Research Summary

Alex Hernandez Juarez

January 15, 2026

Mathematical Linguistics Framework for Real-Time ASL Translation

Overview

I am developing a mathematically rigorous framework for large-vocabulary American Sign Language (ASL) translation, designed to scale to 5,000–10,000 signs while maintaining real-time inference (<200ms latency) on edge devices. This work bridges formal linguistics, machine learning theory, and practical systems engineering.

Problem & Motivation

Current sign language recognition systems face fundamental limitations:

- Treating signs as atomic units prevents compositional generalization
- Ignoring ASL's spatial grammar loses critical referential information
- Black-box neural approaches lack theoretical guarantees for scaling

Technical Contributions

1. Phonological Factorization with Geometric Invariance

I formalize ASL signs as tuples $s = (H, L, O, M, N) \in \Sigma_H \times \Sigma_L \times \Sigma_O \times \Sigma_M \times \Sigma_N$ representing handshape, location, orientation, movement, and non-manual markers. The observation map $\phi : (\mathbb{R}^3)^m \to \mathbb{R}^k$ is designed to be Sim(3)-equivariant, ensuring robustness to signer position and camera angle. Product vector quantization reduces sample complexity from $O(\sqrt{d/n})$ to $O(\sum_j \sqrt{d_j \log k_j / n})$.

2. Spatial Discourse Algebra

ASL grammar relies on spatial variable binding—signers assign referents to locations (loci) in signing space. I develop a probabilistic retrieval framework:

$$p(r \mid g(t)) \propto \exp(-\alpha \angle(g(t), \hat{\ell}_r)) \cdot p(r \mid \text{context})$$

with formal uniqueness guarantees (Lemma 1: angular separation $> 2\tau$ implies $|\Gamma_t(\tau)| \leq 1$) and a uniformly most powerful test for locus assignment vs. retrieval.

3. Non-Associative Morphological Fusion

ASL verbs undergo argument-driven modification (e.g., EAT-TACO). I define a fusion operator $\otimes : \Sigma \times \Sigma \to \Sigma$ and prove it is necessarily non-associative under role-sensitive conditions (Proposition 5), capturing linguistic constraints that sequential models miss.

4. WFST Decoding Cascade

I adapt weighted finite-state transducer composition from speech recognition:

$$H \circ C \circ M \circ D \circ L \circ G$$

where H models observations, C handles coarticulation, M performs morphological fusion, D tracks discourse state, L is the lexicon, and G is a language model. I prove soundness and completeness of the composed transducer and analyze beam search complexity.

5. Information-Theoretic Multimodal Integration

Using mutual information decomposition $I(U;Y) = I(H;Y) + I(F;Y|H) + I(B;Y|H,F)$ and Fano-type bounds, I provide principled guidance for modality selection based on accuracy-cost tradeoffs.

Implementation Status

I have implemented the core pipeline in PyTorch, including MediaPipe-based feature extraction, a BiLSTM encoder with CTC loss, and a three-stage training curriculum. Current work focuses on validating the system on WLASL data before scaling to the full architecture.

Skills Demonstrated

- **Mathematical modeling**: Formal definitions, propositions, and proofs grounded in group theory, probability, and formal language theory
- **Machine learning**: CTC training, multi-task learning, convergence analysis
- **Systems engineering**: Edge deployment constraints, latency optimization, WFST composition
- **Interdisciplinary synthesis**: Integrating ASL linguistics (Stokoe, Liddell, Padden) with computational methods

Research Interests

I am broadly interested in structured approaches to sequence modeling, computational linguistics, and machine learning systems that combine theoretical rigor with practical deployment. I would welcome the opportunity to contribute to research in these areas.

Full paper and implementation available upon request.