

Jan Hajič

Přehled publikací, duben 2008

A Monografie

- [1] Jan Hajič, *KODAS - A Simple Method of Natural Language Interface to a Database*, Explizite Beschreibung der Sprache und automatische Textbearbeitung #XI, Charles University, Prague, 1984, 188 pp.
- [2] Jan Hajič, *NALCOM: A Multilevel NL-interface*, Explizite Beschreibung der Sprache und automatische Textbearbeitung #XV, Charles University, Prague, 1988, 146 pp.
- [3] Jan Hajič, *Disambiguation of Rich Inflection (Computational Morphology of Czech)*, Karolinum, Charles University Press Prague, Prague, Czech Rep., 2004, ISBN 80-246-0282-2, 324 pp.
- [4] Barbora Vidová Hladká, Jan Hajič, Jiří Hana, Jaroslava Hlaváčová, Jiří Mírovský & Jan Votrúbec, *Czech Academic Corpus 1.0 Guide*, Karolinum - Charles University Press, 2007, ISBN 978-80-246-1315-4.

B Kapitoly v monografiích

- [1] Petr Sgall, Alena Appelová, Alla Bémová, Jan Hajič, Zdeněk Kirschner, Květa Králíková & Jarmila Panevová, *Využitie lingvistických metód vo VTEI*, STK, Bratislava, 1984, 115 pp.
- [2] Jan Borota, Jan Hajič, Eva Hajičová, Milena Hnátková, Vladislav Kuboň, Karel Oliva, Jarmila Panevová & Petr Sgall, *Text-And-Inference Based Approach to Question Answering*, Theoretical and Computational Linguistics #3, Institute of Theoretical and Computational Linguistics, FF UK, Charles University, Prague, 1995, Kap. 3, 52–72.
- [3] Alena Böhmová, Jan Hajič, Eva Hajičová & Barbora Hladká, *The Prague Dependency Treebank: A Three-Level Annotation Scenario*, Anne Abeillé (ed.): Building and Using Parsed Corpora, In: Text, Speech and Language Technology #20, Kluwer Academic Publishers, Dordrecht/Boston/London / Springer, 2003, ISBN 1-4020-1335-3. Chapter 7, 103–127.
- [4] Jan Hajič, *Linguistics Meets Exact Sciences*, Susan Schreibmann, Ray Siemens, John Unsworth (eds.): A Companion to Digital Humanities, Blackwell Publishing Ltd., Malden/Oxford/Victoria, 2004, ISBN 1-4051-0321-3. Chapter 7, 79–87.

- [5] Jan Hajič, *Complex Corpus Annotation: The Prague Dependency Treebank*, Mária Šimková (ed.): Insight into Slovak and Czech Linguistics, Nakladatelstvo Veda, Bratislava, Slovakia, 2005, ISBN 80-224-0880-8, in print.
- [6] Jan Hajič, *Treebanks and Tagsets*, Keith Brown (ed.): Encyclopedia of Language and Linguistics, 2nd ed., Elsevier, The Netherlands, 2006, Vol. 13. pp. 109–114. ISBN: 0-08-044299-4.
- [7] Otakar Smrž & Jan Hajič, *The Other Arabic Treebank: Prague Dependencies and Functions*, Ali Farghaly (ed.): Arabic Computational Linguistics: Current Implementations, CSLI Publications, Stanford, CA, USA, 2007.
- [8] Eva Hajičová, Jan Cuřín, Jan Hajič, Ondřej Kučera & Barbora Vidová Hladká, *Jazyk a umělá inteligence: kudy a kam?*, Academia, Praha, 2007, Chapter 14, pp. 272–283. ISBN 978-80-200-1407-2.
- [9] Otakar Smrž, Petr Pajas, Zdeněk Žabokrtský, Jan Hajič, Jiří Mírovský & Petr Němec, *Learning to Use the Prague Arabic Dependency Treebank*, Elabbas Benmamoun (ed.): Perspectives on Arabic Linguistics. Current Issues in Linguistic Theory 289, John Benjamins, Urbana-Champaign, IL, USA, 2007, Vol. XIX. ISBN 978-90-272-4804-6.

#### C1 Pův. věd. práce v zahraničních časopisech

- [1] William Byrne, David Doermann, Martin Franz, Samuel Gustman, Jan Hajič, Douglas Oard, Michael Picheny, Josef Psutka, Bhuvana Ramabhadran, Dagobert Soergel, Todd Ward & Wei-Jing Zhu, “Automatic Recognition of Spontaneous Speech for Access to Multilingual Oral History Archives,” *IEEE Transactions on Speech and Audio Processing* 12, No. 4 (2004), 420–435.
- [2] Jimmy Lin, Craig Murray, Bonnie Dorr, Jan Hajič & Pavel Pecina, “A Cost-effective Lexical Acquisition Process for Large-scale Ontology Translation,” *Multilingual Language Resources and Interoperability, Special Issue of Language Resources and Evaluation* (2008), in print.

#### C2 Pův. věd. práce v domácích časopisech

- [1] Jan Hajič, “Jednoduchý systém kontaktu s bází dat v češtině (A System of Querying a Database in Czech),” *Československá informatika* 12 (1980), 334–338.
- [2] Jan Hajič, “KODAS - A Natural Language Interface to a Simple Database,” *The Prague Bulletin of Mathematical Linguistics* 39 (1983), 65–75.
- [3] Jan Hajič, “Programming a Simple Natural Language Interface (Part 1),” *The Prague Bulletin of Mathematical Linguistics* 43 (1985), 37–62.

