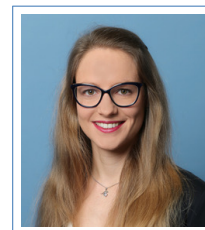


Noëmi Aepli

Curriculum Vitae

✉ naepli@cl.uzh.ch
<https://noe-eva.github.io/>
🐦 @noeminaepli
🌐 noe-eva
in noemi-aepli



- since 2020 **PhD Candidate**, Department of Computational Linguistics, University of Zurich, Switzerland.
Dissertation Topic: *Natural Language Processing for Low-Resourced Language Variations*
Supervisor: Prof. Dr. Rico Sennrich
- 2014 – 2017 **Master Computational Linguistics & Language Technology**, Minor: *General Linguistics*, University of Zurich, Switzerland.
1 semester *Language & Communication Technologies*, University of Saarbrücken, Germany.
Thesis Topic: *Dependency Parsing for Swiss German via Annotation Projection*
- 2010 – 2014 **Bachelor Computational Linguistics & Language Technology**, Minors: *General Linguistics & Sinology*, University of Zurich, Switzerland.
1 semester *Mandarin Chinese*, Communication University of China, Beijing.

Employment

- 2018 – 2020 **Data Scientist**, *Swisscom*, Zurich, Switzerland.
Collaboration in various text & speech processing projects of **Data, Analytics & AI Group**.
- 2011 – 2021 **Teaching Assistant**, *Department of Computational Linguistics*, University of Zurich.
Introduction to Language Data Processing (2021), Voice Analysis (2019), Speech Perception (2019), Text Mining (2018), Quantitative Methods (2016), Introduction to Computational Linguistics (2011)
- 2011 – 2017 **Student Assistant**, *Department of Computational Linguistics & CorpusLab*, University of Zurich.
Collaboration in several NLP projects on Machine Translation & Sentiment Analysis.
- 2015 – 2017 **Freelancing**, *Spi:tech AG*, Zurich, Switzerland.
Transcription of Swiss German speech & German to Swiss German translation.
- 2013, 3 mths **Student Assistant**, *Law Sources Foundation*, University of Zurich.
Developing an XML schema for transcriptions of Swiss law sources.
- 2012, 6 mths **Intern**, *Digital Enterprise Research Institute (DERI)*, Galway, Ireland.
Unit for Natural Language Processing: Information Extraction with GATE and JAPE rules.
- 2009, 1 year **Intern**, *Wegelin & Co. Private Bankers*, St. Gallen, Switzerland.
Clearing & Settlement Department: Transfer of securities, stocks and bonds between banks.

Scholarships

- 2020, 4 years **SNSF Doc.CH Grant** Swiss National Science Foundation funding for my own PhD project.
- 2014, 2 years **Erasmus Mundus** European Masters Program in *Language and Communication Technologies*.
- 2013, 6 mths **Confucius Institute (Hanban)** Chinese Language Studies at the *Communication University of China*, Beijing.
- 2015, 1 mth **Italian Culture Institute Zurich** Italian Language Course at *Accademia Europea di Firenze*.

Skills

- Programming Python (advanced); Prolog, Perl, Java, C++, R (basic)
- Languages Swiss German (native), German, English (C2, CPE), Italian (C1), Chinese (B2, HSK 5), French (B1), Russian (A2), Spanish (A2), Finnish (A1). Also: Latin, Ancient Chinese, Manchu (basic).

Publications (→ *GoogleScholar*)

- 2022 **Aepli, Noëmi** and Rico Sennrich. Improving zero-shot cross-lingual transfer between closely related languages by injecting character-level noise. In *arXiv preprint; accepted at ACL2022*, 2022.
- 2021 Annette Rios, Chantal Amrhein, **Aepli, Noëmi**, and Rico Sennrich. On biasing transformer attention towards monotonicity. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4474–4488, Online, June 2021. Association for Computational Linguistics.
- 2019 Tilia Ellendorff, Lenz Furrer, Nicola Colic, **Aepli, Noëmi**, and Fabio Rinaldi. Approaching SMM4H with merged models and multi-task learning. In *Proceedings of the Fourth Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task*, pages 58–61, Florence, Italy, August 2019. Association for Computational Linguistics.
- 2017 Marcos Zampieri, Shervin Malmasi, Nikola Ljubešić, Preslav Nakov, Ahmed Ali, Jörg Tiedemann, Yves Scherrer, and **Aepli, Noëmi**. Findings of the VarDial evaluation campaign 2017. In *Proceedings of the Fourth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pages 1–15, Valencia, Spain, April 2017. Association for Computational Linguistics.
- 2016 Martin Volk, Chantal Amrhein, **Aepli, Noëmi**, Mathias Müller, and Phillip Ströbel. Building a parallel corpus on the world’s oldest banking magazine. In *Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, pages 288–296, 2016.
- 2016 Simon Clematide, Karina Frick, **Aepli, Noëmi**, and Jean-Philippe Goldman. Crowdsourcing swiss dialect transcriptions for assessing factors in writing variations. *Bochumer Linguistische Arbeitsberichte*, page 62, 2016.
- 2015 Nora Hollenstein and **Aepli, Noëmi**. A Resource for Natural Language Processing of Swiss German Dialects. In *Proceedings of the International Conference of the German Society for Computational Linguistics and Language Technology*, pages 108–109, Duisburg-Essen, Germany, October 2015. German Society for Computational Linguistics and Language Technology.
- 2014 **Aepli, Noëmi**, Ruprecht von Waldenfels, and Tanja Samardžić. Part-of-speech tag disambiguation by cross-linguistic majority vote. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, pages 76–84, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University.
- 2014 Nora Hollenstein and **Aepli, Noëmi**. Compilation of a Swiss German dialect corpus and its application to PoS tagging. In *Proceedings of the First Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects*, pages 85–94, Dublin, Ireland, August 2014. Association for Computational Linguistics and Dublin City University.
- 2013 **Aepli, Noëmi** and Martin Volk. Reconstructing complete lemmas for incomplete german compounds. In *Language Processing and Knowledge in the Web*, pages 1–13. Springer, 2013.

Resources & Software

- Resource **GSW-UD Treebank**, *Universal Dependency Treebank for Swiss German*, First Swiss German
2019 treebank with manual universal dependency annotations.
https://github.com/UniversalDependencies/UD_Swiss_German-UZH
- Resource **NOAH’s Corpus**, First text corpus of written Swiss German manually annotated with part-of-
2011 – 2019 speech tags used for the training of a part-of-speech tagger.
<https://noe-eva.github.io/NOAH-Corpus/>
- Resource **ArchiMob Corpus**, Corpus of spoken Swiss German aligned with transcriptions, annotated with
2016 part-of-speech tags and normalisations.
<https://www.spur.uzh.ch/en/departments/research/textgroup/ArchiMob.html>