

Amos Teo

Linguist. Language Nerd. Data Analyst.

Email: amosbteo@gmail.com

Website: www.amosbteo.com

WORK HISTORY

Language Data Researcher, Amazon	2019-present
Graduate Employee, University of Oregon	2013-2019

PROJECTS & EXPERIENCE

- Alexa Arabic language launch** **2021**
- Improved feature performance to the point of launch-readiness.
 - Conducted deep dive analyses to identify bugs & language patterns to support.
 - Produced Arabic resource materials for team members to use for reference.
 - Managed data volumes for annotation and monitored annotated data quality.
- Linguistic clustering analysis for quantifying differences between languages** **2015-2017**
- Developed a corpus-based method for measuring cross-linguistic typological distances in terms of case marking.
 - Used cluster analysis and dimensionality reduction techniques.
 - Presented results at the 12th Association of Linguistic Typology meeting (Dec 2017).
- Spoken Sümi corpus creation (Nagaland, India)** **2010-2018**
- Recorded 11 hours of speech: 7 hours manually transcribed in Sümi orthography and translated into English.
 - Managed small team of language workers in India; Provided feedback on transcription and translation.
 - Funded by Endangered Languages Documentation Programme (ELDP) and National Science Foundation.
 - Data transcriptions and translations available online at Endangered Language Archive (ELAR).

EDUCATION

- University of Oregon** **2013-2019**
PhD in Linguistics
- Dissertation: "Investigating differential case marking in Sümi using language documentation & experimental methods"
 - Funded by the National Science Foundation (DEL-DDRIG 1723519); UO Doctoral Research Fellowship.
 - Modeled speakers' choice to use probabilistic case marking in an under-resourced language. Skills used: corpus tagging of semantic and pragmatic features; experimental design for speech perception; random forest/recursive partitioning; linear mixed effects models; data visualization; qualitative analysis.
- University of Melbourne** **2008-2010**
MA in Linguistics
- Thesis (published): "A phonological and phonetic description of Sumi, a Tibeto-Burman language of Nagaland"
 - Skills used: fieldwork; compilation of spoken word corpus; analysis of acoustic characteristics of speech, focusing on lexical tone; data visualization.

SKILLS

Technical skills and software:

- Database query & Statistical analysis: SQL, R (e.g. tidy, dplyr, lme4); Python (e.g. numpy, pandas).
- Data visualization: R (ggplot2, ggmap); Python (matplotlib).
- Linguistic data collection and annotation: Praat, EMU Speech Database System; ELAN Annotator (transcription & translation); Toolbox, FLEx (lexical database & morphological parsing); PsychoPy (experiments).

Natural language competencies:

- **English** (N. American & Singapore): native
- **Mandarin**: excellent reading/written/spoken/listening
- **French**: excellent reading/written/spoken/listening
- **Russian**: good reading/written; basic spoken
- **Nepali**: basic reading/written/spoken/listening
- **Assamese**: basic reading/written/spoken/listening

ENGAGEMENT & KNOWLEDGE SHARING

- Talk: "How to construct a language" (3 May 2019) at FLIS Day for high school students. With Charlotte Vaughn.
- UO Linguistics blog posts: [Linguistics in pop culture: Arrival \(film\)](#), [Linguistics in the news – Election season](#).