

Curriculum Vitae



Personal information

First name(s) / Surname(s)

Jan HAJIČ

Address(es)

Nosická 12, 10000 Prague 10, Czech Republic

Telephone(s)

(+420) 221 914 257 (+420) 607 209 212

Fax(es)

(+420) 221 914 306

E-mail, web

hajic@ufal.mff.cuni.cz, <http://ufal.mff.cuni.cz/jan-hajic>

Nationality

Czech

Date of birth

04.11.1960

Gender

Male

Work experience

Dates

2001–

Occupation or position held

Full Professor 2008- (Assistant Prof. 2001-2003, Assoc. Prof. 2003-2007), Vice-director of Institute 2012- (Director of Institute 2003-2012), Institute of Formal and Applied Linguistics, School of Computer Science, Faculty of Mathematics and Physics, Charles University in Prague, Malostranske nam. 25, 11800 Prague 1, Czech Republic

Main activities and responsibilities

Coordination of Research in the Institute, Coordination and management of research teams in EU, U.S. and national projects, Teaching, Research
Director/P.I. of a Large Infrastructure project LINDAT/CLARIN (with 3 national institutes)

Name and address of employer

Charles University in Prague, MFF, Ke Karlovu 3, 12116 Prague 2, CZ

Dates

2016-2019

Occupation or position held

Adjunct Professor, University of Colorado in Boulder, Dept. of Computer Science

Main activities and responsibilities

Research w/Teaching option

Name and address of employer

University of Colorado in Boulder, 1111 Engineering Drive, Boulder, CO 80309-0430 USA

Dates

1999–2000 (on prolonged sabbatical from Charles Univ. in Prague)

Occupation or position held

Assistant Professor, JHU Whiting School of Engineering, Dept. of Computer Science

Main activities and responsibilities

Research and teaching

Name and address of employer

Johns Hopkins University, WSE, 3400 N Charles St, Baltimore, MD 21218, USA

Dates

1993–1999

Occupation or position held

Researcher, Asst. Prof., Institute of Formal and Applied Linguistics, MFF UK

Main activities and responsibilities

Research, National and EU Project management, Teaching

Name and address of employer

Charles University in Prague, MFF, Ke Karlovu 3, 12116 Prague 2, CZ

Dates

1991–1993

Occupation or position held

Visiting Researcher, IBM Thomas J. Watson Research Center, Yorktown Heights,

Main activities and responsibilities

Research (Statistical Machine Translation project Candide, co-funded by DARPA)

Name and address of employer

IBM Research, 1101 Kitchawan Rd., Rt 134, Yorktown Heights, NY 10598, USA

Dates

1984–1991

Occupation or position held

Researcher, Software Unit, Dept. of Documentation

Main activities and responsibilities

Research, management of small research team, machine translation Czech – Russian

Name and address of employer

Research Institute of Mathematical Machines, Lužná 2, 16000 Prague 6, Czech Rep.

Education and training

Dates 1984, 1995, 2003, 2008

Title of qualification awarded RNDr. (MSc.), Dr. (Ph.D.), Assoc. Prof. of Charles Univ., Full Prof. of Charles Univ.

Principal subjects/occupational skills covered Computer Science (RNDr.), Computational Linguistics (Dr., Assoc. Prof., Full Prof.)

Name and type of organisation providing education and training Charles University in Prague, Faculty of Mathematics and Physics

Self-assessment

Common European Framework of Reference for Languages

English, Czech, Slovak**German****French****Russian**

Understanding				Speaking				Writing	
Listening		Reading		Spoken interaction		Spoken production			
C2	Proficient User	C2	Proficient User	C2	Proficient User	C2	Proficient User	C2	Proficient User
B1	Indep. User	B2	Indep. User	A2	Basic user	A2	Basic user	B1	Indep. User
A2	Basic User	B1	Indep. User	A1	Basic user	A1	Basic User	A2	Basic User
B1	Indep. User	B1	Indep. User	A2	Basic User	A1	Basic User	A2	Basic User

