Yiyun Zhao
NLPer · Psycholinguist

Address: 1103 E University Blvd, Tucson, AZ 85721
Email: yiyunzhao@email.arizona.edu
Personal Website

RESEARCH INTEREST

psycholinguistics: linguistic universals, language evolution, miniature artificial language learning, dependency length minimization, language transfer
natural language processing/computational linguistics: model interpretatbility, linguistic probing, semantic parsing, data augmentation

EDUCATION BACKGROUND

Ph.D. in Computational Linguistics, minor in Cognitive Sciences (4.0/4.0)
University of Arizona, Tucson, U.S.A   2017-2022

M.S. in information (4.0/4.0)
University of Arizona, Tucson, U.S.A   2019

M.Sc. in Evolution of Language & Cognition (with distinction)
University of Edinburgh, Edinburgh, U.K.   2016

B.A. in English & Linguistics
Beijing Foreign Studies University, Beijing, China   2015

PRESENTATION PUBLICATION

—In Progress—

- (submitted): Xin Su, **Yiyun Zhao**, Steven Bethard. "A Comparison of Strategies for Source-Free Domain Adaptation"
- (in Prep): **Yiyun Zhao**, Charles Torres, Masha Fedzechkina ."Is Dependency length minimization preference of speakers of different L1s universal and uniform ? "
- w/ Steven Bethard. Simulating artificial language experiments on neural networks. (on-going).
- w/ Xin Su, Steven Bethard. Review on recent advances in Text2Sql (on-going).

—Articles—

- **Yiyun Zhao**, Jian Gang Ngui, Lucy Hall Hartley, Steven Bethard (November 2021)."Do pretrained transformers infer telicity like humans". In:*Proceedings of the SIGNLL conference on Computational Natural Language Learning. (CoNLL 2021)*
- Xin Su, **Yiyun Zhao**, Steven Bethard (November 2021)." The University of Arizona at SemEval-2021 Task 10: Applying Self-training, Active Learning and Data Augmentation to Source-free Domain Adaptation" In: *Proceedings of the 15th International Workshop on Semantic Evaluation.(SemEval 2021)*
- Zhengzhong Liang, **Yiyun Zhao**, Mihai Surdeanu (March 2021)."Using the Hammer Only on Nails: A Hybrid Method for Evidence Retrieval for Question Answering" In: *Proceedings of the 43rd European Conference on Information Retrieval.(ECIR 2021)*

- Egoitz Laparra, Xin Su, **Yiyun Zhao** , Ozlem Uzuner, Timothy Miller, Steven Bethard SemEval-2021 Task 10: Source-Free Domain Adaptation for Semantic Processing. (SemEval 2021)

- **Yiyun Zhao**, Steven Bethard (April 2020). "How does BERTs attention change when you fine-tune? An analysis methodology and a case study in negation scope". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. (ACL 2020)*

- **Yiyun Zhao**, Masha Fedzechkina (Jan 2020). "Learners harmonic preferences are modulated by lexical retrieval difficulty". In: *Proceedings of the 94th Annual Meeting of the Linguistics Society of America. (LSA 2020)*

- Steven Bethard, Egoitz Laparra, Sophia Wang, **Yiyun Zhao**, Ragheb Al-Ghezi, Aaron Lien, Laura Lopez-Hoffman.(2019). Inferring missing metadata from environmental policy texts. In Proceedings of the 3rd Joint SIGHUM Workshop'2019.

—**Conference Presentation** (peer reviewed)—

- 2021. Do pretrained transformers infer telicity like humans? The 2021 SIGNLL conference on Computational Natural Language Learning.

- 2021. Preferences for shorter dependencies in miniature language learning are modulated by the statistics of learners L1. The 34th Conference on Human Sentence Processing (CUNY).

- 2020. Dependency length minimization: an abstract bias or an input-driven preference? The 26th architectures and mechanisms for language processing conference (AMLaP).

- 2020. Learners' harmonic preferences in head ordering are modulated by lexical retrieval difficulty. The 94th Annual Meeting of the Linguistic Society of America, New Orleans, LA.
  (This talk was also presented in Arizona Linguistics Circle 13, Tucson, AZ.)

