

Jan Hajic

CV

Home address: Nosická 12/2389, 100 00 Prague 10, Czech Republic

Work address: ÚFAL MFF UK, Malostranské nám. 25
118 00 Prague 1, Czech Rep., tel: +420 221 914 257

College education: Fac. of Math. and Phys., MFF UK Prague, computer science

Grad. thesis (1984): Natural Language Interface to Robot Control System

Phd thesis (1994): Unification Morphology Grammar

Academic titles: RNDr. awarded 1984, **Dr. (PhD) in Computational Linguistics** 1994

Employment:

1991-now

Institute of Formal and Applied Linguistics (UFAL), Professor, Director
Research in Machine Translation. Text Corpora, Stat. methods
Teaching Computational Linguistics and Computer Science
Consultant for the Institute of the Czech National Corpus

1999-2000 (3 semesters)

Johns Hopkins Univ., Computer Science Dept., CLSP JHU, Assistant Professor
Project "Speech recognition of a Slavic language"

1991-1993, 1994 (20 months total)

IBM Research, Yorktown Heights, USA: visiting scientist
Machine Translation Group/Continuous Speech Recognition Group
Project Candide (Machine Translation: French-English)

1984-1991

Research Institute of Computers, Praha, Czech Rep.
Machine Translation Project Czech-Russian

Visiting and other positions

2002 Jul-Aug

Summer Workshop at CLSP JHU Baltimore. MD, USA
Team leader for "Generation in the context of Machine Translation" project

1998 Jul-Aug

Summer Workshop at CLSP JHU Baltimore. MD, USA
Team leader for "Core NLP technologies – parsing FWO languages"

1990 Feb-Mar, 1991 Mar-Apr

I.S.S.C.O., University of Geneva (Prof. M. King), Multilingual Morphology

Membership

- Member of the **ACL (Association for Computational Linguistics)** since 1984
- Advisory Board member of the European Chapter of the ACL (1995-6)
- Affiliate member of IEEE since 2003
- Member of ISCA since 2003
- Member of ACM since 2005
- Text Encoding Initiative (TEI) Board of Directors member 1998-2002

National Service (current)

- Member of the **Subcommittee for Social Sciences and the Humanities**, Committee for Science, Research and Innovations at the Czech Government, 2010-2012

- Reviewer and panelist for the Czech Grant Agency (since 2002)

Interational Service

Scientific panels

- Science Foundation Ireland, Comp. Linguistics, 2007-
- National Science Foundation, HLT, OISE Programs, USA, 2003-
- European Commission (výzk. projekty), 1997-

Organization of Events and Conferences

- The 48th Annual Meeting of the Association for Computational Linguistics, General Chair, July 2010, Uppsala, Sweden
- Conference on Natural Language Learning, Shared Task (CoNLL ST'09), organization committee chair, Boulder, CO, USA
- The 12th Joint EMNLP-CoNLL Conference, local org. chair, (June 2007), Prague, ČR
- The 45th Annual Meeting of the Association for Computational Linguistics, local org. committee coordinator (June 2007), Prague, ČR
- Treebanks and Linguistic Theories'06, Dec. 2006, Prague, ČR (organization committee chair)
- NSF (USA) PIRE cooperation program (with Johns Hopkins University, Baltimore, MD and Brown University, Providence, RI; workshops in 2006-2010: Prague, ČR, Boulder, CO, USA, Uppsala, Sweden)

Program Committee Chair for Conferences

- TLT'07 (Treebanks and Linguistic Theories), Oslo, Norway, Nov. 2007
- TLT'06 (Treebanks and Linguistic Theories), Prague, ČR, Dec. 2006
- EACL'03 (European ACL Conference), Budapest, Hungary, April 2003
- EMNLP'02 (Empirical Methods in NLP), Philadelphia, PA, USA, July 2002
- Thematic Session on "Parsing inflective and free word order languages", ACL '99, June 1999, College Park, MD, USA

Program Committee Member, Chair of Section)

- EMNLP, Barcelona, Spain, July 2004
- EAMT Workshop, La Valetta, Malta, April 2004
- ACL'02, Philadelphia, PA, USA, July 2002
- EACL'95, Dublin, Ireland, April 1995

Editorial boards

- Prague Bulletin of Mathematical Linguistics, managing editor, 2008-
- Computer Speech and Language, Elsevier, 2010-2011
- Computational Linguistics, Journal of the ACL, 2008-2010

Grants and projects (as a Principal Investigator/Co-PI)

