Tatsuya AOYAMA

Ph.D. Student in Linguistics | Research Assistant @ Georgetown University

in linkedin.com/in/tatsuya-aoyama **☺** github.com/t-aoyam

☐ +1 949 295 4246 **@** ta571@georgetown.edu

⚲ Permanent : 1-877-18 Ogawa-cho, Kodaira City, Tokyo, Japan 1870032

⚲ Current : 1616 18th St NW, Apt 512, Washington, DC 20009

Bio : I am a third year Ph.D. student in Linguistics at Georgetown University, with a primary focus on computational linguistics (see page 3 for relevant coursework). I have experiences working on NLP tasks and research involving English and Japanese texts both in academic and industry settings. I enjoy learning on the job, and my work experience (research assistant, teaching assistant, and NLP engineer/research internship) and academic background (linguistics, NLP, and data science) have equipped me with skills to conduct research, run experiments, visualize the results, and effectively communicate the ideas to other researchers as well as to the general public.

Areas of interest : *distributional semantics, application of NLP techniques to non-native English texts, second language acquisition, learner corpus research, machine learning.*

🎓 EDUCATION

(Exp.) May 2025 August 2020	Ph.D. in Linguistics, GEORGETOWN UNIVERSITY, Washington, D.C. > Computational Linguistics > Adivisors : Dr. Nathan Schneider, Dr. Amir Zeldes > GPA : 4.00/4.00
May 2020 August 2018	M.A. in TESOL, MICHIGAN STATE UNIVERSITY, East Lansing, MI > Advisors : Dr. Sandra C. Deshors (Linguistics), Dr. Kristen M. Johnson (Computer Science) > M.A. Thesis : "A Corpus-Based Multifactorial Analysis of Japanese and Chinese Speakers' English Article Use : Quantifying the Deviation Using MuPDAR" > GPA : 4.00/4.00
May 2018 August 2014	B.A. in Liberal Arts, SOKA UNIVERSITY OF AMERICA, Aliso Viejo, CA > Double Concentration in Humanities and Social and Behavioral Sciences > Advisor : Dr. Tomoko Takahashi (Applied Linguistics) > Senior Thesis : "Japanese ESL Students' Willingness to Communicate in English : The Effects of L2 Self-Confidence, Acculturation, and Motivational Types" > GPA : 3.93/4.00 (Summa Cum Laude)

💼 PROFESSIONAL EXPERIENCES

December 2022 August 2022	Research Assistant, GEORGETOWN UNIVERSITY, Washington, D.C. > Supervisor : Dr. Amir Zeldes
August 2022 May 2022	Summer NLP Research Intern, POSH TECHNOLOGIES, INC., Boston, MA > Supervisor : Rishi Sharma (recommendation) > Evaluated the multinomial classification model behind the chatbot AI to identify challenging classes. > Experimented with different word embeddings and hyperparameters to improve the model. > Introduced label smoothing technique to the model training process, which resulted in a substantial improvement in the model calibration.
May 2022 August 2021	Teaching Assistant, GEORGETOWN UNIVERSITY, Washington, D.C. > LING 001 : Introduction to Linguistics (Fall 2021, Spring 2022) > Graded assignments and quizzes, held weekly office hours, mentored six students on the final project, and led weekly recitation sessions on Fridays.
August 2021 May 2021	Summer Software Engineering Intern, MONEY FORWARD, Tokyo, Japan > Supervisor : Ikuo Kitagishi > Built a rule-based classifier of legal contracts written in Japanese, utilizing syntactic and semantic information. > Satisfied the required accuracy threshold of 80%

