Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models

Junjie Chen* , Xiangheng He**, Yusuke Miyao*

Department of Computer Science, University of Tokyo*
Group on Language, Audio, & Music, Imperial College London**

May 24, 2022

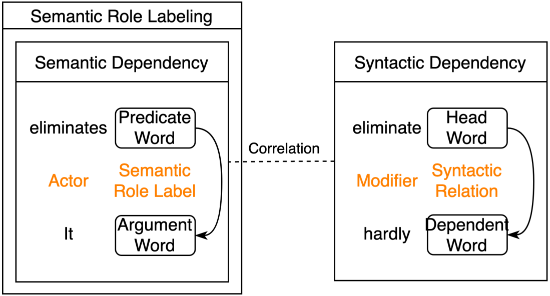
Introduction: Syntax-Aware Semantic Role Labeling

Semantic Role Labeling (SRL)

Extracting semantic dependencies (by modeling a distribution of semantic labels)

Syntactic-Semantic Dependency Correlation

Semantic dependencies are associated with some syntactic dependencies



It will hardly eliminate the pigeon

Introduction: Overview

Research Question

How to **interpret** and **utilize** the Syntactic-Semantic Dependency Correlation

Hypothesis for the Correlation

Semantic label distributions differ by hop patterns

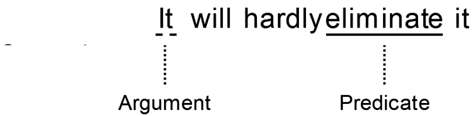
Introduction: Overview

Research Question

How to **interpret** and **utilize** the Syntactic-Semantic Dependency Correlation

Hypothesis for the Correlation

Semantic label distributions differ by hop patterns



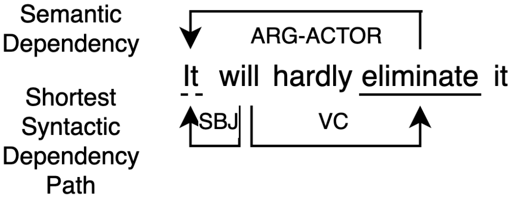
Introduction: Overview

Research Question

How to **interpret** and **utilize** the Syntactic-Semantic Dependency Correlation

Hypothesis for the Correlation

Semantic label distributions differ by hop patterns



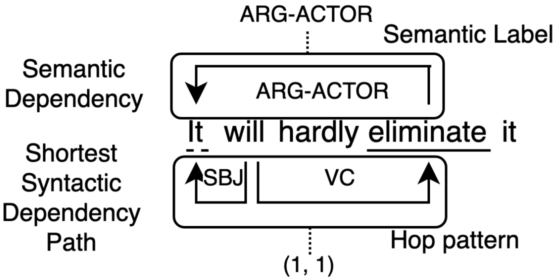
Introduction: Overview

Research Question

How to **interpret** and **utilize** the Syntactic-Semantic Dependency Correlation

Hypothesis for the Correlation

Semantic label distributions differ by hop patterns



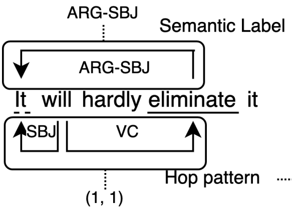
Introduction: Overview

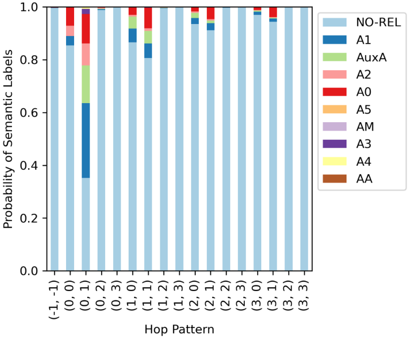
Research Question

How to **interpret** and **utilize** the Syntactic-Semantic Dependency Correlation

Hypothesis for the Correlation

Semantic label distributions differ by hop patterns





Introduction: Shortest Syntactic Dependency Path and Hop Pattern

Shortest Syntactic Dependency Path (SSDP)