- [4] Jan Hajič, “Programming a Simple Natural Language Interface (Part 2),” *The Prague Bulletin of Mathematical Linguistics* 44 (1986), 23—56.
- [5] Jan Hajič, “Semi-automatic Classification of Czech Nouns,” *The Prague Bulletin of Mathematical Linguistics* 65-66 (1996).
- [6] Jan Hajič & Barbora Hladká, “Morfologické značkování korpusu českých textů stochastickou metodou,” *Slovo a slovesnost* 58, 4 (1997), 288–304, AV ČR.
- [7] Jan Hajič, Eva Hajičová, Jarmila Panevová & Petr Sgall, “Syntax v českém národním korpusu,” *Slovo a slovesnost* 59, 3 (1998), 168–177, AV ČR.
- [8] Lukasz Debowski, Jan Hajič & Vladislav Kuboň, “Testing the Limits - Adding a New Language to an MT System,” *The Prague Bulletin of Mathematical Linguistics* 78 (2002), 95—102.
- [9] Jan Hajič & Václav Honetschläger, “Annotation Lexicons: Using the Valency Lexicon for Tectogrammatical Annotation,” *The Prague Bulletin of Mathematical Linguistics* 79–80 (2003), 61—86.

### C3 Pův. věd. práce v mezinárodně recenzovaných sbornících

- [1] Petr Jirků & Jan Hajič, “Inferencing and search for an answer in TIBAQ,” *Abstracts of COLING’82* (1982), 139—141, Prague, July 5th–10th, 1982.
- [2] Jan Hajič, “An MT System for Closely Related Languages,” *Proceedings of the Third Conference of the European Chapter of the Association for Computational Linguistics* (1987), 113—117, Copenhagen, April 1st–3rd 1987.
- [3] Jan Hajič, “Formal Morphology,” *Proceedings of the 12th International Conference on Computational Linguistics* (1988), 222—224, Budapest, August 22nd–27th 1988.
- [4] Jan Hajič, “Morphotactics by Attribute Grammar,” *Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics* (1989), Supplement, 8 pp..
- [5] Jan Hajič & Januš Drózd, “Spelling-checking for Highly Inflective Languages,” *Proceedings of the 13th International Conference on Computational Linguistics III* (1990), 358—360, Helsinki, August 1990.
- [6] Susan Warwick, Jan Hajič & Graham Russell, “Searching on Tagged Corpora: Linguistically Motivated Concordance Analysis,” *Electronic Text Research: Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research* (1990), 10—18, University of Waterloo, 28th–30th October 1990.
- [7] Susan Warwick-Armstrong, Amy Winarske, Graham Russell & Jan Hajič, “Tagging and Alignment of Parallel Texts: Current Status of BCP,” *Proceedings of the 3rd Conference on Applied Natural Language Processing* (1992), 227—228, Trento, March 31st–April 3rd, 1992.

- [8] Alexandr Rosen, Eva Hajičová & Jan Hajič, “Derivation of Underlying Valency Frames from a Learners’ Dictionary,” *Proceedings of the 14th International Conference on Computational Linguistics 2* (1992), 553—559, Nantes, July 22nd–28th 1992.
- [9] Jan Hajič, Alexandr Rosen & Eva Hajičová, “Machine Translation Research in Czechoslovakia,” *META* 37(4) (1992), 802—816.
- [10] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Meredith J. Goldsmith, Jan Hajič, Robert L. Mercer & Surya Mohanty, “But Dictionaries are Data Too,” *Proceedings of the HLT, 1993* (1993), 202—205.
- [11] Jan Hajič & Barbora Hladká, “A Simple Czech and English Probabilistic Tagger: a Comparison,” *Proceedings of the TELRI 1st European Seminar: Language Resources for Language Technology* (ed. by Heike Rettig) (1996), 191–196, Sept. 15-16, 1995.
- [12] Jan Hajič & Barbora Hladká, “Probabilistic and Rule-Based Tagger of an Inflective Language - a Comparison,” *Proceedings of the 5th Conference on Applied Natural Language Processing, Association for Computational Linguistics, March 28–30* (1997), 136–143.
- [13] Jan Hajič & Kiril Ribarov, “Rule-Based Dependencies,” *Proceedings of the Workshop on Empirical Learning of Natural Language Processing Tasks* (1997), 125—136, April 23-25, 1997.
- [14] Jan Hajič & Eva Hajičová, “Syntactic Tagging in the Prague Tree Bank,” *Proceedings of the TELRI 2nd European Seminar: Language Applications for a Multilingual Europe* (ed. by R. Marcinkeviciene and N. Volz) (1997), 55—68, April 17–20, 1997.
- [15] Jan Hajič & Barbora Hladká, “Czech Language Processing - Czech POS Tagging,” *Proceedings of the LREC’98* (1998), 931—936, May 28-30, 1998.
- [16] Jan Hajič & Barbora Hladká, “Tagging Inflective Languages: Predicting Morphological Categories for a Rich, Structured Tagset,” *Proceedings of the COLING/ACL, Aug. 5th-9th, 1998* (1998), 483—490, Association for Computational Linguistics/ICCL.
- [17] Michael Collins, Jan Hajič, Eric Brill, Lance Ramshaw & Christopher Tillmann, “A Statistical Parser for Czech,” *Proceedings of the 37th Meeting of the Association for Computational Linguistics* (1999), 505—512.
- [18] Alevtina Bémová, Jan Hajič, Barbora Hladká & Jarmila Panevová, “Morphological and Syntactic Tagging of the Prague Dependency Treebank,” *Proceedings of the ATALA Workshop, Paris, France* (1999), June 18-19.
- [19] Jan Hajič, “Morphological Tagging: Data vs. Dictionaries,” *Proceedings of the 1st North American Association for Computational Linguistics Chapter Meeting* (2000), 94—101.
- [20] Jan Hajič, Jano Hric & Vladislav Kuboň, “Česílko: machine translation between closely related languages,” *Proceedings of the 6th Conference on Applied Natural Language Processing* (2000), 7—12, Association for Computational Linguistics.

