

# Weiwei Sun

Natural Language and Information Processing Research Group  
Department of Computer Science and Technology  
University of Cambridge

## Education

Doctor of Engineering Sciences, Faculty for Mathematics and Computer Science, Saarland University, September 2009–April 2012, supervised by Prof. Hans Uszkoreit.

M.S. Computer Science, School of Electronics Engineering and Computer Science, Peking University, September 2006–July 2009, supervised by Prof. Zhifang Sui

B.A. Computational and Applied Linguistics, Department of Chinese Language and Literature, Peking University, September 2002–July 2006

B.S. Computer Software, School of Electronics Engineering and Computer Science, Peking University, September 2003–July 2006

## Working Experiences

### *Academic Positions*

University Senior Lecturer ( $\approx$ Associate Professor), Department of Computer Science and Technology, University of Cambridge, since October 2020

Associate Professor, Wangxuan Institute of Computer Technology (formerly Institute of Computer Science & Technology), Peking University, August 2015–September 2020

Assistant Professor, Wangxuan Institute of Computer Technology (formerly Institute of Computer Science & Technology), Peking University, February 2012–July 2015

Adjunct Researcher, Center for Chinese Linguistics, Peking University, August 2017–August 2020

Visiting Associate Professor, Department of Linguistics and Translation, City University of Hong Kong, November 2016–December 2016

### *Other Employment*

Scientific Consultant, Beijing AITC Research Institution Co., Ltd., October 2017–September 2019

Research Assistant, Language Technology Lab, German Research Center for Artificial Intelligence, October 2009–February 2012

Visiting student, Toshiba (China) R&D Center, November 2007–March 2008

## Publications<sup>1</sup>

### Refereed Journal Articles

- ♠ Junjie Cao, Zi Lin, **Weiwei Sun\*** and Xiaojun Wan. A Comparative Analysis of Knowledge-Intensive and Data-Intensive Semantic Parsers. *Computational Linguistics*. (conditional accept)
- ♠♥ **Weiwei Sun**, Yufei Chen, Xiaojun Wan and Meichun Liu. 2019. Parsing Chinese Sentences with Grammatical Relations. *Computational Linguistics*.
- ♠ Xun Zhang, Yantao Du, **Weiwei Sun\***, and Xiaojun Wan. 2016. Transition-based Parsing for Deep Dependency Structures. *Computational Linguistics*.
- ♥ **Weiwei Sun** and Xiaojun Wan. 2016. Towards Accurate and Efficient Chinese Part-of-speech Tagging. *Computational Linguistics*.
- ♥ **Weiwei Sun** and Xiaojun Wan. 2013. Data-driven, PCFG-based and Pseudo-PCFG-based Models for Chinese Dependency Parsing. *Transactions of the Association for Computational Linguistics (TACL)*.

### Refereed Conference Proceedings

- ♣ Yajie Ye and **Weiwei Sun\***. 2020. Exact yet Efficient Graph Parsing, Bi-directional Locality and the Constructivist Hypothesis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Yufei Chen and **Weiwei Sun\***. 2020. Parsing into Variable-in-situ Logico-Semantic Graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ◇♠ Yuanyuan Zhao, **Weiwei Sun\***, Junjie Cao and Xiaojun Wan. 2020. Semantic Parsing for English as a Second Language. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.  
  
Abdul Rafae Khan, Jia Xu and **Weiwei Sun**. 2020. Coding Textual Inputs Boosts the Accuracy of Neural Networks. In *Proceedings of the conference on Empirical Methods on Natural Language Processing (EMNLP)*.
- ♠ Yufei Chen, Yajie Ye and **Weiwei Sun\***. 2019. Peking at MRP 2019: Factorization- and Composition-Based Parsing for Elementary Dependency Structures. In *Proceedings of the Shared Task on Cross-Framework Meaning Representation Parsing at the 2019 CoNLL*.
- ♠ Weimin Lyu, Sheng Huang, Abdul Rafae Khan, Shengqiang Zhang, **Weiwei Sun**, Jia Xu. 2019. CUNY-PKU Parser at SemEval-2019 Task 1: Cross-lingual Semantic Parsing with UCCA. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2019)*.
- ♥ Bowei Zhang, **Weiwei Sun\***, Xiaojun Wan and Zongming Guo. 2019. PKU Paraphrase Bank: A Sentence-Level Paraphrase Corpus for Chinese. In *Proceedings of the Eighth CCF International Conference on Natural Language Processing and Chinese Computing (NLPCC)*.
- ♣ Yajie Ye, **Weiwei Sun\*** and Xiaojun Wan. 2018. Language Generation via DAG Transduction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Yufei Chen, **Weiwei Sun\*** and Xiaojun Wan. 2018. Accurate SHRG-based Semantic Parsing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*.

<sup>1</sup> \*: corresponding author. \_: students directly advised. ♠: parsing to deep linguistic structures. ♣: surface realization. ♥: Chinese language processing. ◇: interlanguages.