- 2020. How does BERTs attention change when you fine-tune? The 58th Annual Meeting of the Association for Computational Linguistics.

- 2018. Synonyms are lost during cultural transmission without an explicit bias against synonyms. Workshop on Emergence of Universals at The Ohio State University. Columbus, OH.

- 2013. The distinctive characteristics of Chinese neologisms. The First Boya Undergraduate Research Forum at Beijing Foreign Studies University. Beijing.

ACADEMIC WORKING EXPERIENCE

—**Research Associate**—
University of Arizona
Computational Language Understanding Lab since 2019 Feb

- Lifestyle Intervention for Ovarian Cancer Enhanced Survival (NIH funded)
  - Visualized patient performance trajectories and identified critical performance changes.
  - Built machine learning baseline models (e.g.,logistic regression models, SVM, decision trees) to predict dropouts from identity features and individual baseline performance.
  - Built neural network models to disentangle intervention style latent variables from the intervention content latent variables.

- eNEPA Harnessing the Power of Big Data Project (NSF-funded)
  - Wrote rule-based models for title-matching and lead agency detection tasks.
  - Built the annotation dataset for title-matching using tf-idf, which was used for the search algorithm in the eNEPA database.
- Temporal Histories of Your Medical Event (THYME) project
  - Probed the negation knowledge before and after fine-tuning the large pre-trained contextualized embedding.
  - Explored the self-attention scores to select substitute words.

Language Evolution, Acquisition & Processing Lab   2018 AUG - 2019 JUN

- Dependency Length Minimization Project
  - designed and implemented artificial language learning experiments
  - wrote python scripts to automatically generate experimental code on FindingFive (json style)
  - wrote python scripts to auto-annotate participants production data
  - analyzed experimental results using linear mixed effect model in R

—**Instructor & Teaching Assistant**—
University of Arizona

- INFO 557 Neural Networks. *Teaching Associate*   2021F
- LING/PHIL/PSY 432 Psychology of Language. *Teaching Associate*   2018F
- LING/PSY/SLHS 341 Language Development. *Instructor*   2017F
- LING 150 Language in the World. *Teaching Assistant*   2017F,2018S

INDUSTRIAL WORKING EXPERIENCE

—**Applied Scientist Intern**—
AWS AI Labs, Amazon, USA. JUN 2021 - SEP 2021

- Researched the current Text2SQL methods, classified and compared current system design.
- Built a data synthesizing pipeline that extracted SQL templates, synthesized new SQL and natural language pairs using pretrained transformers with various data filtering.
- Tested the data augmentation design using different training paradigms with different synthesizing ratio and iteration steps.

—**Data Scientist Intern**—
FindingFive, USA. July 2020 - August 2020

- Optimized the queries to retrieve users records in MongoDB database
- Built statistical models to predict online participants behavior using logistic regression, rule-based sentiment analysis and topic modeling in Python and R
- Summarized the results and wrote up an analysis report

AWARDS	University of Arizona, *GPSC travel grant* $655	2019
GRANTS	University of Arizona, Linguistics Department *travel grant* $200	2018, 2020
	The First Boya Undergraduate Research Forum, *2nd Prize for the Best Presentation*	
	of the Linguistics Section	2013
	Beijing Foreign Studies University, *Liu Shimu Scholarship*	2012

LANGUAGE
SKILLS

Programming Languages: Python, R, SQL, MongoDB
Natural Languages: Mandarin(native), English(fluent), Spanish(basics)
NLP tools: Pytorch, Keras, Scikit-learn, Pandas, SpaCy, NLTK
Others: AWS tools, Mechanical Turk, FindingFive

REFERENCES

Dr. Steven Bethard, *Associate Professor of Information*
Mailing: School of Information, 1103 E. 2nd St. Tucson, AZ 86721
Email: bethard@email.arizona.edu
Phone: (520) 621-5223

Dr. Masha Fedzechkina, *Assistant Professor of Linguistics*
Mailing: Communication 114E, Department of Linguistics, Tucson, AZ 86721
Email: mfedzech@email.arizona.edu

Dr. Mike Hammond, *Professor of Linguistics*
Mailing: Douglass 308, Department of Linguistics, Tucson, AZ 86721
Email: hammond@email.arizona.edu
Phone: (520) 621-5759