

Chihiro Taguchi

Updated July 28, 2024

田口 智大

✉ Email: ctaguchi@nd.edu

🐙 GitHub: [ctaguchi](https://github.com/ctaguchi)

🌐 LinkedIn: <https://www.linkedin.com/in/ctaguchi>

💻 Website: <https://ctaguchi.github.io>

Education

- | | | |
|-------------------|---|---------------------------|
| 08/2022 – 07/2027 | University of Notre Dame | IN, United States |
| | PhD in Computer Science and Engineering. | |
| 09/2021 – 08/2022 | University of Edinburgh | Edinburgh, United Kingdom |
| | MSc by Research in Linguistics with Distinction. | |
| 09/2021 – 08/2022 | Nara Institute of Science and Technology | Nara, Japan |
| | MA in Engineering. | |
| 09/2018 – 06/2019 | School of Oriental and African Studies, University of London | London, United Kingdom |
| | Exchange program. | |
| 04/2015 – 09/2019 | Keio University | Tokyo, Japan |
| | Bachelor of Laws (LL.B.). | |

Training

- | | | |
|-------------------|--|-------------------|
| 06/2023 – 07/2023 | Linguistic Society of America (LSA) Institute | MA, United States |
| 05/2023 – 06/2023 | Summer Language Abroad Program (Quechua) , University of Notre Dame | Quito, Ecuador |
| 02/2018 | 3rd Miyako Language Documentation Workshop | Okinawa, Japan |

Experience

- | | |
|-------------------|---|
| 12/2023 – 05/2024 | Program Committee , 2nd Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI) at LREC-COLING 2024 |
| 04/2021 – 09/2021 | Research Assistant , National Institute of Informatics (Japan) |

Projects

- | | |
|-------------------|--|
| 04/2024 – 03/2028 | 科研費基盤研究 (C) ジンポー語口承資料の資源化に関する研究 (Grants-in-Aid for Scientific Research: “Digitization of Jinghpaw oral literature”), Japan Society for the Promotion of Science (JSPS).
PI: Dr. Keita Kurabe. |
| 04/2024 – 03/2027 | ユーラシア諸言語における談話の研究 (Studies on the grammar of discourse in Eurasian languages), Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
PI: Dr. Norikazu Kogura. |
| 09/2022 – Present | Language Documentation with an AI Helper . National Science Foundation. PI: Dr. Antonis Anastasopoulos. |
| 01/2023 | Building a multilingual speech-to-IPA system . SLT-Code Hackathon, IEEE. |
| 08/2021 – 06/2022 | Universal Dependencies: Tatar NMCTT treebank . CICP, Nara Institute of Science and Technology. |

Publications

- 2024 **Chihiro Taguchi**, David Chiang. Language Complexity and Speech Recognition Accuracy: Orthographic Complexity Hurts, Phonological Complexity Doesn't. In ACL (main, to appear).
- 2024 Yuhi Matogawa, Yusuke Sakai, Taro Watanabe, **Chihiro Taguchi**. Japanese Rule-based Grapheme-to-phoneme Conversion System and Multilingual Named Entity Dataset with International Phonetic Alphabet. In SIGMORPHON at NAACL.
- 2024 **Chihiro Taguchi**, Jefferson Saransig, Dayana Velásquez, David Chiang. KILLKAN: The Automatic Speech Recognition Dataset for Kichwa with Morphosyntactic Information.
- 2024 Tatsuya Aoyama, **Chihiro Taguchi**, Nathan Schneider. J-SNACS: Adposition and Case Supersenses for Japanese Joshi. In LREC-COLING.
- 2024 Jonathan Washington, Çağrı Çöltekin, Furkan Akkurt, Bermet Chontaeva, Soudabeh Eslami, Gulnura Jumalieva, Aida Kasieva, Aslı Kuzgun, Büşra Marşan, **Chihiro Taguchi**. Strategies for the Annotation of Pronominalised Locatives in Turkic Universal Dependency Treebanks. In MWE-UD at LREC-COLING.
- 2024 **Chihiro Taguchi**. Non-discourse-configurationality in Imbabura Kichwa. In the *Proceedings of the Linguistic Society of America*.
- 2023 **Chihiro Taguchi**, Yusuke Sakai, Parisa Haghani, David Chiang. Universal Automatic Phonetic Transcription into the International Phonetic Alphabet. In INTERSPEECH.
- 2023 **Chihiro Taguchi**, David Chiang. Introducing Morphology in Universal Dependencies Japanese. At the 6th Workshop on the Universal Dependencies.
- 2022 **Chihiro Taguchi**. Mermaid Construction in Lexical Functional Grammar. In the *Proceedings of LFG'22*.
- 2022 **Chihiro Taguchi**, Sei Iwata, Taro Watanabe. Universal Dependencies Treebank for Tatar: Incorporating Intra-Word Code-Switching Information. In the EURALI workshop at LREC.
- 2021 **Chihiro Taguchi**, Yusuke Sakai, Taro Watanabe. Transliteration for Low-Resource Code-Switching Texts: Building an Automatic Cyrillic-to-Latin Converter for Tatar. In the CALCS workshop at NAACL.

