

Freda Shi (a.k.a., Haoyue Shi)

Assistant Professor, University of Waterloo
200 University Ave W., Waterloo, ON, Canada, N2L 3G1

Faculty Member, Vector Institute
661 University Ave., Toronto, ON, Canada, M5G 1M1

E-mail: fhs@uwaterloo.ca

Last updated: Sunday 16th June, 2024

Research Interests

Computational linguistics, natural language processing, and machine learning: compositional semantics, grounded language acquisition, unsupervised and semi-supervised representation learning, structured prediction, narrative understanding, and information theory for natural language processing.

Education

Toyota Technological Institute at Chicago, Chicago, IL, USA 2018-2024
Ph.D. in Computer Science (Ph.D. thesis of distinction; Master's degree awarded Sept. 2020)
Thesis: Learning Language Structures through Grounding
Advisors: Kevin Gimpel and Karen Livescu
Thesis Committee: Kevin Gimpel, Karen Livescu, Roger Levy and Luke Zettlemoyer

Peking University, Beijing, China 2013-2018
B.S. in Intelligence Science and Technology (Computer Science Track), *summa cum laude*
Minor in Sociology
Thesis: On Multi-Sense Word Embeddings via Matrix Factorization and Matrix Transformation
Advisor: Junfeng Hu

Non-Degree Academic Experience:

Visiting Student with Roger P. Levy 2024
Massachusetts Institute of Technology, Cambridge, MA, USA.

Visiting Student with Samuel R. Bowman 2017
New York University, New York City, NY, USA.

Visiting Student with Alexander G. Hauptmann 2016
Carnegie Mellon University, Pittsburgh, PA, USA

Appointments

Assistant Professor, University of Waterloo, Ontario, ON, Canada 2024-

Faculty Member, Vector Institute, Toronto, ON, Canada 2024-

Selected Honors and Awards

Thesis of Distinction, Toyota Technological Institute at Chicago 2024

Highlighted Reviewer, ICLR 2022

Google Ph.D. Fellowship (\approx USD \$220,000 for tuition and stipend in 3 years) 2021

Finalist, Facebook Ph.D. Fellowship 2021

Nomination for the Best Paper Award, ACL-IJCNLP (*with L. Zettlemoyer and S. I. Wang*) 2021

Nomination for the Best Paper Award, ACL (*with J. Mao, K. Gimpel and K. Livescu*) 2019

Best Undergraduate Dissertation Award, School of EECS, Peking University 2018

Robin Lee Scholarship (top 2 out of 400, CNY ¥20,000), Peking University	2016
WeTech Qualcomm Global Scholarship (USD \$5,000)	2016
Arawana Scholarship (top 4 out of 400, CNY ¥10,000), Peking University	2015
Gold Medalist (<i>with T. Li and C. Mo</i>), ACM-ICPC Chengdu Site	2013

Research Internships

Google Brain (Hybrid Internship), Waterloo, ON, Canada → Chicago, IL, USA Host: Denny Zhou	Jun. 2022-Dec. 2022
Meta (Facebook) AI Research (Remote Internship), Seattle, WA, USA Mentors and collaborators: Luke Zettlemoyer, Sida Wang, Daniel Fried, and Marjan Ghazvininejad	Aug. 2021-Dec. 2021
Facebook AI Research (Remote Internship), Seattle, WA, USA Mentors: Sida Wang and Luke Zettlemoyer	Jun. 2020-Dec. 2020
ByteDance AI Lab , Beijing, China Mentors: Hao Zhou and Lei Li	Mar. 2018-Aug. 2018
Megvii (Face++) Research , Beijing, China Mentors: Yuning Jiang and Jian Sun	Oct. 2017-Mar. 2018
Microsoft Research Asia , Beijing, China Mentors: Zhongyuan Wang and Jun Yan	Sep. 2016-Feb. 2017

Engineering Internships

4th Paradigm Inc. , Beijing, China Mentors: Weiwei Tu and Yuqiang Chen	Mar. 2017-Jun. 2017
Google Inc. , Beijing, China Mentors: Xiaoyi Ren and Jie Mao	Jul. 2015-Dec. 2015

Teaching Experience

Instructor at Toyota Technological Institute at Chicago and the University of Chicago TTIC 31190 Natural Language Processing Co-Instructor: Jiawei (Joe) Zhou	Autumn 2023
Guest Lecturer at the University of Chicago MPCS 53113 Natural Language Processing Instructor: Amitabh Chaudhary	Summer 2021
Teaching Assistant at Toyota Technological Institute at Chicago TTIC 31220 Unsupervised Learning and Data Analysis Instructor: Karen Livescu	Winter 2021
Teaching Assistant at School of EECS, Peking University Practice of Programming in C&C++ Instructor: Wei Guo	Spring 2018
Programming & Algorithms (MOOC on Coursera) Instructor: Wei Guo	Fall 2016
Practice of Programming in C&C++ Instructor: Jiaying Liu	Spring 2015
Volunteer Lecturer in Mathematics Rongxian High School Summer Camp, Guangxi, China	Summer 2014

Referred Conference Publications

*: Equal contribution.