- [21] Ondřej Cikhart & Jan Hajič, “Word Sense Disambiguation of Czech Texts,” *Proceedings of the Text, Speech and Dialogue 1999, Lecture Notes in Artificial Intelligence* (1999), 109—114, Springer Verlag, Vol. 1692.
- [22] William Byrne, Jan Hajič, Pavel Ircing, Frederick Jelinek, Sanjeev Khudanpur, John McDonough, Nino Peterek & Josef Psutka, “Large Vocabulary Speech Recognition for Read and Broadcast Speech in Czech,” *Proceedings of the Text, Speech and Dialogue 1999, Lecture Notes in Artificial Intelligence* (1999), 235—240, Springer Verlag, Vol. 1692.
- [23] William Byrne, Jan Hajič, Pavel Krbec, Pavel Ircing & Josef Psutka, “Morpheme Based Language Models for Speech Recognition of Czech,” *Proceedings of the Text, Speech and Dialogue 2000, Lecture Notes in Artificial Intelligence* (2000), 211—216, Springer Verlag, Vol. 1902.
- [24] Jan Hajič, Petr Pajas & Barbora Hladká, “The Prague Dependency Treebank: Annotation Structure and Support,” *IRCS Workshop on Linguistic Databases, Philadelphia, PA, Dec. 11–13* (2001), 105—114.
- [25] Eva Hajičová, Jan Hajič, Martin Holub, Veronika Řezníčková, Petr Pajas, Barbora Vidová Hladká & Petr Sgall, “The Current Status of the Prague Dependency Treebank,” *Proceedings of the Text, Speech and Dialogue 2001, Lecture Notes in Artificial Intelligence* (2001), 11—20, Springer Verlag, Vol. 2166.
- [26] Jan Hajič, “Statistické modelování a automatická analýza přirozeného jazyka (morfolgie, syntax, překlad),” *Slovenčina a čeština v počítačovom spracovaní (SLOVKO)*, ed. A. Jarošová (2001), 11—33, 26.-27.10.2001, ISBN 80-224-0692-9.
- [27] Douglas Oard, Dina Demner-Fushman, Jan Hajič, Bhuvana Ramabhadran, Samuel Gustman, William J. Byrne, Dagobert Soergel, Bonnie Dorr, Philip Resnik & Michael Picheny, “Cross-Language Access to Recorded Speech in the MALACH Project,” *Proceedings of the Text, Speech and Dialogue 2002, Lecture Notes in Artificial Intelligence* (2002), 57—64, Springer Verlag, Vol. 2448.
- [28] Josef Psutka, Pavel Ircing, Josef V. Psutka, Vlasta Radová, William J. Byrne, Jan Hajič, Samuel Gustman & Bhuvana Ramabhadran, “Automatic Transcription of Czech Language Oral History in the MALACH Project: Resources and Initial Experiments,” *Proceedings of the Text, Speech and Dialogue 2002, Lecture Notes in Artificial Intelligence* (2002), 253—260, Springer Verlag, Vol. 2448.
- [29] Jan Hajič, Pavel Krbec, Pavel Květoň, Karel Oliva & Vladimír Petkevič, “Serial Combination of Rules and Statistics: a Case Study in Czech Tagging,” *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics* (2001), 260—267.
- [30] Josef Psutka, Pavel Ircing, Josef V. Psutka, Vlasta Radová, William J. Byrne, Jan Hajič, Jiří Mírovský & Samuel Gustman, “Large Vocabulary ASR for Spontaneous Czech in the MALACH Project,” *Poster in Proceedings of the 8th European Conference on Speech Communication and Technology* (2003), 1821—1824.