- The shortest path connecting predicates and arguments
- A salient feature for utilizing syntactic information in SRL [7]

Hop Pattern

- Tracks the transition from predicates to arguments
- Denoted as (α, β)
  - α: the number of dependent-to-head transitions
  - β: the number of head-to-dependent transitions

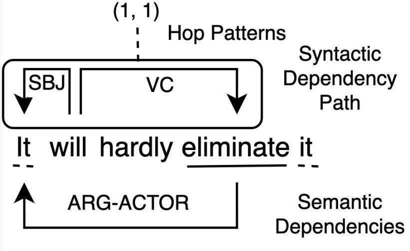


Figure: Example illustrating SSDP and hop patterns

Introduction: Shortest Syntactic Dependency Path and Hop Pattern

Shortest Syntactic Dependency Path (SSDP)

- The shortest path connecting predicates and arguments
- A salient feature for utilizing syntactic information in SRL [7]

Hop Pattern

- Tracks the transition from predicates to arguments
- Denoted as (α, β)
  - α: the number of dependent-to-head transitions
  - β: the number of head-to-dependent transitions

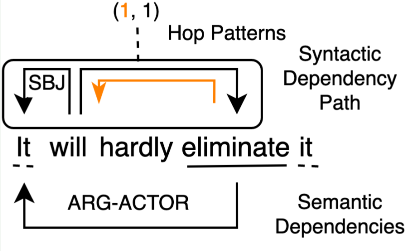


Figure: Example illustrating SSDP and hop patterns

Introduction: Shortest Syntactic Dependency Path and Hop Pattern

Shortest Syntactic Dependency Path (SSDP)

- The shortest path connecting predicates and arguments
- A salient feature for utilizing syntactic information in SRL [7]

Hop Pattern

- Tracks the transition from predicates to arguments
- Denoted as (α, β)
  - α: the number of dependent-to-head transitions
  - β: the number of head-to-dependent transitions

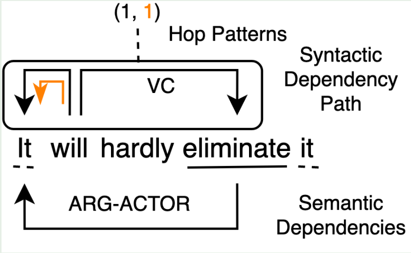


Figure: Example illustrating SSDP and hop patterns

Introduction: Shortest Syntactic Dependency Path and Hop Pattern

Shortest Syntactic Dependency Path (SSDP)

- The shortest path connecting predicates and arguments
- A salient feature for utilizing syntactic information in SRL [7]

Hop Pattern

- Tracks the transition from predicates to arguments
- Denoted as (α, β)
  - α: the number of dependent-to-head transitions
  - β: the number of head-to-dependent transitions

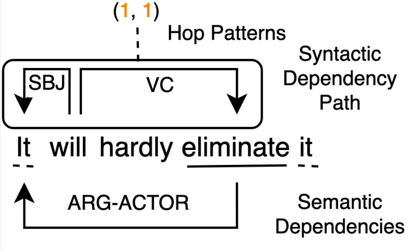


Figure: Example illustrating SSDP and hop patterns

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

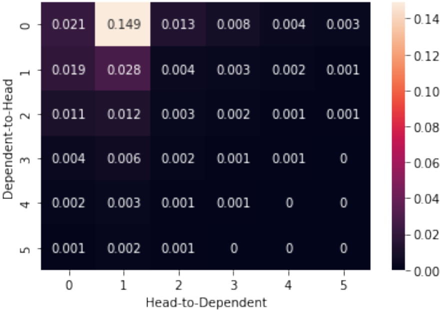


Figure: Mutual information gain of hop patterns up to (5,5)

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

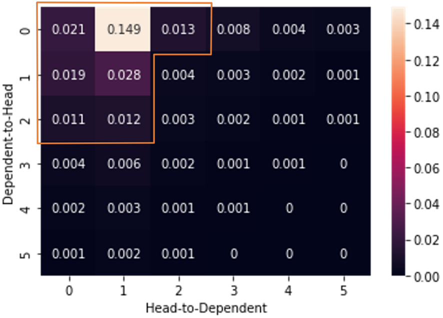


Figure: Mutual information gain of hop patterns up to (5,5)

