

Atsuki YAMAGUCHI

Email: atsuki.yamaguchi@gmail.com

Website: <https://gucci-j.github.io/about/>

GitHub: <https://github.com/gucci-j>

EDUCATION

- SEP 2023-MAR 2027
(expected) **Ph.D. in Computer Science, The University of Sheffield, Sheffield, United Kingdom**
Research topic: Efficient cross-lingual transfer for generative large language models
Advisors: Prof. Nikolaos Aletras and Prof. Aline Villavicencio
- SEP 2019-SEP 2020 **Master of Science in Computer Science with Speech and Language Processing (with Distinction), The University of Sheffield, Sheffield, United Kingdom**
Thesis: "Dementia Detection from Narrative Speech Using Multitask Learning and Mixup"
Advisor: Prof. Aline Villavicencio
- APR 2015-MAR 2019 **Bachelor of Engineering in Computer and Information Sciences (Hons.), Tokyo University of Agriculture and Technology, Tokyo, Japan**
Thesis: "Generating Emoji with Conditional Variational Autoencoders and Word Embedding" | Advisor: Prof. Katsuhide Fujita
GPA: 3.63 / 4.00 | Earned 176 credits (128 credits are required for graduation)
- FALL 2017 **Exchange Semester at Electronics and Communications Engineering, De La Salle University, Manila, Philippines**
GPA: 2.82 / 4.00 | Earned 14 credits

WORK HISTORY

- APR 2022-SEP 2023 **Researcher**
Research and Development Group, Hitachi, Ltd., Tokyo, Japan
Conducted natural language processing-related research and application development with several business partners. Typical projects include multimodal information extraction and efficient language modeling. The responsibilities include writing patents and scientific papers. Those in this position also supervise interns and write research proposals.
- JAN 2021-MAR 2022 **Associate Researcher**
Research and Development Group, Hitachi, Ltd., Tokyo, Japan
Worked on natural language processing-related projects with several business partners. Typical projects include analysing stock exchange documents using information extraction techniques and automatic meeting minutes generation. The responsibilities include writing patents and scientific papers.
- SEP 2020-MAY 2021 **Research Intern**
The University of Sheffield, Sheffield, United Kingdom
Worked on building efficient language representation models in the research group led by Prof. Nikolaos Aletras.
- MAY-JUN 2018 **Teaching Practice**
Tokyo Denki University Junior and Senior High School, Tokyo, Japan
Worked for three weeks and taught Information Technology to first-year and fourth-year students.
- JAN-FEB 2018 **Computer Engineer (Intern)**
JTEKT Philippines Corporation (JPC), Batangas, Philippines
Worked in general affairs (IT section) for one month. Maintained internal server systems.

SCHOLARSHIPS

SEP 2023-MAR 2027	University of Sheffield EPSRC Doctoral Training Partnership Scholarship Sponsored by: The University of Sheffield, UK Research & Innovation (UKRI) Details: Covers full international tuition fees, a stipend at the standard UKRI rate, and research support and training grant.
SEP 2023-AUG 2026	JASSO Student Exchange Support Program (Graduate Scholarship for Degree Seeking Students) Sponsored by: Japan Student Services Organization (JASSO) Details: Covers a monthly stipend of ¥118,000 and potential tuition fees support of up to ¥3,000,000/year.
SEPTEMBER 2019	International Postgraduate Taught Merit Scholarship (£5,650) Sponsored by: The University of Sheffield Description: 100 new international students were nominated from all applications.
APR 2019-DEC 2020	Carlit Holdings Scholarship (¥30,000/month) Sponsored by: Carlit Holdings Co., Ltd. Description: Five students were nominated from all applications.
AUG 2016-MAR 2019	Kawamura Foundation Scholarship (¥30,000/month) Sponsored by: Kawamura Scholarship Foundation Description: Five students were nominated from all applications.
APR 2016-MAR 2019	Outstanding Student Scholarship (¥100,000/year) Sponsored by: Tokyo University of Agriculture and Technology Description: The top few percent of students in each department were nominated based on GPA.
AUG 2017-FEB 2018	JASSO Student Exchange Support Program (¥70,000/month) Sponsored by: Japan Student Services Organization (JASSO)

AWARDS & CERTIFICATES

JULY 2022	Best System Paper Honorable Mention The 16th International Workshop on Semantic Evaluation (SemEval-2022) Paper title: Hitachi at SemEval-2022 Task 10: Comparing Graph- and Seq2Seq-based Models Highlights Difficulty in Structured Sentiment Analysis
MARCH 2019	Outstanding Student Award Awarded by: Tokyo University of Agriculture and Technology Award criterion: The best overall performance in the BEng Degree in the Department of Computer and Information Sciences.
MARCH 2019	High School Teacher's License (Mathematics & Information) Licensed by: Tokyo Metropolitan Board of Education
JULY 2018	IELTS® Overall 7.0 / 9.0 Issued by: British Council