Organisational skills and competences**Service**

- Managing small and mid-size, hierarchical teams, incl. international research teams in projects (up to 70 researchers / students)
- Organization of events (conferences, workshops up to 1000 participants)
- META-NET, Executive Board Chair (2015-)
- Member of the Standing Coordinating Committee for Technical Centers (Clarin ERIC EU infrastructure) (since 2012, now 2nd term)
- Scientific Council member MFF CU (2012-), FEL CTU (2015-) CIIRC CTU (2015-)
- Member of Technical Advisory Board Dariah DE and Clarin DE (Germany, since 2013)
- Member of the Research Council of the Technology Agency of the Czech Republic (technology grant and project funding agency), 2011-2019 (VR TACR)
- Member of the SSH Area Committee of the Council for Research, Development and Innovations of the Government of the Czech Republic, 2009-2014 (OK SHV RVVI)
- Member of the Security Applications Panel of the Expert Committee on "Priorities 2030", Strategic research and development program of the Czech Government for 2012-2030
- Coling 2014, Program Chair (with J-I Tsujii), ICCL member since 2012
- ACL Annual Meeting (Conference), General Chair, Uppsala, 2010
- Editorial Board member, Computer Speech and Language, Elsevier, 2009-2016
- Editorial Board member, Computational Linguistics Journal (by ACL), 2008-2012
- ACL Annual Meeting (Conference), Local Org. Committee Coordinator, Prague 2007
- Editorial Board Member, Managing editor, Editor in Chief, The PBML, 2004-now
- Text Encoding Initiative Consortium Board of Directors member, founding member 1999-2002
- Program Committee (Co)-chair of conferences: 1999, 2002, '03, '06, '07, '12, '14, '15, '17, '18
- Program Committee Area Chair of conferences in 1995, 2002-2015 (total: 13)
- Organization or co-organization of 34 workshops and smaller events, including spring/winter schools by the VMC, Prague, between 1995 and 2017
- European ACL Board member, 1995-1996
- Czech National Corpus founding member, 1990, Advisory Board member (2016-)

Projects managed

- PI or Co-PI of (last 10 years only)
- EU projects QT21 (2015-18), QTLeap (13-16), HimL (15-18), PARSEME (COST, 13-16), Euromatrix (06-09), Companions (06-10), EuromatrixPlus (09-12, Co-PI for 09-10), T4ME Net (10-13), Faust (10-13), Khresmoi (10-14), Cracker (15-17)
 - US projects MALACH (01-08), PIRE (05-10), Mellon Foundation (16-18), part in 3 more
 - National grants and projects: LINDAT/CLARIN (10-15, 16-19), Center for Computational Linguistics (05-11), Ministry of Education and M. of Culture (7 projects, 12-19)

Other

- Team leader of 2 JHU Summer Workshop 2-months projects, Baltimore, MD, USA (1998, 2002)
- Research visits (short and mid-term) at Univ. of Colorado, Univ. of Pennsylvania; Hamburg, Saarbuecken, Stuttgart (Germany). ISSCO/Univ. of Geneva (Switzerland), Kytoto Univ., Japan

Publications: 190 publications with 6400+ citations, h-index 37 (according to Google Scholar; last 5 years: h-index 29). For selected publications, see below.