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

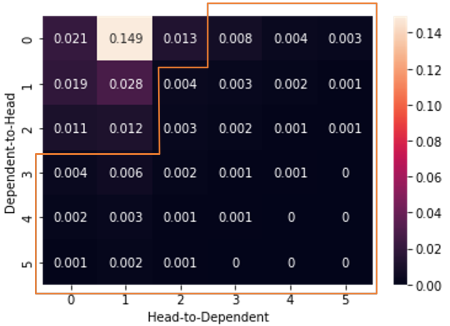


Figure: Mutual information gain of hop patterns up to (5,5)

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

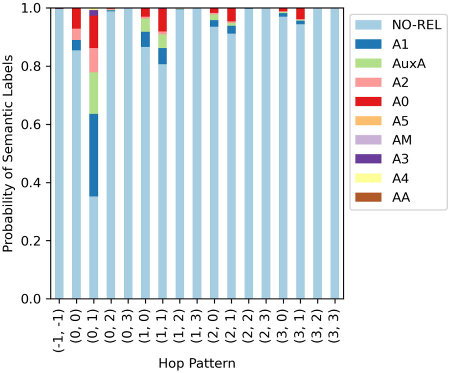


Figure: Mutual information gain of hop patterns up to (5,5)

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

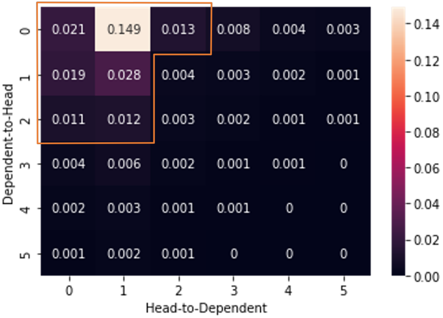


Figure: Mutual information gain of hop patterns up to (5,5)

Motivation

Variation in Semantic Label Distributions

Hop patterns have diverse mutual information gains with semantic label distributions

1. (0, 1) has the highest information gain
2. short SSDPs have diverse non-zero information gains
3. long SSDPs have near-zero information gains

- 1 and 2 confirms **the variation in semantic label distributions**
- 3 confirms that **long SSDPs have similar semantic label distributions**

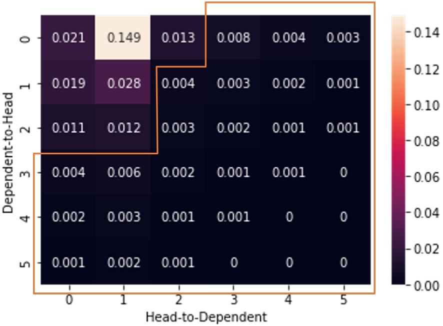


Figure: Mutual information gain of hop patterns up to (5,5)

Proposal: Separate Estimation for Different Semantic Label Distributions

We propose a mixture model-based semantic parser that

- Decouples semantic label distributions using component distributions
- Clusters hop patterns with similar semantic label distributions using mixture weights

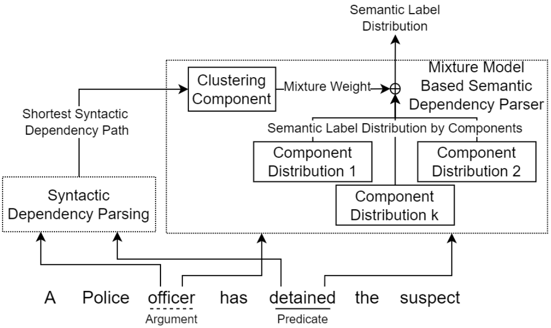


Figure: Overview of the mixture-model based semantic parser

Proposal: Separate Estimation for Different Semantic Label Distributions

We propose a mixture model-based semantic parser that

- Decouples semantic label distributions using component distributions
- Clusters hop patterns with similar semantic label distributions using mixture weights

