

# Tao Yu

---

CONTACT INFORMATION	<a href="https://taoyds.github.io/">https://taoyds.github.io/</a> 272 Elm Street, New Haven, CT 06511	<a href="mailto:tao.yu@yale.edu">tao.yu@yale.edu</a> 801-493-9166
RESEARCH INTERESTS	Natural Language Processing and Deep Learning. More specifically, I am currently interested in building conversational natural language interfaces (NLIs) that can help humans explore and reason over data in any application (e.g., relational databases and mobile apps) in a robust and trusted manner.	
EDUCATION	<b>Yale University</b> , New Haven, CT  Ph.D. Computer Science - Natural Language Processing 08/2017 - 05/2021 expected <ul style="list-style-type: none"><li>• Advisors: <a href="#">Dragomir R. Radev</a></li><li>• Committee: <a href="#">Kathleen McKeown</a>, <a href="#">Luke Zettlemoyer</a>, <a href="#">Robert Frank</a>, and <a href="#">Marynel Vázquez</a></li></ul> <b>Columbia University</b> , SEAS, New York, NY  M.S. Data Science GPA: 3.82/4.0 09/2015 - 05/2017 <ul style="list-style-type: none"><li>• Advisors: <a href="#">Owen Rambow</a> and <a href="#">Kathleen McKeown</a></li></ul> <b>University of Utah</b> , Salt Lake City, UT  B.S. Mathematics and Economics GPA: 3.86/4.0 01/2012 - 05/2015 <ul style="list-style-type: none"><li>• Advisors: <a href="#">Jingyi Zhu</a> and <a href="#">Peter Philips</a></li></ul>	
PUBLICATIONS & PAPERS IN PREPARATION	<ol style="list-style-type: none"><li>1. SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing <a href="#">Tao Yu</a>, Rui Zhang, Oleksandr Polozov, Chris Meek, Ahmed Hassan Awadallah In <i>Proceedings of International Conference on Learning Representations (ICLR)</i>, 2021</li><li>2. GraPPa: Grammar-Augmented Pre-Training for Table Semantic Parsing <a href="#">Tao Yu</a>, Chien-Sheng Wu, Xi Victoria Lin, Bailin Wang, Yi Chern Tan, Xinyi Yang, Dragomir Radev, Richard Socher, Caiming Xiong. In <i>Proceedings of International Conference on Learning Representations (ICLR)</i>, 2021</li><li>3. QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization Ming Zhong, Da Yin, <a href="#">Tao Yu</a>, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, Dragomir Radev In <i>Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL)</i>, 2021</li><li>4. DART: Open-Domain Structured Data Record to Text Generation Linyong Nan, Dragomir Radev, Rui Zhang, Amrit Rau, Abhinand Sivaprasad, Chiachun Hsieh, Nazneen Fatema Rajani, Xiangru Tang, Aadit Vyas, Neha Verma, Pranav Krishna, Yangxiaokang Liu, Nadia Irwanto, Jessica Pan, Faiaz Rahman, Ahmad Zaidi, Murori Mutuma, Yasin Tarabar, Ankit Gupta, <a href="#">Tao Yu</a>, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, Nazneen Fatema Rajani In <i>Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL)</i>, 2021</li><li>5. Semantic Evaluation for Text-to-SQL with Distilled Test Suites Ruiqi Zhong, <a href="#">Tao Yu</a>, Dan Klein In <i>Proceedings of Empirical Methods in Natural Language Processing (EMNLP)</i>, 2020</li><li>6. Did You Ask a Good Question? A Cross-Domain Question Intention Classification Benchmark for Text-to-SQL Yusen Zhang, Xiangyu Dong, Shuaichen Chang, <a href="#">Tao Yu</a>, Peng Shi, Rui Zhang In <i>Interactive and Executable Semantic Parsing Workshop at EMNLP</i>, 2020</li></ol>	

7. CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases  
Tao Yu, Rui Zhang, He Yang Er, Suyi Li, Eric Xue, Bo Pang, Xi Victoria Lin, Yi Chern Tan, Tianze Shi, Zihan Li, Youxuan Jiang, Michihiro Yasunaga, Sungrok Shim, Tao Chen, Alexander Fabbri, Zifan Li, Luyao Chen, Yuwen Zhang, Shreya Dixit, Vincent Zhang, Caiming Xiong, Richard Socher, Walter Lasecki, Dragomir Radev  
 In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2019
8. Editing-based SQL Query Generation for Cross-Domain Context-Dependent Questions  
 Rui Zhang, Tao Yu, He Yang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, Dragomir Radev  
 In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2019
9. SPaRC: Cross-Domain Semantic Parsing in Context  
Tao Yu, Rui Zhang, Michihiro Yasunaga, Yi Chern Tan, Xi Victoria Lin, Suyi Li, Heyang Er, Irene Li, Bo Pang, Tao Chen, Emily Ji, Shreya Dixit, David Proctor, Sungrok Shim, Jonathan Kraft, Vincent Zhang, Caiming Xiong, Richard Socher, Dragomir Radev  
 In *Proceedings of Association for Computational Linguistics (ACL)*, 2019
10. Twitter Sentiment in New York City Parks as Measure of Well-being  
 Richard A Plunz, Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen McKeown, Tao Yu, Laura Ugucioni, Maria Paola Sutto  
 In *Landscape and Urban Planning*, 2019
11. Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task  
Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, Dragomir Radev  
 In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2018
12. SyntaxSQLNet: Syntax Tree Networks for Complex and Cross Domain Text-to-SQL Task  
Tao Yu, Michihiro Yasunaga, Kai Yang, Rui Zhang, Dongxu Wang, Zifan Li, Dragomir Radev  
 In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP)*, 2018
13. TypeSQL: Knowledge-based Type-Aware Neural Text-to-SQL Generation  
Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, Dragomir Radev  
 In *Proceedings of North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018
14. Cross-Lingual Sentiment Transfer with Limited Resources  
 Mohammad Sadegh Rasooli, Noura Farra, Axinia Radeva, Tao Yu, and Kathleen McKeown  
 In *Machine Translation*, 2017
15. The Columbia-GWU System at the 2016 TAC KBP BeSt Evaluation  
 Owen Rambow, Tao Yu, Axinia Radeva, Sardar Hamidian, Alexander R. Fabbri, Debanjan Ghosh, Christopher Hidey, Tianrui Peng, Mona Diab, Kathleen McKeown, Smaranda Muresan  
 In *NIST TAC KBP Workshop*, 2016

