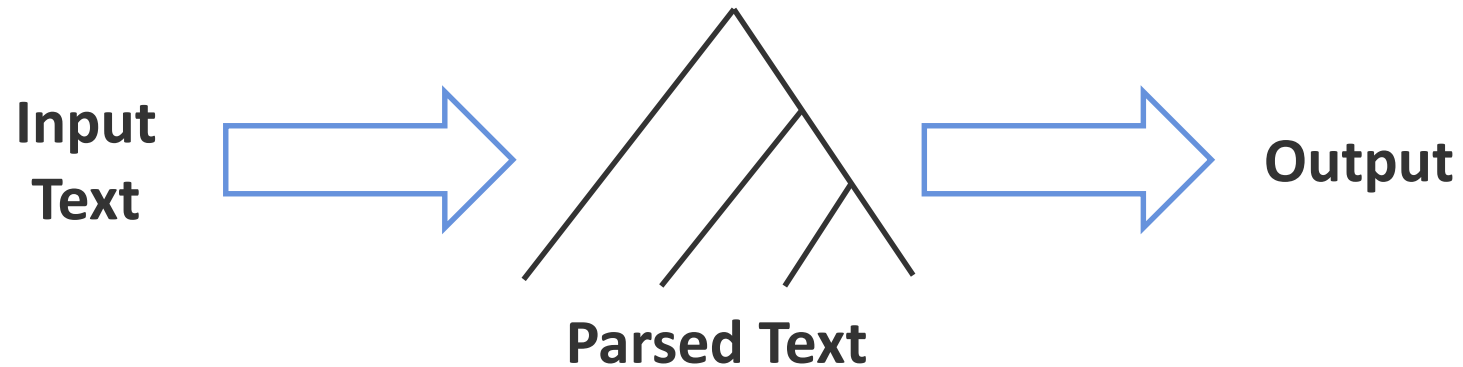
Thank you!

Code and models:

Chart + BERT: `parser.kitaev.io`
In-Order RNNG + BERT: `github.com/dpfried/rnng-bert`

High Accuracy Parsers Benefit NLP Systems

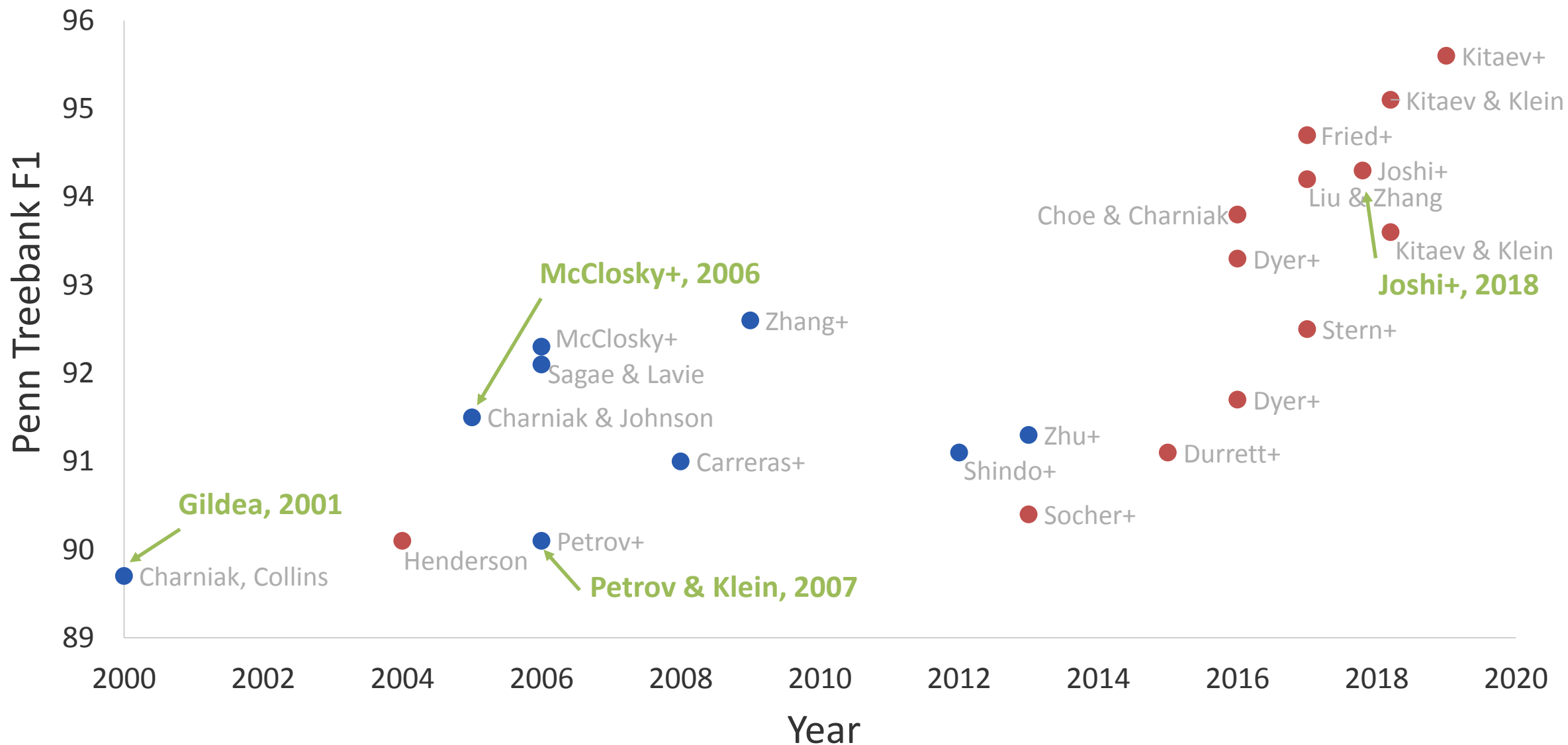**Input Text** → → **Parsed Text** → → **Output**

