Siyu Liang

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

Research interests

My research interests are computational linguistics, phonetics, field work, as well as historical linguistics and typology. I am particularly interested in low-resource speech NLP, language technology, and quantitative modeling of speech.

Education

2021 – Present	**University of Washington** – Seattle, WA Ph.D. in Computational Linguistics Advisors: Professors 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: Professor Elizabeth Zsiga

Publications

Siyu Liang, Nicolas Ballier, Gina-Anne Levow, Richard Wright. 2025 (accepted). 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.

Siyu Liang, Gina-Anne Levow. 2025. Breaking the Transcription Bottleneck: Fine-tuning ASR Models for Extremely Low-Resource Fieldwork Languages . In *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 *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*.

Other Presentations

Trent Ukasick, **Siyu Liang**. 2025 (accepted). Interaction of Laryngeal Contrast and f_0 in Gyegu Tibetan: Acoustic Evidence for a Fortis–lenis Split. In *Acoustical Society of America (189th ASA/ASJ Joint Meeting)*, Honolulu, Hawaii, USA. (Poster presentation)

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 *56th International Conference on Sino-Tibetan Languages and Linguistics (ICSTLL 56)*. (Oral presentation)

Siyu Liang. 2021. Incomplete Vowel Assimilation in Tehrani Persian. In *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. Benchmarking ASR models based on proprietary dataset.
September 2023 - Present	TurkicNLP at UW
	PI: Dr. Talant Mawkanuli (University of Washington)
	Building multilingual speaking dictionaries for Turkic languages. Providing mentorship on data annotation, NLP pipeline, and digital humanities practices.

Teaching

Spring 2026	LING 471: Computational Methods for Linguists, University of Washington (Instructor of Record, scheduled)
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

Summer 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

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*

Service

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

2023 – 2024	Treasurer, *Linguistic Society of the University of Washington (LSUW)*

Technical skills

Programming languages
Proficient in: Python.
Familiar with: 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: Nobiin, Chaghatay, Tuvan, Nepali, Mongolian, etc.