Raksit T. Lau-Preechathammarach

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

Summary

Ph.D linguist with expertise applying data science, machine learning, and statistical methods to analyze language. Proficient in Python, R, and command line tools. Independently motivated, strong communicator, and passionate about continually learning new skills. Specific strengths include:

- Development and implementation of predictive models to study populations of individuals
- Scripting for data wrangling, machine learning and statistical analysis, and data visualization
- Management of research teams in developing annotation pipelines for databases

Education

University of California Berkeley, Ph.D., Linguistics — expected August 2022
Harvard University, M.A., Linguistics — 2015
Yale University, B.A., Linguistics — 2012

Selected projects and roles

Speech perception and production experiments — UC Berkeley, 2018-present

- Scripted synthetic speech in Praat to create stimuli for speech perception & production studies
- Addressed issues of variable rating quality by using principles of interrater reliability
- Designed survey instruments and experiments, taking into account accessibility and usability principles, with participants in rural Thailand who had no familiarity with experimental settings
- Leveraged machine learning methods in Python and R to interpret highly covarying data
- Demonstrated significant effects of bilingualism on language change
- Published in *Linguistics Vanguard* (2022) [link], manuscript submitted to *Laboratory Phonology* (2022) [link], and presented results at 5 conference venues

Neural net modeling of language change — Harvard | UC Berkeley, 2014-2018

- Employed webscraping of APIs to create corpus for modeling language change
- Trained neural networks in Python to carry out simulations of language change
- Yielded results showing that the state of Latin was predictive of changes that would occur
- Published in *Proceedings of the 28th Annual UCLA Indo-European Conference* (2018) [link] and presented results at 3 conference venues

Data collection & analysis of endangered languages — Harvard | UC Berkeley, 2013-present

- Collected data from and created speech corpora and databases for 5 endangered languages
- Leveraged forced alignment, command line tools, and Python scripting to develop annotation pipeline to expedite audio file segmentation and manual transcription
- Coordinated with various stakeholders to create and pilot website for revitalization of an endangered language and tools for language learning
- Carried out acoustic analyses and analyzed underlying processes in language change
- Managed and trained 9 teams of 23 research assistants in annotation, scripting, and data analysis, resulting in 7 journal articles, 2 book articles, & 9 conference presentations [CV]

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

Skills

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

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

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

Languages: *Native:* English, Thai; *Advanced:* Cantonese, Japanese, Korean, Mandarin, Spanish, French