

Jan Hajič

Životopis

Květen 2007

Osobní údaje:

Narozen 4.11.1960 v Praze, r.č. 601104/0981
Bydliště: Nosická 12/2389, 100 00 Praha 10
Rodinný stav: ženatý, 2 děti
Kontakt: hajic@ufal.mff.cuni.cz, tel. 607 209 212

Vzdělání:

1994 - Dr. (I-3 Matematická lingvistika)
MFF UK, Praha
Doktorská práce: Unification Morphology Grammar
1988 - 1994 MFF UK, ext. aspirantura/doktorské studium, spec. I-3 Matematická lingvistika
1984 - RNDr. (Samočinné počítače a programování)
MFF UK, Praha
1979 - 1984 MFF UK, obor TKMITS, spec. Samočinné počítače a programování
abs. s vyznamenáním (červený diplom)
Diplomová práce: Jednoduchý systém řízení robota v přirozeném jazyce
1975 - 1979 Gymnázium W. Piecka, Praha 2 (přírodovědná větev, spec. na matematiku)

Zaměstnání:

1991-nyní
Karlova Univerzita v Praze
postupně: asistent, odborný asistent, docent (jmenován k 1.5.2003),
profesorské řízení ukončeno, jmenování podzim 2007
Ústav formální a aplikované lingvistiky (ÚFAL MFF UK)
ředitel ÚFAL od 1.2.2003, nyní druhé období od 1.2.2007
1999-2000 (3 semestry)
Johns Hopkins University, Baltimore, MD, USA
Visiting Assistant Professor, Computer Science dept., JHU;
zároveň v "Center for Language and Speech Processing", JHU.
1991-1993, 1994 (20 měsíců celkem)
IBM T.J.Watson Research Center, Yorktown Heights, NY, USA
Visiting Scientist
Machine Translation Group/Continuous Speech Recognition Group
Projekt "Candide" (Strojový překlad, DARPA: French-English).
1984-1991
Výzkumný ústav matematických strojů, Praha
Výzkumný pracovník, projekt: česko-ruský strojový překlad

Významné krátké pobyty a stáže:

2006 květen-červen **Univ. of Colorado, Boulder, CO, USA**
Práce na projektu "Vícejazyčný predikátový a valenční slovník přirozeného jazyka"
Projekt KONTAKT ME752
2002 červenec-srpen **Johns Hopkins University**
Vedoucí týmu, Summer Workshop at CLSP JHU Baltimore. MD, USA
Projekt: "NL Generation in the Context of Machine Translation"

1998 červenec–srpen **Johns Hopkins University**

Vedoucí týmu, Summer Workshop at CLSP JHU Baltimore. MD, USA

Projekt: "Core NLP Technologies – Parsing Free-Word Order Languages"

1990 únor–březen, 1991 březen–duben **ISSCO, Univ. de Genève**

Visiting Professor

Další profesní aktivity:

Ústav českého národního korpusu FF UK

Externí spolupracovník (Český národní korpus), 1996–nyní

Zakládající člen občanského sdružení pro Český národní korpus, 1990–1996

Board of Directors (člen za ACL):

Text Encoding Initiative Consortium, 1999-2002;

Steering Committee (člen za ACL):

Text Encoding Initiative, 1998-1999;

Advisory Board (člen)

European Chapter of the Association for Computational Linguistics (ACL), 1995-1996.

Programový výbor mez. konference (předseda, spolupředseda)

- TLT'07 (Treebanks and Linguistic Theories), Oslo, Norsko, listopad 2007
- TLT'06 (Treebanks and Linguistic Theories), Praha, prosinec 2006
- EACL'03 (European ACL Conference), Budapešť, Maďarsko, duben 2003
- EMNLP'02 (Empirical Methods in NLP), Philadelphia, PA, USA, červenec 2002
- Thematic Session on "Parsing inflective and free word order languages", ACL '99, June 1999, College Park, MD, USA

Programové výbory mez. konferencí (člen, předseda sekce)

- EMNLP, Barcelona, Španělsko, červenec 2004
- EAMT Workshop, La Valetta, Malta, duben 2004
- Konference ACL'02, Philadelphia, PA, USA, červenec 2002
- Konference Evropské sekce ACL'95, Dublin, Irsko, duben 1995

Programové výbory mezinárodních a domácích konferencí (recenzent)

- Workshop on Linguistic Annotation (4th LAW), Praha, červen 2007
- Konference TSD'07, Brno, ČR, září 2006
- Konference EMNLP'07, Praha, ČR, červen 2006
- Konference ACL'06, Sydney, Austrálie, červenec 2006
- Konference NAACL/HLT'06, New York, NY, USA, červen 2006
- Konference Znalosti'06, Hradec Králové, ČR, únor 2006
- Konference Treebanks and Linguistic Theories '05, Barcelona, Španělsko, prosinec 2005