- LINDAT-Clarin
Language resources infrastructure project, Ministry of Education, Youth and Sports of the CR, LM2010013, 2010-2015
Budget (EUR): 4,650,000
- Khresmoi
EU 7th FP large Integrated Project #IST-257528, 2010-2014
Budget (EUR): 810,000
- META-NET
EU 7th FP Network of Excellence #IST-249119, 2010-2013
Budget (EUR): 320,000
- Faust
EU 7th FP STREP project #IST-247762, 2010-2013
Budget (EUR): 530,000
- EuroMatrixPlus
EU 7th FP STREP project #IST-231720, 2009-2012
Budget (EUR): 680,000
- Companions
EU 6th FP Integrated project #IST-034434, 2006-2010
Budget (EUR): 778,000
- EuroMatrix
EU 6th FP STREP project #IST-034291, 2006-2009
Budget (EUR): 620,000
- Repräsentace významu a automatické porozumění přirozenému jazyku
MŠMT ME838, 2006-2010
Budget (EUR): 148,000
- Tektogramatický popis jazyka pro rozpoznávání mluvené řeči a strojový překlad
GAČR GA405/06/0589, 2006-2008
Budget (EUR): 120,000
- Centrum počítačové lingvistiky (Integrované centrum počítačového zpracování přirozeného jazyka)
Fundamental research core grant, MŠMT, LC536, 2005-2011
Budget (EUR): 2,120,000
- Od jazyka ke znalostem a sémantickému webu
Program “Informatio society” (NPV TP2), GAAV, 1ET201120505, 2005-2009
Budget (EUR): 820,000
- International Collaboration: Meaning Representation for Language Understanding
Projekt NSF (OISE) USA (#0530118), 2005-2009
Budget (EUR): 0 (cf. GA405/06/0589 and ME838 for travel support).
- Vícejazyčný valenční a predikátový slovník přirozeného jazyka
Project with NSF USA (OISE) a MŠMT (1P05ME752), 2005-2007
Budget (EUR): 75,000
- Machine Translation with a Structural Transfer
Spol. projekt NSF USA a MŠMT (ME642/03), 2003-2005
Budget (EUR): 85,000
- Přenos anotace mezi blízkými jazyky, program VTS MŠMT
Common project SAV SR (Slovakia) and MŠMT, project No. 130, 2004-2005
Budget (EUR): 8,000
- Velké jazykové korpusy a jejich automatická analýza
GAČR GA405/03/0913, 2003-2005
Budget (EUR): 275,000

- Identifikace mluvčího
“Post-workshop award”, Workshop’02 at Johns Hopkins University, 2003-2004
Supervision for: David Klusáček, student MFF UK
Budget (EUR): 20,000
- Data preparation grant, WS’02/Johns Hopkins University
NSF grant #IIS-0121285, 2002
Budget (EUR): 30,000
- Projekt MALACH - Multilingual Access to Oral History Archive
NSF grant #IIS-0122466, subcontractor for JHU, 2001-2006
Budget (EUR): 210,000
- Speech recognition of a Slavic Language (Czech)
Joint project, NSF USA a MŠMT (ME293/98), 1998-2001
Budget (EUR): 75,000
- Metody kombinace značkování
“Post-workshop award”, Workshop’98 at Johns Hopkins University, 1998-1999
Supervision for: Barbora Hladká, student ÚFAL MFF UK
Budget (EUR): 20,000
- EU Project INCO-Copernicus: CEGLEX (COP 1032)
Lexical resources for Central and Eastern European Languages, 1995-1996
Budget (EUR): 30,000
- Data preparatio for WS’98/Johns Hopkins University
NSF grant #IIS-9732388, 1998
Budget (EUR): 38,000
- EU Research Project: STEEL (PL1113)
Bilingual terminology extraction from corpora, 1997-1999
Budget (EUR): 50,000
- Český tagger (Czech POS Tagger)
Grantová agentura Univerzity Karlovy, grant GAUK č. 39/94, 1994-1997
Budget (EUR): 5,000

Bibliography selection

- J. Lin, G. C. Murray, B. J. Dorr, J. Hajic and P. Pecina: A Cost-effective Lexical Acquisition Process for Large-scale Thesaurus Translation. *Language Resource and Evaluation*, ISSN 1574-020X. Springer, Netherlands. 2009, in press.
- J. Hajic et al.: The Prague Dependency Treebank 2.0. CDROM. ISBN 1-58563-370-4. LDC Cat. No. LDC2006T01. Linguistics Data Consortium. Philadelphia, PA, USA. 2006.
- W. J. Byrne, D. Doermann, M. Franz, S. Gustman, J. Hajic, D. W. Oard, M. Picheny, J. V. Psutka, B. Ramabhadran, D. Soergel, T. Ward, W. Zhu: Automatic Recognition of Spontaneous Speech for Access to Multilingual Oral History Archives. *IEEE Transactions on Speech and Audio Processing* 4(12). ISSN 1063-6676. 420-435. 2004.
- A. Bohmova, J. Hajic, E. Hajicova, B. Hladka: The Prague Dependency Treebank. Chapter in: Abeille, Anne (ed.), *Treebanks: Building and Using Parsed Corpora*, Kluwer Academic Publishers, Dordrecht/Boston/London, 103-127. 2003.
- J. Hajic, J. Panevova, Z. Uresova, A. Bemova, V. Kolarova, P. Pajas: PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*. ISSN 1651-0267. 57-68. 2003.