Selected Recent Publications

- Zeman Daniel, Popel Martin, Straka Milan, Hajič Jan, Nivre Joakim, Ginter Filip, et al.: CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies. In: *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, ACL, USA, ISBN 978-1-945626-70-8, pp. 1-19, 2017
- Nivre Joakim et al.: Universal Dependencies 2.1. Data/software, LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, Praha, Czechia, <http://hdl.handle.net/11234/1-2515>, Nov 2017
- Mikulová Marie, Mirovský Jiří, Nedoluzhko Anna, Pajas Petr, Štěpánek Jan, Hajič Jan: PDTSC 2.0 - Spoken Corpus with Rich Multi-layer Structural Annotation. In: *Lecture Notes in Computer Science, Text, Speech, and Dialogue 20th International Conference, TSD 2017*, pp. 129-137, 2017
- Urešová Zdeňka, Fučíková Eva, Hajičová Eva: CzEngClass – Towards a Lexicon of verb Synonyms with Valency linked to Semantic Roles. In: *Jazykovedný časopis / Journal of Linguistics*, Vol. 68, No. 2, Slovak Academic Press, ISSN 0021-5597, pp. 364-371, 2017
- Hajič Jan, Fučíková Eva, Šindlerová Jana, Urešová Zdeňka: Verb Argument Pairing in Czech-English Parallel Treebank. In: *GLOBALEX 2016: Lexicographic Resources for Human Language Technology*, pp. 1-7, 2016
- Hajič Jan, Hajičová Eva, Mikulová Marie, Mirovský Jiří: Prague Dependency Treebanks. *Handbook on Linguistic Annotation*, Springer Verlag, Berlin, Germany, 200 pp., Dec 2016
- Nivre Joakim, de Marneffe Marie-Catherine, Ginter Filip, Goldberg Yoav, Hajič Jan, Manning Christopher, McDonald Ryan, Petrov Slav, Pyysalo Sampo, Silveira Natalia, Tsarfaty Reut, Zeman Daniel: Universal Dependencies v1: A Multilingual Treebank Collection. Accepted for publication in: *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association, Paris, France, 2016
- Oepen Stephan, Kuhlmann Marco, Miyao Yusuke, Zeman Daniel, Cinková Silvie, Flickinger Dan, Hajič Jan, Ivanova Angelina, Urešová Zdeňka: Towards Comparability of Linguistic Graph Banks for Semantic Parsing. Accepted for publication in: *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, European Language Resources Association, Paris, France, 2016
- Przepiórkowski Adam, Hajič Jan, Hajnec Elzbieta, Urešová Zdeňka: *Phraseology in two Slavic valency dictionaries: limitations and perspectives*. International Journal of Lexicography, Oxford University Press, Oxford, UK, ISSN 0950-3846, 2016
- Bojar Ondřej, Hajič Jan: *MT for Industry from Researchers' Point of View*. Contributed talk, IBM MT Workshop 4th TLC 2015, IBM TLS EMEA, Warsaw, Poland, Mar 2015
- Dušek Ondřej, Fučíková Eva, Hajič Jan, Popel Martin, Šindlerová Jana, Urešová Zdeňka: Using Parallel Texts and Lexicons for Verbal Word Sense Disambiguation. In: *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, Uppsala University, Uppsala, Sweden, ISBN 978-91-637-8965-6, pp. 82-90, 2015
- Fučíková Eva, Hajič Jan, Šindlerová Jana, Urešová Zdeňka: Czech-English Bilingual Valency Lexicon Online. In: *14th International Workshop on Treebanks and Linguistic Theories (TLT 2015)*, IPIPAN, Warszawa, Poland, ISBN 978-83-63159-18-4, pp. 61-71, 2015