- [31] Pavel Krbec, Petr Podveský & Jan Hajič, “Combination of a Hidden Tag Model and a Traditional N-Gram Model: A Case Study in Czech Speech Recognition,” *Proceedings of the 8th European Conference on Speech Communication and Technology* (2003), 2289—2291.
- [32] Josef Psutka, Pavel Ircing, Josef V. Psutka, Vlasta Radová, William J. Byrne, Veera Venkataramani, Jan Hajič & Samuel Gustman, “Towards Automatic Transcription of Spontaneous Czech Speech in the MALACH Project,” *Proceedings of the Text, Speech and Dialogue 2003, Lecture Notes in Artificial Intelligence* (2003), 214–219, Springer Verlag, Vol. 2807.
- [33] Josef Psutka, Ilja Iljuchin, Pavel Ircing, Josef V. Psutka, Václav Trejbal, William J. Byrne, Jan Hajič & Samuel Gustman, “Building LVCSR System for Transcription of Spontaneously Pronounced Russian Testimonies in the MALACH Project: Initial Steps and First Results,” *Poster in Proceedings of the Text, Speech and Dialogue 2003, Lecture Notes in Artificial Intelligence* (2003), 327—332, Springer Verlag, Vol. 2807.
- [34] Jan Hajič & Zdeňka Urešová, “Linguistic Annotation: from Links to Cross-Layer Lexicons,” *Proceedings of the 2nd International Workshop on Treebanks and Linguistic Theories* (2003), 69—80.
- [35] Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Rezníčková & Petr Pajas, “PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation,” *Proceedings of the 2nd International Workshop on Treebanks and Linguistic Theories* (2003), 57—68.
- [36] Jan Hajič, Vladislav Kuboň & Petr Homola, “A Simple Multilingual Machine Translation System,” *Proceedings of Machine Translation Summit IX* (2003), 157—164, New Orleans, Sept. 2003.
- [37] Jan Hajič, Josef Psutka, Pavel Ircing, William Byrne, Jiří Mírovský, Bhuvana Ramabhadran, Samuel Gustman, Josef V. Psutka & Vlasta Radová, “Language Model Data Selection for Czech ASR in the MALACH Project,” *Proceedings of ICASSP (CDROM)* (2003), , Scheduled for Hong Kong, China, April 6–10.
- [38] Josef Psutka, Jan Hajič & William J. Byrne, “The Development of ASR for Slavic Languages in the MALACH Project,” *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2004), 749—752, Montreal, Canada, May 17–21.
- [39] Martin Čmejrek, Jan Cuřín, Jiří Havelka, Jan Hajič & Vladislav Kuboň, “Prague Czech-English Dependency Treebank: Syntactically Annotated Resources for Machine Translation,” *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)* (2004), 1597—1600, Lisboa, Portugal, May 26–28.
- [40] Jan Hajič, Martin Holub, Marie Hučínová, Martin Pavlík, Pavel Pecina, Pavel Straňák & Pavel Šidák, “Validating and Improving the Czech WordNet via Lexico-Semantic Annotation of the Prague Dependency Treebank,” *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)* (2004), Lisboa, Portugal, May 26–28.

- [41] Josef Psutka, Pavel Ircing, Jan Hajič, Vlasta Radová, Josef V. Psutka & William Byrne, “Issues in annotation of the Czech spontaneous speech corpus in the MALACH project,” *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC)* (2004), 607—610, Lisboa, Portugal, May 26–28.
- [42] Otakar Smrž, Petr Pajas, Zdeněk Žabokrtský, Jan Hajič, Jiří Mírovský & Petr Němec, “Learning to Use the Prague Arabic Dependency Treebank,” *Elabbas Benmamoun (ed.): Proceedings of Annual Symposium on Arabic Linguistics (ALS-19)* (2005), Urbana, IL, USA, April 1–3.
- [43] Martin Čmejrek, Jan Cuřín, Jan Hajič & Jiří Havelka, “Prague Czech-English Dependency Treebank: Resource for Structure-based MT,” *Proceedings of the 10th European Association for Machine Translation Conference* (2005), 73—78, ISBN 963-9206-04-0, Budapest, Hungary, May 30-31.
- [44] Josef Psutka, Pavel Ircing, Vlasta Radová, Josef V. Psutka, Jan Hajič, William Byrne & Jiří Mírovský, “Automatic transcription of Czech, Russian, and Slovak Spontaneous Speech in the MALACH Project,” *Proceedings of Eurospeech* (2005), 1349—1352, Lisboa, Portugal, Sept. 4–8.
- [45] Ryan McDonald, Fernando Pereira, Kiril Ribarov & Jan Hajič, “Non-Projective Dependency Parsing using Spanning Tree Algorithms,” *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HTL/EMNLP)* (2005), 523–530, Vancouver, BC, Canada, Oct. 6–8.
- [46] Jan Hajič, Otakar Smrž, Tim Buckwalter & Hubert Jin, “Feature-based Tagger of Approximations of Functional Arabic Morphology,” *Proceedings of Treebanks and Linguistic Theories* (2005), 53—64, ISBN 84-475-2992-4, Barcelona, Spain, Dec. 9–10.
- [47] Ondřej Bojar & Jan Hajič, “Extracting Translation Verb Frames,” *Proceedings of Modern Approaches in Translation Technologies at the Recent Advances in NLP* (2005), 2—6, ISBN 954-90906-9-8, Borovetz, Bulgaria, Sept. 24–25.
- [48] Craig Murray, Bonnie Dorr, Jimmy Lin, Pavel Pecina & Jan Hajič, “Leveraging Recurrent Phrase Structure in Large-scale Ontology Translation,” *Proceedings of the 11th Annual conference of the European Association for Machine Translation (EAMT)* (2006), 1—10, Oslo, Norway. ISBN 82-7368-294-3.
- [49] Craig Murray, Bonnie Dorr, Jimmy Lin, Pavel Pecina & Jan Hajič, “Leveraging Reusability: Cost-Effective Lexical Acquisition for Large-scale Ontology Translation,” *Proceedings of the 21th International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL)* (2006), 945—952, ISBN 1-932432-69-8.
- [50] Václav Novák & Jan Hajič, “Perspectives of Turning Prague Dependency Treebank into a Knowledge Base,” *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC)* (2006), 439—442, Genova, Italy, ISBN 2-9517408-2-4.

