Lanni Bu

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

Education

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

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

Leiden University — Leiden, Netherlands
Summer School in Linguistics — 07/2023

University of Crete — Crete, Greece
Summer School in Linguistics — 07/2024

Research Interests

Broadly, my research interests include:

- Understanding where human and LLM generalization diverge.
- Developing computational models that approximate human real-time language processing.
- Computational models of discourse, with a focus on salience.

Research Experiences

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

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

Publications and Preprints

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

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).

Ongoing Project

Lanni Bu, Xiulin Yang, Christian Clark, Ethan Wilcox. *What Transformer Attention Mechanism Provides the Best Fit for Human Reading Times?*. Submitted to the 39th Annual Conference on Human Sentence Processing (under review).

Course Project

Dependency-Based and Constituency-Based Inductive Biases for BabyLM: Comparison and Analysis. Final Project for Empirical Methods in Natural Language Processing, 2024.

Relevant Coursework

Computational Linguistics

- Natural Language Processing; Empirical Natural Language Processing; Information Structure & Language;
- Computational Corpus Linguistics

Linguistics

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

Computational & Mathematics

- Advanced Mathematics; Discrete Mathematics; Data Structures and Algorithms; Probability Theory
- Machine Learning

Skills

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