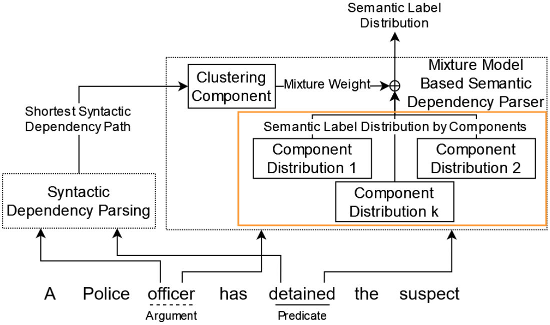


Figure: Overview of the mixture-model based semantic parser

Proposal: Separate Estimation for Different Semantic Label Distributions

We propose a mixture model-based semantic parser that

- Decouples semantic label distributions using component distributions
- Clusters hop patterns with similar semantic label distributions using mixture weights

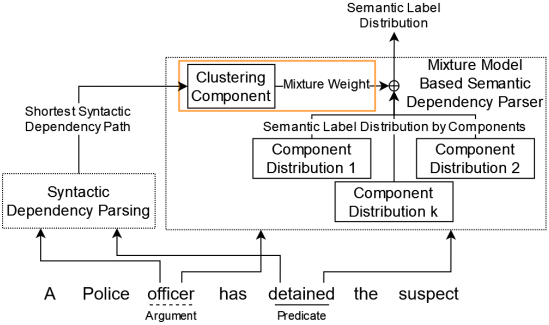


Figure: Overview of the mixture-model based semantic parser

Related Works

Correlation as Co-occurrence Biases [4]

Semantic dependencies prefer short syntactic dependency paths over the long ones

Problems

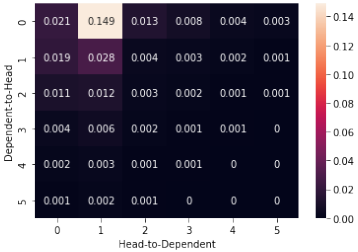
- Modeling the correlation **regardless of** semantic labels
- Utilizing the correlation by **language-dependent** heuristics

Our Contributions

- Modeling the correlation **with respect to** semantic labels
- The proposed approach is **language-independent**

Results: Syntactic-Semantic Dependency Correlation

Semantic label distributions of diffrent hop patterns have unique properties



α \ β	0	1	2	3
0	3	1	0	0
1	2	2	0	0
2	2	2	0	0
3	2	0	0	0
4	0	0	0	0
5	0	0	0	0

(a) FastText

α \ β	0	1	2	3
0	2	0	4	4
1	2	1	4	4
2	3	3	4	4
3	4	4	4	4
4	4	4	4	4
5	4	4	4	4

(b) ELMo

α \ β	0	1	2	3
0	4	0	2	2
1	4	3	2	2
2	4	3	2	2
3	4	4	2	2
4	2	2	2	2
5	2	2	2	2

(c) BERT

Table: Component assignments extracted from models

Results: Improvement over Baselines

Our Mixture-Model (MM) method outperforms baseline methods over input embeddings and languages in the CoNLL-2009 dataset [3]

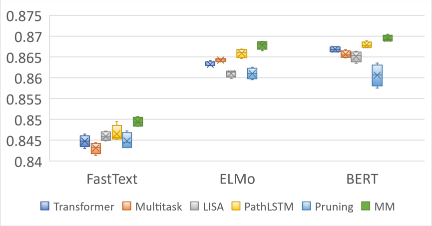


Figure: English LAS comparison on input embeddings

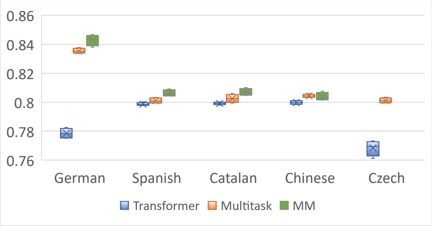


Figure: LAS comparison on five languages using FastText

Results: Improvement over Baselines

MM outperforms baseline methods over input embeddings and languages in the CoNLL-2009 dataset [3]