- [51] Jan Hajič & Eva Hajičová, “Some of Our Best Friends are Statisticians,” *Proceedings of the 10th International Conference on Text, Speech and Dialogue. Lecture Notes in Artificial Intelligence (LNAI/LNCS)* (2007), 2—11, Pilsen, Czech Republic, Vol. 4629, No. XVII. ISBN 978-3-540-74627-0. ISSN 0302-9743.
- [52] Drahomíra Spoustová, Jan Hajič, Jan Votrubec, Pavel Krbec & Pavel Květoň, “The Best of Two Worlds: Cooperation of Statistical and Rule-Based Taggers for Czech,” *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing, at the 45th Meeting of the Association for Computational Linguistics* (2007), 67—74, Prague, Czech Republic.

#### C4 Pův. věd. práce v domácích recenzovaných sbornících

- [1] Jan Hajič & Eva Hajičová, “Jednoduchý systém kontaktu s bází dat v češtině (A System of Querying a Database in Czech),” *Proceedings of the SOFSEM’80* (1980), 356—358, Bílý Kříž, November 23th–December 5th, 1980.
- [2] Petr Sgall & Jan Hajič, “Kontakt s bází dat v češtině (Querying a Database in Czech),” in *Využitie lingvistických metód vo VTEI (The Use of Linguistic Methods in Information Services) #*, Slovak Technical Society, Bratislava, 1984, 33—65.
- [3] Jan Borota & Jan Hajič, “Metoda MOZAIKA (The Method MOSAIC),” *Proceedings of the Moderní programování (Modern programming)* (1984), 75—83, Trutnov, Czechoslovakia, May 27th–June 4th 1984.
- [4] Jan Hajič, Zdeněk Kirschner & Alexandr Rosen, “Některé problémy aplikací strojového překladu (On Some Problems of Machine Translation Applications),” *Proceedings of the Využití lingvistických přístupů v informatice (Using the Linguistic Approach in Information Retrieval)* (1986), 20—25, June 9th–10th 1986.
- [5] Jan Hajič & Vladimír Zvelebil, “Komunikační interface v češtině (Communication Interface in Czech),” *Proceedings of the Využití lingvistických přístupů v informatice (Using the Linguistic Approach in Information Retrieval)* (1986), 37—43, June 9th–10th 1986.
- [6] Jan Hajič & Karel Oliva, “Projekt česko–ruského strojového překladu (The Czech To Russian Machine Translation Project),” *Proceedings of the SOFSEM’86* (1986), 202—205, Liptovský Ján, November 23th–December 5th 1986.
- [7] Jan Hajič, “RUSLAN - a MT from Czech to Russian,” *Proceedings of the AI’87* (1987), ČSVTS Prague, April 1987.
- [8] Jan Hajič, “Formální morfologie (Formal Morphology),” *Proceedings of the AI’88* (1988), 225—232, ČSVTS Prague, March 8th–10th, 1988.
- [9] Jan Hajič, “Strojový překlad v Japonsku (Machine translation activities in Japan),” *Proceedings of the AI’89* (1989), ČSVTS Prague, April 24th–26th, 1989.

- [10] Januš Drózd & Jan Hajič, “Kontrola českého pravopisu na PC (Czech spelling-checker on a PC),” *Proceedings of the AI’90* (1990), 215—222, ČSVTS Prague, April 12th–14th, 1990.
- [11] Jan Hajič, Alexandr Rosen, Eva Hajičová & Hana Skoumalová, “MATRACE - MACHine TRANslation between Czech and English,” *Proceedings of the The IBM AI Projects Seminar* (1992), 75—82, November 18th–19th 1992.
- [12] Jan Hajič, “Počítačová lingvistika,” *Proceedings of the EurOpen’97* (1997), 44—59, 23.6.-25.6.1997.
- [13] Jan Hajič, “Building a Syntactically Annotated Corpus: The Prague Dependency Treebank,” *Festschrift for Jarmila Panevová* (1998), 106—132, Karolinum, Charles University.
- [14] Jan Hajič & Vladislav Kuboň, “Tagging as a Key To Successful MT,” *MIS 2003* (2003), 56—65, MATFYZPRESS.

## E1 Výzkumné zprávy

- [1] Peter Brown, Robert Mercer, Vincent DellaPietra, Stephen DellaPietra, Jan Hajič, John Gillet & Meredith Goldsmith, “MT Project Candide: Report for DAPRA Evaluation,” Newton, Cambridge, MA, 1993, IBM Research.
- [2] Vladimír Petkevič, ed., “Počítačový fond češtiny,” výroční zpráva pro GAČR, 1993, Prague, 1993.
- [3] Vladimír Petkevič, ed., “Počítačový fond češtiny,” výroční zpráva pro GAČR, 1994, Prague, 1994.
- [4] Vladimír Petkevič, ed., “Počítačový fond češtiny (1993-1995),” výroční zpráva pro GAČR, 1995, Prague, 1995.
- [5] Vladimír Petkevič, ed., “Programové nástroje pro zprac. textu,” výroční zpráva pro GAČR, 1995, Prague, 1995.
- [6] Eva Hajičová, ed., “Projekt MATRACE,” výroční zpráva pro GAČR, 1993, Prague, 1993.
- [7] Eva Hajičová, ed., “Projekt MATRACE,” výroční zpráva pro GAČR, 1994, Prague, 1994.
- [8] Eva Hajičová, ed., “Projekt MATRACE (1993-1995),” závěrečná zpráva pro GAČR, Prague, 1995.
- [9] Jan Hajič, ed., “Český tagger,” Dílčí zpráva projektu (GAUK 39/94), Prague, 1994.
- [10] Jan Hajič, ed., “Český tagger,” Dílčí zpráva projektu (GAUK 39/94), Prague, 1995.
- [11] Jan Hajič, ed., “Český tagger,” Závěrečná zpráva projektu (GAUK 39/94), Prague, 1996.
- [12] Jan Hajič, “Textual Corpora in Linguistics: Annotation and Comparison,” Průběžná zpráva pro RSS/HESP (Soros Foundation), Prague, 1996.

