

## RNDr. Ondřej Bojar

Date of Birth: 7<sup>th</sup> March 1979 in Prague  
Address: U Lesa 12, Praha 4, CZ-142 00, Czech Republic  
E-mail, Web: bojar@ufal.mff.cuni.cz; <http://www.cuni.cz/~obo>

### Education:

since 2003 PhD. study at Institute of Formal and Applied Linguistics (ÚFAL), MFF UK  
1997-2003 Bachelor and Master degree (summa cum laude), Charles University in Prague, Faculty of Mathematics and Physics (MFF UK)  
Master thesis: Automatic extraction of lexico-syntactic information from corpora  
1997-1999 Parallel study of Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague; finished 4 semesters  
1993-1997 High School Zborovská, Graduation Exam, best results

### Other Experience:

2006-2007 Twelve-month research visit at CSSE, University of Melbourne, tutor  
Jul-Aug 2006 Language Engineering Workshop (Machine Translation Team) at JHU, Baltimore (6 weeks)  
since 2004 Teaching assistant at both MFF UK and Czech Technical University  
Oct-Nov 2005 Two-month research stay at RWTH Aachen University  
2000-2005 Programming and analysis, Internet protocols.  
Internet Info, Ltd., <http://www.iinfo.cz/>  
2003-2004 Six-month study and research stay at University of Saarland, Saarbrücken  
1995-2003 Teacher, Computer courses (grades 8 and 9), Primary School Fr. Plamínkové, Prague 7  
1996-1999 Created and maintained web pages at <http://www.cestina.cz/>  
1996-1997 Co-founder and vice-chairman of KPPM, the Czech Macintosh User Group  
1994-1997 Publications in the Czech Macworld (not in the list of publications below)

### Involved in Projects:

since Sep 2006 EuroMatrix (Machine translation between European languages)  
since spring 2004 VALLEX (Valency Lexicon of Czech Verbs)  
since Jul 2004 PCEDT (The Prague Czech-English Dependency Treebank)  
2003-2005 Machine translation project of economical texts from Czech to English  
2000-2002 Software project The ENTs—A simulation of natural environment with human-like computer-driven agents, <http://ufal.mff.cuni.cz/~bojar/enti>

### Other Skills:

- English (fluent, Certificate in Advanced English), German (fluent, Zentrale Mittelstufenprüfung)
- programming languages Mercury, Perl, Prolog, PHP, SQL (excellent knowledge), C, C++, Java (good)
- extensive programming experience with Unix (and previously Mac OS and DOS)
- extensive programming experience with TCP/IP, Internet

### Publications:

#### Refereed

Ondřej Bojar, Silvie Cinková, and Jan Ptáček. 2007. Towards English-to-Czech MT via Tectogrammatical Layer. In *Proceedings of the Sixth International Workshop on Treebanks and Linguistic Theories (TLT 2007)*, Bergen, Norway.

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open source toolkit for statistical machine translation. In *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.

