CURRICULUM VITAE

Yoshihiro (Yoshi) MATSUURA

松浦 義広

Email: nguekwang@gmail.com Website: nguekwang.github.io/en/ GitHub: github.com/nguekwang

Research topics

- Language universals and typology: what all human languages share, and how they differ
- Cross-linguistic semantics, with formal semantics as the theory and computational methods as the tools
- Measuring similarity between languages and dialects (phonology, lexicon)
- Computational and phylogenetic linguistics
- Writing systems and multilingual computing (input methods, internationalization)

Current research

What similarity is, and how the properties of a language can be described objectively. I explore whether such methods can reach beyond spoken human language to written language, sign languages, what may count as language in beings other than Homo sapiens, constructed languages and formal languages.

Publications & Presentations

- 言語間の類似度の定量的な算出 (A Quantitative Analysis of Cross-Linguistic Similarity) (2026) (bachelor's thesis, unpublished)

Projects

- **linguasim**: A Python toolkit for measuring language similarity (in development)
- **Lang Log**: An interactive atlas of the world's languages, with corpora and data for low-resource languages (in development)
- **phrase-marker-gen**: A tool for drawing syntax trees in generative grammar (planned)

Outlook

- **Dialect or language**: use similarity measures to propose criteria for when a variety counts as a dialect and when as a language
- **Syntactic similarity**: measure how close syntax is, alongside sounds and vocabulary, and combine them into one index
- **Second-language learning**: test whether languages closer to one's own are easier to learn
- **A GUI tool**: build a tool anyone can use to compute similarity

Background

- Born in December 2003 in Hiroshima, Japan
- Graduated, Hiroshima Municipal Motomachi High School (general course, standard track), March 2022
- B.A. in Spanish, School of Foreign Studies, The University of Osaka, March 2026

Languages

- **First language (L1)**: Japanese
- **Working languages**: English (CEFR C1) · Spanish (CEFR B1, B.A. major)
- **Languages under study**: Korean · French · Russian · Chinese · Arabic

Keywords

- **Language comparison**: Language similarity and distance · Linguistic typology and language universals · Phylogenetic linguistics · Language contact · Language and nationalism
- **Theoretical linguistics**: Saussurean linguistics · Phonology · Formal semantics and cross-linguistic semantics · Generative grammar · Cognitive linguistics · Analytic philosophy and Wittgenstein

- **Objects of study**: Romance languages (Ibero-Romance) · Medieval Castilian · Mozarabic (Andalusi Romance) · Ladino (Judaeo-Spanish) · Writing systems · Morse code · Sign languages · Constructed languages · Formal languages
- **Mathematics and data**: Computational linguistics and natural language processing · Information theory · Probability theory and statistics · Vibrations and waves · Signal processing · Data visualization
- **Technology**: Internationalization (i18n) · Unicode and character encoding · Input methods · Digital typesetting · Linux and operating system kernels

---

Last updated: November 2026 · nguekwang.github.io/en/