- Workshop Linguistically Interpreted Corpora, at the 2nd International Conf. on Natural Language Processing (IJCNLP'05) Jeju Island, Korea, říjen 2005
- Konference Text, Speech and Dialogue '05, Karlovy Vary, ČR, září 2005
- Konference RANLP'05, Borovetz, Bulharsko, září 2005
- Konference ACL'05, Ann Arbor, MI, USA, červen 2005
- Konference Znalosti'05, Vysoké Tatry, Slovensko, únor 2005
- Konference Treebanks and Linguistic Theories '04, Tuebingen, Německo, prosinec 2004
- Konference Coling'04, Ženeva, Švýcarsko, srpen 2004
- Konference LREC'04, Lisabon, Portugalsko, květen 2004
- Konference NAACL/HLT'04, Boston/Cambridge, MA, USA, květen 2004
- Konference RANLP'03, Borovetz, Bulharsko, září 2003
- Konference EMNLP'03, Sapporo, Japonsko, červenec 2003
- Konference ACL'03, Sapporo, Japonsko, červenec 2003
- NLPXML-2002 (použití XML v NLP), Taipei, Taiwan, srpen 2002
- Coling 2002, Taipei, Taiwan, srpen 2002
- LREC 2002 (Language Resources), Las Palmas, Španělsko, květen 2002
- Konference Severoamerické sekce Association for Computational Linguistics (NAACL)'01, Pittsburgh, PA, USA, červen 2001
- Konference ACL'00, Seattle, WA, USA, duben 2000
- Konference Applied Natural Language Processing (ANLP)'00, Seattle, WA, USA, duben 2000
- Konference ACL '99, College Park, MD, USA, červen 1999
- Konference Evropské sekce ACL '99, Bergen, Norsko, červen 1999

Recenzní výbory a panely (člen)

- National Science Foundation, HLT, OISE Programs, USA, 2003–nyní
- Grantová agentura České republiky, 1996–nyní
- European Commission (výzk. projekty), 1997–nyní

Organizace konferencí:

- The 12th Joint EMNLP-CoNLL Conference, předseda org. výboru (červen 2007), Praha, ČR
- The 45th Annual Meeting of the Association for Computational Linguistics (červen 2007), Praha, ČR (člen org. výboru)
- Treebanks and Linguistic Theories'06, prosinec 2006, Praha, ČR (předseda org. výboru)

Cizí jazyky:

- Plyně anglicky.
- Pasivně i aktivně německy, rusky, slovensky.
- Pasivně francouzsky, polsky.

Individuální granty, ceny a stipendia

- Člen kolektivu pracovníků Českého národního korpusu odměněného stříbrnou medailí UK, 2001
- Research Support Scheme, Open Society Fund (Soros Foundation), Corpora: Annotation and Comparison, grant 1996-1998
- Open Information Society of Japan, na 1 měsíc návštěva japonských univerzit a průmyslových pracovišť v oboru, 1988

Odpovědný řešitel projektu nebo odpovědný řešitel české části projektu (Principal Investigator (PI) / Co-PI) – granty a projekty

- Companions
EU 6.RP Integrated project #IST-034434, 2006-2010
Celkový spravovaný fin. objem grantu: cca Kč 24000 tis.
- EuroMatrix
EU 6.RP STREP project #IST-034291, 2006-2008
Celkový spravovaný fin. objem grantu: cca Kč 16700 tis.
- Reprezentace významu a automatické porozumění přirozenému jazyku
MŠMT ME838, 2006-2010
Celkový spravovaný fin. objem grantu: Kč 4063 tis.
- Tektogramatický popis jazyka pro rozpoznávání mluvené řeči a strojový překlad
GAČR GA405/06/0589, 2006-2008
Celkový spravovaný fin. objem grantu: Kč 3400 tis.
- Centrum počítačnické lingvistiky (Integrované centrum počítačového zpracování přirozeného jazyka)
Projekt Centra základního výzkumu, MŠMT, LC536, 2005-2009
Celkový spravovaný fin. objem grantu: Kč 53616 tis.
- Od jazyka ke znalostem a sémantickému webu
Projekt Informační společnosti (NPV TP2), GAAV, 1ET201120505, 2005-2009
Celkový spravovaný fin. objem grantu: Kč 16061 tis.
- International Collaboration: Meaning Representation for Language Understanding
Projekt NSF (OISE) USA (#0530118), 2005-2009
Celkový spravovaný fin. objem grantu: Kč 0 tis (viz GA405/06/0589).
- Vícejazyčný valenční a predikátový slovník přirozeného jazyka
Spol. projekt NSF USA a MŠMT (1P05ME752), 2005-2007
Celkový spravovaný fin. objem grantu: Kč 1800 tis.
- Machine Translation with a Structural Transfer
Spol. projekt NSF USA a MŠMT (ME642/03), 2003-2005
Celkový spravovaný fin. objem grantu: Kč 2200 tis.
- Přenos anotace mezi blízkými jazyky, program VTS MŠMT
Spol. projekt SAV SR a MŠMT, projekt č. 130, 2004-2005
Celkový spravovaný fin. objem grantu: Kč 150 tis.
- Velké jazykové korpusy a jejich automatická analýza
GAČR GA405/03/0913, 2003-2005
Celkový spravovaný fin. objem grantu: Kč 6400 tis.
- Identifikace mluvčího
“Post-workshop award”, Workshop’02 at Johns Hopkins University, 2003-2004
Uděleno pro: David Klusáček, student MFF UK
Celkový spravovaný fin. objem grantu: Kč 480 tis.
- Příprava dat pro WS’02/Johns Hopkins University
NSF grant #IIS-0121285, 2002
Celkový spravovaný fin. objem grantu: Kč 700 tis.

