

# 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) 2017/18 - 2019/20
<b>Saarland University + Groningen University</b> MSc. in Language Science and Technology (fully funded scholarship) Double degree programme Thesis Knowing What's Important: Automatic Knowledge Base Article Splitting & Filtering	(2 years) 2020/21 - exp. 2021/22

## SELECTED ACADEMIC PROJECTS AND PUBLICATIONS

[Google Scholar](#)

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
Paper on machine translation with eyetracking <i>Collaborator</i>	in writing
Paper on machine translation with EEG <i>Collaborator</i>	in writing
Paper on efficient knowledge base use <i>Main author</i>	in writing

## TECHNICAL KNOWLEDGE

<b>Programming</b>	Python, JS/TS, Rust, R, C/C++
<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) (9 months) 2021
- University of Saarland (Germany)
  - Information retrieval efficiency through dimensionality reduction ([code](#))
  - Language modelling with 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
  - Coordination among 2 other tutors
- [Institute of Formal and Applied Linguistics](#) (student research assistant) 2018-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
  - Brno bez vizuálního smogu, MrFox, Velab, Intimedcare

## ACADEMIC MISC.

---

- Department of Formal and Applied Linguistics 2019-2020
- NER Presentation at NKÚ (supreme audit office)
  - Department presentation at Open Days
  - ELITR project coordination at a hackaton [UniHack](#)
- [Deep Molecule](#) Quantitative Structure-Property Relationships 2020-present
- Experimenting with bioinformatics, predicting molecule key temperature points using DL

## 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
- Misc. projects present
- Small projects either for convenience, out of professional interest, or as a hobby
  - Code hosted at [GitHub](#)
- Game Jams 2015-2020
- Various game jams, mostly Ludum Dare
  - Several [games](#) programmed and presented in limited time
- Kasiopea 2017-2019
- Organization of Kasiopea, an annual coding competition for talented highschool students