

## PERSONAL INFORMATION

**Ondřej Bojar**

 Za Pohořelcem 15, Praha 6, 169 00, Czech Republic

 +420 724003635  +420 951554276

 [bojar@ufal.mff.cuni.cz](mailto:bojar@ufal.mff.cuni.cz)

 <http://ufal.mff.cuni.cz/ondrej-bojar>

 Skype [ondrej.bojar](#)

Gender Male | Date of birth 7 March 1979

Nationality Czech



## WORK EXPERIENCE

- 2004 – Present **Charles University in Prague, Faculty of Mathematics and Physics**  
Research assistant; since 2010 assistant professor  
Participating in research projects, teaching
- 2004 – Present **Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering**  
Teaching assistant
- 2007 **University of Melbourne; Australia**  
Tutor for the course Data on the Web
- 2000–2005 **Internet Info, s.r.o; Czech Republic**  
Internet Analyst; programming and analysis, Internet protocols
- 1995–2003 **Primary School Fr. Plamínkové, Praha 7; Czech Republic**  
Teacher, Computer courses (grades 8 and 9)
- 1996–1997 **Appleklub / KPPM**  
Co-founder and vice-chairman of the Czech Macintosh User Group
- 1994–1997 **Technical Journalist**  
Articles in the Czech Macworld (not in the list of publications below)

## EDUCATION AND TRAINING

- 2003–2008 **PhD – Thesis Title: 'Exploiting Linguistic Data in Machine Translation'** ISCED-97 6 /  
ISCED-2011 840  
Charles University in Prague, Faculty of Mathematics and Physics  
Thesis supervisor: Doc. RNDr. Vladislav Kuboň, Ph.D.
- 1997–2003 **Master degree (summa cum laude, RNDr.)** ISCED-97 5A /  
ISCED-2011 740

Charles University in Prague, Faculty of Mathematics and Physics  
Thesis supervisor: Doc. RNDr. Vladislav Kuboň, Ph.D.

### 1997–1999 Parallel study in master programme; finished 4 semesters

Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering

#### STAYS ABROAD

- 2010 Two-month research stay at RWTH Aachen University, Germany
- 2006–2007 Twelve-month research visit at CSSE, University of Melbourne, Australia
- Jul-Aug 2006 Language Engineering Workshop (Machine Translation Team) at JHU, Baltimore, USA (6 weeks)
- 2005 Two-month research stay at RWTH Aachen University, Germany
- 2003-2004 Six-month study and research stay at University of Saarland, Saarbrücken, Germany

#### PERSONAL SKILLS

Mother tongue Czech

#### Other languages

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken interaction	Spoken production	
English	C2	C2	C2	C2	C2
Certificate in Advanced English					
German	B1	B2	B2	B2	B2
Zentrale Mittelstufenprüfung					

Levels: A1/A2: Basic user – B1/B2: Independent user – C1/C2: Proficient user  
[Common European Framework of Reference \(CEF\) level](#)

**Organisational skills**

- organizing workshops (~100 participants): MT Marathon in 2009, 2013, 2015, 2016, 2018
- organizing conferences (~100 participants): EAMT 2017
- managing local contributions to EU grants, see below
- supervising student school team projects

**Computer skills**

- programming languages Perl, Mercury, Prolog, PHP, SQL (excellent knowledge), Python, C, C++, Java (good)
- extensive programming experience with Unix (and previously Mac OS and DOS)
- extensive experience with cluster computing, big data processing, parallelization
- extensive programming experience with TCP/IP, Internet

#### INVOLVEMENT IN PROJECTS

- to start 2019 EU H2020 RIA ELITR (European Live Translator)  
Coordinator
- to start 2019 EU H2020 RIA Bergamot (MT in Browser)  
Principal Investigator of the partner Charles University
- since 2018 GAČR: Multi-Lingual Machine Translation  
Principal Investigator
- 2015–2018 EU H2020 IA HimL (Health in my Language)  
Leader of the partner Charles University
- 2015–2018 EU H2020 RIA QT21 (Quality Translation 21)  
Leader of the partner Charles University

