


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 13.6.2023
University of Turku, Department of Mathematics and Statistics
Thesis: *Explainable AI in NLP and Image Recognition* (in Finnish, grade 5)

M.S. (Tech.) in Information and Communication Technology, Health Technology exp. Spring 2025
University of Turku, Department of Computing
Thesis: *Zeroshot Redaction of Personally Identifiable Information using Large Language Models* (in progress)

B.S. in Mathematics 27.3.2020
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

Project Researcher September 2023 — December 2024
TurkuNLP, University of Turku

- 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 — September 2021
TurkuNLP, University of Turku

- 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 — May 2022
Nordic Semiconductor

- 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

- A. Myntti, E. Henriksson, V. Laippala, S. Pyysalo 2025
Register Always Matters: Analysis of LLM Pretraining Data Through the Lens of Language Variation
Preprint
<https://arxiv.org/abs/2504.01542>
- L. Burchell et al. 2025
An Expanded Massive Multilingual Dataset for High-Performance Language Technologies
Preprint
<https://arxiv.org/abs/2503.10267>
- 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
Preprint
<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. Rönqvist, 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
Short presentation
- NLP4DH workshop in EMNLP 2024, Miami, USA** November 2024
Intersecting Register and Genre: Exploring the Contents of Web-Crawled Corpora
Poster