- [13] Jan Hajič & Hana Skoumalová, “Czech Core Lexicon,” Report, project Copernicus CEGLEX (EU 1032/96), D10-T5.3, Prague, 1996.
- [14] Jan Hajič, “Using the Czech Core Lexicon in an Application Context,” Report, project Copernicus CEGLEX (EU 1032/96), D13-T8.2, Prague, 1996.
- [15] Eva Hajičová, ed., “Projekt Formální reprezentace jazykových struktur (1996-1998),” průběžná zpráva pro GAČR, Prague, 1996.
- [16] Eva Hajičová, ed., “Komplexní projekt Čeština ve věku počítačů (1996-2001),” průběžná zpráva pro GAČR, Prague, 1996.
- [17] Eva Hajičová, ed., “Projekt podpory výzkumu na VŠ VS96151 Laboratoř pro zpracování jazykových dat (1996-2000),” průběžná zpráva pro MŠMT, Prague, 1996.
- [18] Jan Hajič, Eva Hajičová & Alexandr Rosen, “Formal Representation of Language Structures,” *TELRI Newsletter* 3 (1996), 12–19, June 1996.
- [19] Eva Hajičová, ed., “Projekt Formální reprezentace jazykových struktur (1996-1998),” průběžná zpráva pro GAČR, Prague, 1997.
- [20] Eva Hajičová, ed., “Komplexní projekt Čeština ve věku počítačů (1996-2001),” průběžná zpráva pro GAČR, Prague, 1997.
- [21] Eva Hajičová, ed., “Projekt podpory výzkumu na VŠ VS96151 Laboratoř pro zpracování jazykových dat (1996-2000),” průběžná zpráva pro MŠMT, Prague, 1997.
- [22] Eva Hajičová, ed., “Projekt Formální reprezentace jazykových struktur (1996-1998),” závěrečná zpráva pro GAČR, Prague, 1998.
- [23] Eva Hajičová, ed., “Komplexní projekt Čeština ve věku počítačů (1996-2001),” průběžná zpráva pro GAČR, Prague, 1998.
- [24] Eva Hajičová, ed., “Projekt podpory výzkumu na VŠ VS96151 Laboratoř pro zpracování jazykových dat (1996-2000),” průběžná zpráva pro MŠMT, Prague, 1998.
- [25] Jan Hajič, ed., “Projekt KONTAKT ME293 Rozpoznávání mluveného slovanského jazyka,” průběžná zpráva pro MŠMT, Prague, 1998.
- [26] Jan Hajič, ed., “Projekt EU INCO STEEL PL1113, set of intermediate reports ,” Tagging terminological dictionaries, WP2.1, WP2.2, Prague, 1998.
- [27] Jan Hajič, Eric Brill, Michael Collins, Barbora Hladká, Doug Jones, Cynthia Kuo, Lance Ramshaw, Oren Schwartz, Christopher Tillmann & Daniel Zeman, “Core Natural Language Processing Technology Applicable to Multiple Languages ,” *CLSP Technical Reports*, Baltimore, MD (1998).
- [28] Jan Hajič & Barbora Hladká, “Error-driven Tagging for a Rich, Structured Tagset Based on an Exponential Model,” *Technical Report 4* (1997), ÚFAL MFF UK.
- [29] Alla Bémová, Eva Buráňová, Jan Hajič, Jiří Kárník, Petr Pajas, Jarmila Panevová, Jan Štěpánek & Zdeňka Uřešová, “Příručka pro anotátory (Annotator’s Handbook). Draft.,” *Technical Report 3* (1997), ÚFAL MFF UK.

- [30] Jan Hajič, Jason Eisner, Gerald Penn, Owen Rambow, Drago Radev, Martin Čmejrek, Yuan Ding, Terry Koo & Kristen Parton, “Natural Language Generation in the Context of Machine Translation,” *CLSP Technical Reports*, Baltimore, MD (2002).
- [31] Jiří Hana, Daniel Zeman, Jan Hajič, Hana Hanová, Barbora Hladká & Emil Jeřábek, “Manuál pro morfologickou anotaci pro Pražský závislostní korpus 2.0,” *Technická zpráva ÚFAL TR-2005-27* (2005), ISSN 1214-5521, 55 pp..
- [32] Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Ivona Kučerová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá & Zdeněk Žabokrtský, “Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka,” *Technická zpráva ÚFAL TR-2005-28* (2005), ISSN 1214-5521, 1185 pp..