- Projekt MALACH - Multilingual Access to Oral History Archive
NSF grant #IIS-0122466, 2001-2006
Celkový spravovaný fin. objem grantu: Kč 5600 tis.
- Speech recognition of a Slavic Language (Czech)
Spol. projekt NSF USA a MŠMT (ME293/98), 1998-2001
Celkový spravovaný fin. objem grantu: Kč 1800 tis.
- Metody kombinace značkování
“Post-workshop award”, Workshop’98 at Johns Hopkins University, 1998-1999
Uděleno pro: Barbora Hladká, doktorandka ÚFAL MFF UK
Celkový spravovaný fin. objem grantu: Kč 800 tis.
- EU Project INCO-Copernicus: CEGLEX (COP 1032)
Lexical resources for Central and Eastern European Languages, 1995-1996
Celkový spravovaný fin. objem grantu: Kč 750 tis.
- Příprava dat pro WS’98/Johns Hopkins University
NSF grant #IIS-9732388, 1998
Celkový spravovaný fin. objem grantu: Kč 950 tis.
- EU Research Project: STEEL (PL1113)
Bilingual terminology extraction from corpora, 1997-1999
Celkový spravovaný fin. objem grantu: Kč 1200 tis.
- Český tagger
Grantová agentura Univerzity Karlovy, grant GAUK č. 39/94, 1994-1997
Celkový spravovaný fin. objem grantu: Kč 125 tis.

Účast na řešení grantů a projektů

Celkem 31 grantů a projektů (vč. započtení grantů a projektů výše uvedených; podrobný seznam viz příloha s publikacemi):

- 7 grantů GAČR (z toho jeden Komplexní projekt),
- 1 projekt GAAV
- 11 projektů MŠMT,
- 3 granty GAUK,
- 1 grant OSF/RSS,
- 3 granty Evropské Unie,
- 5 grantů a projektů v USA.

Vybraná pozvání do zahraničí (pozvané přednášky, tutoriály)

- “Syntax meets Semantics”, Invited talk, Symposium on Parallel Treebanks, Univ. of Stockholm, září 2006
- “The Prague Dependency Treebank and Valency Annotation”, Invited talk, University of Colorado, červen 2006
- “Prague Dependency Treebank and Valency”, Invited tutorial, RANLP’05 (Borovetz, Bulharsko, září 2005)
- “What’s new in Prague Dependency Treebank”, Invited talk, Columbia University, New York, červen 2005
- “Rich Linguistic Annotation and Machine Translation”, Invited talk, LINC’03 (Associated with EACL’03, Budapešť, Maďarsko, duben 2003)

- “Prague Dependency Treebank and Other Languages”, Invited talk, Swedish National Corpus Conference, Vaxjo, Švédsko, listopad 2002
- “Možnosti použití Pražského závislostního korpusu pro projekt Slovenského národního korpusu”, pozvaná přednáška, Jazykovedný ústav Ľudovíta Štúra, Slovenská akadémia vied, Bratislava, Slovensko, prosinec 2002
- “Tectogrammatical Representation: Towards a Minimal Transfer In Machine Translation”, Invited talk, TAG+6 Conference, Benátky, Itálie, květen 2002
- “The Prague Dependency Treebank”, AT&T Research, Florham Park, NJ, USA, červen 1999
- “Tagging morphologically rich languages”, HCRC seminar, Univ. of Edinburgh, Skotsko, Velká Británie, březen 1999
- “Corpora annotation and lexicography”, TELRI seminar, Univ. of Birmingham/COBUILD, Anglie, Velká Británie, květen 1997
- “Probabilistic and Rule-Based Tagger of an Inflective Language - a Comparison”, CLSP seminar, Johns Hopkins Univ., Baltimore, MD, USA, březen 1997

Pedagogická činnost

Výuka, vč. výuky v zahraničí, na letních školách apod., a výsledky ve vědecké výchově studentů a doktorandů viz příloha “Výkaz pedagogické činnosti”.

Publikace

Viz příloha.