- ♠♥ Yufei Chen, Yuanyuan Zhao, **Weiwei Sun\*** and Xiaojun Wan. 2018. Pre- and In-parsing Models for Neural Empty Category Detection. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ◇ Zi Lin, Yuguang Duan, Yuanyuan Zhao, **Weiwei Sun\*** and Xiaojun Wan. 2018. Semantic Role Labeling for Learner Chinese: The Importance of Syntactic Parsing and L2-L1 Parallel Data. In *Proceedings of the conference on Empirical Methods on Natural Language Processing (EMNLP)*.
- ♠ Yufei Chen, Sheng Huang, Fang Wang, Junjie Cao, **Weiwei Sun\*** and Xiaojun Wan. 2018. Neural Maximum Subgraph Parsing for Cross-Domain Semantic Dependency Analysis. In *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*.
- ◇ Yuanyuan Zhao, Nan Jiang, **Weiwei Sun\*** and Xiaojun Wan. 2018. Overview of the NLPCC 2018 Shared Task: Grammatical Error Correction. In *Proceedings of the Seventh CCF International Conference on Natural Language Processing and Chinese Computing*.
- ◇ Yuguang Duan, Zi Lin, **Weiwei Sun\***. 2018. Is Argument Structure of Learner Chinese Understandable: A Corpus-Based Analysis. In *Proceedings of the 2018 International Conference on Bilingual Learning and Teaching*.
- ♠ **Weiwei Sun**, Junjie Cao and Xiaojun Wan. 2017. Semantic Dependency Parsing via Book Embedding. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Junjie Cao, Sheng Huang, **Weiwei Sun\*** and Xiaojun Wan. 2017. Parsing to 1-Endpoint-Crossing, Pagenumber-2 Graphs. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Junjie Cao, Sheng Huang, **Weiwei Sun\*** and Xiaojun Wan. 2017. Quasi-Second-Order Parsing for 1-Endpoint-Crossing, Pagenumber-2 Graphs. In *Proceedings of the conference on Empirical Methods on Natural Language Processing (EMNLP)*.
- ♠♥ **Weiwei Sun**, Yantao Du and Xiaojun Wan. 2017. Parsing for Grammatical Relations via Graph Merging. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL)*.
- ♠ Xun Zhang, **Weiwei Sun\*** and Xiaojun Wan. 2017. The Covert Helps Parse the Overt. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL)*.
- ♠ Yufei Chen, Junjie Cao, **Weiwei Sun\*** and Xiaojun Wan. 2017. Peking at EPE 2017: A Comparison of Tree Approximation, Transition-based and Maximum Subgraph Models for Semantic Dependency Analysis. In *Proceedings of the 2017 Shared Task on Extrinsic Parser Evaluation*.
- ♠ Yantao Du, **Weiwei Sun\*** and Xiaojun Wan. 2015. A Data-driven, Factorization Parser for CCG Dependency Structures. In *Proceedings of the conference of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Yantao Du, Fan Zhang, Xun Zhang, **Weiwei Sun\*** and Xiaojun Wan. 2015. Peking: Building Semantic Dependency Graphs with a Hybrid Parser. 2015 In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*.
- ♠♥ **Weiwei Sun**, Yantao Du, Xin Kou, Shuoyang Ding and Xiaojun Wan. 2014. Grammatical Relations in Chinese: GB-Grounded Extraction and Data-Driven Parsing. In *Proceedings of the conference of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♠ Yantao Du, Fan Zhang, **Weiwei Sun** and Xiaojun Wan. 2014. Peking: Profiling Syntactic Tree Parsing Techniques for Semantic Graph Parsing. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*.

