Tao Yu

CONTACT INFORMATION	https://taoyds.github.io/ Rm 204E Chow Yei Ching Building, Pokfulam, Hong Kong	tyu@cs.hku.hk +852-9768-6811

ACADEMIC POSITION

University of Hong Kong, Hong Kong 08/2021 - *Present*

Assistant Professor in Computer Science, Co-director of the HKU NLP Group

University of Washington, Seattle, WA 09/2021 - 08/2022

Postdoctoral Fellow in UW NLP Group

Working with Noah Smith, Luke Zettlemoyer, and Mari Ostendorf

EDUCATION

Yale University, New Haven, CT

Ph.D. Computer Science - Natural Language Processing 08/2017 - 05/2021

- Advisors: Dragomir R. Radev
- Committee: Kathleen McKeown, Luke Zettlemoyer, Robert Frank, Marynel Vázquez

Columbia University, New York, NY

M.S. Data Science   GPA: 3.82/4.0 09/2015 - 05/2017

- Advisors: Owen Rambow and Kathleen McKeown

University of Utah, Salt Lake City, UT

B.S. Mathematics and Economics   GPA: 3.86/4.0 01/2012 - 05/2015

PUBLICATIONS & PAPERS IN PREPARATION

1. Zhoujun Cheng*, Tianbao Xie*, Peng Shi, Chengzu Li, Rahul Nadkarni, Yushi Hu, Caiming Xiong, Dragomir Radev, Mari Ostendorf, Luke Zettlemoyer, Noah A Smith, **Tao Yu**, Binding Language Models in Symbolic Languages, **Preprint 2022**

2. Hongjin Su, Jungo Kasai, Chen Henry Wu, Weijia Shi, Tianlu Wang, Jiayi Xin, Rui Zhang, Mari Ostendorf, Luke Zettlemoyer, Noah A. Smith, **Tao Yu**, Selective Annotation Makes Language Models Better Few-Shot Learners, **Preprint 2022**

3. 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, **Tao Yu**, UnifiedSKG: Unifying and Multi-Tasking Structured Knowledge Grounding with Text-to-Text Language Models, **EMNLP 2022**

4. Jiacheng Ye*, Jiahui Gao*, Qintong Li, Hang Xu, Jiangtao Feng, Zhiyong Wu, **Tao Yu**, Lingpeng Kong, ZeroGen: Efficient Zero-shot Learning via Dataset Generation, **EMNLP 2022**

5. Yushi Hu, Chia-Hsuan Lee, Tianbao Xie, **Tao Yu**, Noah A. Smith, Mari Ostendorf, In-Context Learning for Few-Shot Dialogue State Tracking, **Findings of EMNLP 2022**

6. Jiacheng Ye, Jiahui Gao, Zhiyong Wu, Jiangtao Feng, **Tao Yu**, Lingpeng Kong, ProGen: Progressive Zero-shot Dataset Generation via In-context Feedback, **Findings of EMNLP 2022**

7. Qi Liu, Zihuiwen Ye, **Tao Yu**, Phil Blunsom, Linfeng Song, Augmenting Multi-Turn Text-to-SQL Datasets with Self-Play, **Findings of EMNLP 2022**

8. Yiru Chen, Ryan Li, Austin Mac, Tianbao Xie, **Tao Yu**, Eugene Wu, NL2INTERFACE: Interactive Visualization Interface Generation from Natural Language Queries, *IEEE Visualization Conference NLVIZ Workshop*, 2022

9. Ziming Mao*, Chen Henry Wu*, Ansong Ni, Yusen Zhang, Rui Zhang, **Tao Yu**, Budhaditya Deb, Chenguang Zhu, Ahmed H Awadallah, Dragomir Radev, DYLE: Dynamic Latent Extraction for Abstractive Long-Input Summarization, **ACL 2022**

10. Yusen Zhang, Ansong Ni, **Tao Yu**, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, Dragomir Radev, An Exploratory Study on Long Dialogue Summarization: What Works and What's Next, **Findings of EMNLP 2021**

11. Ansong Ni, Zhangir Azerbayev, Mutethia Mutuma, Troy Feng, Yusen Zhang, **Tao Yu**, Ahmed Hassan Awadallah, Dragomir Radev, SummerTime: Text Summarization Toolkit for Non-experts, **EMNLP 2021**