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]
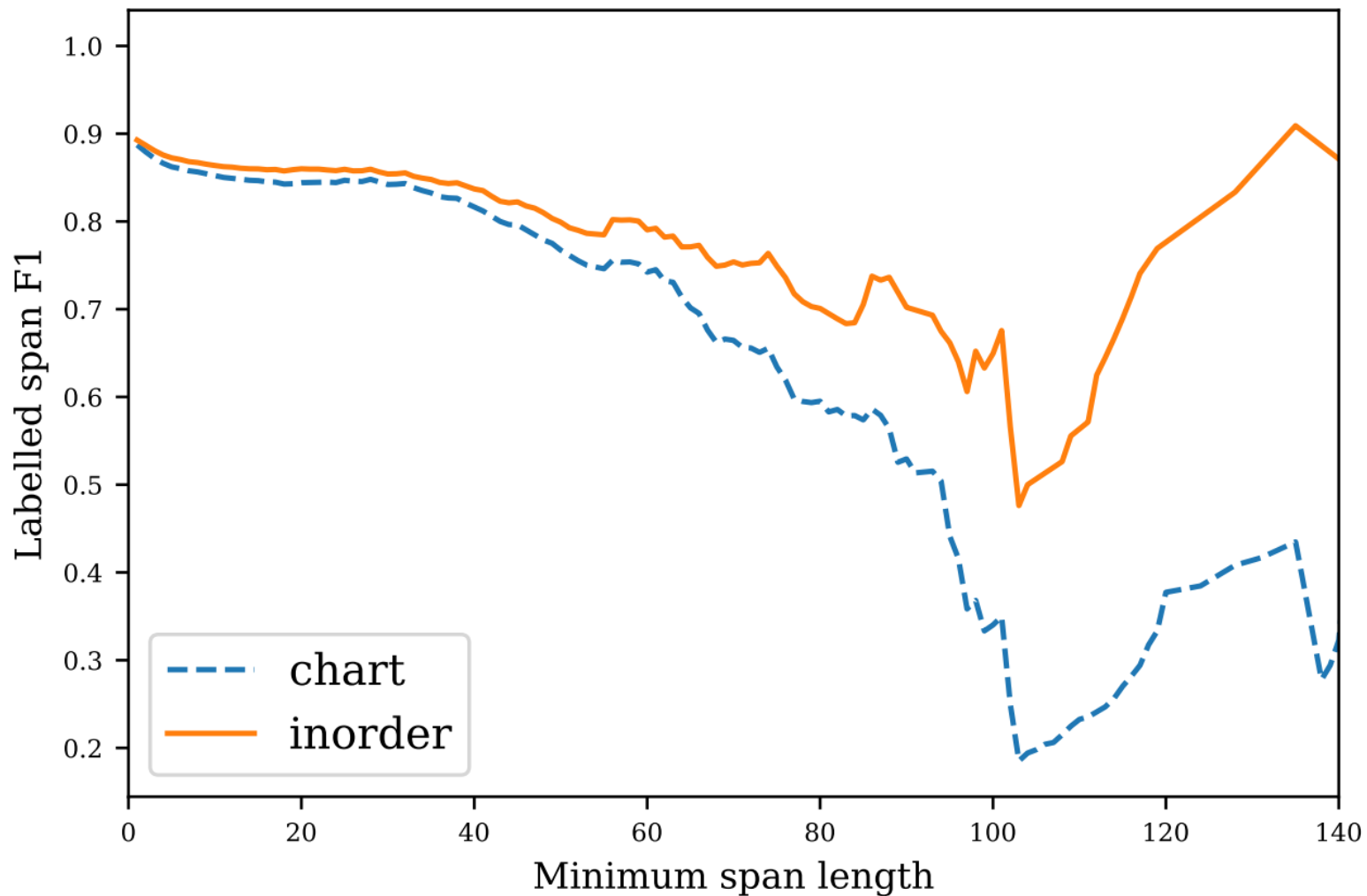
Past Evaluations of Generalization

Structure Helps with Larger Spans

F1 by minimum span length, on Genia corpus

Effects of Pre-Trained Representations

Chinese Corpora

In-Domain | **Out-of-Domain**

% Relative Error Reductions:

In-Order RNNG | +Word Embeddings | +BERT