Lanni Bu

Email: lb1437@georgetown.edu | **Phone:** +1 (202) 386 1989 | **Website:** https://lanni-ni.github.io/

Education

Georgetown University — Washington, DC
M.S. Computational Linguistics — 09/2024 - 05/2026

Qingdao University — Qingdao, Shandong
B.A. Language and Big Data — 09/2020 - 07/2024

Leiden University — Leiden, Netherlands
Summer school — 07/2023

University of Crete — Crete, Greece
Summer school — 07/2024

Research Interests

Broadly, my research focuses on computational linguistics and natural language processing, especially on discourse and cognitive modeling. I am particularly interested in using computational methods to investigate linguistic and cognitive questions.

Language model evaluation

– Designing linguistically motivated benchmarks to probe LLMs' discourse and structural understanding.

Cognitive modeling of language comprehension

– Studying model–human differences in comprehension and developing cognitively inspired modeling architectures.

Research Experiences

Georgetown University — 2024 – Now
Corpus Linguistics Lab (Corpling), PI: Amir Zeldes

Georgetown University — 2024 – Now
Psycholinguistics, Information, and Computation Lab (PICoL), PI: Ethan Gotlieb Wilcox

Working Experiences

Tyche Partners — Los Altos, California
Data Intern — Summer 2023

– Assisted with data organization and cleaning using Python and Excel.

Publications and Preprints

Lanni Bu, Lauren Levine, Amir Zeldes. *DiscoTrack: A Multilingual LLM Benchmark for Discourse Tracking*. Preprint, 2025. Submitted to EACL (under review) [paper]

Xiulin Yang, Zhuoxuan Ju, **Lanni Bu**, Zoey Liu, Nathan Schneider. UD-English-CHILDES: A Collected Resource of Gold and Silver Universal Dependencies Trees for Child Language Interactions. *Proceedings of the Eighth Workshop on Universal Dependencies (UDW, SyntaxFest 2025)*. [paper] [resource]

Other Projects

Zhuoxuan Ju, **Lanni Bu**, Dagny Whall, Vattana Chan. *Dependency-Based and Constituency-Based Inductive Biases for BabyLM: Comparison and Analysis*. Manuscript, 2024. [paper] [poster]

Relevant Coursework

Linguistics

- Introduction to Linguistics; Syntax; Semantics; Phonetics and Phonology
- Corpus Linguistics; Experimental Psycholinguistics; Historical Linguistics

Computational & Mathematics

- Advanced mathematics; Discrete Mathematics; Data Structure basis; Probability Theory
- Machine Learning

Computational Linguistics

- Natural Language Processing; Empirical Natural Language processing; Information Stucture & Lang; Computational Corpus Linguistics

Skills

Language: Mandarin Chinese (Native), English (Proficient)
Coding: Python (PyTorch, Transformers, pandas), R(lme4, tidyverse, ggplot2), LaTeX, Hugging Face