- Hajič Jan, Hajičová Eva, Mikulová Marie, Mirovský Jiří, Panevová Jarmila, Zeman Daniel: Deletions and node reconstructions in a dependency-based multilevel annotation scheme. In: *Lecture Notes in Computer Science*, Vol. 9041, *16th International Conference on Computational Linguistics and Intelligent Text Processing*, Springer, Berlin / Heidelberg, ISBN 978-3-319-18111-0, ISSN 0302-9743, pp. 17-31, 2015
- Nivre Joakim et al.: *Universal Dependencies 1.2.*, LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague, Praha, Czechia, <http://hdl.handle.net/11234/1-1548>, Nov 2015
- Oepen Stephan, Kuhlmann Marco, Miyao Yusuke, Zeman Daniel, Cinková Silvie, Flickinger Dan, Hajič Jan, Urešová Zdeňka: SemEval 2015 Task 18: Broad-Coverage Semantic Dependency Parsing. In: *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-40-2, pp. 915-926, 2015
- Straka Milan, Hajič Jan, Straková Jana, Hajič, jr. Jan: Parsing Universal Dependency Treebanks using Neural Networks and Search-Based Oracle. In: *14th International Workshop on Treebanks and Linguistic Theories (TLT 2015)*, IPIPAN, Warszawa, Poland, ISBN 978-83-63159-18-4, pp. 208-220, 2015
- Urešová Zdeňka, Dušek Ondřej, Fučíková Eva, Hajič Jan, Šindlerová Jana: Bilingual English-Czech Valency Lexicon Linked to a Parallel Corpus. In: *Proceedings of the The 9th Linguistic Annotation Workshop (LAW IX 2015)*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-47-1, pp. 124-128, 2015
- Urešová Zdeňka, Fučíková Eva, Hajič Jan, Šindlerová Jana: *CzEngVallex - Czech English Valency Lexicon*. Data/software, Charles University in Prague, Prague, Czech Republic, <https://lindat.mff.cuni.cz/en/services#CzEngVallex>, Oct 2015
- Choi Jinho, de Marneffe Marie-Catherine, Dozat Timothy, Ginter Filip, Goldberg Yoav, Hajič Jan, Manning Christopher, McDonald Ryan, Nivre Joakim, Petrov Slav, Pyysalo Sampo, Silveira Natalia, Tsarfaty Reut, Zeman Daniel: *Universal Dependencies*. Stanford, Google, Uppsala, Prague, Turku, Tel Aviv, 95 pp., Oct 2014
- De Smedt Koenraad, Hinrichs Erhard, Meurers Detmar, Skadiņa Inguna, Pedersen Bolette Sandford, Navarretta Costanza, Bel Núria, Lindén Krister, Lopatková Markéta, Hajič Jan, Andersen Gisle, Lenkiewicz Przemyslaw: CLARA: A New Generation of Researchers in Common Language Resources and Their Applications. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, European Language Resources Association, Reykjavik, Iceland, ISBN 978-2-9517408-8-4, pp. 2166-2174, 2014
- Dušek Ondřej, Hajič Jan, Urešová Zdeňka: Verbal Valency Frame Detection and Selection in Czech and English. In: *The 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-14-3, pp. 6-11, 2014
- Oepen Stephan, Kuhlmann Marco, Zeman Daniel, Miyao Yusuke, Flickinger Dan, Hajič Jan, Ivanova Angelina, Zhang Yi: SemEval 2014 Task 8: Broad-Coverage Semantic Dependency Parsing. In: *Proceedings of the Eighth International Workshop on Semantic Evaluation (SemEval 2014)*, Dublin City University, Dublin, Ireland, ISBN 978-1-937284-96-1, pp. 63-72, 2014

