

ANNA BJÖRKLUND

Bay Area, CA
(407)-922-4943
aebjorklund@gmail.com

aebjorklund.github.io
github.com/aebjorklund
linkedin.com/in/aebjorklund

EXPERIENCE

- 2024 – present **PhD dissertation: Methods in low-resource language revitalization** **UC Berkeley**
Build fine-tuned multilingual vs from-scratch TTS models for low-resource indigenous language Wintu (Fastspeech2, PyTorch) • Collect, transcribe, align, & preprocess archival Wintu speech data for model use • Train custom Montreal Forced Aligner model for model alignment, incl. scripting for OCR/OCR post-processing and pronunciation dictionary creation (Python)
- 2023 – present **Senior Data Science Fellow/Data Science Fellow** **UC Berkeley D-Lab**
Design and lead workshops for deep learning (vanilla NNs, CNNs) with Keras, machine learning (scikit-learn), data wrangling & visualization (Python, R), web scraping, and web APIs • Provide technical and quantitative support for cross-functional social science researchers
- 2020 – present **Tribal linguist** **Paskenta Band of Nomlaki Indians/Wintu Tribe of Northern California**
Analyze all linguistic aspects of indigenous Californian languages Nomlaki and Wintu (phonetics, morpho-syntax, pragmatics, etc) from archival materials (text & audio) • Leverage linguistic analysis to create curriculum for/teach regular language revitalization classes for tribal members • Identify gaps in linguistic knowledge/materials and work as a team with tribal members to develop solutions that serve revitalization needs • Clearly communicate linguistic topics to learners with diverse levels of experience • Work with tribal leadership to identify best research practices
- 2023 **Data Science for Social Justice Fellow** **UC Berkeley D-Lab**
Conducted NLP analysis (e.g. bag-of-words, TF-IDF, sentiment analysis, word embedding) on Reddit (r/languagelearning) data. Found significant user association between non-native speech accent and negative sentiment; recommended community guidelines centering comprehension over nativism to subreddit moderators
- 2022 **Graduate Student Researcher** **California Language Archive (CLA)**
Organized and catalogued archival documents across native Californian languages for public reference • Connected visiting Californian tribal groups to relevant diverse linguistic material; worked with tribal groups to interpret archival materials and reclaim them for indigenous use

EDUCATION

- 2019 – present **PhD candidate in linguistics** **University of California, Berkeley**
Dissertation (in progress): Methods in low-resource language revitalization
- 2021 **MA linguistics** **University of California, Berkeley**
Thesis: Nomlaki Vowel Quality and Duration: An Archival Examination. Collected, transcribed, and hand-aligned Nomlaki data from archival materials. Designed and conducted statistical and computational analysis (k-means clustering, LDA, etc) of data using scikit-learn, rstatix, etc.
- 2015 – 2019 **BA linguistics; BA Chinese** **University of Florida**
Thesis: Perceptual Bias In Lexical Tone Perception. Collected, designed and conducted statistical and computational analysis of Chinese Mandarin experimental data (SPSS)

LINGUISTIC & RESEARCH SKILLS

- Speech transcription & annotation
- Data wrangling & manipulation
- Statistical & text analysis
- Experimental design
- Hypothesis testing

COMPUTATIONAL LANGUAGES

- **Python**: scikit-learn, SciPy, matplotlib, parselmouth
- **Praat, ELAN**
- **Keras, Pytorch**
- **R**: tidyverse, rstatix
- **SQL**

NATURAL LANGUAGES

English: native
Chinese Mandarin: ~HSK 5
Swedish: heritage
Nomlaki/Wintu: working knowledge