LANGUAGES

JAPANESE: Native | ENGLISH: Fluent

PROGRAMMING SKILLS

Intermediate Knowledge: MySQL, Elasticsearch, HTML, CSS, JavaScript, Flask, Jinja
Professional Knowledge: Python (PyTorch, TensorFlow, Keras, Scikit-learn), C, Java

RESEARCH INTERESTS

Cross-lingual Transfer, Language Modeling, Natural Language Understanding, Natural Language Processing

PUBLICATIONS & PRESENTATIONS

[Google Scholar](#)

[Semantic Scholar](#)

International Conferences & Workshops (refereed)

1. Terufumi Morishita, **Atsuki Yamaguchi**, Gaku Morio, Hikaru Tomonari, Osamu Imaichi and Yasuhiro Sogawa, “[JFLD: A Japanese Benchmark for Deductive Reasoning based on Formal Logic](#),” The 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), May 2024.
2. Terufumi Morishita, Yuta Koreeda, **Atsuki Yamaguchi**, Gaku Morio, Osamu Imaichi and Yasuhiro Sogawa, “[CHICOT: A Developer-Assistance Toolkit for Code Search with High-Level Contextual Information](#),” The 38th AAAI Conference on Artificial Intelligence (AAAI-24), February 2024. (Demonstration track)
3. Takuro Fujii, Koki Shibata, **Atsuki Yamaguchi**, Terufumi Morishita and Yasuhiro Sogawa, “[How do different tokenizers perform on downstream tasks in scriptio continua languages?: A case study in Japanese](#),” The 61st Annual Meeting of the Association for Computational Linguistics: Student Research Workshop (ACL-SRW 2023), July 2023.
4. **Atsuki Yamaguchi**, Hiroaki Ozaki, Terufumi Morishita, Gaku Morio and Yasuhiro Sogawa, “[How does the task complexity of masked pretraining objectives affect downstream performance?](#),” Findings of the Association for Computational Linguistics: ACL 2023, July 2023.
5. Terufumi Morishita, Gaku Morio, **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[Learning Deductive Reasoning from Synthetic Corpus based on Formal Logic](#),” Fortieth International Conference on Machine Learning (ICML 2023), Hawaii, USA, July 2023.
6. Masaya Tsunokake, **Atsuki Yamaguchi**, Yuta Koreeda, Hiroaki Ozaki and Yasuhiro Sogawa, “[Hitachi at SemEval-2023 Task 4: Exploring Various Task Formulations Reveals the Importance of Description Texts on Human Values](#),” The 17th International Workshop on Semantic Evaluation (SemEval-2023), July 2023.
7. Yuta Koreeda, Ken-ichi Yokote, Hiroaki Ozaki, **Atsuki Yamaguchi**, Masaya Tsunokake and Yasuhiro Sogawa, “[Hitachi at SemEval-2023 Task 3: Exploring Cross-lingual Multi-task Strategies for Genre and Framing Detection in Online News](#),” The 17th International Workshop on Semantic Evaluation (SemEval-2023), July 2023.
8. **Atsuki Yamaguchi**, Gaku Morio, Hiroaki Ozaki and Yasuhiro Sogawa, “[Hitachi at SemEval-2022 Task 2: On the Effectiveness of Span-based Classification Approaches for Multilingual Idiomaticity Detection](#),” The 16th International Workshop on Semantic Evaluation (SemEval-2022), July 2022.
9. Gaku Morio, Hiroaki Ozaki, **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[Hitachi at SemEval-2022 Task 10: Comparing Graph- and Seq2Seq-based Models Highlights Difficulty in Structured Sentiment Analysis](#),” The 16th International Workshop on Semantic Evaluation (SemEval-2022), July 2022.
10. **Atsuki Yamaguchi**, George Chrysostomou, Katerina Margatina and Nikolaos Aletras, “[Frustratingly Simple Pretraining Alternatives to Masked Language Modeling](#),” The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021), Online, November 2021. (Short paper)
11. **Atsuki Yamaguchi**, Gaku Morio, Hiroaki Ozaki, Ken-ichi Yokote and Kenji Nagamatsu, “[Team Hitachi @ AutoMin 2021: Reference-free Automatic Minuting Pipeline with Argument Structure Construction over Topic-based Summarization](#),” First Shared Task on Automatic Minuting @ Interspeech 2021, September 2021.
12. **Atsuki Yamaguchi**, Kosui Iwasa and Katsuhide Fujita, “[Dialogue Act-based Breakdown Detection in Negotiation Dialogues](#),” The 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021), Online, April 2021. (Long paper)

