

AMANDA MYNTTI

PhD Student, TurkuNLP, University of Turku

✉ amanda.a.myntti@utu.fi ⓒ github.com/mmanteli

BASIC INFORMATION

Full name: Amanda Anni Augusta Myntti

Citizenship: Finnish

Language proficiency: Finnish (native), English (fluent), French (good), Swedish (basic), Chinese (beginner)

EDUCATION

M.S. in Mathematics

June 2023

University of Turku, Department of Mathematics and Statistics

Thesis: *Explainable AI in NLP and Image Recognition* (in Finnish, grade 5/5)

M.S. (Tech.) in Information and Communication Technology, Health Technology

June 2025

University of Turku, Department of Computing

Thesis: *Zeroshot Redaction of Personally Identifiable Information using Large Language Models* (grade 5/5)

B.S. in Mathematics

March 2020

University of Turku, Department of Mathematics and Statistics

Thesis: *The Derivative of Hardy-Littlewood Maximal Function*

ACADEMIC EXPERIENCE

PhD student, Doctoral Program in Technology

January 2025

— Present

TurkuNLP, University of Turku

Studying the structure and interpretability of LLM induced embedding spaces.

Project Researcher

September 2023

TurkuNLP, University of Turku

— December 2024

- Exploring web register universals through multilingual embeddings, their connection to genres, explaining predictions of register classification models.
- Developing a pipeline to compile question-answering datasets for low-resource languages.
- Training LLMs, curating and filtering datasets, masking Personally Identifiable Information in datasets used to train LLMs for European languages ([HPLT](#)).

Research Assistant

May 2021

TurkuNLP, University of Turku

— September 2021

- Training register classification models for multilingual web data.
- Implementing a pipeline to explain the classifier's predictions and extract keywords representing the register classes. Developing metrics to evaluate the results.

TEACHING EXPERIENCE

Teaching assistant

Fall 2021

Department of Mathematics and Statistics, University of Turku

- Hosted exercise sessions for the course Analysis 1

WORK EXPERIENCE

Algorithm design student

April 2021

Nordic Semiconductor

— May 2022

- Simulating fading in 5G channels in different conditions, verifying and evaluating simulation results.

FUNDING AND GRANTS

AI-DOC 1000 PhD pilot, full funding, 3 years

2024

Finnish Doctoral Program Network in Artificial Intelligence

PUBLICATIONS

- S. Oepen et al. 2025
HPLT 3.0: Very Large-Scale Multilingual Resources for LLM and MT. Mono- and Bi-lingual Data, Multilingual Evaluation, and Pre-Trained Models
Preprint
<https://doi.org/10.48550/arXiv.2511.01066>
- A. Myntti, E. Henriksson, V. Laippala, S. Pyysalo 2025
Register Always Matters: Analysis of LLM Pretraining Data Through the Lens of Language Variation
Second Conference on Language Modeling (COLM 2025)
<https://openreview.net/forum?id=FqXXtSZWEZ>
- L. Burchell et al. 2025
An Expanded Massive Multilingual Dataset for High-Performance Language Technologies
Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL 2025)
<https://aclanthology.org/2025.acl-long.854/>
- A. Myntti, L. Repo, E. Freyermuth, A. Kanner, V. Laippala, E. Henriksson 2024
Intersecting Register and Genre: Exploring the Contents of Web-Crawled Corpora
Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities at EMNLP 2024
<https://aclanthology.org/2024.nlp4dh-1.38/>
- E. Henriksson, A. Myntti, S. Hellström, S. Erten-Johansson, A. Eskelinen, L. Repo, V. Laippala 2024
From Discrete to Continuous Classes: A Situational Analysis of Multilingual Web Registers with LLM Annotations
Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities at EMNLP 2024
<https://aclanthology.org/2024.nlp4dh-1.30/>
- E. Henriksson, A. Myntti, V. Laippala 2024
Using deep learning to examine cross-linguistic similarities of registers
Digital Research Data and Human Sciences (DRDHum) 2024
- E. Henriksson, A. Myntti, A. Eskelinen, S. Erten-Johansson, S. Hellström, V. Laippala 2024
Automatic register identification for the open web using multilingual deep learning
Natural Language Processing (in print)
<https://arxiv.org/abs/2406.19892>
- A. Eskelinen, A. Myntti, E. Henriksson, S. Pyysalo, V. Laippala. 2024
Building Question-Answer Data Using Web Register Identification
Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)
<https://aclanthology.org/2024.lrec-main.234>
- S. Rönnqvist, A.-J. Kyröläinen, A. Myntti, F. Ginter, V. Laippala. 2022
Explaining Classes through Stable Word Attributions
Findings of the Association for Computational Linguistics: ACL 2022
<https://aclanthology.org/2022.findings-acl.85/>

PRESENTATIONS

- DHNB 2024 in Reykjavik, Iceland** May 2024
Intersecting Register and Genre: Exploring the Contents of Web-Crawled Corpora
Oral presentation
- NLP4DH workshop in EMNLP 2024, Miami, USA** November 2024
Intersecting Register and Genre: Exploring the Contents of Web-Crawled Corpora
Poster
- COLM 2025, Montreal, Canada** October 2025
Register Always Matters: Analysis of LLM Pretraining Data Through the Lens of Language Variation
Poster