

Haruki Sakajo

✉ sakajoharuki@gmail.com

🌐 <https://sj4x.net/>

Employment History

Research Assistant, The National Institute of Advanced Industrial Science and Technology.

2025.07 – present

- Interpreting and analyzing Large-scale Vision Language Models based on internal representations.

Researcher, Nara Institute of Science and Technology.

2025.02 – 2025.06

- Developed a domain-specific lightweight vision language model to improve cost efficiency.

Software Engineer, Nagase Brothers Inc.

2020.06 – present

Nagase Brothers Inc. is an education-focused company that provides services, including online feedback and correction support for high school students.

- Led a front-end team of 4 engineers and designed applications.
- Improved maintainability and reduced long-term costs by migrating existing applications to React, a modern and stable framework.
- Integrated multiple legacy applications into a unified interface, reducing manual workload.

Education

PhD, Nara Institute of Science and Technology.

2025.10 — present

M.E. Computer Science, Nara Institute of Science and Technology.

2024.04 – 2025.09

- Thesis title: Exploring Language Models' Understanding of Vowel Articulation: From a Visual Information Perspective

B.A. Linguistics, Kyoto University.

2020.04 – 2024.03

- Thesis title: Morphological Analysis of Manchu Based on Grammatical Forms and Phonology.

Research Publications

Conference Proceedings

- [1] **Haruki Sakajo**, Yusuke Ide, Justin Vasselli, Yusuke Sakai, Yingtao Tian, Hidetaka Kamigaito, and Taro Watanabe. “Dictionaries to the Rescue: Cross-Lingual Vocabulary Transfer for Low-Resource Languages Using Bilingual Dictionaries”. In: *Findings of the Association for Computational Linguistics: ACL 2025*. Ed. by Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar. Vienna, Austria: Association for Computational Linguistics, July 2025, pp. 25963–25976. ISBN: 979-8-89176-256-5.  URL: <https://aclanthology.org/2025.findings-acl.1333/>.

- [2] **Haruki Sakajo**, Yusuke Sakai, Hidetaka Kamigaito, and Taro Watanabe. “Tonguescape: Exploring Language Models Understanding of Vowel Articulation”. In: *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Ed. by Luis Chiruzzo, Alan Ritter, and Lu Wang. Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 12605–12619. ISBN: 979-8-89176-189-6.  URL: <https://aclanthology.org/2025.naacl-long.627/>.
- [3] Justin Vasselli, **Haruki Sakajo**, Arturo Martínez Peguero, Frederikus Hudi, and Taro Watanabe. “Leveraging Dictionaries and Grammar Rules for the Creation of Educational Materials for Indigenous Languages”. In: *Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP)*. Ed. by Manuel Mager, Abteen Ebrahimi, Robert Pugh, Shruti Rijhwani, Katharina Von Der Wense, Luis Chiruzzo, Rolando Coto-Solano, and Arturo Oncevay. Albuquerque, New Mexico: Association for Computational Linguistics, May 2025, pp. 112–118. ISBN: 979-8-89176-236-7.  URL: <https://aclanthology.org/2025.americasnlp-1.13/>.
- [4] 坂上 温紀, Justin Vasselli, 井手 佑翼, 坂井 優介, Yingtao Tian, 上垣外 英剛, and 渡辺 太郎. “低資源言語のための辞書を用いた言語間語彙転移”. In: *言語処理学会第 31 回年次大会*. 2025.  URL: https://www.anlp.jp/proceedings/annual_meeting/2025/pdf_dir/Q4-20.pdf.
- [5] 坂上 温紀, 坂井 優介, 上垣外 英剛, and 渡辺 太郎. “満洲語古典語における母音調和の計算言語学的考察”. In: *言語処理学会第 31 回年次大会*. 2025.  URL: https://www.anlp.jp/proceedings/annual_meeting/2025/pdf_dir/E3-3.pdf.
- [6] 北野 雄士, 西田 悠人, 坂上 温紀, 上垣外 英剛, and 渡辺 太郎. “独立成分分析による事前学習済み多言語モデルの層を横断した単語埋め込み表現の分析”. In: *言語処理学会第 31 回年次大会*. 2025.  URL: https://www.anlp.jp/proceedings/annual_meeting/2025/pdf_dir/P2-15.pdf.

Journal Articles

- [7] **Haruki Sakajo**, Yusuke Sakai, Hidetaka Kamigaito, and Taro Watanabe. “Vowel Articulation and Tongue Position in Vision Language Models”. In: *自然言語処理 32.3* (2025), pp. 859–885.  DOI: 10.5715/jnlp.32.859.

Miscellaneous Experience

Awards and Achievements

Outstanding, YANS2024.

Project

The Central Eurasian World in the Manchu-Language Archives 2025 – 2028
Type AI Research Project at the Institute for Research in Humanities, Kyoto University
<https://manju-sibe.com/>

- Digitize and construct datasets in the Manchu language.
- Conduct historical research focusing on deciphering unstudied documents.

Explaining Inner Workings of Multimodal Language Models using Internal Representations

2025 July – 2026 March

KAKUSEI Project (3,000,000 JPY)
<https://kakusei.aist.go.jp/r7/>

Skills

Languages	• Japanese (native), English (academic)
Coding	• TypeScript, JavaScript, Python, Rust, Dart
Web Dev	• React, Next.js, Vue.js, HTML, CSS, JavaScript, TypeScript