

## 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. It involves studies in semantic parsing, dialogue systems, human computer interaction, and robotics.	
ACADEMIC POSITION	Assistant Professor, CS, The University of Hong Kong Co-directing <a href="#">HKU NLP Group</a> Postdoctoral Fellow, CSE, University of Washington Working with <a href="#">Noah Smith</a> and <a href="#">Mari Ostendorf</a>	08/2021 - <i>Present</i>  09/2021 - 06/2022
EDUCATION	<b>Yale University</b> , New Haven, CT Ph.D. Computer Science - Natural Language Processing • Advisors: <a href="#">Dragomir R. Radev</a> • Committee: <a href="#">Kathleen McKeown</a> , <a href="#">Luke Zettlemoyer</a> , <a href="#">Robert Frank</a> , <a href="#">Marynel Vázquez</a>  <b>Columbia University</b> , New York, NY M.S. Data Science GPA: 3.82/4.0 • Advisors: <a href="#">Owen Rambow</a> and <a href="#">Kathleen McKeown</a>  <b>University of Utah</b> , Salt Lake City, UT B.S. Mathematics and Economics GPA: 3.86/4.0 • Advisors: <a href="#">Jingyi Zhu</a> and <a href="#">Peter Philips</a>	08/2017 - 05/2021  09/2015 - 05/2017  01/2012 - 05/2015
PUBLICATIONS & PAPERS IN PREPARATION	<ol style="list-style-type: none"><li>1. Tianbao Xie, Chen Henry Wu, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida I. Wang, Victor Zhong, Bailin Wang, Chengzu Li, Connor Boyle, Ansong Ni, Ziyu Yao, Dragomir Radev, Caiming Xiong, Lingpeng Kong, Rui Zhang, Noah A. Smith, Luke Zettlemoyer, <b>Tao Yu</b>, <a href="#">UnifiedSKG: Unifying and Multi-Tasking Structured Knowledge Grounding with Text-to-Text Language Models</a>, <b>Preprint 2022</b></li><li>2. Yusen Zhang, Ansong Ni, <b>Tao Yu</b>, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, Dragomir Radev, <a href="#">An Exploratory Study on Long Dialogue Summarization: What Works and What's Next</a>, the Conference of the Empirical Methods in Natural Language Processing, <b>Findings of EMNLP 2021</b></li><li>3. Ansong Ni, Zhangir Azerbayev, Mutethia Mutuma, Troy Feng, Yusen Zhang, <b>Tao Yu</b>, Ahmed Hassan Awadallah, Dragomir Radev, <a href="#">SummerTime: Text Summarization Toolkit for Non-experts</a>, the Conference of the Empirical Methods in Natural Language Processing, <b>EMNLP 2021</b></li><li>4. Heather Lent, Semih Yavuz, <b>Tao Yu</b>, Tong Niu, Yingbo Zhou, Dragomir Radev, Xi Victoria Lin, <a href="#">Testing Cross-Database Semantic Parsers Using Canonical Utterances</a>, Evaluation &amp; Comparison of NLP Systems Workshop at EMNLP 2021, <b>Best Paper Award</b></li><li>5. Chang Shu, Yusen Zhang, Xiangyu Dong, Peng Shi, <b>Tao Yu</b>, Rui Zhang, <a href="#">Logic-Consistency Text Generation from Semantic Parses</a>, the Annual Meeting of the Association for Computational Linguistics, <b>Findings of ACL 2021</b></li></ol>	

