

VILÉM ZOUHAR

vilem.zouhar@gmail.com | vilda.net

EDUCATION

Charles University in Prague (3 years) 2017/18 - 2019/20

BSc. in Computer Science (+ graduate-level classes)

Specialization in Computational linguistics

Thesis [Enabling Outbound Machine Translation](#)

Saarland University + Groningen University (2 years) 2020/21 - exp. 2021/22

MSc. in Language Science and Technology (funded scholarship)

Double degree programme

Thesis Knowing What's Important: Automatic Knowledge Base Article Splitting & Filtering

SELECTED ACADEMIC PROJECTS AND PUBLICATIONS

[Google Scholar](#)

Neural Machine Translation Quality and Post-Editing Performance [EMNLP 2021](#)

Vilém Zouhar, Ondřej Bojar, Martin Popel, Aleš Tamchyna

Providing Backtranslation Improves Users Confidence in MT, Not Quality [NAACL 2021](#)

Vilém Zouhar, Michal Novák, Matúš Žilínek, Ondřej Bojar, Mateo Obregón, Robin L. Hill, Frédéric Blain, Marina Fomicheva, Lucia Specia, Lisa Yankovskaya

Artefact Retrieval: Overview of NLP Models with Knowledge Base Access [AKBC CSKB 2021](#)

Vilém Zouhar, Marius Mosbach, Debanjali Biswas, Dietrich Klakow

Sampling and Filtering of Neural Machine Translation Distillation Data [NAACL SRW 2021](#)

Vilém Zouhar

Leveraging Neural Machine Translation for Word Alignment [PBML 116](#)

Vilém Zouhar, Daria Pylypenko

WMT20 Document-Level Markable Error Exploration [WMT20](#)

Vilém Zouhar, Tereza Vojtěchová, Ondřej Bojar

Extending Ptakopět for MT User Interaction Experiments [PBML 115](#)

Vilém Zouhar, Michal Novák

Outbound Translation User Interface Ptakopet: A Pilot Study [LREC 2020](#)

Vilém Zouhar, Ondřej Bojar

A Collection of Machine Learning Exercises 2018/2019

50 pages of ML tasks in R; full version available per request (used as [teaching material](#))

Awarded Student Faculty Grant at MFF Charles University

Paper on machine translation with eyetracking in writing

Collaborator

Paper on machine translation with EEG in writing

Collaborator

Paper on efficient knowledge base use in writing

Main author

Paper on partial answers as artefacts for classification in writing

Main author

TECHNICAL KNOWLEDGE

LANGUAGE PROFICIENCY

Programming Python, JS/TS, Rust, C/C++, R

Toolkits PyTorch, Scikit, Marian NMT

Misc. UNIX (e.g. GPU cluster), Matplotlib, LaTeX
Always trying to learn more

Czech Native

English C2 (*iBT TOEFL 118/120*)

German B2 (*in development*)

Polish A1

WORK EXPERIENCE

[Spoken Language Systems](#) group (student research assistant) (14 months) 2021-2022

- University of Saarland (Germany)
- Information retrieval efficiency through dimensionality reduction ([code](#))
- Language modelling with an external source of information

Neural Networks Implementation and Application class (tutor) winter semester 2021

Statistical Natural Language Processing class (tutor) summer semester 2021

- University of Saarland (Germany)
- Weekly tutorials for students
- Preparation of the [SNLP class material](#), [NN class material](#) and the final exam
- Designing and grading weekly assignments and the final project

[Institute of Formal and Applied Linguistics](#) (student research assistant) (2 years+) 2019-present

- Charles Univeristy (Czech Republic)
- Machine translation related projects
- [Bergamot project](#) (in-browser MT)
- Psycholinguistic project consultation
- Miscellaneous research tasks

Previo (intern software dev) (3 months) summer 2018

- Development of multilayer CMS using JS, PHP, Zend and MySQL

BIM Project (intern software dev) (3 months) summer 2017

- Development of plugins for the ArchiCAD suite with C++/Boost and C#

Web development 2015-2017

- Participation in several commercial website projects using the PHP/JS/HTML/CSS stack

ACADEMIC MISC.

[CSRR](#): Workshop on Commonsense Representation and Reasoning ACL 2022

- Organizer

Department of Formal and Applied Linguistics 2019-present

- NER Presentation at NKÚ (supreme audit office)
- Department presentation at Open Days
- ELITR project coordination at a hackathon [UniHack](#)

[MachineTranslate.org](#): open guide to machine translation 2021-present

- Contributor of articles

Smaller projects

- [Pandemic Crisis Communication](#): Automatic Classification of Interviews With Experts 2021
- [Fact Learning](#) with Adaptive Color Palette: Effect of Stimuli-Independent Hints 2021
- [Hyperparameters of RNN Architectures](#): for POS Tagging using Surface-Level BERT Embeddings 2021
- [Deep Molecule](#): Quantitative Structure-Property Relationships 2021
- [SlowAlign](#): IBM model-based word aligned with extra features and heuristics 2020
- [SlowAlign Displayer](#): Quick online word-alignment visualization tool 2020
- [MosQEto](#): Machine translation quality estimation data synthesis 2019
- Other small projects either for convenience, out of professional interest, or as a hobby, hosted at [GitHub](#)

EXTRA-CURRICULAR

Academic senate 2018-2020

- Member of the Academic Senate at Charles University Faculty of Mathematics and Physics
- Participation in Faculty meetings, communicating with students, introductory summer camp

Game Jams 2015-2020

- Several [games](#) programmed and presented in limited time, mostly Ludum Dare

Kasiopea 2017-2019

- Organization of [Kasiopea](#), an annual coding competition for talented high school students