Xiaoxi Luo

✉ x25luo@uwaterloo.ca | 📞 +1 437 996 6032 | https://xiaoxi-luo-cl.github.io/

Research Interests: natural language processing, computational linguistics, interpretability, multi-lingual NLP, diachronic language change, computational historical linguistics

EDUCATION

University of Waterloo, Master of Mathematics in Computer Science

    Supervisor: Freda Shi   Sept. 2025 – Aug. 2026 (Expected)

- **Selected Courses**: Causal Inference (randomized and observational experiments, causal discovery algorithms), Deep Learning Theory (neural network dynamics, scaling limits)

Peking University, B.S. in Data Science   Sept. 2019 – Jul. 2024

- **Overall GPA**: 86/100
- **Courses in Mathematics**: Mathematical Analysis, Advanced Algebra, Probability Theory and Statistics, Foundations of Machine Learning, Applied Stochastic Processes, Optimization Methods, Bayesian Theory and Computation
- **Courses in Linguistics**: Modern Chinese Language, Ancient Chinese, Historical Chinese Syntax, Historical Chinese Phonology, Theoretical Linguistics, Historical Linguistics

PUBLICATION

Xiaoxi Luo and Weiwei Sun. *Phonetic Reconstruction of the Consonant System of Middle Chinese via Mixed Integer Optimization*. In: *Transactions of the Association for Computational Linguistics* (2025).

Shuyu Wu*, Ziqiao Ma*, **Xiaoxi Luo***, Yidong Huang, Josue Torres-Fonseca, Freda Shi, Joyce Chai. *The Mechanistic Emergence of Symbol Grounding in Language Models*. Under Review.

                                                (*: Equal Contribution)

RESEARCH EXPERIENCES

Revisiting Sound-Meaning Systematicity: A Fine-Grained Computational Analysis

                                           Aug. 2025 – Present

Advisor: Dr. Freda Shi, David R. Cheriton School of Computer Science, University of Waterloo

- Re-evaluating the degree of systematicity in the lexicon through a large-scale multi-lingual study (English, Mandarin, French, Hindi).
- Proposed a fine-grained computational pipeline that mitigates key confounders (etymology, morphology) and integrates a distinctive feature-aware phonemic distance metric with BERT-based semantic representations.
- Constructed a strictly controlled, monomorphemic Chinese word list, manually filtering historical cognates to enable precise analysis of sound-meaning correlations.

The Mechanistic Emergence of Symbol Grounding in Language Models   June 2025 – Oct. 2025

Advisor: Dr. Freda Shi, David R. Cheriton School of Computer Science, University of Waterloo

- Conducted an interpretability study on the emergence of symbol grounding in LLMs (GPT-2) and VLMs (LLaVA) during training.
- Localized grounding to middle layers (7-9) by analyzing ground-to-symbol attention flow and layer-wise representation probing with Tuned Lens.
- Visualized attention patterns to validate the hypothesized "gather-and-aggregate" mechanism, and devised an automated filtering criterion to identify candidate heads, facilitating subsequent causal analysis.

Reconstruction of the Consonant System of Middle Chinese via Optimization   Feb. 2023 – May 2024

Advisor: Dr. Weiwei Sun, Department of Computer Science and Technology, University of Cambridge
- Formalized phonetic reconstruction of Middle Chinese as a Mixed Integer Programming problem, optimizing for consistency with historical evidence (ancient rhyme dictionaries) and modern dialect reflexes.
- Tailored a mainstream distinctive feature set (based on generative phonology) and proposed a compact Chinese-specific set for precise phonetic representation of phonemes.
- Designed the evaluation framework for this unsupervised task, including synthetic data experiments (93.2% recovery rate on Latin consonant system) and held-out validation.
- Achieved an Adjusted Mutual Information (AMI) of 0.8148 with philologists' categorical reconstruction, showing a high consistency with philological consensus.

Historical Syntax and Lexical Semantics Research Based on Annotated Corpus of Pre-Qin Chinese

Mar. 2021 – Apr. 2023

Advisor: Prof. Yonghai Shao, Department of Chinese Language and Literature, Peking University
- Led the data processing group to construct a large-scale annotated corpus of Pre-Qin Chinese, providing technical support for linguistic analysis.
- Trained a parser on Old Chinese treebanks to automate syntactic annotation on raw historical texts.
- Constructed a WordNet-style lexical network for Old Chinese.
- Designed an algorithm for identifying cognates among two Sino-Tibetan languages.

SELECTED PROJECT

Variational Inference for Hierarchical Dirichlet Process Hidden Markov Model Jun. 2024
Advisor: Dr. Cheng Zhang, School of Mathematical Sciences, Peking University
- Conducted a detailed derivation of Gibbs sampling and variational inference methods for HDP-HMM.
- Implemented the variational inference algorithm and validated its performance on synthetic data.
- Applied both algorithms to part-of-speech tagging using the Penn Treebank and CHILDES datasets.

SKILLS
- **Programming Languages and Tools**: Python, C/C++, R, Linux, Git, LaTeX
- **Languages**: Mandarin (native), English (fluent), classical Chinese (fluent reading)

TEACHING EXPERIENCE

Cambridge Centre for International Research, *Remote Teaching Assistant* Mar. 2024 – Sept. 2025
- Mentored 20+ high school students from around the world to complete independent research papers in ML & NLP, providing guidance on code implementation, mathematical derivation, and academic writing.
- Supervised a paper titled *Error Analysis for POS Tagging of Hindi-English Code-Mixed Data*, accepted by the 2025 AMLDS conference.

WORK EXPERIENCE

Minimax, *LLM Engineer Intern* Mar. 2025 – May 2025
Worked on the post-training alignment of the MiniMax-M1 model. Constructed high-quality instruction sets, and experimentally verified optimal data mixing strategies to enhance model alignment and instruction-following capabilities.

SELECTED AWARDS AND HONORS

Vector Scholarships in Artificial Intelligence ($17,500 CAD)	May 2025
International Master's Award of Excellence	Dec. 2024
YuanPei Young Scholars (9/331)	Mar. 2024
Award for Academic Excellence	Oct. 2023
Outstanding Freshman Scholarship	Sep. 2019