Raksit T. Lau-Preechathammarach

raksittp@gmail.com https://rakxsit.github.io/
https://www.linkedin.com/in/raksitlp/ https://github.com/rakxsit/

Education

Ph.D. in Linguistics (expected August 2022); University of California, Berkeley 2015-present
M.A. in Linguistics; Harvard University 2013-2015
B.A. in Linguistics; Yale University 2007-2012

Summary

Ph.D linguist with 7+ years experience applying data science, machine learning, and statistical methods to solve questions about language. Specializes in spoken language processing, acoustics, language change, and multilingualism. Proficient in Python, R, and command line scripting. Strong communicator and technical writer. Independently motivated and passionate about collaboration.

Skills

Data machine learning, annotation schema creation & implementation, data analysis & visualization, corpus analysis, statistical modeling, webscraping, experiment design, language documentation, survey methodology, audio recording

Technical Python, SQL, R, Jupyter, Regex, Praat, OpenSesame, Montreal Forced Aligner, ELAN, LaTeX, Command line, Git, XML, HTML, CSS, Javascript/JSON, FLEx

Statistical Hypothesis testing, dimensionality reduction (principal component analysis, linear discriminant analysis), linear & logistic mixed-effects regression models, classification algorithms, correlation analysis, descriptive statistics, t-test

Languages
Native: English, Thai
Advanced: Cantonese, Japanese, Korean, Mandarin, Spanish, French
Basic: Arabic, Khmer, Vietnamese
Fieldwork: Kuy, Yaeyaman, Dunan, So, Thavung

Experience

Doctoral Researcher | UC Berkeley 2015-present

- Employed webscraping of APIs and machine learning algorithms to create and analyze corpora of high- & low-resource languages, resulting in 3 journal articles and 5 conference presentations
- Applied neural networks to carry out computational simulations of language change, resulting in 1 journal article and 2 conference presentations
- Scripted synthetic speech and normalized using principles of interrater reliability to create experimental stimuli for a project resulting in 2 journal articles and 5 conference presentations
- Designed survey instruments and experiments, taking into account accessibility and usability principles, with participants in rural Thailand who had no familiarity with experimental settings, and analyzed using dimensionality reduction methods to interpret highly covarying data
- Developed annotation pipeline to expedite audio file segmentation and manual transcription of endangered languages, leveraging forced alignment, command line tools, and Python scripting
- Communicated results from model and experimental data to wider scientific and non-scientific audiences through 8 journal articles, 2 book chapters, and 13 conference presentations
- Collected data from 5 endangered languages, outlined annotation principles, and created speech corpora and databases for analysis of low-resource languages

- ❖ Coordinated with various stakeholders to create and pilot website for revitalization of an endangered language and tools for language learning
- ❖ Managed and trained 9 teams of 23 research assistants in annotation, scripting, and data analysis and visualization to create an endangered language speech corpus and a conference presentation
- ❖ Collaborated on 7 cross-functional teams to assess problems and implement solutions

Course Designer and Instructor | UC Berkeley 2016-present

- ❖ Developed course and assessment materials and evaluated student learning for 8 courses
- ❖ Designed curricula for 2 courses and led discussion sections for 6 courses
- ❖ Guided students in data analysis and original research design using language data

Freelance Translator 2010-2015

- ❖ Translated 2 test preparation books for TOPIK (Test of Proficiency in Korean) and 2 Korean language textbooks from Korean into English for Darakwon Publishing Company
- ❖ Translated 2 Korean language textbooks from Korean into English for Sogang University Korean Language Education Center

Research Assistant 2009

- ❖ Developed and implemented data processing pipeline for transcribing unlabeled infant babbles for phonetic analysis with Haskins Laboratories
- ❖ Analyzed infants' eye movements to determine their understanding of communication and language with NYU Infant Cognition and Communication Lab