

# VILÉM ZOUHAR

vilem.zouhar@gmail.com | vilda.net

## EDUCATION

---

	<b>ETH Zürich</b> PhD in Computer Science in <a href="#">Mrinmaya Sachan's lab</a> Focus: machine translation of scientific, textbook and literary texts	2022-present
	<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>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

## TECHNICAL KNOWLEDGE

<b>Programming</b>	Python, JS/TS, Rust, C/C++, R
<b>Toolkits</b>	PyTorch, Scikit, Numpy, Huggingface, Marian NMT, Matplotlib
<b>Misc.</b>	Linux (long-term user, GPU cluster etc), visualization, typesetting (pandoc, LaTeX)

## 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>Others</b>	bits of random languages (linguistic curiosity)

## TEACHING

---

Neural Networks Implementation and Application class (tutor)	winter semester 2021
Statistical Natural Language Processing class (tutor)	summer semester 2021
<ul style="list-style-type: none"><li>- University of Saarland (Germany)</li><li>- Weekly tutorials for students</li><li>- Preparation of the <a href="#">SNLP class material</a>, <a href="#">NN class material</a> and the final exam</li><li>- Designing and grading weekly assignments and the final project</li></ul>	

## WORK EXPERIENCE

---

<a href="#">Spoken Language Systems</a> group (student research assistant)	(14 months) 2021-2022
<ul style="list-style-type: none"><li>- University of Saarland (Germany)</li><li>- Information retrieval efficiency through dimensionality reduction (<a href="#">code</a>)</li><li>- Language modelling with an external source of information</li></ul>	
<a href="#">Institute of Formal and Applied Linguistics</a> (student research assistant)	(3 years) 2019-2022
<ul style="list-style-type: none"><li>- Charles University (Czech Republic)</li><li>- Machine translation related projects</li><li>- <a href="#">Bergamot project</a> (in-browser MT)</li><li>- Psycholinguistic project consultation</li><li>- Miscellaneous research tasks</li></ul>	
Previo (intern software dev)	(3 months) summer 2018
<ul style="list-style-type: none"><li>- Development of multilayer CMS using JS, PHP, Zend and MySQL</li></ul>	
BIM Project (intern software dev)	(3 months) summer 2017
<ul style="list-style-type: none"><li>- Development of plugins for the ArchiCAD suite with C++/Boost and C#</li></ul>	
Web development	2015-2017
<ul style="list-style-type: none"><li>- Participation in several commercial website projects using the PHP/JS/HTML/CSS stack</li></ul>	

## SELECTED ACADEMIC PROJECTS AND PUBLICATIONS

Google Scholar

Fusing Sentence Embeddings Into LSTM-based Autoregressive Language Models <i>Vilém Zouhar, Marius Mosbach, Dietrich Klakow</i>	In review
Knowledge Base Index Compression via Dimensionality and Precision Reduction <i>Vilém Zouhar, Marius Mosbach, Miaoran Zhang, Dietrich Klakow</i>	ACL Spa-NLP 2022
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>	In review
Neural Machine Translation Quality and Post-Editing Performance <i>Vilém Zouhar, Ondřej Bojar, Martin Popel, Aleš Tamchyna</i>	EMNLP 2021
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>	NAACL 2021
Artefact Retrieval: Overview of NLP Models with Knowledge Base Access <i>Vilém Zouhar, Marius Mosbach, Debanjali Biswas, Dietrich Klakow</i>	AKBC CSKB 2021
Sampling and Filtering of Neural Machine Translation Distillation Data <i>Vilém Zouhar</i>	NAACL SRW 2021
Leveraging Neural Machine Translation for Word Alignment <i>Vilém Zouhar, Daria Pylypenko</i>	PBML 116
WMT20 Document-Level Markable Error Exploration <i>Vilém Zouhar, Tereza Vojtěchová, Ondřej Bojar</i>	WMT20
Extending Ptakopět for MT User Interaction Experiments <i>Vilém Zouhar, Michal Novák</i>	PBML 115
Outbound Translation User Interface Ptakopet: A Pilot Study <i>Vilém Zouhar, Ondřej Bojar</i>	LREC 2020
A Collection of Machine Learning Exercises <i>50 pages of ML tasks in R; full version available per request (used as teaching material)</i> <i>Awarded Student Faculty Grant at MFF Charles University</i>	2018/2019

## SERVICE

Reviewing: CoNLL 2022, ACL-IJCNLP 2022	
CSRR: Workshop on Commonsense Representation and Reasoning (organizer)	ACL 2022
Institute of Formal and Applied Linguistics, Charles University	2019-2022
- NER Presentation at NKÚ (supreme audit office)	
- Department presentation at Open Days	
- ELITR project coordination at a hackathon UniHack	

## ACADEMIC MISC.

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
<ul style="list-style-type: none"><li>- Member of the Academic Senate at Charles University Faculty of Mathematics and Physics</li><li>- Participation in Faculty meetings, communicating with students, introductory summer camp</li></ul>	
Game jams	2015-present
<ul style="list-style-type: none"><li>- Several <a href="#">games</a> programmed and presented in limited time, mostly Ludum Dare</li></ul>	
Kasiopea	2017-2019
<ul style="list-style-type: none"><li>- Organization of <a href="#">Kasiopea</a>, an annual coding competition for talented high school students</li></ul>	
Electric guitar	2021-present
<ul style="list-style-type: none"><li>- <a href="#">Random Strum Pattern Generator</a></li></ul>	