

VILÉM ZOUHAR

vilem.zouhar@gmail.com | vilda.net

EDUCATION

	ETH Zürich PhD in Computer Science in Mrinmaya Sachan's lab Thesis & focus TBD	2022-present
	Saarland University + Groningen University MSc. in Language Science and Technology (funded scholarship) Double degree programme Thesis Shrinking Knowledge Base Size: Dimension Reduction, Splitting & Filtering	(2 years) 2020 - 2022
	Charles University in Prague BSc. in Computer Science (+ graduate-level classes) Specialization in Computational linguistics Thesis Enabling Outbound Machine Translation	(3 years) 2017 - 2020

TECHNICAL KNOWLEDGE

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

LANGUAGE PROFICIENCY

Czech	Native
English	C2 (<i>iBT TOEFL 118/120</i>)
German	B2 (<i>in development</i>)
Others	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">- 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	

WORK EXPERIENCE

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

SELECTED ACADEMIC PROJECTS AND PUBLICATIONS

Google Scholar

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

ACADEMIC MISC.

CSRR: Workshop on Commonsense Representation and Reasoning (organizer)	ACL 2022
Department of Formal and Applied Linguistics	2019-2022
<ul style="list-style-type: none">- 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
<ul style="list-style-type: none">- 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-present
<ul style="list-style-type: none">- Several games programmed and presented in limited time, mostly Ludum Dare	
Kasiopea	2017-2019
<ul style="list-style-type: none">- Organization of Kasiopea, an annual coding competition for talented high school students	
Electric guitar	2021-present
<ul style="list-style-type: none">- Random Strum Pattern Generator	