

Kaveri K. Sheth

Updated June 12, 2026

Postdoctoral Researcher

Laboratoire de Sciences Cognitives et Psycholinguistique · ENS, EHESS, CNRS, PSL University · Paris, France

ksheth2019@gmail.com

Research Overview

Computational linguist and developmental language researcher working at the intersection of AI and the science of child language learning. Expertise in how infants acquire language from sparse, noisy, multi-modal input; in building speech-processing pipelines and benchmarks from long-form, ecological audio recordings across typologically diverse languages; and in open-science tooling for the research community. Publishing in both machine learning venues (Interspeech) and high-impact cognitive science journals.

Education

University of Washington, Seattle, WA

Sept 2020 – June 2025

PhD in Linguistics. Advisors: Drs. Naja Ferjan Ramírez, Patricia Kuhl, and Gina-Anne Levow

University of Washington, Seattle, WA

Sept 2020 – June 2024

MS in Computational Linguistics.

Emory University, Atlanta, GA

Aug 2015 – May 2019

BA in Psychology and Linguistics; BA in Spanish. *Summa cum laude*

Professional Experience

Postdoctoral Researcher, École Normale Supérieure, CNRS, Paris

Oct 2025 – Present

- Conduct research on language development across diverse cultures and languages using data science, automatic speech recognition, and open-science methods for reproducibility on large-scale, long-form naturalistic recordings.
- Lead development and cross-linguistic validation of a rule-based system for detecting child-directed vs. overheard speech in long-form recordings—directly addressing how to characterize the ecological input children actually learn from.
- Co-developed context-aware child-directed speech detection models and a framework for deriving machine learning benchmarks from long-form recordings (both published at Interspeech 2026), enabling principled comparison of human and machine learning from similar input.

Chief Executive Officer, ELSI (nonprofit startup)

Jan 2025 – Present

- Lead operations of a nonprofit sustaining multimodal and open research infrastructure for the language development community: funding, hiring, project management, and client relations.
- Convey the goals and subsequent research utilizing ELSI to non-technical audiences to secure funding

**Doctoral Researcher, Language Development & Processing Lab
and I-LABS, University of Washington**

Sept 2020 – June 2025

- Led longitudinal corpus analyses of infant-directed speech (parentese), quantifying how the quantity, syntactic, and lexical properties of caregiver input shape early language outcomes — published in *Language Learning and Development* and *Journal of Child Language*.
- Studied second-language learning interventions in early education settings, isolating effects of age, dosage, and socioeconomic status on learning from limited input.

Computational Linguist Intern, Smart Information Flows Technologies (SIFT) Jan – Mar 2023

- Supported human-subjects research on DARPA-style projects (HUGO, VICTOR) at the intersection of language technology and human behavior.

Data Science Intern, ReThink Media

June – Sept 2022

- Built and debugged a Python NLP pipeline (spaCy, NeuralCoref, pandas, NumPy) applying deep learning and machine learning techniques to track gendered citation patterns in news media at scale.

Publications & Manuscripts

Context-aware child-directed speech detection from long-form recordings.

Charlot, T., Kunze, T., **Sheth, K.K.** & Cristia, A. *Interspeech*. 2026.

Deriving benchmarking from long-form recordings: Challenges and opportunities.

Sheth, K.K., Borst, L., Kunze, T., Levachin, M., Rassenen, O., Tsuji, S., & Cristia, A. *Interspeech*. 2026.

From academic tool to community infrastructure: A call for Indigenous partnership in speech data governance.

Sheth, K.K. & Christian, S. *Interspeech*. 2026.

A rule-based approach for identifying child-directed and overheard speech in long-form recordings: Validation across typologically diverse datasets.

Sheth, K.K., Tey, K., Borst, L., Bergelson, E., Casillas, M., Rowland, C., Soderstrom, M., *et al.* *In Preparation*.

A longitudinal corpus analysis of the quantity, syntactic and lexical characteristics of maternal and paternal parentese.

Sheth, K.K. & Ferjan Ramírez, N. *Language Learning and Development*., 2025.

Parentese in infancy predicts 5-year language complexity and conversational turns.

Ferjan Ramírez, N., Weiss, Y., **Sheth, K.K.**, & Kuhl, P.K. *Journal of Child Language*., 2023.

Second language learning in infant education centers in Madrid: Effects of age, dosage, and socioeconomic status.

Ferjan Ramírez, N., Weiss, Y., **Sheth, K.K.**, & Kuhl, P.K. *Int. J. Environmental Research and Public Health*. 2023.

Conference Talks and Posters

A rule-based approach for identifying key child-directed speech and overheard speech in long-form recordings: Validation and performance across typologically diverse datasets.

Sheth, K.K., Tey, K., Borst, L., Bergelson, E., Bourree, A., Casillas, M., Rowland, C., Scaff, C., Soderstrom,

M. & Cristia, A.

Talk to be presented at the International Congress for Infant Studies (ICIS). Panama City, Panama. (July 2026).

Paternal and Maternal Parentese: A Longitudinal Corpus Analysis.

Sheth, K.K. & Ferjan Ramirez, N.

Poster presented at the International Congress for Infant Studies (ICIS). Ottawa, Canada. (July 2022).

Paternal and Maternal Parentese and Infant Vocabularies, A Longitudinal Corpus Analysis.

Sheth, K.K. & Ferjan Ramírez, N.

Poster presented at the International Congress for Infant Studies (ICIS). Talk presented at the Boston University Conference on Language Development (BUCLD), Boston, MA. (November, 2021).

Teaching and Mentorship

Teaching Assistant, University of Washington

2020 – 2022

- LING 200 (Introduction to Linguistic Thought)
- LING 580 (Naturalistic Audio Recordings of Child Language)
- LING 233 (Language and Society)

Undergraduate Research Mentor, University of Washington

2020 – 2025

- Supervised and mentored 5 undergraduate volunteers in the Language Development and Processing Lab (June-Sept 2021)
- Supervised and mentored 10 undergraduate volunteers in the Language Development and Processing Lab (Sept 2021-June 2022)
- Supervised Research Assistants and Interns at the Institute for Learning and Brain Sciences (Sept 2022-June 2025)

Honors & Scholarships

Paula Menyuk Award, BUCLD

2021

Top Scholar Award for the Recruitment of Outstanding Graduate Students, UW

2020

Independent Research Grant, Emory University

2018

Skills

Programming: R; Python; PyTorch.

Languages: English (native); Spanish (B1); French (A2)