Talks

- 06/2024 "KILLKAN: The Automatic Speech Recognition Dataset for Kichwa with Morphosyntactic Information." Poster presentation at the AmericasNLP workshop at NAACL.
- 04/2024 "Kichwa Meets Language Technologies: Automatic Speech Recognition for Kichwa." Oral presentation at the Quechua Alliance Annual meeting.
- 01/2024 "El conjunto de datos de reconocimiento automático de voz para kichwa." Oral presentation at the Taller: Tecnologías Digitales y Lenguas Indígenas.
- 01/2024 "Non-Discourse-Configurationality in Imbabura Kichwa." Poster presentation at the Linguistic Society of America Annual Meeting.
- 12/2023 "Building an Automatic Speech Recognition Dataset and Model for an Extremely Low-Resource Language." Oral presentation at Notre Dame NL+ Seminar.
- 11/2023 "Reconocimiento automático del habla de kichwa para su inclusión y empoderamiento en el ciberespacio." Oral presentation at the VI Seminario Internacional Revitalizando Ando.
- 09/2023 "UD-Tatar NMCTT Treebank: Issues in annotation across Turkic UD." Oral presentation at the UD Turkic Workshop.
- 08/2023 "Grammaticalization of modal nominal predicates in Tatar." Oral presentation at the International Conference on Turkish Linguistics.
- 07/2023 "Bridging Natural Language Processing and Descriptive Linguistics with Universal Dependencies." Oral presentation at the Field Linguistics Workshop: Grammatical Studies Workshops 23 "Grammatical Studies and Digital Humanities (2)".
- 07/2023 "Japanese gapless relativization: The syntax-prosody interface to semantics." Poster presentation at LFG23: the 28th International Lexical Functional Grammar Conference.

- 07/2023 “Incorporating AI-based speech transcription into language documentation: A case study of Imbabura Kichwa.” Poster presentation at the LSA Institute.
- 10/2022 “Cross-linguistic analysis of the Mermaid Constructions in LFG.” Oral presentation at the 33rd South of England LFG Meetings.
- 09/2022 “Universal Dependencies Tatar for Code-Switching Detection.” Oral presentation at the Notre Dame NL+ Seminar.
- 07/2022 “Mermaid Construction in Lexical Functional Grammar.” Oral presentation at LFG22: the 27th International Lexical Functional Grammar Conference.
- 07/2022 “Consistent Grammatical Annotation of Turkic Languages for more universal Universal Dependencies.” Oral presentation at the Workshop on Computational Linguistics on East Asian languages.
- 03/2022 “Mermaid Construction as raising with a nominal predicate.” Oral presentation at the 74th Language Lunch At Edinburgh.
- 10/2021 “Text Processing of Open-Access Jinghpaw Data on Google Colaboratory.” Oral presentation at the Follow-up Meeting on Intensive Language Course 2019 (Jinghpaw) / The 3rd Meeting on Kachin Studies.

— Honors and Awards

- 06/2024 **Conference Travel Grant**, Kellogg Institute for International Studies, University of Notre Dame.
- 05/2024 **Leadership Advancing Socially Engaged Research (LASER) Award**, University of Notre Dame.
- 04/2024 **Doctoral Student Research Grant**, Kellogg Institute for International Studies, University of Notre Dame.
- 01/2024 **Conference Presentation Grant**, Graduate Student Government, University of Notre Dame.
- 06/2023 **Conference Travel Grant**, Kellogg Institute for International Studies, University of Notre Dame.
- 05/2023 **Professionalization Grant**, Kellogg Institute for International Studies, University of Notre Dame.
- 03/2023 **委員特別賞** (Committee’s Honorable Mention), Association for Natural Language Processing.
- 03/2023 **Conference Presentation Grant**, Graduate Student Government, University of Notre Dame.
- 03/2023 **Summer Language Abroad Grant**, Center for the Study of Languages and Cultures, University of Notre Dame.
- 09/2021 **JASSO Scholarship for Graduate-level Study Abroad**, Ministry of Education, Culture, Sports, Science and Technology, Japan.
- 06/2021 **CICP NAIIST Multilingual Corpus**, Nara Institute of Science and Technology.
- 04/2019 **Grand Prix Winner** at the 7th International Olympiad of the Tatar Language and Literature, Ministry of Education and Science, Republic of Tatarstan, Russia.
- 09/2018 **Exchange Program Scholarship**, Gyomu-Super Japan Dream Foundation
- 11/2014 **Honorable Mention** at the Keio University Academic Writing Contest, Keio University.

— Skills

- Programming:** Python, R, C, Bash, \LaTeX , HTML/CSS
- Software/libraries:** PyTorch, Tensorflow, Flask, Git, AWS, Docker, ELAN, Praat
- Languages:** Japanese (native), English (fluent), Spanish (advanced), Russian (intermediate), Tatar (intermediate), Kichwa (intermediate), Korean (intermediate), Mandarin Chinese (intermediate), French (intermediate), German (intermediate)

— Other academic affiliations

Kellogg Institute Doctoral Student Affiliate
 Association of Computational Linguistics (ACL)
 Linguistic Society of America (LSA)
 言語処理学会 [Association for Natural Language Processing]
 日本言語学会 [Linguistic Society of Japan]