## E2 Softwarové produkty

- [1] Jan Hajič, “Morfologický analyzátor pro flexivní jazyky,” 1988-2000, Použito pro Český národní korpus.
- [2] Jan Hajič, Hana Skoumalová & Januš Drózd, “Morfologický slovník češtiny,” 1988-2000, Použito pro Český národní korpus.
- [3] Jan Hajič & Barbora Hladká, “Tagger češtiny,” 1994-2000, Použito pro Český národní korpus.
- [4] Jan Hajič & Hana Skoumalová, “Malý morfologický slovník češtiny ve tvaru ”Two-level morphology”,” 1996-1997, Použito pro Morphologic, Inc. (Maďarsko) a IBM.
- [5] Jan Hajič, Hana Skoumalová & Barbora Hladká, “Morfologický slovník češtiny a tagger pro Xerox Language Tools,” 1996-1998, Použito pro Xerox Research Center Europe.
- [6] Jan Hajič, Hana Skoumalová, Jan Hric & Ondřej Cikhart, “Systém lemmatizace a lematizační slovník češtiny a slovenštiny,” 1991-2000, Použito pro BYLL Software.
- [7] Jan Hajič, Vladislav Kuboň & Jan Hric, “Překlad z češtiny do slovenštiny,” 1999-2000.
- [8] Jan Hajič, Eva Hajičová, Petr Pajas, Jarmila Panevová, Petr Sgall & Barbora Vidová Hladká, *The Prague Dependency Treebank. CDROM Catalog No. LDC2001T10*, Linguistic Data Consortium, Univ. of Pennsylvania, Philadelphia, PA, USA, 2001, ISBN 1-58563-212-0.
- [9] Jan Cuřín, Martin Čmejrek, Jiří Havelka, Jan Hajič, Vladislav Kuboň & Zdeněk Žabokrtský, *Prague Czech-English Dependency Treebank v 1.0. CDROM Catalog No. LDC2004T25*, Linguistic Data Consortium, Univ. of Pennsylvania, Philadelphia, PA, USA, 2004, ISBN 1-58563-321-6.

- [10] Jan Hajič, Otakar Smrž, Petr Zemánek, Petr Pajas, Jan Šnidauf, Emanuel Beška, Jakub Kráčmar & Kamila Hassanová, *Prague Arabic Dependency Treebank v 1.0. CDROM Catalog No. LDC2004T23*, Linguistic Data Consortium, Univ. of Pennsylvania, Philadelphia, PA, USA, 2004, ISBN 1-58563-319-4.

## F Zvané přednášky na význ. konferencích

- [1] Jan Hajič, “Multilanguage Morphology with Broad Coverage,” *ISSCO 21th Anniversary Conference SwAN’21* (1993), ISSCO, Geneva, September 5th–9th, 1993.
- [2] Jan Hajič, “Machine Translation in Czechoslovakia,” Luxembourg, European Commission, 1995, Association for Machine Translation (AMTA).
- [3] Jan Hajič, “Tectogrammatical Representation: Towards a Minimal Transfer in Machine Translation,” *Proceedings of the Sixth International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+6)*, Venice, Italy (2002), Università di Venezia.
- [4] Jan Hajič, “Prague Dependency Treebank and Other Languages,” *Swedish National Corpus Conference*, Vaxjo, Sweden, Nov. 2002 (2002).
- [5] Jan Hajič, “Možnosti použití Pražského závislostního korpusu pro projekt Slovenského národního korpusu,” *Jazykovedný ústav Ľudovíta Štúra, Slovenská akadémia vied*, Bratislava, Dec. 2002 (2002).
- [6] Jan Hajič, “The Tectogrammatical Level of the Prague Dependency Treebank,” *Linguistically Interpreted Corpora ’03, Workshop at the EACL’03*, Budapest, May 2003 (2003).

## I Disertační práce

- [1] Jan Hajič, “Softwarový systém pro analýzu českých pokynů řídicích robota (A Software System for an Analysis of Robot Controlling Commands in Czech),” Dept. of CS, Faculty of Mathematics and Physics, Charles University, RNDr. Thesis, Prague, 1984, 195 pp.
- [2] Jan Hajič, “Unification Morphology Grammar,” Faculty of Mathematics and Physics, Charles University, PhD Thesis, Prague, 1994, 110 pp.
- [3] Jan Hajič, “Disambiguation of Rich Inflection (Computational Morphology of Czech),” Faculty of Mathematics and Physics, Charles University, habilitation, Prague, 2002, 334 pp.