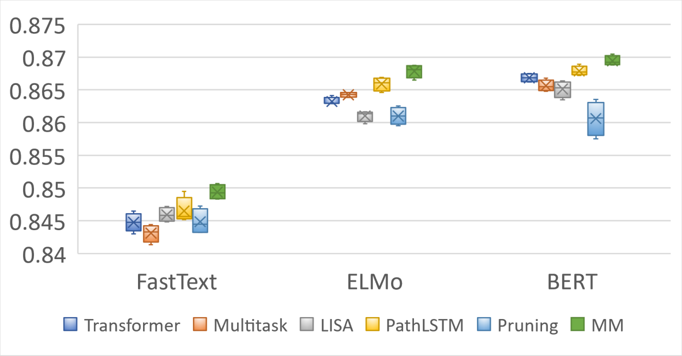


Figure: English LAS comparison on input embeddings

Results: Improvement over Baselines

MM outperforms baseline methods over input embeddings and languages in the CoNLL-2009 dataset [3]

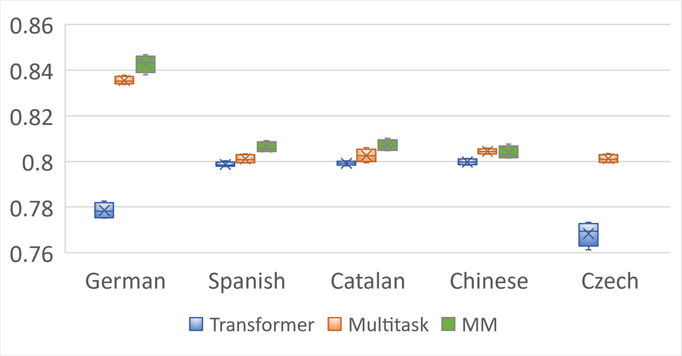


Figure: LAS comparison on five languages using FastText

Results: Improvement on Short-distance Semantic Dependencies

MM improves on short-distance semantic dependencies

- MM is the only method improving over Transformer on short-distance semantic dependencies (FastText and ELMo)
- MM retains the performance advantage of syntax-aware methods on long-distance semantic dependencies

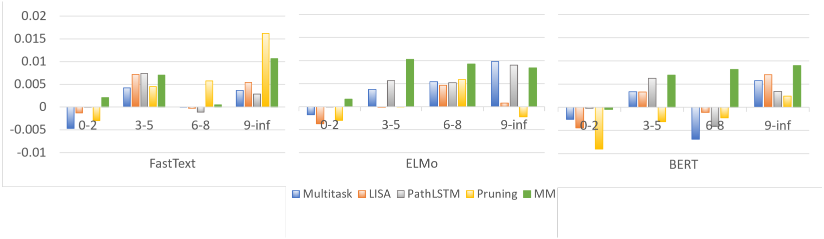


Figure: Relative LAS of MM and baselines to Transformer method

Results: Improvement on Short-distance Semantic Dependencies

MM improves on short-distance semantic dependencies

- MM is the only method improving over Transformer on short-distance semantic dependencies (FastText and ELMo)
- MM retains the performance advantage of syntax-aware methods on long-distance semantic dependencies

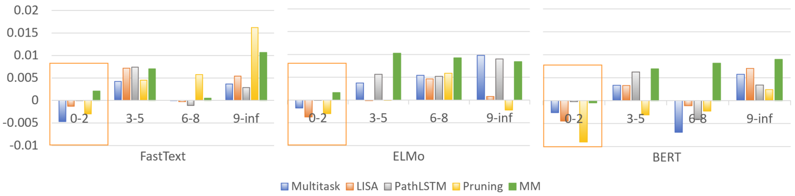


Figure: Relative LAS of MM and baselines to Transformer method

Results: Improvement on Short-distance Semantic Dependencies

MM improves on short-distance semantic dependencies

- MM is the only method improving over Transformer on short-distance semantic dependencies (FastText and ELMo)
- MM retains the performance advantage of syntax-aware methods on long-distance semantic dependencies

