XIULIN YANG

Washington, D.C., United States | xiulin.yang.compling@gmail.com | xiulinyang.github.io | LinkedIn | GitHub

RESEARCH INTEREST

I'm broadly interested in cognitive science-inspired NLP and computational modeling of language learning in both monolingual, multilingual, and cross-linguistic settings. My thesis focuses on improving the data sample efficiency of language models through introducing cognitively motivated inductive biases by modifying tokenization, pretraining objectives, and model architecture.

EDUCATION

Georgetown University Expected 2028
PhD in Computational Linguistics — GPA 3.97/4.00

University of Groningen & Saarland University 2023
Erasmus Mundus Joint Master's in Language and Communication Technologies — GPA 8.5/10 (RuG), 1.2/4.0 (UdS)

University of Oxford 2021
MSt in Linguistics, Philology and Phonetics — Distinction

Shandong University 2020
B.A. in English Linguistics (Liangshiqiu Pilot Programme) — 90.71/100

PROJECTS IN PROGRESS

Agent **Yang, X.**, & Hu, J. Communication with silence. *(ongoing project)*.

Tokenization Tokenization survey *(ongoing collaborative project)*.

LM Bu, L., **Yang, X.**, Clark, C., & Wilcox, E. (2026). What Transformer Attention Mechanism Provides the Best Fit for Human Reading Times? *under review*.

Multilingual **Yang, X.**, Arnett, C., & Wilcox, E. (2026). Autoregressive Transformer Models Are Systemically Biased Toward English-like Languages. *Pre-registered study in progress; EMNLP under review*.

Benchmarking **Yang, X.**, Bisazza, A., Wilcox, E., & Schneider, N. (2026). A Unified Assessment of the Poverty of the Stimulus Argument for Neural Language Models. *TACL, R&R, under review*.

Multilingual **Yang, X.**, Wilcox, E., & Arnett, C. (2026). Apples to Apples? Towards Comparable Crosslingual Language Model Evaluation. *EMNLP 2026 under review*.

SELECTED PAPERS

ACL 2026 **Yang, X.**, Getz, H., & Wilcox, E. (2026). Function Words as Statistical Cues for Language Learning. *Proceedings of ACL 2026*. [paper] [code] [model]

CoNLL 2026 Padovani, F., **Yang, X.**, Jumelet, J., & Bisazza, A. (2026). A Syntactic Annotation Toolkit for Child–Adult InTeractions (CAIT). *CoNLL 2026* [paper] [code].

ACL 2025 **Yang, X.**, Aoyama, T., Yao, Y., & Wilcox, E. (2025). Anything Goes? A Crosslinguistic Study of (Im)possible Language Learning in LMs. *Proceedings of ACL 2025 (Volume 1: Long Papers)*. [paper] [code]

COLING 2025 **Yang, X.** (2025). Language Models at the Syntax–Semantics Interface: A Case Study of the Long-Distance Binding of Chinese Reflexive *ziji*. *Proceedings of COLING 2025*. [paper] [code]

EMNLP 2024 **Yang, X.**, Groschwitz, J., Koller, A., & Bos, J. (2024). Scope-enhanced Compositional Semantic Parsing for DRT. *Proceedings of EMNLP 2024*. [paper] [code]

COLING 2024 Otto, C., Groschwitz, J., Koller, A., **Yang, X.**, & Donatelli, L. (2024). A Corpus of German Abstract Meaning Representation (DeAMR). *Proceedings of LREC-COLING 2024*. [paper]

RESEARCH & WORKING EXPERIENCE

Visiting Graduate Scholar, GLINT Lab, Johns Hopkins University 05/2026 – present
Host: Prof. Jennifer Hu

- Studying implicit coordination and convention formation through human–human, human–agent, and LLM–LLM simulations of a cooperative card game.
- Built the full experimental stack: JS/Firebase web game, Python simulator with Bayesian, rule-based, and random agent types, Prolific data collection pipeline, and JHU server deployment.

AIGC Algorithm Engineer Intern, Kunlun Tech 04/2024 – 07/2024
Advisor: Dr. Yuqiang Xie · LLM

– Built a self-refinable LLM pipeline and finetuned LLMs for controllable story generation from summaries, and for NPC-profile-based summarization.
– Developed a domain-specific lexicon for NSFW content evaluation to support storyline design.

Research Assistant, Georgetown University (GUM) 09/2023 – 05/2024
Advisors: Prof. Nathan Schneider & Prof. Amir Zeldes · Computational Semantics
– Built a silver semantic layer within the PropBank framework for the GUM corpus, including verb-sense disambiguation and semantic role labeling.

Research Assistant, Saarland University 08/2022 – 04/2023
PI: Prof. Alexander Koller · LLM, Coreference
– Trained an end-to-end model for coreference resolution in recipes, covering zero and explicit anaphora, to support downstream AMR parsing (avg. F1 0.68 on explicit coreference).
– Evaluated zero-anaphora resolution using few-shot prompting with GPT-3.5, reaching a BLEU score of 0.30.

INVITED TALKS

Boston University, TinLab 11.2025
"LLMs and Poverty of the Stimulus"

Shandong University 07.2025
"Language Models as Models of Languages"

Georgetown University, Learning and Development Lab 04.2025
"Multilingual Impossible Language Learning"

Shandong University 07.2024
"Large Language Models as Approaches to Linguistics Research"

TEACHING & SERVICE

Teaching Assistant, Intro to Natural Language Processing 2026 Spring

Teaching Assistant, Empirical Methods in NLP (Graduate) 2025 Spring

Reviewer — ARR (Feb, Apr, Oct 2025; Jan, Mar, May 2026), SRW 2025 2025

Maintainer, AMR Bibliography 01.2024

GRANTS & AWARDS

Outstanding Reviewer Award, EMNLP 2025 2025

Merit-Based Graduate School Scholarship, Georgetown University 2023 – 2026
USD 45,000 per year

George Wolf Prize, University of Oxford 07.2022
Awarded for the best performance in the MSt in Linguistics, Philology and Phonetics

Erasmus Mundus Language and Communication Program Scholarship 04.2021
University of Groningen & Saarland University — ~EUR 48,000; sole Chinese scholarship recipient, 2021–2022

Outstanding Graduate of Shandong University 07.2020
Awarded to the top 15% of the university

Scholarship for Outstanding Students, Shandong University 2017 – 2019
~CNY 6,000 total; awarded to the top 15% of the university

Merit Student, School of Foreign Languages, Shandong University 2017 – 2019
Awarded to the top 20% of the college

LANGUAGES & SKILLS

Languages	Chinese (native), English (advanced), French (conversational), Dutch & German (basic)
Programming	Python, Java, JavaScript, LaTeX, R, Prolog, SQL
Tools	PyTorch, Hugging Face Transformers, pandas, scikit-learn, wandb, Linux, Git, CUDA