- 2015–2018 EU H2020 CSA CRACKER (Cracking the Language Barrier)  
Team member
- 2012–2015 EU FP7 CSA MosesCore (Support for the Moses translation system)  
Principal Investigator of the partner Charles University
- 2011–2013 GAČR: Czech in the Machine Translation Era  
Team member
- 2010–2012 GAČR PGS: Combining Phrase-Based and Deep Syntactic Machine Translation  
Principal Investigator
- 2009–2012 EU FP7 Strep EuroMatrixPlus (Bringing Machine Translation for European Languages to the User)  
Leader of the partner Charles University and Principal Investigator for the co-financed part 7E11051 (2011–2012)
- 2006–2009 EU FP6 Strep EuroMatrix (Machine translation between European languages)  
Team member
- 2005–2006 GAUK Czech-English Machine Translation Dictionary  
Principal Investigator
- since 2004 PCEDT (The Prague Czech-English Dependency Treebank)  
Team member
- 2004–2005 VALLEX (Valency Lexicon of Czech Verbs)  
Team member
- 2003–2005 Machine translation project of economical texts from Czech to English  
Team member
- 2000–2002 Software project The ENTs—A simulation of natural environment with human-like computer-driven agents  
Team member

#### AWARDS

---

- 2011 1. Meta Prize to the Moses team  
<http://www.meta-net.eu/meta-prize>

#### SELECTED PUBLICATIONS

---

- Ondřej Bojar. 2017. Strojový překlad. In Petr Karlík, Marek Nekula, and Jana Pleskalová, editors, *Nový encyklopedický slovník češtiny*. Nakladatelství Lidové noviny, Praha, Czechia.
- Ondřej Bojar, 2015. *Machine translation*, chapter 13, pages 323–347. Oxford Handbooks in Linguistics. Oxford University Press, Oxford, UK.
- Ondřej Bojar. 2012. *Čeština a strojový překlad (Czech Language and Machine Translation)*, volume 11 of *Studies in Computational and Theoretical Linguistics*. ÚFAL, Praha, Czechia.
- Ondřej Cífk a Ondřej Bojar. 2018. Are BLEU and Meaning Representation in Opposition? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1362–1371. Association for Computational Linguistics, Association for Computational Linguistics.
- Matthias Huck, Aleš Tamchyna, Ondřej Bojar, and Alexander Fraser. 2017. Producing Unseen Morphological Variants in Statistical Machine Translation. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 369–375, Stroudsburg, PA, USA. Universitat Politècnica de València, Association for Computational Linguistics.
- Tom Kocmi and Ondřej Bojar. 2017. LanideNN: Multilingual Language Identification on Character Window. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 927–936, Valencia, Spain, April. Association for Computational Linguistics.
- Tom Kocmi and Ondřej Bojar. 2016. SubGram: Extending Skip-gram Word Representation with Substrings. In Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Text, Speech, and Dialogue: 19th International Conference, TSD 2016*, number 9924 in Lecture Notes in Computer Science, pages 182–189, Cham / Heidelberg / New York / Dordrecht / London. Masaryk University, Springer International Publishing.

- Ondřej Bojar, Yvette Graham, Amir Kamran, and Miloš Stanojević. 2016. Results of the WMT16 Metrics Shared Task. In Ondřej Bojar and et al., editors, *Proceedings of the First Conference on Machine Translation (WMT). Volume 2: Shared Task Papers*, volume 2, pages 199–231, Stroudsburg, PA, USA. Association for Computational Linguistics, Association for Computational Linguistics.
- Alexandra Birch, Omri Abend, Ondřej Bojar, and Barry Haddow. 2016. HUME: Human UCCA-Based Evaluation of Machine Translation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1264–1274, Austin, Texas, November. Association for Computational Linguistics.
- Ondřej Bojar, Ondřej Dušek, Tom Kocmi, Jindřich Libovický, Michal Novák, Martin Popel, Roman Sudarikov, and Dušan Variš. 2016. CzEng 1.6: Enlarged Czech-English Parallel Corpus with Processing Tools Dockered. In Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Text, Speech, and Dialogue: 19th International Conference, TSD 2016*, number 9924 in Lecture Notes in Computer Science, pages 231–238, Cham / Heidelberg / New York / Dordrecht / London. Masaryk University, Springer International Publishing.
- Ondřej Bojar, Miloš Ercegovčević, Martin Popel, and Omar Zaidan. 2011. A Grain of Salt for the WMT Manual Evaluation. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 1–11, Edinburgh, Scotland, July. Association for Computational Linguistics.
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In *ACL 2007, Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic, June. Association for Computational Linguistics.

In Prague, October 30, 2018,

RNDr. Ondřej Bojar, Ph.D.

---

Charles University, Faculty of Mathematics and Physics,  
Institute of Formal and Applied Linguistics,  
Malostranské nám. 25, 118 00 Prague, Czech Republic,  
e-mail: bojar@ufal.mff.cuni.cz.