Sara Court

sarakcourt@gmail.com | (614) 230-6018 | Columbus, OH | [LinkedIn](#) | [GitHub](#)

EDUCATION

The Ohio State University   **Aug 2020 – Present**
Ph.D. Linguistics   *Columbus, OH*
Responsible Applications of LLMs for Low-Resource Machine Translation   Expected graduation: May 2026

The Ohio State University   **Aug 2025**
M.A. Linguistics   *Columbus, OH*

University of Florida   **Aug 2013**
B.A. Linguistics with Mathematics minor   *Gainesville, FL*
Bilingual Processing of Number Agreement in English, Spanish, and Arabic

TECHNICAL SKILLS

- **Machine learning and generative AI:** PyTorch, TensorFlow, Transformers, Large Language Models (LLMs)
- **Data analysis and collaboration:** Python3, Pandas, Matplotlib, Bash/Unix, GitHub, GPU cluster computing
- **Front-end development:** Some experience with Javascript, Vue.js, TensorFlow.js, Django, and SQL

PROJECTS

- Retrieval-augmented (RAG) pipeline for bias mitigation in low-resource machine translation using LLMs
- Active-learning and front-end interface development for semi-supervised human-in-the-loop data processing
- Transformers models for text generation and analysis of typologically diverse human languages
- Python tools for local and API-based music library management and analysis

RESEARCH EXPERIENCE

Graduate Researcher   **Aug 2020 – Present**
Department of Linguistics, The Ohio State University   *Columbus, OH*
- Implement novel machine translation pipelines leveraging large language models (LLMs) to address data sparsity in collaboration with speakers of underrepresented languages
- Lead a team exploring active learning and human-in-the-loop workflows for automated data annotation and analysis, resulting in peer-reviewed publications and an invited talk at the University of Helsinki
- Develop Transformer-based AI models of morphological inflection, contributing to NSF-funded research on the role of analogy in the mental lexicon
- Designed curriculum, implemented web interface, and taught statistics for CogSciDIY, a semester-long language science outreach program for approximately 50 high school students in Columbus, OH and New York, NY
- Publish and present machine learning research at top-tier peer-reviewed conferences in computational linguistics (SIGMORPHON and ComputEL at ACL; WMT at EMNLP; SCiL; COLING), language documentation (LD&A), and morphology (IMM; MMM)

Research Intern   **May 2023 - Aug 2024**
Speech Technology and Research (STAR) Lab, SRI International   *Menlo Park, CA*
May 2024 – Aug 2024
- Analyzed OCR finetuning experiments across multiple languages and conditions to enhance multilingual support and create visualizations to share outcomes with clients
- Innovated an error-tracking tool for a speech translation pipeline to optimize data flow and error analysis in large-scale and multi-stage processing environments
- Refactored HuggingFace code to experiment with Whisper ASR model architecture for improved efficiency

May 2023 – Aug 2023
- Developed and analyzed transformer models for Mandarin-English machine translation and English ASR
- Translated and curated a custom in-house test dataset (~900 items) for Mandarin-English machine translation
- Initiated and drafted onboarding documentation to assist future hires with technical setup and lab best practices

Research Assistant   **Mar 2020 – Aug 2020**
Department of Linguistics, National Taiwan University   *Taipei, Taiwan*
- Co-authored COLING publication on computational methods for Mandarin morphological analysis

Research Fellow   **Oct 2013 – Dec 2013**
Innovative Networking in Infrastructure for Endangered Languages (INNET)   *Cologne, Germany and Nijmegen, Netherlands*
  hosted by the University of Cologne & The Language Archive (TLA) at the Max Planck Institute for Psycholinguistics
- Awarded one of two government fellowships funded by CLARIN ERIC and the European Commission to help build collaborative digital infrastructure for language research, documentation, analysis, and multimedia archives
- Participated in the 2013 INNET workshop for regional archivists to discuss server virtualization, language archiving software solutions, the CLARIN European infrastructure project for language resources, and the creation of language archive portals using the WordPress content management system

Research Assistant   **Aug 2011 – Aug 2013**
Department of Linguistics, University of Florida   *Gainesville, FL*
- Developed stimuli, conducted experiments, and analyzed results of eye-tracking and self-paced reading tasks culminating in a self-directed honors thesis using psycholinguistic methods to investigate the effects of multilingualism on plural agreement in English

SELECTED PUBLICATIONS

- **Sara Court** and Micha Elsner. 2024. Shortcomings of LLMs for Low-Resource Translation: Retrieval and Understanding are Both the Problem. To appear in Proceedings of the Ninth Conference on Machine Translation (WMT24). Miami, Florida.