May 2021 October 2020	Research Assistant, GEORGETOWN UNIVERSITY, Washington, D.C. ❯ Supervisor : Dr. Nathan Schneider ❯ Conducted a research project on the first language effect on a sequence tagger.
October 2020 August 2020	Research Assistant, GEORGETOWN UNIVERSITY, Washington, D.C. ❯ Supervisor : Dr. Alison Mackey ❯ Assisted with editorial works, such as updating references with more recent studies and adding new sub-sections on quantitative measures in Applied Linguistics.
July 2019 June 2019	Summer Research Intern, CENTER FOR APPLIED LINGUISTICS, Washington, D.C. ❯ Supervisors : Samantha Musser, Dr. Megan Montee ❯ Assisted research and development of a pragmatics test to assess English for Heritage Language Speakers program, as a part of a conference presentation by Dr. Montee at ACTFL 2019. ❯ Assisted research and development of coding scheme for task-based English test items, as a part of a conference presentation given by Dr. Montee at MwALT 2019. ❯ Featured on the university news. [website]
December 2019 August 2018	Teaching Assistant, MICHIGAN STATE UNIVERSITY, East Lansing, MI ❯ Supervisor : Mariko Kawaguchi ❯ Served as an instructor of record for 2 sections of undergraduate-level introductory Japanese classes (JPN-101 and JPN-102). ❯ Taught classes, developed class materials, held office hours, and graded exams.

✎ PUBLICATIONS & PRESENTATIONS

Aoyama, T. and Schneider, N. (2022). Probe-Less Probing of BERT's Layer-Wise Linguistic Knowledge with Masked Word Prediction. North American Chapter of the Association for Computational Linguistics Student Research Workshop (NAACL-SRW) 2022. [paper]

Aoyama, T. and Schneider, N. (2022). Revisiting Layer-Wise Linguistic Knowledge with Masked Word Prediction. Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL) 2022, Temple University, Philadelphia : PA.

Aoyama, T. and Takahashi, T. (2020). International Students' Willingness to Communicate in English as a Second Language : Effects of L2 Self-Confidence, Acculturation, and Motivational Types. Journal of International Students 10(3). [paper]

Aoyama, T. (2020). A Corpus-based Multifactorial Analysis of Japanese and Chinese Speakers' English Article Use : Quantifying the Deviation Using MuPDAR. M.A. Thesis, Michigan State University. [paper]

Aoyama, T. "Universal Grammar and Second Language Acquisition"
   ❯ Presented as an invited graduate-level guest lecture at LLT 860 (Second Language Acquisition) taught by Dr. Sandra C. Deshors at Michigan State University, February 13, 2020.

Aoyama, T. "A Corpus-based Multifactorial Analysis of Japanese and Chinese Learners' English Article Use : Quantifying the Deviation using MuPDAR," Paper presented at the Second Language Research Forum (SLRF2020), Vanderbilt University, Nashville : TN.
   ❯ Presented online synchronously due to COVID-19.

Takahashi, T. and **Aoyama, T.** "Japanese ESL Students' Willingness to Communicate in English : The Effects of L2 Self-Confidence, Acculturation, and Motivational Types."
   ❯ Co-authored with and presented by Tomoko Takahashi, The Asian Conference on Language (ACL2020), Tokyo, Japan, March 30, 2020.

🖥 RESEARCH PROJECTS

JAPANESE ADPOSITIONAL SUPERSENSE JANUARY 2022 - PRESENT
   ❯ Supervisor : Dr. Nathan Schneider
   ❯ Identifies Japanese particles, *joshi*, that qualify as adpositions; develops an annotation scheme for their supersense by extending English Semantic Network of Adposition and Case Supersenses (SNACS); and annotates a Japanese corpus.

DIFFICULTY JUDGMENT OF C-TEST WITH BERT JULY 2021 - DECEMBER 2021
   ❯ Supervisor : Dr. Nathan Schneider
   ❯ Aims at predicting the difficulty level of batches of C-test by measuring how well BERT solves the problems.

Search Engine with Inverted Index
<div align="right">August 2021 - December 2021</div>

> Supervisor : Dr. Nazli Goharian
> Built a search engine from scratch with inverted index as a semester-long course project for COSC-488 (Information Retrieval) with virtually no reliance on external libraries.
> Experimented with various similarity metrics, such as Cosine, BM25, and Language Model with Dirichlet Smoothing, and various query modification algorithms, such as query expansion with pseudo-relevance feedback and query reduction.
> Outperformed a competitive ElasticSearch baseline. [code]

First Language Effect on Lexical Semantic Recognition Sequence Tagger
<div align="right">Oct 2020 - May 2021</div>

> Supervisor : Dr. Nathan Schneider
> Examined the performance of prepositional super-sense tagging by Lexical Semantic Recognition (LSR) tagger on English texts written by native Spanish, French, German, and English speakers extracted from Reddit international corpora.
> Conducted randomized experiments to test if the differences in the accuracy scores could be attributed to the differences in the distribution of super-sense tags within prepositions.

