Siyu Liang

liangsy@uw.edu · siyuliang.github.io · +1 2022627554
3940 Benton Ln NE, Room 407, Seattle, WA 98105, United States

Education

2021 – Present   **University of Washington** – Seattle, WA
Ph.D. in Computational Linguistics
Advisors: Drs. Gina-Anne Levow & Richard Wright

2025   **University of Washington** – Seattle, WA
M.S. in Computational Linguistics

2020   **Georgetown University** – Washington, DC
B.A. in Linguistics with minors in Persian & Computer Science
Advisor: Dr. Elizabeth Zsiga

Publications

Siyu Liang, Nicolas Ballier, Gina-Anne Levow, Richard Wright. 2026 (accepted). The Limits of Data Scaling: Sub-token Utilization and Acoustic Saturation in Multilingual ASR. In *Proceedings of the 15th International Conference on Language Resources and Evaluation (LREC 2026)*, European Language Resources Association (ELRA), Palma de Mallorca, Spain.

Michael Scott, **Siyu Liang**, Alicia Wassink, Gina-Anne Levow. 2026 (accepted). A Sociophonetic Analysis of Racial Bias in Commercial ASR Systems Using the Pacific Northwest English Corpus. In *Proceedings of the 15th International Conference on Language Resources and Evaluation (LREC 2026)*, European Language Resources Association (ELRA), Palma de Mallorca, Spain.

Siyu Liang, Talant Mawkanuli, Gina-Anne Levow. 2026 (accepted). Hybrid Neural-LLM Pipeline for Morphological Glossing in Endangered Language Documentation: A Case Study of Jungar Tuvan. In *Proceedings of the 5th Workshop on NLP Applications to Field Linguistics (FieldMatters)*, Association for Computational Linguistics, Rabat, Morocco.

Siyu Liang, Zhaxi Zerong. 2026 (accepted). The Tonogenesis Continuum in Tibetan: A Computational Investigation. In *Proceedings of the 6th International Workshop on Computational Approaches to Language Change (LChange)*, Association for Computational Linguistics, Rabat, Morocco.

Siyu Liang, Nicolas Ballier, Gina-Anne Levow, Richard Wright. 2025. Beyond WER: Probing Whisper's Sub-token Decoder Across Diverse Language Resource Levels. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Suzhou, China. SAC HIGHLIGHT AWARD

Siyu Liang, Gina-Anne Levow. 2025. Breaking the Transcription Bottleneck: Fine-tuning ASR Models for Extremely Low-Resource Fieldwork Languages . In *Proceedings of the 4th Workshop on NLP Applications to Field Linguistics (FieldMatters)*, Association for Computational Linguistics, Vienna, Austria.

Siyu Liang, Gina-Anne Levow. 2025. Tone in Perspective: A Computational Typological Analysis of Tone Function in ASR. In *Proceedings of the 7th Workshop on Research in Computational Linguistic Typology and Multilingual NLP (SIGTYP)*, Association for Computational Linguistics, Vienna, Austria.

Siyu Liang. 2020. Documenting Eynu: A Case Study of Language Contact. In *Proceedings of the 43rd Annual Penn Linguistics Conference*, Philedephia, USA.

Other Presentations

Siyu Liang. 2025. Acoustic Correlates of Performed Sexual Orientation in Thai Media. In *New Ways of Analysing Variation Asia Pacific (NWAV-AP 8)*, Singapore. (Oral presentation)

Siyu Liang. 2023. A Phonological Account of Intervocalic Voicing in Yangon Burmese. In *the 56th International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL 56)*. (Oral presentation)

Siyu Liang. 2021. Incomplete Vowel Assimilation in Tehrani Persian. In *the 4th Conference of Central Asian Language and Linguistics (ConCALL-4)*. (Oral presentation)

Other Projects

September 2024 - Present	SocioLinx: PI: Dr. Alicia Beckford Wassink (University of Washington) A large-scale sociophonetic analysis of ASR errors for speech by underrepresented regional and ethnic groups of English speakers in the US. Designing linguistically informed benchmark for ASR models based on proprietary dataset. Developing speech models to handle dialectical variation in English.
September 2023 - Present	TurkicNLP at UW PI: Dr. Talant Mawkanuli (University of Washington) Building multilingual speaking dictionaries for Turkic languages. Providing mentorship for undergraduate and graduate students on data annotation, NLP pipeline, and digital humanities practices.

Teaching

Spring 2026	LING 471: Computational Methods for Linguists, University of Washington (Instructor of Record)
Winter 2024	LING 572: Advanced Statistical Methods in Natural Language Processing, University of Washington (Teaching Assistant)
Autumn 2023 Autumn 2025	LING 570: Shallow Processing Techniques for Natural Language Processing, University of Washington (Teaching Assistant)
Spring 2023	LING 471: Computational Methods for Linguists, University of Washington (Teaching Assistant, guest lecture on Statistics Basics for NLP)
Spring 2022 Autumn 2021	LING 200: Introduction to Linguistics, University of Washington (Teaching Assistant, with teaching duties twice weekly)
Winter 2022	LING 461: Syntax I (Grader)

Industry experience

07-11.2020	**ByteDance** – Beijing, China NLP Engineer Intern Building a FrameNet-based intent framework for large-scale dialog systems. Developing an extendable symbol set for multilingual NLP tasks.

Honors and scholarships

2025   GPSS Travel Award ($500, University of Washington)

2024   Imre Boba Summer Research Fellowship ($1000, University of Washington)

2021   ConCALL-4 Travel Award ($500)

2019   Royden B. Davis Fellowship ($5000, Georgetown University)

2018   Lisa J. Raines Fellowship ($5000, Georgetown University)

Talks and tutorials

2025   Analyzing Whisper Across Diverse Language Resource Levels
Workshop on Promoting Fairness for Under-represented Languages in Multilingual LLMs, May 26 2025, Université Paris Cité and Online

2025   Languages of the Silk Road (Guest lecture)
MELC 268: Silk Road and Globalization, Spring 2025, University of Washington

2024   Tibetan Music (Guest lecture)
MELC 269: Music Cultures of the Silk Road, Winter 2024, University of Washington

Mentorship

2024 – 2025   Ezra Mason, Linguistics Undergraduate Research Apprenticeship (LURAP) on *Automatic Segmentation of Tuvan*

2024 – 2025   Melody Vahadi, M.S. project on *OCR for Chaghatay*

Professional Service

Reviewer, *LREC 2026, ICML 2026*

Service

2025   Editor, *Proceedings of the 43rd West Coast Conference on Formal Linguistics*

Technical skills

Programming languages
Proficient: Python
Familiar: C++, Java, JavaScript, R

Software
ELAN, Praat

Natural Languages
Proficient: Mandarin, English
Advanced: Persian, French (reading)
Intermediate: Thai, Tibetan, Vietnamese, Indonesian, Japanese
Elementary: Burmese, Uyghur, Uzbek, Arabic, Amharic, Turkish, Russian, Portuguese, Swahili
Experience: Chaghatay, Nepali, Mongolian, etc.
Fieldwork: Nobiin, Eynu, Persian, Tibetan, Tuvan

Last updated: 02.2026