Pecina Pavel, Dušek Ondřej, Goeuriot Lorraine, Hajič Jan, Hlaváčová Jaroslava, Jones Gareth J.F., Kelly Liadh, Leveling Johannes, Mareček David, Novák Michal, Popel Martin, Rosa Rudolf, Tamchyna Aleš, Uřešová Zdeňka: *Adaptation of machine translation for multilingual information retrieval in medical domain*. In: *Artificial Intelligence in Medicine*, Vol. 61, No. 3, Elsevier, ISSN 0933-3657, pp. 165-185, Jul 2014

Pletneva Natalia, Uřešová Zdeňka, Altman Jean-Jacques, Vinay Nicolas Postel, Degoulet Patrice, Hajič Jan, Boyer Célia: *Observations and Lessons Learnt from Non Health Professionals Evaluating a Health Search Engine*. In: *25th Conference of the European Federation of Medical Informatics (MIE)*, IOS PRESS, Amsterdam, Netherlands, ISBN 978-1-61499-432-9, ISSN 0926-9630, pp. 940-944, 2014

Straková Jana, Straka Milan, Hajič Jan: *Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition*. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-00-6, pp. 13-18, 2014

Uřešová Zdeňka, Dušek Ondřej, Hajič Jan, Pecina Pavel: *Multilingual Test Sets for Machine Translation of Search Queries for Cross-Lingual Information Retrieval in the Medical Domain*. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, European Language Resources Association, Reykjavík, Iceland, ISBN 978-2-9517408-8-4, pp. 3244-3247, 2014

Uřešová Zdeňka, Šindlerová Jana, Fučíková Eva, Hajič Jan: *Verb-Noun Idiomatic Combinations in a Czech-English Dependency Corpus*. Institute for Language and Speech Processing of the Athena Research Center, Athens, Greece, Mar 2014

Xue Nianwen, Bojar Ondřej, Hajič Jan, Palmer Martha, Uřešová Zdeňka, Zhang Xiuhong: *Not an Interlingua, But Close: Comparison of English AMRs to Chinese and Czech*. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, European Language Resources Association, Reykjavík, Iceland, ISBN 978-2-9517408-8-4, pp. 1765-1772, 2014

Zeman Daniel, Dušek Ondřej, Mareček David, Popel Martin, Ramasamy Loganathan, Štěpánek Jan, Žabokrtský Zdeněk, Hajič Jan: *HamleDT: Harmonized Multi-Language Dependency Treebank*. In: *Language Resources and Evaluation*, Vol. 48, No. 4, Springer Netherlands, Dordrecht, Netherlands, ISSN 1574-020X, pp. 601-637, Dec 2014

Aswani Niraj et al.: *Khresmoi Professional: Multilingual Semantic Search for Medical Professionals*. In: *Proceedings of the ACM SIGIR Workshop on Health Search and Discovery: Helping Users and Advancing Medicine*, Microsoft Research, Cambridge, UK, pp. 31-34, 2013

Hajič Jan, Bohnet Bernd, Nivre Joakim, Boguslavsky Igor, Farkas Richárd, Ginter Filip: *Joint Morphological and Syntactic Analysis for Richly Inflected Languages*. In: *Transactions of the Association for Computational Linguistics*, Vol. 1, Association for Computational Linguistics, ISSN 2307-387X, pp. 415-428, 2013

Smrž Otakar, Bielický Viktor, Smržová Iveta, Pajas Petr, Žabokrtský Zdeněk, Zeman Daniel, Kráčmar Jakub, Saleh Shadi, Hajič Jan, Zemánek Petr: *Prague Arabic Dependency Treebank 1.5*, UFAL MFF UK, <http://ufal.mff.cuni.cz/padt>, 2013

Straková Jana, Straka Milan, Hajič Jan: *A New State-of-The-Art Czech Named Entity Recognizer*. In: *Lecture Notes in Computer Science*, Vol. 8082, *Text, Speech and Dialogue: 16th International Conference, TSD 2013. Proceedings*, Springer Verlag, Berlin / Heidelberg, ISBN 978-3-642-40584-6, ISSN 0302-9743, pp. 68-75, 2013

Uřešová Zdeňka, Fučíková Eva, Hajič Jan, Šindlerová Jana: *An Analysis of Annotation of Verb-Noun Idiomatic Combinations in a Parallel Dependency Corpus*. In: *The 9th Workshop on Multiword Expressions (MWE 2013)*, Association for Computational Linguistics, Atlanta, Georgia, USA, ISBN 978-1-937284-47-3, pp. 58-63, 2013

Hajič Jan, Hajičová Eva, Panevová Jarmila, Sgall Petr, Bojar Ondřej, Cinková Silvie, Fučíková Eva, Mikulová Marie, Pajas Petr, Popelka Jan, Semecký Jiří, Šindlerová Jana, Štěpánek Jan, Toman Josef, Uřešová Zdeňka, Žabokrtský Zdeněk: *Announcing Prague Czech-English Dependency Treebank 2.0*. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, European Language Resources Association, İstanbul, Turkey, ISBN 978-2-9517408-7-7, pp. 3153-3160, 2012