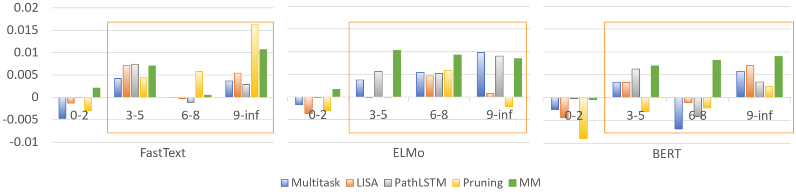


Figure: Relative LAS of MM and baselines to Transformer method

Results: Comparison with STOA

MM performs competitively with state-of-the-art methods

- MM is the best performant on the GloVe and the ELMo embeddings on the WSJ section
- MM performs best on the ELMo and the BERT embeddings on the Brown section

	WSJ			Brown		
GloVe	P	R	F1	P	R	F1
[9]	88.73	89.83	89.28	82.46	83.2	**82.82**
[5]	87.8	88	87.9	77	76.8	76.9
[4]	89.7	89.3	89.5	81.9	76.9	79.3
[7]	88.1	85.3	86.7	76.9	73.8	75.3
MM	91.03	90.13	**90.58**	80.59	79.21	79.83
MM (FastText)	91.16	90.19	90.71	83.93	82.64	83.28
ELMo	P	R	F1	P	R	F1
[5]	90.5	92.1	91.3	81.7	81.9	81.8
[1]	91.7	90.8	91.2	83.2	81.9	82.5
[6]	-	-	91	-	-	82.2
[2]	-	-	91.1	-	-	82.7
MM	92.21	91.45	**91.82**	86.51	85.30	**85.90**
BERT	P	R	F1	P	R	F1
[8](base)	92.1	91.9	92	85.6	84.7	85.1
[8](large)	92.4	92.3	**92.4**	85.7	85.8	85.7
[9]	91.21	91.19	91.2	85.65	86.09	85.87
MM	92.33	91.77	92.05	87.00	85.98	**86.32**

Conclusion

In this paper, we

- Interpreted the syntactic-semantic dependency correlation by **the variation in semantic label distributions**
- Learned the correlation with a mixture model-based semantic parser
- Produced a SRL method that
  - **excels in predicting short-distance semantic dependencies**
  - performs competitively with state-of-the-art methods

References I

Rui Cai and Mirella Lapata.
Semi-supervised semantic role labeling with cross-view training.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1018–1027, Hong Kong, China, November 2019. Association for Computational Linguistics.

Xinchi Chen, Chunchuan Lyu, and Ivan Titov.
Capturing argument interaction in semantic role labeling with capsule networks.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages

References II

5415–5425, Hong Kong, China, November 2019. Association for Computational Linguistics.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang.
The CoNLL-2009 shared task: Syntactic and semantic dependencies in multiple languages.
In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18, Boulder, Colorado, June 2009. Association for Computational Linguistics.

References III

Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai.
Syntax for semantic role labeling, to be, or not to be.
In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2061–2071, Melbourne, Australia, July 2018. Association for Computational Linguistics.

Zuchao Li, Shexia He, Junru Zhou, Hai Zhao, Kevin Parnow, and Rui Wang.
Dependency and span, cross-style semantic role labeling on propbank and nombank.
CoRR, abs/1911.02851, 2019.

References IV

Chunchuan Lyu, Shay B. Cohen, and Ivan Titov.
Semantic role labeling with iterative structure refinement.
In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1071–1082, Hong Kong, China, November 2019. Association for Computational Linguistics.

Michael Roth and Mirella Lapata.
Neural semantic role labeling with dependency path embeddings.
In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1192–1202, Berlin, Germany, August 2016. Association for Computational Linguistics.

References V

Peng Shi and Jimmy Lin.
Simple BERT models for relation extraction and semantic role labeling.
CoRR, abs/1904.05255, 2019.

Junru Zhou, Zuchao Li, and Hai Zhao.
Parsing all: Syntax and semantics, dependencies and spans.
In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4438–4449, Online, November 2020. Association for Computational Linguistics.