RESEARCH &  
INDUSTRY  
EXPERIENCE

**Yale University**, New York, NY  
Graduate Student Research – NLP

08/2017 - Present

- Working on dialog-based robust and reliable semantic parsing/text-to-SQL systems.
- Open Text-to-SQL Challenge Series: [Spider](#), [SPaRC](#), [CoSQL](#).

**Microsoft Research**, Redmond, WA  
NLP Research Intern

06/2020 - 08/2020

- Context representation in conversational semantic parsing via pre-training

Mentors: [Ahmed Hassan Awadallah](#), [Oleksandr Polozov](#), and [Chris Meek](#)

**Salesforce Research**, Palo Alto, CA

NLP Research Intern

06/2019 - 08/2019

- Language model pre-training for table semantic parsing

Mentors: [Victoria Lin](#), [Caiming Xiong](#), and [Richard Socher](#)

**Samsung Research America**, Mountain View, CA

NLP Research Intern

06/2018 - 08/2018

**Columbia CCLS & NLP Group**, New York, NY

Graduate Research Assistant

05/2016 - 10/2016

- Developed systems for DARPA Text Analysis Conference (TAC) 2016 Belief and Sentiment Evaluation, and achieved the best performances on all 12 English and Spanish tasks.

Mentors: [Owen Rambow](#) and [Kathleen McKeown](#)

**AIG Science**, New York, NY

App Development Intern

01/2016 - 03/2016

PROFESSIONAL  
SERVICES

**Organizing Committee Member**

The Workshop on Interactive and Executable Semantic Parsing at EMNLP 2020

**Program Committee Member**

The Conference on Empirical Methods in Natural Language Processing (EMNLP), 2020

The 28th International Conference on Computational Linguistics (COLING), 2020

The Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (AAACL-IJCNLP), 2020

The Conference of the Association for Computational Linguistics (ACL), 2020

The Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019

The North American Chapter of the Association for Computational Linguistics (NAACL), 2019, 2021

AWARDS

Yale Nathan Hale Associates Scholarship Award

2020

Yale Conference Travel Fellowship

2018

Yale PhD Fellowship

2017

Connie and Robert White Honor Roll Scholarship

2014

Mathematics Department Scholarship

2014

International Student and Scholar Services Scholarship

2014

NSF Research Experience for Undergraduates (REU) Fellowship

2014

TEACHING  
EXPERIENCE

Natural Language Processing (CPSC 577), Yale University

2019

Database Systems (CPSC 537), Yale University

2019

Introduction to Natural Language Processing (CS 4705), Columbia University

2017

Natural Language Processing (Coursera)

2016

PROGRAMMING & OTHER SKILLS Proficient in Python, PyTorch, TensorFlow, R, Java, Matlab, SQL. Familiar with C, C++, JavaScript, Theano, Hadoop, Spark. FAA Student Pilot.

REFERENCES

**Dragomir Radev**

Professor, Yale University  
Email: [dragomir.radev@yale.edu](mailto:dragomir.radev@yale.edu)  
Cell: 646-420-1367

**Kathleen McKeown**

Professor, Columbia University  
Email: [kathy@cs.columbia.edu](mailto:kathy@cs.columbia.edu)  
Cell: 212-939-7118

**Luke Zettlemoyer**

Professor, University of Washington  
Email: [lsz@cs.washington.edu](mailto:lsz@cs.washington.edu)  
Cell: 206-685-1227

**Owen Rambow**

Professor, Stony Brook University  
Email: [owen.rambow@gmail.com](mailto:owen.rambow@gmail.com)  
Cell: 917-679-5265

**Caiming Xiong**

Managing Director, Salesforce Research  
Email: [cxiong@salesforce.com](mailto:cxiong@salesforce.com)  
Cell: 716-435-9316

**Ahmed Hassan Awadallah**

Principal Research Manager, Microsoft Research AI  
Email: [hassanam@microsoft.com](mailto:hassanam@microsoft.com)