1. **Freda Shi**, Kevin Gimpel, Karen Livescu. 2024. Structured Tree Alignment for Evaluation of (Speech) Constituency Parsing. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*.

2. Danlu Chen, **Freda Shi**, Aditi Adagarwal, Jacobo Myerston, Taylor Berg-Kirkpatrick. 2024. Comparing Visual and Textual Representations of Ancient Logographic Writing Systems for NLP. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL)*.
3. **Freda Shi***, Xinyun Chen*, Kanishka Misra, Nathan Scales, David Dohan, Ed Chi, Nathanael Schärli, Denny Zhou. 2023. Large Language Models Can Be Easily Distracted by Irrelevant Context. In *Proceedings of the Fortieth International Conference on Machine Learning (ICML)*.
4. **Freda Shi***, Mirac Suzgun*, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder, Denny Zhou, Dipanjan Das, Jason Wei. 2023. Language models are multilingual chain-of-thought reasoners. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*.
5. Daniel Fried*, Armen Aghajanyan*, Jessy Lin, Sida Wang, Eric Wallace, **Freda Shi**, Ruiqi Zhong, Wen-tau Yih, Luke Zettlemoyer, Mike Lewis. 2023. InCoder: A Generative Model for Code Infilling and Synthesis. In *Proceedings of the Eleventh International Conference on Learning Representations (ICLR)*.
6. **Freda Shi**, Daniel Fried, Marjan Ghazvininejad, Luke Zettlemoyer, Sida I. Wang. 2022. Natural Language to Code Translation with Execution. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
7. **Freda Shi**, Kevin Gimpel, Karen Livescu. 2022. Substructure Distribution Projection for Zero-Shot Cross-Lingual Dependency Parsing. In *Proceedings of the Joint Conference of the 60th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL)*.
8. Vikram Gupta, **Haoyue Shi**, Kevin Gimpel, Mrinmaya Sachan. 2022. Deep Clustering of Text Representations for Supervision-Free Probing of Syntax. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*.
9. Jiayuan Mao, **Haoyue Shi**, Jiajun Wu, Roger Levy, Joshua B. Tenenbaum. 2021. Grammar-Based Grounded Lexicon Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
10. **Haoyue Shi**, Luke Zettlemoyer, Sida I. Wang. 2021. Bilingual Lexicon Induction via Unsupervised Bitext Construction and Word Alignment. In *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP; Best Paper Nominee)*.
11. **Haoyue Shi**, Karen Livescu, Kevin Gimpel. 2021. Substructure Substitution: Structured Data Augmentation for NLP. In *Findings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Findings of ACL-IJCNLP)*.
12. **Haoyue Shi**, Karen Livescu, Kevin Gimpel. 2020. On the Role of Supervision in Unsupervised Constituency Parsing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
13. **Haoyue Shi***, Jiayuan Mao*, Kevin Gimpel, Karen Livescu. 2019. Visually Grounded Neural Syntax Acquisition. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL; Best Paper Nominee)*.
14. **Haoyue Shi**, Hao Zhou, Jiaze Chen, Lei Li. 2018. On Tree-Based Neural Sentence Modeling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
15. **Haoyue Shi***, Jiayuan Mao*, Tete Xiao*, Yuning Jiang, Jian Sun. 2018. Learning Visually-Grounded Semantics from Contrastive Adversarial Samples. In *Proceedings of the 27th International Conference on Computational Linguistics (COLING)*.
16. **Haoyue Shi**, Jia Chen, Alexander G. Hauptmann. 2017. Joint Saliency Estimation and Matching using Image Regions for Geo-Localization of Online Video. In *Proceedings of the 2017 ACM International Conference on Multimedia Retrieval (ICMR)*.

Conference Publications Referred by Abstract

17. **Haoyue Shi**, Xihao Wang, Yuqi Sun, Junfeng Hu. 2018. Constructing High Quality Sense-specific Corpus and Word Embedding via Unsupervised Elimination of Pseudo Multi-sense. In *Proceedings of the 11th Language Resources and Evaluation Conference (LREC)*.
18. Shan Xu, **Haoyue Shi**, Xiaohui Duan, Tiangang Zhu, Peihua Wu, Dongyue Liu. 2016. Cardiovascular Risk Prediction Method Based on Test Analysis and Data Mining Ensemble System. In *Proceedings of the 2016 IEEE International Conference on Big Data Analysis*.

Referred Workshop Publications

19. Cheng-I Jeff Lai*, **Freda Shi***, Puyuan Peng*, Yoon Kim, Kevin Gimpel, Shiyu Chang, Yung-Sung Chuang, Saurabhchand Bhati, David Cox, David Harwath, Yang Zhang, Karen Livescu, James Glass. 2023. Audio-Visual Neural Syntax Acquisition. In *Proceedings of the 2023 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*.
20. Shubham Toshniwal, **Haoyue Shi**, Bowen Shi, Lingyu Gao, Karen Livescu, Kevin Gimpel. 2020. A Cross-Task Analysis of Text Span Representations. In *Proceedings 4th Workshop on Representation Learning for NLP*.
21. Yuqi Sun, **Haoyue Shi**, Junfeng Hu. 2018. Implicit Subjective and Sentimental Usages in Multi-sense Word Embeddings. In *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.
22. **Haoyue Shi**, Caihua Li, Junfeng Hu. 2016. Real Multi-Sense or Pseudo Multi-Sense: An Approach to Improve Word Representation. In *Proceedings of the 1st Workshop on Computational Linguistics for Linguistic Complexity*.