- **Sara Court**, Andrea D. Sims, and Micha Elsner. 2023. Analogy in Contact: Modeling Maltese Plural Inflection. In *Proceedings of the Society for Computation in Linguistics:* Vol. 6, Article 6, pages 35-46. Amherst, Massachusetts.

- Micha Elsner and **Sara Court**. 2022. OSU at SigMorphon 2022: Analogical Inflection With Rule Features. In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 220–225, Seattle, Washington. Association for Computational Linguistics.

- Maria Copot, **Sara Court**, Noah Diewald, Stephanie Antetomaso, and Micha Elsner. 2022. A Word-and-Paradigm Workflow for Fieldwork Annotation. In *Proceedings of the Fifth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 159–169, Dublin, Ireland. Association for Computational Linguistics.

- Yu-Hsiang Tseng, Shu-Kai Hsieh, Pei-Yi Chen, and **Sara Court**. 2020. Computational Modeling of Affixoid Behavior in Chinese Morphology. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2879–2888, Barcelona, Spain (Online). International Committee on Computational Linguistics.

AWARDS & HONORS

- **FLAS Fellowship, Quechua,** *U.S. Department of Education*   **Summer 2022**
- **Honorable Mention,** *National Science Foundation Graduate Research Fellowship Program (GRFP)*   **Spring 2022**
- **Huayu Enrichment Scholarship,** *Taiwan (R.O.C.) Ministry of Education*   **Fall 2016**
- **Fellowship with INNET Project,** *CLARIN ERIC & European Commission*   **Fall 2013**

LEADERSHIP & SERVICE

President   **Aug 2023 – Aug 2024**
Student Linguistics Society (SLA)   *Columbus, OH*
- Awarded $1,000 grant from the university's Office of Diversity, Equity, Inclusion, and Justice (DEIJ) to design and implement the SLA's "hidden curriculum" mentorship program for incoming graduate students, including workshops on Imposter Syndrome, mental health, and professionalization skills
- Established SLA Teams SharePoint to streamline communication and resource sharing among grad students
- Lead revisions to optimize department infrastructure and operations resulting in new communication pipelines and more efficient collaborations among Linguistics graduate students and faculty

Vice President   **Aug 2022 – Aug 2023**
Student Linguistics Society (SLA)   *Columbus, OH*
- Created cloud-based shared teaching materials repository to support GTA instructors of undergraduate courses
- Organized recruitment weekend and prepared informational welcome materials for incoming 1st year students

Workshop Facilitator   **January 2024 – May 2024**
Center for Healthy Families   *Columbus, OH*
- Designed and delivered a 2-hour workshop on AI basics, digital literacy, and career paths in data and tech with a focus on empowering underrepresented communities in Franklin County
- Shared resources and facilitated discussions around personal narratives, obstacles, and empowerment in order to break down barriers and improve diversity of representation in the AI, software, data, and technology workforce

Organizing Committee Member   **Jan 2021 – Aug 2021**
AIMM5: 5th American International Morphology Meeting   *Virtual / Remote*
- Coordinated the conference CFP and program development and designed custom GatherTown space for virtual poster sessions and socializing during the COVID-19 pandemic

WORK EXPERIENCE

Freelance Copy Editor and Translator   **Aug 2011 – Present**
- Translate (Mandarin-English) and edit (English) academic and commercial publications in Linguistics, Healthcare, Education, Technology, and Business
- Deliver web copy, marketing materials, PhD theses, and book manuscripts while communicating with multiple stakeholders and punctually meeting short deadlines and fast turnarounds

Director of Communications   **Jan 2014 – May 2015**
Allogy   *Orlando, FL & Nairobi, Kenya*
- Led a team of writers and designers to update company websites and draft white papers, conference proposals, grant applications, and multimedia web content

Digital Content Editor   **May 2012 – Jan 2014**
Allogy   *Orlando, FL & Nairobi, Kenya*
- Collaborated with local partners to format education and healthcare materials using HTML and Allogy's CMS

NATURAL LANGUAGES

- English (native)
- Mandarin Chinese (fluent)
- Spanish (conversational)
- French (conversational)
- Modern Standard Arabic (intermediate)
- Maltese (intermediate)
- Quechua (intermediate