6. **Tao Yu**, Rui Zhang, Oleksandr Polozov, Chris Meek, Ahmed Hassan Awadallah, [SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing](#), International Conference on Learning Representations, **ICLR 2021**
7. **Tao Yu**, Chien-Sheng Wu, Xi Victoria Lin, Bailin Wang, Yi Chern Tan, Xinyi Yang, Dragomir Radev, Richard Socher, Caiming Xiong, [GraPPa: Grammar-Augmented Pre-Training for Table Semantic Parsing](#), International Conference on Learning Representations, **ICLR 2021**
8. Ming Zhong, Da Yin, **Tao Yu**, Ahmad Zaidi, Mutethia Mutuma, Rahul Jha, Ahmed Hassan Awadallah, Asli Celikyilmaz, Yang Liu, Xipeng Qiu, Dragomir Radev, [QMSum: A New Benchmark for Query-based Multi-domain Meeting Summarization](#), the Conference of the North American Chapter of the Association for Computational Linguistics, **NAACL 2021**
9. 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, **Tao Yu**, Yi Chern Tan, Xi Victoria Lin, Caiming Xiong, Richard Socher, Nazneen Fatema Rajani, [DART: Open-Domain Structured Data Record to Text Generation](#), the Conference of the North American Chapter of the Association for Computational Linguistics, **NAACL 2021**
10. Ruiqi Zhong, **Tao Yu**, Dan Klein, [Semantic Evaluation for Text-to-SQL with Distilled Test Suites](#), the Conference of the Empirical Methods in Natural Language Processing, **EMNLP 2020**
11. Yusen Zhang, Xiangyu Dong, Shuaichen Chang, **Tao Yu**, Peng Shi, Rui Zhang, [Did You Ask a Good Question? A Cross-Domain Question Intention Classification Benchmark for Text-to-SQL](#) In *Interactive and Executable Semantic Parsing Workshop at EMNLP*, 2020
12. **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, [CoSQL: A Conversational Text-to-SQL Challenge Towards Cross-Domain Natural Language Interfaces to Databases](#), the Conference of the Empirical Methods in Natural Language Processing, **EMNLP 2019**
13. Rui Zhang, **Tao Yu**, He Yang Er, Sungrok Shim, Eric Xue, Xi Victoria Lin, Tianze Shi, Caiming Xiong, Richard Socher, Dragomir Radev, [Editing-based SQL Query Generation for Cross-Domain Context-Dependent Questions](#), the Conference of the Empirical Methods in Natural Language Processing, **EMNLP 2019**
14. **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, [SPaC: Cross-Domain Semantic Parsing in Context](#), the Annual Meeting of the Association for Computational Linguistics, **ACL 2019**
15. Richard A Plunz, Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen McKeown, **Tao Yu**, Laura Ugucioni, Maria Paola Sutto, [Twitter Sentiment in New York City Parks as Measure of Well-being, Landscape and Urban Planning 2019](#)
16. **Tao Yu**, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, Dragomir Radev, [Spider: A Large-Scale Human-Labeled Dataset for Complex and Cross-Domain Semantic Parsing and Text-to-SQL Task](#), the Conference of the Empirical Methods in Natural Language Processing, **EMNLP 2018**

17. **Tao Yu**, Michihiro Yasunaga, Kai Yang, Rui Zhang, Dongxu Wang, Zifan Li, Dragomir Radev, [SyntaxSQLNet: Syntax Tree Networks for Complex and Cross Domain Text-to-SQL Task](#), the Conference of the Empirical Methods in Natural Language Processing, **EMNLP 2018**
18. **Tao Yu**, Zifan Li, Zilin Zhang, Rui Zhang, Dragomir Radev, [TypeSQL: Knowledge-based Type-Aware Neural Text-to-SQL Generation](#), the Conference of the North American Chapter of the Association for Computational Linguistics, **NAACL 2018**
19. Mohammad Sadegh Rasooli, Noura Farra, Axinia Radeva, **Tao Yu**, and Kathleen McKeown, [Cross-Lingual Sentiment Transfer with Limited Resources](#), **Machine Translation 2017**
20. Owen Rambow, **Tao Yu**, Axinia Radeva, Sardar Hamidian, Alexander R. Fabbri, Debanjan Ghosh, Christopher Hidey, Tianrui Peng, Mona Diab, Kathleen McKeown, Smaranda Muresan, [The Columbia-GWU System at the 2016 TAC KBP BeSt Evaluation](#), **NIST TAC KBP Workshop 2016**