Patents

23. Hao Zhou, Lei Li, Jiaze Chen, **Haoyue Shi**. 2019. Method and Device for Generating Information. CN201910105241.

Conference and Workshop Presentations without Proceedings

1. **Haoyue Shi**, Jiayuan Mao, Kevin Gimpel, Karen Livescu. 2019. Visually Grounded Neural Syntax Acquisition. Talk, Midwest Speech and Language Days, Chicago, IL, USA, May 2-3.

Invited Talks

1. *Towards Computational Multilingualism with Large Language Models*. Boston University, May 2024.
2. *Computational Multilingualism in the Era of Large Language Models*. Vector NLP Workshop, February 2024.
3. *Learning Syntactic Structures from Visually Grounded Text and Speech*. University of Michigan, October 2023.
4. *Learning Language Structures through Grounding*. Peking University, September 2023.
5. *Learning Language Structures through Grounding*. University of Toronto, August 2023.
6. *Language Models Are Multilingual Chain-of-Thought Reasoners*. Translate Theory Reading Group, Google AI, October 28th, 2021.
7. *Naturally Supervised Parsing: Assumptions, Methods and Evaluation*. Yahoo! NYC Remote Research Seminar, April 15th, 2021.
8. *Visually Grounded Neural Syntax Acquisition*. Remote Seminar, Carnegie Mellon University, August 19th, 2020.
9. *Structures in Natural Language: How to learn it and how to use it?* Remote NLP Seminar, University of Alberta, May 4th, 2020.
10. *Visually Grounded Neural Syntax Acquisition*. NLP Seminar, Peking University, Decemember 24th, 2019.

Open-Sourced Project Contributions

Implementations that accompany to the publications listed above are open-sourced if permitted, and are not listed below.

1. NLTK

A suite of open source Python modules, data sets, and tutorials.
<https://nltk.org>

2. NL-Augmenter

A general-purpose data augmentation framework for NLP.
<https://github.com/GEM-benchmark/NL-Augmenter>

3. Multimodal concreteness score estimator

Implementation of the paper *Quantifying the Visual Concreteness of Words and Topics in Multimodal Datasets* (Hessel et al., 2018).
<https://github.com/victorssilva/concreteness>

4. Structured self-attentive sentence embeddings

Implementation of the paper *A Structured Self-Attentive Sentence Embedding* (Lin et al., 2017).
<https://github.com/explorerfreda/structured-self-attentive-sentence-embedding>

Skills

Programming Languages:

- Proficient: C/C++, Python(2/3), MATLAB, Pascal, HTML/CSS
- Capable: C#, SCOPE, JavaScript, Java, Scala, Mathematica, Bash, SQL

Natural Languages:

Mandarin (native), English (fluent), classical Chinese (advanced reading & writing), Cantonese (intermediate listening & speaking), Japanese (intermediate), German (beginner), Hebrew (beginner), Spanish (beginner)

Tools & Frameworks: Vim, Caffe, Torch, PyTorch, GDB, Git, L^AT_EX, CMake, Visual Studio, ssh

Service

Area Chair for

- COLM, 2024;
- Language Grounding to Vision, Robotics and Beyond, ACL 2023;
- Machine Learning for NLP, EMNLP 2023.

Reviewer for Conferences and Journals in

- Computational Linguistics and Natural Language Processing: TACL (2023–2025), ACL (2019–2022), ACL Rolling Review (2020–2024), COLING (2020, 2022), EACL (2022), EMNLP (2020–2022), LREC (2020), NAACL (2021), NLPCC (2020, 2021);
- Machine Learning: JMLR (2023), TPAMI (2022), ICLR (2020–2024), ICML (2020–2024), NeurIPS (2020–2024);
- Artificial Intelligence: IJCAI (2021), AAI (2019, *secondary to Hao Zhou*), UAI (2023, *secondary to Lili Mou*);
- Computer Vision: ViGiL Workshop (2021), CVPR (2020, *secondary to Jiayuan Mao*);
- Robotics: ICRA (2024).

Co-Organizer of UChicago-TTIC NLP Reading Group, 2022–2023.

Student Member of the TTIC Student Admission Committee, 2021–2022.

Co-Organizer of TTIC Student Workshop, 2020.

Student Representative at TTIC, 2020–2021.

Student Co-Chair of the Women at TTIC group and **coordinator** with UChicago Graduate Women in CS, 2019–2022.

Peer Mentor for new students at TTIC, 2019–2020, 2022–2023.

Chief of the PKU Guqin Society, 2016–2017.