- Ondřej Bojar. 2007. English-to-Czech factored machine translation. In *Proceedings of the Second Workshop on Statistical Machine Translation*, pages 232–239, Prague, Czech Republic, June. Association for Computational Linguistics.
- Ondřej Bojar and Zdeněk Žabokrtský. 2006. CzEng: Czech-English Parallel Corpus, Release version 0.5. *Prague Bulletin of Mathematical Linguistics*, 86:59–62.
- Václava Benešová and Ondřej Bojar. 2006. Czech Verbs of Communication and the Extraction of their Frames. In *Text, Speech and Dialogue: 9th International Conference, TSD 2006*, volume LNAI 3658, pages 29–36. Springer Verlag, September.
- Ondřej Bojar and Magdalena Prokopová. 2006. Czech-English Word Alignment. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1236–1239. ELRA.
- Ondřej Bojar, Evgeny Matusov, and Hermann Ney. 2006. Czech-English Phrase-Based Machine Translation. In *FinTAL 2006*, volume LNAI 4139, pages 214–224, Turku, Finland, August. Springer.
- Ondřej Bojar, Jiří Semecký, and Václava Benešová. 2005. VALEVAL: Testing VALLEX Consistency and Experimenting with Word-Frame Disambiguation. *Prague Bulletin of Mathematical Linguistics*, 83:5–17.
- Markéta Lopatková, Ondřej Bojar, Jiří Semecký, Václava Benešová, and Zdeněk Žabokrtský. 2005. Valency Lexicon of Czech Verbs VALLEX: Recent Experiments with Frame Disambiguation. In Václav Matoušek, Pavel Mautner, and Tomáš Pavelka, editors, *Text, Speech and Dialogue: 8th International Conference, TSD 2005, Karlovy Vary, Czech Republic, September 12-15, 2005. Proceedings*, volume LNAI 3658, pages 99–106. Springer Verlag, September.
- Ondřej Bojar, Cyril Brom, Milan Hladík, and Vojtěch Toman. 2005. The Project ENTs: Towards Modelling Human-like Artificial Agents. In Peter Vojtáš, Mária Bieliková, Bernadette Charron-Bost, and Ondrej Sýkora, editors, *SOFSEM 2005 Communications*, pages 111–122. Society for Computer Science, January.
- Ondřej Bojar, Petr Homola, and Vladislav Kuboň. 2005. Problémy recyklování systému automatického překladu. In Peter Vojtáš, editor, *ITAT 2005 Information Technologies – Applications and Theory*, pages 335–344, Košice, Slovakia, September. University of P. J. Šafařík.
- Ondřej Bojar, Petr Homola, and Vladislav Kuboň. 2005. Problems Of Reusing An Existing MT System. In *IJCNLP 2005 - Companion Volume to the Proceedings of Conference including Posters/Demos and Tutorial Abstracts*, pages 181–186, October.
- Ondřej Bojar and Jan Hajič. 2005. Extracting Translation Verb Frames. In Walther von Hahn, John Hutchins, and Christina Vertan, editors, *Proceedings of Modern Approaches in Translation Technologies, workshop in conjunction with Recent Advances in Natural Language Processing (RANLP 2005)*, pages 2–6. Bulgarian Academy of Sciences, September.
- Ondřej Bojar. 2005. Budování česko-anglického slovníku pro strojový překlad. In Peter Vojtáš, editor, *ITAT 2005 Information Technologies – Applications and Theory*, pages 201–211, Košice, Slovakia, September. University of P. J. Šafařík.
- Ondřej Bojar, Petr Homola, and Vladislav Kuboň. 2005. An MT System Recycled. In *Proceedings of MT Summit X*, pages 380–387, September.
- Ondřej Bojar. 2004. Problems of Inducing Large Coverage Constraint-Based Dependency Grammar for Czech. In *Constraint Solving and Language Processing, CSLP 2004*, volume LNAI 3438, pages 90–103, Roskilde University, September. Springer.
- Ondřej Bojar. 2004. Czech Syntactic Analysis Constraint-Based, XDG: One Possible Start. *Prague Bulletin of Mathematical Linguistics*, 81:43–54.
- Ondřej Bojar. 2004. Automated Extraction of Lexico-Syntactic Information. In Jana Šafránková, editor, *WDS'04 Proceedings of Contributed Papers: Part I - Mathematics and Computer Sciences*, pages 211–217, Prague, June 15–18. Charles University, Matfyzpress.
- Ondřej Bojar. 2003. Towards Automatic Extraction of Verb Frames. *Prague Bulletin of Mathematical Linguistics*, 79–80:101–120.
- Ondřej Bojar. 2003. Building Subcorpora Suitable for Extraction of Lexico-Syntactic Information. In *Proceedings of the Student Session, ESSLLI*, August.
- Ondřej Bojar. 2003. AX - Systém pro automatizovanou extrakci lexikálně-syntaktických údajů. In *MIS 2003*, pages 15–24. MATFYZPRESS, January 18–25, 2003.

## Other

- Ondřej Bojar and Magdalena Prokopová. 2007. Czech-English Machine Translation Dictionary. Technical report, ÚFAL MFF UK, Prague, Czech Republic.
- Ondřej Bojar. 2006. Strojový překlad: zamyšlení nad účelností hloubkových jazykových analýz. In *MIS 2006*, pages 3–13, Josefův Důl, Czech Republic, January. MATFYZPRESS.
- Philipp Koehn, Marcello Federico, Wade Shen, Nicola Bertoldi, Ondřej Bojar, Chris Callison-Burch, Brooke Cowan, Chris Dyer, Hieu Hoang, Richard Zens, Alexandra Constantin, Christine Moran, and Evan Herbst. 2006. Open Source Toolkit for Statistical Machine Translation: Factored Translation Models and Confusion

Network Decoding. Technical report, Johns Hopkins University, Center for Speech and Language Processing. in prep.

Ondřej Bojar, Jiří Semecký, Shravan Vasishth, and Ivana Kruijff-Korbayová. 2004. Processing noncanonical word order in Czech. In *Proceedings of Architectures and Mechanisms for Language Processing, AMLaP 2004*, pages 91–91, Université de Provence, September 16-18.

Ondřej Bojar, Cyril Brom, Milan Hladík, Mikuláš Vejlupek, Vojtěch Toman, and David Voňka. 2003. ENTI – Simulátor přirozeného prostředí lidského světa. In *MIS 2003*, pages 3–14. MATFYZPRESS, January 18–25, 2003.

Prague, November 6, 2007.

Ondřej Bojar