Automatic Generation of Memorable and Secure Passwords
<div align="right">April 2021 - May 2021</div>

> Built a program that automatically generates passwords that balance memorability and security, as a semester-long group project for COSC-572 (Empirical Methods for Natural Language Processing).
> Utilized semantic information of user-defined keywords to enhance the memorability of passwords.
> Evaluated the security level of the generated password through intrinsic (e.g., cross-entropy) and extrinsic (e.g., how long it takes to crack the password) evaluations.
> Developed an algorithm to incorporate user defined keywords into the password. [code]

Psychopathy Prediction from Twitter Activity Data
<div align="right">Feb 2021 - May 2021</div>

> Aims to predict users with a high psychopathic tendency using an extremely unbalanced data ($\approx$3% true, $\approx$97% false).
> Utilizes various supervised and unsupervised statistical learning methods to find patterns and ways to classify the users based on their Twitter activity data.

Corpus-Based Investigation of the Use of Passives in Contemporary Written Japanese
<div align="right">Aug 2020 - Dec 2020</div>

> Examined the relationship between verb subcategorization and the corresponding frequenecy of passivization.
> Showed that the frequency order can be explained by the cross-linguistic markedness hypothesized in Noun Phrase Accessibility Hierarchy (NPAH).

☰ Skills

Programming	Proficient : Python, R | Elementary : JavaScript, CSS
NLP & ML	Tensorflow, Keras, Pytorch, AllenNLP, scikit-learn, nltk, STANZA, spaCy
Miscellaneous	LaTeX, RMarkdown, Statcrunch, SPSS, ElasticSearch, GitHub
Languages	Japanese–Native, **English–Fluent (TOEFL iBT : 115/120)**, Mandarin–Limited (HSK 5 : 273/300)

🏆 Honors & Awards

Graduate Student Travel Grant
<div align="right">Summer 2022</div>

> $250 travel grant for the presentation at MASC-SLL at Temple University, Philadelphia, PA.

Graduate Student Stipend
<div align="right">AY 2020 - 2025</div>

Research Funding
<div align="right">Oct 2019</div>

> Departmental $300 funding at Michigan State University

TA Tuition Pool
<div align="right">Spring 2019, Fall 2019</div>

> $2,500 funding for extra courses at Michigan State University.

Graduate Student Stipend
<div align="right">AY 2018 - 2020</div>

Merit Award
<div align="right">AY 2016 - 2017</div>

> $10,000 scholarship given to top 3 students annually at Soka University of America.

Dean's List
<div align="right">AY 2014 - 2018</div>

Soka Merit Scholarship
<div align="right">AY 2014 - 2018</div>

📘 Relevant Courseworks

Computational Linguistics

LING-362	Intro NLP	4.0	GU
LING-367	Computational Corpus Linguistics	4.0	GU
LING/COSC-572	Empirical NLP	4.0	GU
COSC-488	Information Retrieval	4.0	GU
LING-504	Machine Learning for Linguistics	4.0	GU
LING/COSC-672	Advanced Semantic Representation	4.0	GU
LING-765	Computational Discourse Modeling	4.0	GU
LING-429	Computational Grammar Formalism	*	GU
LING/COSC-469	Dialogue Systems	*	GU

Linguistics

LIN-434	Intro Syntax	4.0	MSU
LIN-834	Syntax I	4.0	MSU
LING-411	Phonology I	4.0	GU
LING-431	Semantics and Pragmatics I	4.0	GU

Data Science / Math

MATH-101	Statistics	4.0	SUA
MATH-121	Computer Science	4.0	SUA
LLT-873	Statistics with R	4.0	MSU
ANLY-511	Probabilistic Modeling and Statistical Computing	4.0	GU
MATH-642	Statistical Learning	4.0	GU
ANLY-555	Data Structure and Algorithm	*	GU
	Linear Algebra		Coursera
	Multivariate Calculus		Coursera

*Notes : * indicates currently enrolled ; SUA = Soka University of America ; MSU = Michigan State University ; GU = Georgetown University*