13. **Atsuki Yamaguchi** and Katsuhide Fujita, “[Breakdown Detection in Negotiation Dialogues](#),” Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), New York, USA, February 2020. (Student Abstract)
14. **Atsuki Yamaguchi** and Katsuhide Fujita, “Breakdown Detection in Negotiation Dialogues,” European Natural Language Processing Summit (EurNLP) 2019, London, United Kingdom, October 2019. (Acceptance rate: 30.5%; Non-archival)
15. **Atsuki Yamaguchi** and Melvin Cabatuan, “[Generative Model Based Frame Generation of Volcanic Flow Video](#),” IEEE HNICEM 2017, Manila, Philippines, December 2017.

Regional Conferences

1. Terufumi Morishita, **Atsuki Yamaguchi**, Gaku Morio, Osamu Imaichi and Yasuhiro Sogawa, “Acquiring Generalizable Reasoning Ability through a Large-Scale Logical Corpus with Inductively Diverse Examples,” The 38th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2024), Shizuoka, Japan, May 2024. (*In Japanese*)
2. Terufumi Morishita, **Atsuki Yamaguchi**, Gaku Morio, Masaya Tsunokake, Hikaru Tomonari, Osamu Imaichi and Yasuhiro Sogawa, “[Proposal for a Japanese Logical Reasoning Benchmark: JFLD](#),” The 30th Annual Meeting of the Association for Natural Language Processing (NLP 2024), Kobe, Japan, March 2024. (*In Japanese*)
3. **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[Open Information Extraction from Document Images Using VQA Datasets](#),” The 37th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2023), Kumamoto, Japan, June 2023. (*In Japanese*)
4. Terufumi Morishita, Gaku Morio, **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[How do synthetic deduction corpora enhance language models?](#),” The 37th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2023), Kumamoto, Japan, June 2023. (*In Japanese*)
5. Takuro Fujii, Koki Shibata, **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[Do different Japanese tokenizers perform differently on downstream tasks?](#),” The 29th Annual Meeting of the Association for Natural Language Processing (NLP 2023), Okinawa, Japan, March 2023. (*In Japanese*)
6. Terufumi Morishita, Gaku Morio, **Atsuki Yamaguchi** and Yasuhiro Sogawa, “[Learning Deductive Reasoning from Synthetic Corpus based on Formal Logic](#),” The 29th Annual Meeting of the Association for Natural Language Processing (NLP 2023), Okinawa, Japan, March 2023. (*In Japanese*)
7. **Atsuki Yamaguchi**, Kosui Iwasa and Katsuhide Fujita, “[Predicting Issue Weights and Detecting Breakdowns in Negotiation Dialogue](#),” IEICE SIG-AI, Hokkaido, Japan, July 2019. (*In Japanese*)
8. **Atsuki Yamaguchi** and Katsuhide Fujita, “[Generating Emoji with Conditional Variational Autoencoders and Word Embedding](#),” The 33rd Annual Conference of the Japanese Society for Artificial Intelligence (JSAI 2019), Niigata, Japan, June 2019. (*In Japanese*)

Preprints

1. **Atsuki Yamaguchi**, Aline Villavicencio and Nikolaos Aletras, “[Vocabulary Expansion for Low-resource Cross-lingual Transfer](#),” arXiv preprint: 2406.11477, June 2024.
2. **Atsuki Yamaguchi**, Aline Villavicencio and Nikolaos Aletras, “[An Empirical Study on Cross-lingual Vocabulary Adaptation for Efficient Language Model Inference](#),” arXiv preprint: 2402.10712, February 2024.
3. **Atsuki Yamaguchi** and Terufumi Morishita, “[appjsonify: An Academic Paper PDF-to-JSON Conversion Toolkit](#),” arXiv preprint: 2310.01206, October 2023.

ACADEMIC ACTIVITIES

Membership Association for Computational Linguistics (ACL)

Service

- 2024 **Communications/Support Team Volunteer** – ACL Rolling Review (ARR)
Reviewer – ACL Rolling Review (ARR), EACL 2024 (System demonstration track)
Student Volunteer – EACL 2024
- 2023 **Reviewer** – ACL Rolling Review (ARR), NLLP 2023 (Workshop @ EMNLP 2023), SICon 2023 (Workshop @ ACL 2023)
Secondary Reviewer – SemEval 2023
- 2022 **Reviewer** – SemEval 2022