

Jonne Sälevä

jonnesaleva@brandeis.edu

www.jonnesaleva.com

Summary

Highly driven 4th year Ph.D. candidate in Computer Science with extensive experience in machine learning and NLP. Proven track record of taking research projects from initial ideas on a whiteboard to implementation and publication. Special focus on multilingual NLP, large language models, machine translation and statistical model evaluation. Seeking an industrial post-PhD opportunity to apply my skills and experience in a practical setting, while continuing to innovate and push the boundaries of the field.

Education

Brandeis University

Ph.D. in Computer Science, GPA: 3.93/4.00

Waltham, MA

August 2025 (expected)

⌘ Advisor: Constantine Lignos, Broadening Linguistic Technologies Lab

⌘ Topics: multilingual NLP, machine translation, information extraction, large language models

Brandeis University

M.S. in Computer Science, GPA: 3.93/4.00

Waltham, MA

May 2021

Harvard University

A.B. in Statistics, GPA: 3.64/4.00

Cambridge, MA

May 2017

Language Citation in Yiddish.

Work experience

Google LLC

Student Researcher

Remote / Cambridge, MA

May – December 2023

⌘ Researched statistical methods to identify out-of-distribution inputs to large, decoder-only language models. Applied the methods to generative natural language tasks such as open and closed-book question answering. (Supervisor: Dong Lin).

USC Information Sciences Institute

Visiting Research Assistant

Marina del Rey, CA

Summer 2022

⌘ Conducted research on linguistic heritage-aware transfer learning for low-resource language modeling and named entity recognition, with applications to diasporic languages such as Yiddish and Ladino.

⌘ Work done as part of the [ISI NLP Internship Program](#). (Supervisor: Jonathan May).

Broadening Linguistic Technologies Lab

Graduate Research Assistant

Waltham, MA

October 2019 – present

⌘ Researched the impact of randomized subword tokenization algorithms on the robustness of low-resource neural machine translation models. Findings published at the [EACL 2023 Insights from Negative Results workshop](#) (**best paper award**).

⌘ Built morphology-aware subword segmentation system to improve NMT performance on morphologically rich languages, with experiments on Nepali, Sinhala and Kazakh. Findings published at the [EACL 2021 Student Research Workshop](#).

⌘ For a complete list of publications, see my Google Scholar profile at <https://tinyurl.com/jonne-google-scholar>.

Think-A-Move Ltd.

Natural Language Processing Engineer

Beachwood, OH

November 2018 – August 2019

⌘ Developed offline knowledge-intensive dialogue systems for the Telemedicine and Advanced Technology Research Center (TATRC) and medics in the U.S. Army Ranger Battalion using Python and SQL. Deployed system on an Android end user device.

Haave Inc.

NLP Data Scientist

Denver, CO

February 2018 – November 2018

⌘ Migrated eDiscovery search engine from monolith to microservice architecture using Flask, NLTK, scikit-learn, and MongoDB. Achieved a 1600% performance increase compared to systems based on conventional LSA embeddings.

⌘ Engineered cloud pipelines for word embedding APIs and server-side rendering of visualizations using Flask, HTML and CSS.

Skills & Interests

Technical: Python (Huggingface, PyTorch, fairseq, scikit-learn, NumPy, pandas, PyMC, Flask, requests, etc.). SQL. Shell scripting. Cloud computing (AWS, Azure) and general Linux system administration. Some C and Java.

Natural languages: Finnish (native), English, German, Swedish, Yiddish, Esperanto

Interests: Noise and invariance as regularization. Bayesian statistical modeling. Learning foreign languages.