#### RESEARCH & INDUSTRY EXPERIENCE

**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](#)

#### PROFESSIONAL SERVICES

**Organizing Committee Member**

[SUKI: Workshop on Structured and Unstructured Knowledge Integration](#) at **NAACL 2022**  
[The Workshop on Interactive and Executable Semantic Parsing](#) at **EMNLP 2020**

**Program Committee Member**

**ACL Rolling Review**, **ACL**: (2020, 2021, 2022), **EMNLP**: (2019, 2020, 2021), **ICLR**: (2022), **NAACL**: (2019, 2021), **COLING**: (2020), **AACL-IJCNLP**: (2020), and other various conferences and workshops.

#### 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
TALKS & PRESENTATIONS	Invited Talk, ElementAI/ServiceNow Research	Feb 2022
	Invited Talk, NLP Reading Group, Google Research	Apr 2021
	Invited Talk, CS, The University of Hong Kong	Mar 2021
	Invited Talk, CS, The National University of Singapore	Mar 2021
	Invited Talk, CS, The University of Wisconsin-Madison	Mar 2021
	Invited Talk, CS, Simon Fraser University	Mar 2021
	Invited Talk, CSE, The University of Minnesota Twin Cities	Mar 2021
	Invited Talk, CS, City University of Hong Kong	Feb 2021
	Invited Talk, CS, Fudan University	Feb 2021
	Invited Talk, CS, Southern University of Science and Technology	Jan 2021
	Invited Poster, AI Breakthroughs Workshop, Microsoft Research Redmond	Sep 2019
STUDENTS	<a href="#">Tianbao Xie</a> (HKU PhD, 2022, co-advised with <a href="#">Lingpeng Kong</a> )	
	<a href="#">Jiacheng Ye</a> (HKU PhD, 2022, co-advised with <a href="#">Lingpeng Kong</a> )	
	<a href="#">Yiheng Xu</a> (HKU PhD, 2022, co-advised with <a href="#">Lingpeng Kong</a> and <a href="#">Benjamin Kao</a> )	
	<a href="#">Xueliang Zhao</a> (HKU PhD, 2022, co-advised with <a href="#">Lingpeng Kong</a> )	
	<a href="#">Chang Ma</a> (HKU PhD, 2022, co-advised with <a href="#">Lingpeng Kong</a> )	
	<a href="#">Ming Zhong</a> (Summer Intern 2020, Fudan MS → UIUC PhD)	
	<a href="#">Da Yin</a> (Summer Intern 2020, PKU BS → UCLA PhD)	
	<a href="#">Yusen Zhang</a> (Summer Intern 2020, Emory MS → PSU PhD)	
	<a href="#">Michihiro Yasunaga</a> (Project Student, 2018-19, Yale BS → Stanford PhD)	
	<a href="#">Suyi Li</a> (Project Student, 2019, Yale MS → HKUST PhD)	
REFERENCES	<a href="#">Kai Yang</a> (Project Student, 2018, Yale MS → Google SE)	
	<a href="#">James Ma</a> (Project Student, 2018, Yale BS → LinkedIn DS)	
	<a href="#">Dragomir Radev</a> Professor, Yale University Email: <a href="mailto:dragomir.radev@yale.edu">dragomir.radev@yale.edu</a> Cell: 646-420-1367	
	<a href="#">Kathleen McKeown</a> Professor, Columbia University Email: <a href="mailto:kathy@cs.columbia.edu">kathy@cs.columbia.edu</a> Cell: 212-939-7118	
	<a href="#">Luke Zettlemoyer</a> Professor, University of Washington Email: <a href="mailto:lsz@cs.washington.edu">lsz@cs.washington.edu</a> 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)