- ♡ **Weiwei Sun**, Xiaochang Peng and Xiaojun Wan. 2013. Capturing Long-distance Dependencies in Sequence Models: A Case Study of Chinese Part-of-speech Tagging. In *International Joint Conference on Natural Language Processing (IJCNLP)*.
- Minwei Feng, **Weiwei Sun**, and Hermann Ney. 2012. Semantic Cohesion Model for Phrase-based SMT. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING)*.
- ♡ **Weiwei Sun** and Hans Uszkoreit. 2012. Capturing Paradigmatic and Syntagmatic Lexical Relations: Towards Accurate Chinese Part-of-Speech Tagging. In *Proceedings of the conference of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♡ **Weiwei Sun** and Xiaojun Wan. 2012. Reducing Approximation and Estimation Errors for Chinese Lexical Processing with Heterogeneous Annotations. In *Proceedings of the conference of the 50th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- ♡ **Weiwei Sun** and Jia Xu. 2011. Enhancing Chinese Word Segmentation with Using Unlabeled Data. In *Proceedings of the conference on Empirical Methods on Natural Language Processing (EMNLP)*.
- ♡ **Weiwei Sun**. 2011. A Stacked Sub-word Model for Joint Chinese Word Segmentation and Part-of-Speech Tagging. In *Proceedings of the conference of the 49th Annual Meeting of the Association for Computational Linguistics (ACL)*.
- Jia Xu and **Weiwei Sun**. 2011. Generating virtual Parallel Corpus: A Compatibility Centric Method. In *Proceedings of the Machine Translation Summit XIII*.
- ♡ **Weiwei Sun**. 2010. Word-based and Character-based Word Segmentation Models: Comparison and Combination. In *Proceedings of the 23rd International Conference of Computational Linguistics (COLING)*.
- ♡ **Weiwei Sun**. 2010. Improving Chinese Semantic Role Labeling with Rich Syntactic Features. In *Proceedings of the conference of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), short papers*.
- ♡ **Weiwei Sun**. 2010. Semantics-Driven Shallow Parsing for Chinese Semantic Role Labeling. In *Proceedings of the conference of the 48th Annual Meeting of the Association for Computational Linguistics (ACL), short papers*.
- ♡ **Weiwei Sun**, Rui Wang, and Yi Zhang. 2010. Discriminative Parse Reranking for Chinese with Homogeneous and Heterogeneous Annotations. In *Proceedings of the first CIPS-SIGHAN Joint Conference on Chinese Language Processing*.
- ♡ **Weiwei Sun**, Zhifang Sui, Meng Wang, and Xin Wang. 2009. Chinese Semantic Role Labeling with Shallow Parsing. In *Proceedings of the conference on Empirical Methods on Natural Language Processing (EMNLP)*.
- Weiwei Sun**, Zhifang Sui and Meng Wang. 2009. Prediction of Thematic Rank for Structured Semantic Role Labeling. In *Proceedings of the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics (ACL) and the 4th International Joint Conference on Natural Language Processing (IJCNLP), short papers*.
- ♡ **Weiwei Sun** and Zhifang Sui. 2009. Chinese Function Tag Labeling. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information and Computation (PACLIC)*.
- Weiwei Sun**, Zhifang Sui, and Haifeng Wang. 2008. Prediction of Maximal Projection for Semantic Role Labeling. In *Proceedings of the 22nd International Conference of Computational Linguistics (COLING)*.

### *Papers in Chinese*

- ♡ Lulu Wang, **Weiwei Sun** and Yulin Yuan. 2014. On automatic Interpretation and Pattern Alternation of Chinese Ba-Construction. *Computer Engineering and Applications*.
- ♡ Meng Wang, Shiwen Yu, Huiming Duan, and **Weiwei Sun**. 2008. Quantitative Research on Grammatical Characteristics of Noun in Contemporary Chinese. *Journal of Chinese Information Processing*.
- ♡ Tieliang Ji, **Weiwei Sun**, and Zhifang Sui. 2007. The Acquisition of Chinese Verbs' Subcategorization Lexion Based on Linguistic Theory and Statistical Algorithm. *Journal of Chinese Information Processing*.

### *Refereed Book Chapters*

- ♡ **Weiwei Sun**. Heterogeneous Annotations Help Reduce Approximation and Estimation Errors. *Chinese Language Resources: Data Collection, Linguistic Analysis, Annotation, and Language Processing*. Springer Volume on Language Resources and Processing.

### *Presentations*

- ♠ **Weiwei Sun**. 2019. Empty Categories Help Parse the Overt. In *Proceedings of the Society for Computation in Linguistics*.

## Shared Tasks

CoNLL 2019: Cross-Framework Meaning Representation Parsing. Best performing system for Elementary Dependency Structures. **team leader**.

SemEval 2015, Task 18: Broad-coverage Semantic Dependency Parsing. Rank 1st for English of in-domain evaluation (close track), rank 1st for Chinese, rank 1st for Czech of out-of-domain evaluation (close track). **team leader**.

SemEval 2014, Task 8: Broad-coverage Semantic Dependency Parsing. Rank 1st. **team leader**.

## Teaching

### *University of Cambridge*

L90: Overview of Natural Language Processing

### *Peking University*

Empirical Methods in Natural Language Processing

Programming and Linguistic Analysis

Language, Logic and Computation

Introduction to Formal Syntax

Syntactic Theory and Grammar Engineering

Compiler

## *International Summer School of University of International Business and Economics*

Introduction to Java Programming

### *Tutorial*

Graph-Based Meaning Representations: Design and Processing. ACL 2019. with Prof. Stephan Oepen and Prof. Alexander Koller. 245 people had registered for ACL.

## Professional Service and Membership

Area Chair:

ACL 2020 (syntax and parsing)

ACL 2019 (tagging, chunking, syntax and parsing)

EMNLP-IJCNLP 2019 (tagging, chunking, syntax and parsing)

CCL 2019 (linguistics and cognitive science)

ACL 2018 (tagging, chunking, syntax and parsing)

Evaluation Co-chair: NLPCC 2019

Regular PC Members of ACL, EMNLP, NAACL, COLING, IJCNLP, IJCAI, AAAI.

Organizer of NLPCC 2018 Shared Task: Grammatical Error Correction.

NLPCC = The CCF International Conference on Natural Language Processing and Chinese Computing

CCL = The China National Conference on Computational Linguistics

## Language Skills

Native in Mandarin

Fluent in English, Cantonese

## Activities

Captain of Women's Soccer Team of Peking University, May 2005–September 2007.

Vice Captain of Women's Soccer Team of Peking University, September 2004–May 2005.

Last updated: October 7, 2020