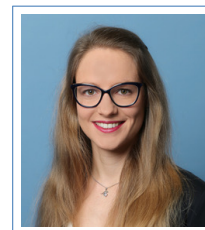


Noëmi Aepli

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*
1 semester at the University of California, Berkeley, at the Speech and Computation Lab
- 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 the **Data, Analytics & AI Group**.
- 2011 – 2021 **Co-Teaching & Teaching Assistant**, *Department of Computational Linguistics*.
Introduction to Language Data Processing (2021), Seminar on Automatic Speech and Text Analysis (2020), 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); various ML & DL frameworks.
- 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).

Resources

- 2023 **Swiss German Part-of-Speech Tagging Models**, published on the Huggingface Model Hub..
https://huggingface.co/noeminaepli/swiss_german_stts_pos_model
https://huggingface.co/noeminaepli/swiss_german_pos_model
- 2023 **SID4LR data set**, VarDial 2023 shared task: Slot and Intent Detection for Low-Resource Languages. Extension of the xSID-0.4 corpus with Neapolitan and Swiss German.
<https://bitbucket.org/robvander/sid4lr/src/master>
- 2022 **ITDI data set**, VarDial 2022 Shared Task: Identification of Languages and Dialects of Italy.
https://github.com/noe-eva/ITDI_2022
- 2019 **GSW-UD Treebank**, Universal Dependency Treebank for Swiss German: First Swiss German treebank with manual universal dependency annotations.
https://github.com/UniversalDependencies/UD_Swiss_German-UZH
- 2011 – 2019 **NOAH's Corpus**, First text corpus of written Swiss German manually annotated with part-of-speech tags used for the training of a part-of-speech tagger.
<https://noe-eva.github.io/NOAH-Corpus/>
- 2016 **ArchiMob Corpus**, Corpus of spoken Swiss German aligned with transcriptions, annotated with part-of-speech tags and normalisations.
<https://www.spur.uzh.ch/en/departments/research/textgroup/ArchiMob.html>

Publications (→ *Google Scholar*)

- 2023 **Aepli, Noëmi**, Çağrı Çöltekin, Rob van der Goot, Tommi Jauhiainen, Mourhaf Kazzaz, Nikola Ljubešić, Kai North, Barbara Plank, Yves Scherrer, and Marcos Zampieri. Findings of the VarDial evaluation campaign 2023. In *Proceedings of the Tenth Workshop on NLP for Similar Languages, Varieties and Dialects*, Dubrovnik, Croatia, May 2023. Association for Computational Linguistics.
- 2022 **Aepli, Noëmi** and Rico Sennrich. Improving zero-shot cross-lingual transfer between closely related languages by injecting character-level noise. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 4074–4083, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- 2022 **Aepli, Noëmi**, Antonios Anastasopoulos, Adrian Chifu, William Domingues, Fahim Faisal, Mihaela Găman, Radu Tudor Ionescu, and Yves Scherrer. Findings of the VarDial evaluation campaign 2022. In *Proceedings of the Ninth Workshop on NLP for Similar Languages, Varieties and Dialects*, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics (ICCL).
- 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.