12. Heather Lent, Semih Yavuz, **Tao Yu**, Tong Niu, Yingbo Zhou, Dragomir Radev, Xi Victoria Lin, Testing Cross-Database Semantic Parsers Using Canonical Utterances, Evaluation & Comparison of NLP Systems Workshop at EMNLP 2021, **Best Paper Award**

13. Chang Shu, Yusen Zhang, Xiangyu Dong, Peng Shi, **Tao Yu**, Rui Zhang, Logic-Consistency Text Generation from Semantic Parses, **Findings of ACL 2021**

14. **Tao Yu**, Rui Zhang, Oleksandr Polozov, Chris Meek, Ahmed Hassan Awadallah, SCoRe: Pre-Training for Context Representation in Conversational Semantic Parsing, **ICLR 2021**

15. **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, **ICLR 2021**

16. 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, **NAACL 2021**

17. 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, **NAACL 2021**

18. Ruiqi Zhong, **Tao Yu**, Dan Klein, Semantic Evaluation for Text-to-SQL with Distilled Test Suites, **EMNLP 2020**

19. 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

20. **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, **EMNLP 2019**

21. 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, **EMNLP 2019**

22. **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, SParC: Cross-Domain Semantic Parsing in Context, **ACL 2019**

23. Richard A Plunz, Yijia Zhou, Maria Isabel Carrasco Vintimilla, Kathleen Mckeown, **Tao Yu**, Laura Uguccioni, Maria Paola Sutto, Twitter Sentiment in New York City Parks as Measure of Well-being, **Landscape and Urban Planning 2019**

24. **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, **EMNLP 2018**

25. **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, **EMNLP 2018**

26. **Tao Yu**, Zifan Li, Zilin Zhang, Rui Zhang, Dragomir Radev, TypeSQL: Knowledge-based Type-Aware Neural Text-to-SQL Generation, **NAACL 2018**

27. Mohammad Sadegh Rasooli, Noura Farra, Axinia Radeva, **Tao Yu**, and Kathleen McKeown, Cross-Lingual Sentiment Transfer with Limited Resources, **Machine Translation 2017**

28. 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**

PROFESSIONAL SERVICES

Organizing Committee Member

ACL 2023
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, 2022), **ICLR**: (2022), **NAACL**: (2019, 2021), **NeurlPS**: (2022), **COLING**: (2020), **AACL-IJCNLP**: (2020), and other various conferences and workshops.

GRANTS

Amazon Research Award ($40,000 + $80,000 AWS Credits) 2022

STUDENTS & COLLABORATORS

Tianbao Xie (HKU PhD, 2022)

Hongjin Su (HKU PhD, 2022)

Yiheng Xu (HKU PhD, 2022, co-advised with **Lingpeng Kong**)

Jiacheng Ye (HKU PhD, 2022, co-advised with **Lingpeng Kong**)

Zhoujun Cheng (Intern 2022, SJTU BS/MS)

Chen Henry Wu (Intern 2022, PKU BS → CMU MS)

Yushi Hu (Student Collaborator, 2022, UW PhD)

Weijia Shi (Student Collaborator, 2022, UW PhD)

Jungo Kasai (Student Collaborator, 2022, UW PhD)

Ming Zhong (Summer Intern 2020, Fudan MS → UIUC PhD)

Da Yin (Summer Intern 2020, PKU BS → UCLA PhD)

Yusen Zhang (Summer Intern 2020, Emory MS → PSU PhD)

Michihiro Yasunaga (Project Student, 2018-19, Yale BS → Stanford PhD)

Suyi Li (Project Student, 2019, Yale MS → HKUST PhD)

Kai Yang (Project Student, 2018, Yale MS → Google SE)

James Ma (Project Student, 2018, Yale BS → LinkedIn DS)

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

Talks & Presentations

Invited Talk, AllState	Jun 2022
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 Poster, AI Breakthroughs Workshop, Microsoft Research Redmond	Sep 2019

Awards

HKU-100 Scholar	2021
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	**Introduction to Data Science with Python**, University of Hong Kong	2022
	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