## J Účast na řešení grantů

- [1] 1993-1995, "MATRACE (Machine Translation from Czech to English)," GAČR 405/93/2173, odp. řeš. Eva Hajičová.
- [2] 1993-1995, "Počítačový korpus češtiny I," GAČR 405/93/0218, odp. řeš. Vladimír Petkevič.
- [3] 1994-1996, "Český tagger," GAUK 39/94, odp. řeš. Jan Hajič.
- [4] 1994-1996, "Central European Genelex ," CEGLEX (EU proj. COPERNICUS 1032), (CZ) odp. řeš. Jan Hajič.
- [5] 1995-1997, "Počítačový korpus češtiny II," GAČR 405/95/0190, odp. řeš. Eva Hajičová.
- [6] 1995-1997, "Textual Corpora in Linguistics: Annotation and Comparison," RSS/HESP (OSF) grant 195/1995, odp. řeš. Jan Hajič.
- [7] 1996-1998, "Formální reprezentace jazykových struktur," GAČR 405/96/0198, odp. řeš. Eva Hajičová.
- [8] 1996-2000, "Laboratoř pro zpracování jazykových dat," MŠMT Projekt podpory výzkumu na VŠ VS96151, odp. řeš. Eva Hajičová.
- [9] 1996-2000, "Český národní korpus," MŠMT Projekt podpory výzkumu na VŠ VS96139, odp. řeš. František Čermák.
- [10] 1996-2001, "Čeština ve věku počítačů (Komplexní projekt)," GAČR 405/96/K214, odp. řeš. Eva Hajičová.
- [11] 1997-1999, "Developing Specialized Translation/Foreign Language Understanding Tools for Eastern European Languages," STEEL (EU proj. INCO-COPERNICUS PL1113), (CZ) odp. řeš. Jan Hajič.
- [12] 1998, "Workshop'98: Language Engineering for Students and Professionals Integrating Research and Education: Core NLP Technologies: Parsing of Czech," NSF IIS-9732388, odp. řeš. Frederick Jelinek, CLSP Johns Hopkins Univ., MD, USA.
- [13] 1998-1999, "Workshop'98: Post-workshop research award to B. Hladká: Tagger Combination: A Method How to Get Better Results for an Inflective Language," NSF IIS-9732388(sub), odp. řeš. Jan Hajič, řeš. Barbora Vidová Hladká.
- [14] 1998-2001, "Speech recognition of a Slavic Language (Czech)," NSF USA, spol. s MŠMT ČR program KONTAKT ME293/98, odp. řeš. Fred Jelinek, JHU, MD, USA.
- [15] 1998-2001, "Rozpoznávání mluveného slovanského jazyka (češtiny)," MŠMT KONTAKT ME293/98 (spol. s NSF, USA), odp. řeš. Jan Hajič.
- [16] 1998-2000, "Automatic Generation of Instructions in Languages of Eastern Europe," AGILE (EU proj. INCO-COPERNICUS PL961104), (CZ) odp. řeš. Eva Hajičová.
- [17] 1999-2003, "Databázové a lingvistické systémy," MŠMT Výzkumné záměry, odp. řeš. Jaroslav Pokorný.

- [18] 1999-2000, "AutoLex - Source Language Analysis for Low-density Languages," Dept. of Defense, USA, odp. řeš. D. Yarowsky, CS Dept. Johns Hopkins Univ..
- [19] 2001-2006, "MALACH," NSF USA, CO-PI of subcontract; odp. řešitel Sam Gustman, VHF USA.
- [20] 2002, "Workshop'02: Language Engineering for Students and Professionals Integrating Research and Education: Generation in MT," NSF USA, IIS, odp. řeš. Frederick Jelinek, CLSP Johns Hopkins Univ., MD, USA.
- [21] 2002-2003, "Speaker Identification," NSF USA, PI; Post-workshop'02 Award to Mgr. David Klusáček.
- [22] 2003-2005, "Velké jazykové korpusy a jejich automatická analýza," GAČR 405/03/0913, Řešitel; spolupráce s pěti dalšími partnery.
- [23] 2003-2005, "Statický strojový překlad se strukturálním transferem," MŠMT KONTAKT ME642/03 (spol. s NSF, USA), odp. řeš. Jan Hajič.
- [24] 2004-2005, "Přenos anotace mezi blízkými jazyky," MŠMT VTS Slovensko 130, odp. řeš. Jan Hajič.
- [25] 2005-2007, "Vícejazyčný valenční a predikátový slovník přirozeného jazyka," MŠMT KONTAKT 1P05ME752, odp. řeš. Jan Hajič.
- [26] 2005-2006, "Projekt Malach - spolupráce s USA," MŠMT KONTAKT 1P05ME786, odp. řeš. Josef Psutka.
- [27] 2005-2009, "International Collaboration: Meaning Representation for Language Understanding," NSF USA (OISE), #0530118, odp. řeš. Frederick Jelinek.
- [28] 2005-2009, "Od jazyka ke znalostem a sémantickému webu," GAAV Informační společnost NPV TP2 1ET201120505, odp. řeš. Jan Hajič.
- [29] 2005-2009, "Centrum kopmutační lingvistiky," MŠMT ČR Centra základního výzkumu LC536, odp. řeš. Jan Hajič.
- [30] 2006-2008, "Tektogramatický popis jazyka pro rozpoznávání mluvené řeči a strojový překlad," GAČR 405/06/0589, odp. řeš. Jan Hajič.
- [31] 2005-2010, "Metody, algoritmy a systémy informatiky," MŠMT ČR Výzkumné záměry MSM0021620838, odp. řeš. Jan Kratochvíl.

### K Popularizační články, recenze

- [1] Jan Hajič, "Kontrola překlepů (Spelling checking). In Czech.," *Computerworld (spec. příloha)*, Prague (1996), 51-52.
- [2] Jan Hajič, "Co je to tagging (What is tagging). In Czech.," *Computerworld (spec. příloha)*, Prague (1996), 51-52.
- [3] Jan Hajič, "Vyhledávání informací (Information retrieval.) In Czech.," *Computerworld (spec. příloha)*, Prague (1996), 51-52.
- [4] R. Sedgewick, "Algorithms, 2nd ed. (review)," *The Prague Bulletin of Mathematical Linguistics*, Prague (1990).