Zinan Liang 梁梓楠 (Tsinam Leung)

liangzinan21@hotmail.com | zinan.liang@u.nus.edu | +65 85382257
Website: tsinamleung.github.io ORCID: 0009-0003-0226-0412

EDUCATION

National University of Singapore | Singapore 2025 – Present

MA (Research) in Linguistics

Thesis: *The Sketch Grammar of Malaysian Lui Chiew (Leizhou Min)*

Supervisor: Nala Lee

Chengyi College, Jimei University | Xiamen, Fujian, PRC June 2021

BEng, Network Engineering

RESEARCH INTERESTS

- Digital humanities approaches to language documentation and revitalization.
- Exploring and adapting language technologies for low-resource languages.
- Typological variation and language contact in Southern Sinitic and Southeast Asian languages.

RESEARCH EXPERIENCE

Malaysian Lui Chiew (Leizhou Min) 2025 – Present

MA Research and Fieldwork | National University of Singapore

- Conducted fieldwork in Tiang Dua, Melaka, and Muar, Johor, in July 2026 with 11 consultants, collecting wordlists, elicited sentences, narratives and conversation.
- Prepared a preliminary description of phonology, word classes and syntax, recording variation among the consultants.
- Used ELAN for time-aligned annotation and Praat for acoustic analysis of tones and rhyme duration.
- Compared recorded patterns with Leizhou Min, Hainan Min and Malaysian Hokkien descriptions to assess regional variation and possible contact explanations.

TypeDuck (Cantonese IME with Multilingual Dictionary) Sept 2022 – Dec 2024

Developer / Research Assistant | PI: Dr. Chaak Ming Lau, EduHK

- Developed a Cantonese IME and multilingual dictionary to support Cantonese learners from diverse ethnic backgrounds.
- Built and deployed the IME algorithm server and implemented the dictionary interface.
- Filtered obsolete or rare terms from the IME dictionary.

Chinese Newspaper Corpus System October 2020

Developer | PI: Prof. Sze Wing Tang, CUHK

- Developed a scraping and data-extraction system for online Chinese newspapers across Asian countries.

ACADEMIC & OPEN-SOURCE CONTRIBUTIONS

Cantonese Computational Linguistics Infrastructure Development Workgroup (CanCLID) 2020 – Present

Core Contributor

- Collected and curated Guangzhou Cantonese text, and trained a language model on Guangzhou and Hong Kong corpora to build a Cantonese/Mandarin classifier.
- Led the localization of Mozilla Common Voice into Cantonese, including UI translation, corpus collection/refinement, and audio recording verification.
- Worked on the Cantonese IME algorithm and word prediction.

Lingnaam Jyutjam (嶺南粵音) 2017 – Present

Core Contributor

- Developed Jyutdict.org, a Pan-Cantonese Dictionary web application recording Yue pronunciations across Guangdong and Guangxi.
- Participated in discussions of Yue classification and Early-Qing Cantonese phonology.
- Collected and analysed colloquial Cantonese vocabulary and used Python to parse and filter transcripts for dictionary work.
- Implemented tone-sandhi transcription for Jyutping++, an extended romanization scheme for Yue varieties used in dictionaries and input methods, including Bakhoi and Hamzau resources.

PUBLICATIONS & PRESENTATIONS

Liang, Zinan and Ching-Hei Yim. (June 2026). "A typology of 'finger' as 'little hand': a shared morphological strategy in Western Yue-speaking Region." *32nd Annual Conference of the International Association of Chinese Linguistics (IACL-32)*, Guangzhou.

Tsinam Leung. (June 2025). "The Establishing Function of "箇" (ge) in Southern Chinese Dialects." *The Eighth Forum on Cantonese Linguistics (FoCaL-8)*, The University of Hong Kong.

Liang, Zinan and Huang, Junxin. (November 2024). "Man1 in Ngzhau Cantonese: A Study of Language Contact between Yuehai and Goulou Yue." *Buckeye East Asian Linguistics*, vol. 9, p. 163-179. (Presented at *The 27th International Conference on Yue Dialect*, Sept. 2024).

HUANG Junxin, **LIANG Zinan**. (June 26, 2021). "The Construction and Implementation of the Jyutping++." *The 21st Workshop on Cantonese*, Hong Kong.

COMPUTATIONAL METHODS & DATA SKILLS

- **Data & NLP Workflows:** Corpus collection, web scraping, data cleaning, parsing, metadata organization, and language-model-based classification for Cantonese/Mandarin data.
- **Programming & Tools:** C++, Python (pandas, regular expressions), PHP/MySQL, Git/GitHub.
- **Linguistics Tools:** ELAN (time-aligned annotation), Praat (acoustic analysis), lexical data preparation for dictionaries and input methods.
- **ASR workflow:** Used speech recognition to locate passages, then checked and transcribed cited forms by listening to the original recordings.
- **Languages:** Cantonese, Dianbai Min (native); Mandarin (fluent); English (professional); Taiwan Hokkien (conversational); Malay, Indonesian, Japanese (basic).