

# VILÉM ZOUHAR

vilem.zouhar@gmail.com | vilda.net

## EDUCATION

<b>Charles University in Prague</b> BSc. in Computer Science (+ graduate-level classes) Specialization in Computational linguistics Thesis <a href="#">Enabling Outbound Machine Translation</a>	(3 years) 2017 - 2020
<b>Saarland University + Groningen University</b> MSc. in Language Science and Technology (funded scholarship) Double degree programme Thesis <a href="#">Shrinking Knowledge Base Size: Dimension Reduction, Splitting &amp; Filtering</a>	(2 years) 2020 - 2022
<b>ETH Zürich</b> PhD with Mrinmaya Sachan	(expected) 2022

## SELECTED ACADEMIC PROJECTS AND PUBLICATIONS

[Google Scholar](#)

Knowledge Base Index Compression via Dimensionality and Precision Reduction <i>Vilém Zouhar, Marius Mosbach, Miaoran Zhang, Dietrich Klakow</i>	<a href="#">ACL Spa-NLP 2022</a>
EMMT: A simultaneous eye-tracking, 4-electrode EEG and audio corpus for multi-modal reading and translation scenarios <i>Sunit Bhattacharya, Věra Kloudová, Vilém Zouhar, Ondřej Bojar</i>	<a href="#">In review</a>
Neural Machine Translation Quality and Post-Editing Performance <i>Vilém Zouhar, Ondřej Bojar, Martin Popel, Aleš Tamchyna</i>	<a href="#">EMNLP 2021</a>
Providing Backtranslation Improves Users Confidence in MT, Not Quality <i>Vilém Zouhar, Michal Novák, Matúš Žilinc, Ondřej Bojar, Mateo Obregón, Robin L. Hill, Frédéric Blain, Marina Fomicheva, Lucia Specia, Lisa Yankovskaya</i>	<a href="#">NAACL 2021</a>
Artefact Retrieval: Overview of NLP Models with Knowledge Base Access <i>Vilém Zouhar, Marius Mosbach, Debanjali Biswas, Dietrich Klakow</i>	<a href="#">AKBC CSKB 2021</a>
Sampling and Filtering of Neural Machine Translation Distillation Data <i>Vilém Zouhar</i>	<a href="#">NAACL SRW 2021</a>
Leveraging Neural Machine Translation for Word Alignment <i>Vilém Zouhar, Daria Pylypenko</i>	<a href="#">PBML 116</a>
WMT20 Document-Level Markable Error Exploration <i>Vilém Zouhar, Tereza Vojtěchová, Ondřej Bojar</i>	<a href="#">WMT20</a>
Extending Ptakopět for MT User Interaction Experiments <i>Vilém Zouhar, Michal Novák</i>	<a href="#">PBML 115</a>
Outbound Translation User Interface Ptakopet: A Pilot Study <i>Vilém Zouhar, Ondřej Bojar</i>	<a href="#">LREC 2020</a>
A Collection of Machine Learning Exercises <i>50 pages of ML tasks in R; full version available per request (used as <a href="#">teaching material</a>) Awarded Student Faculty Grant at MFF Charles University</i>	2018/2019

## TECHNICAL KNOWLEDGE

<b>Programming</b>	Python, JS/TS, Rust, C/C++, R
<b>Toolkits</b>	PyTorch, Scikit, Marian NMT
<b>Misc.</b>	UNIX (e.g. GPU cluster), Matplotlib, LaTeX Always trying to learn more

## LANGUAGE PROFICIENCY

<b>Czech</b>	Native
<b>English</b>	C2 ( <i>iBT TOEFL 118/120</i> )
<b>German</b>	B2 ( <i>in development</i> )
<b>Polish</b>	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) (3 years) 2019-2022

- 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 (organizer) ACL 2022

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 (contributor) 2021-present

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