Zeman Daniel, Mareček David, Popel Martin, Ramasamy Loganathan, Štěpánek Jan, Žabokrtský Zdeněk, Hajič Jan: *HamleDT: To Parse or Not to Parse?*. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, European Language Resources Association, İstanbul, Turkey, ISBN 978-2-9517408-7-7, pp. 2735-2741, 2012

Hajič Jan, Pajas Petr, Ircing Pavel, Romportl Jan, Peterek Nino, Spousta Miroslav, Mikulová Marie, Grüber Martin, Legát Milan: *Pražská databáze mluvené češtiny*. Data/software, Institute of Formal and Applied Linguistics, Prague, <http://ufal.mff.cuni.cz/pdts1.0>, Feb 2011

Mareček David, Popel Martin, Ramasamy Loganathan, Štěpánek Jan, Zeman Daniel, Žabokrtský Zdeněk, Hajič Jan: *HamleDT - HARmonized Multi-LanguagE Dependency Treebank*. Data/software, Institute of Formal and Applied Linguistics, MFF UK, <http://ufal.mff.cuni.cz/hamledt>, 2011

Zeman Daniel, Popel Martin, Mareček David, Ramasamy Loganathan, Štěpánek Jan, Žabokrtský Zdeněk, Hajič Jan: *From the Jungle to a Park: Harmonizing Dependency Treebanks of 30 Languages*. Talk, Seminář ÚFAL, Univerzita Karlova v Praze, Praha, Czechia, Oct 2011

Hajič Jan, Ciaramita Massimiliano, Johansson Richard, Kawahara Daisuke, Martí Maria Antònia, Màrquez Lluís, Meyers Adam, Nivre Joakim, Padó Sebastian, Štěpánek Jan, Straňák Pavel, Surdeanu Mihai, Xue Nianwen, Zhang Yi: *The CoNLL-2009 Shared Task: Syntactic and Semantic Dependencies in Multiple Languages*. In: *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL): Shared Task*, Association for Computational Linguistics, Boulder, CO, USA, ISBN 978-1-932432-29-9, pp. 1-18, 2009

Lin Jimmy, Murray Craig G., Dorr Bonnie J., Hajič Jan, Pecina Pavel: *A Cost-effective Lexical Acquisition Process for Large-scale Thesaurus Translation*. In: *Language Resources and Evaluation*, Vol. 43, No. 1, Springer Netherlands, ISSN 1574-020X, pp. 27-40, Mar 2009

Spoustová Drahomíra, Hajič Jan, Raab Jan, Spousta Miroslav: *Semi-Supervised Training for the Averaged Perceptron POS Tagger*. In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Athina, Greece, ISBN 978-1-932432-16-9, pp. 763-771, 2009

Hajič Jan, Panevová Jarmila, Hajičová Eva, Sgall Petr, Pajas Petr, Štěpánek Jan, Havelka Jiří, Mikulová Marie, Žabokrtský Zdeněk, Ševčíková-Razímová Magda, Uřešová Zdeňka: *Prague Dependency Treebank 2.0*. Software prototype, Linguistic Data Consortium, Philadelphia, PA, USA, ISBN 1-58563-370-4, <http://www ldc.upenn.edu>, LDC Catalog No. LDC2006T01, sJul 2006

McDonald Ryan, Pereira Fernando, Ribarov Kiril, Hajič Jan: *Non-Projective Dependency Parsing using Spanning Tree Algorithms*. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Vancouver, BC, Canada, ISBN 1-932432-55-8, pp. 523-530, 2005

Hajič Jan, Collins Michael, Ramshaw Lance, Tillmann Christoph: *A Statistical Parser for Czech*. In: *ACL'99*, Maryland, USA, p. 505-512, 1999