SƠ YẾU LÝ LỊCH

Yoshihiro (Yoshi) MATSUURA

松浦 義広

Email: nguekwang@gmail.com Trang web: nguekwang.github.io/vi/ GitHub: github.com/nguekwang

Trang này được dịch máy.

Chủ đề nghiên cứu

- Phổ quát ngôn ngữ và loại hình học: điểm chung của mọi ngôn ngữ và sự khác biệt giữa chúng
- Ngữ nghĩa học liên ngôn ngữ, với ngữ nghĩa học hình thức làm lý thuyết và phương pháp tính toán làm công cụ
- Đo độ tương đồng giữa các ngôn ngữ và phương ngữ (âm vị học, từ vựng)
- Ngôn ngữ học tính toán và ngôn ngữ học phát sinh chủng loại
- Hệ thống chữ viết và điện toán đa ngôn ngữ (bộ gõ, quốc tế hóa)

Nghiên cứu hiện tại

Thế nào là tương đồng, và làm thế nào để mô tả khách quan các đặc tính của ngôn ngữ. Tôi tìm hiểu khả năng áp dụng những phương pháp này không chỉ cho ngôn ngữ nói của con người mà cả ngôn ngữ viết, ngôn ngữ ký hiệu, những gì có thể coi là ngôn ngữ ở các thực thể ngoài Homo sapiens, ngôn ngữ nhân tạo và ngôn ngữ hình thức.

Công bố và báo cáo

- 言語間の類似度の定量的な算出 (A Quantitative Analysis of Cross-Linguistic Similarity) (2026) (khóa luận tốt nghiệp, không công bố)

Dự án

- **linguasim**: Bộ công cụ Python để đo độ tương đồng giữa các ngôn ngữ (đang phát triển)
- **Lang Log**: Bản đồ tương tác về các ngôn ngữ trên thế giới, kèm kho ngữ liệu và dữ liệu về các ngôn ngữ ít tài nguyên (đang phát triển)
- **phrase-marker-gen**: Công cụ vẽ cây cú pháp theo ngữ pháp tạo sinh (đang lên ý tưởng)

Định hướng

- **Phương ngữ hay ngôn ngữ**: dựa trên các thước đo tương đồng, đề xuất tiêu chí khi nào một biến thể được coi là phương ngữ và khi nào là ngôn ngữ
- **Tương đồng cú pháp**: đo cả mức gần gũi về cú pháp bên cạnh ngữ âm và từ vựng, rồi gộp lại thành một chỉ số
- **Thụ đắc ngôn ngữ thứ hai**: kiểm chứng xem ngôn ngữ càng gần tiếng mẹ đẻ thì có càng dễ học hay không
- **Công cụ có giao diện đồ họa**: xây dựng công cụ để ai cũng tính được độ tương đồng

Tiểu sử

- Sinh tháng 12 năm 2003 tại Hiroshima, Nhật Bản
- Tốt nghiệp THPT, Trường THPT Thành phố Hiroshima Motomachi (ban phổ thông, chương trình tiêu chuẩn), tháng 3 năm 2022
- Cử nhân tiếng Tây Ban Nha, Khoa Ngoại ngữ, Đại học Osaka, tháng 3 năm 2026

Ngôn ngữ

- **Bản ngữ (L1)**: Tiếng Nhật
- **Ngôn ngữ làm việc**: Tiếng Anh (CEFR C1) · Tiếng Tây Ban Nha (CEFR B1, chuyên ngành cử nhân)
- **Ngôn ngữ đang học**: Tiếng Hàn · Tiếng Pháp · Tiếng Nga · Tiếng Trung · Tiếng Ả Rập

Từ khóa

- **So sánh ngôn ngữ**: Độ tương đồng và khoảng cách giữa các ngôn ngữ · Loại hình học ngôn ngữ và phổ quát ngôn ngữ · Ngôn ngữ học phát sinh chủng loại · Tiếp xúc ngôn ngữ · Ngôn ngữ và chủ nghĩa dân tộc
- **Ngôn ngữ học lý thuyết**: Ngôn ngữ học Saussure · Âm vị học · Ngữ nghĩa học hình thức và ngữ nghĩa học liên ngôn ngữ · Ngữ pháp tạo sinh · Ngôn ngữ học tri nhận · Triết học phân tích và Wittgenstein
- **Đối tượng nghiên cứu**: Các ngôn ngữ Rôman (các ngôn ngữ Rôman Iberia) · Tiếng Castilla trung đại · Tiếng Mozarab (Rôman Andalus) · Tiếng Ladino (Do Thái - Tây Ban Nha) · Hệ thống chữ viết · Mã Morse · Ngôn ngữ ký hiệu · Ngôn ngữ nhân tạo · Ngôn ngữ hình thức
- **Toán học và dữ liệu**: Ngôn ngữ học tính toán và xử lý ngôn ngữ tự nhiên · Lý thuyết thông tin · Lý thuyết xác suất và thống kê học · Dao động và sóng · Xử lý tín hiệu · Trực quan hóa dữ liệu
- **Công nghệ**: Quốc tế hóa (i18n) · Unicode và mã hóa ký tự · Bộ gõ · Dàn trang kỹ thuật số · Linux và nhân hệ điều hành

---

Cập nhật lần cuối: tháng 11 năm 2026 · nguekwang.github.io/vi/