

David Mareček

assistant professor

Malostranské nám. 25, 118 00 Prague, Czechia

+420 728 772 349

✉ marecek@ufal.mff.cuni.cz

🌐 <http://ufal.mff.cuni.cz/david-marecek>



Education

- 2012 **Ph.D in Computer Science**, *Faculty of Mathematics and Physics, Charles University in Prague*,
Doctoral thesis: Unsupervised Dependency Parsing.
- 2008 **Mgr. (equiv. level of M.S.) in Computer Science**, *Faculty of Mathematics and Physics, Charles University in Prague*,
Master thesis: Automatic Alignment of Tectogrammatical Trees from Czech-English Parallel Corpus.
- 2006 **Bc. in Computer Science**, *Faculty of Mathematics and Physics, Charles University in Prague*,
Bachelor thesis: Novelizátor zákonů (in Czech).

Current position

since 2019 **Assistant Professor**, *Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, Prague, Czechia.*

Research interests

interpretation of deep neural networks, machine translation, dependency parsing, unsupervised and semi-supervised machine learning

Projects

- 2018–present **LSD**, *Linguistic Structure representation in Deep neural networks*, Czech Science Foundation grant, project investigator.
- 2015–2018 **HiML**, *Health in My Language: creating usable, reliable, fully automatic translation of public health information*, H2020, team member.
- 2012–2015 **HamleDT**, *Harmonized Multi-Language Dependency Treebank. A compilation of dependency treebanks transformed to the same annotation style*, team member.
- 2011–2016 **UDP**, *Unsupervised dependency parsing from raw texts without using annotated treebanks*, Czech Science Foundation grant, project investigator.
- 2010–2013 **FAUST**, *developing machine translation systems which respond rapidly and intelligently to user feedback*, FP7, CUNI team scientific leader.
- 2010–2012 **Depfix**, *a system for automatic post-editing of English-to-Czech machine translation outputs*, project investigator.
- 2008–2011 **Treex**, *a highly modular NLP software system primarily aimed at Machine Translation*, team member.

Study and research stays abroad

- 2019 **University of Helsinki**, *visiting researcher at Language Technology group*, Helsinki, Finland, 4 months.
- 2018 **University of Copenhagen**, *NLP research group CoAStAL*, Copenhagen, Denmark, 2 weeks.
- 2010 **DFKI**, *Deutsches Forschungszentrum fuer Kuenstliche Intelligenz*, Saarbrücken, Germany, 2 weeks.
- 2009 **ESSLLI**, *European Summer School in Logic, Language and Information*, Bordeaux, France, 2 weeks.

Teaching

2012–present **Unsupervised machine learning in NLP (formerly “Selected problems in machine learning”)**, *Charles University, Czechia*, winter terms.

Academic service

Reviewing conference papers ACL (2017, 2018, 2019), NAACL (2016, 2019), TSD (2015), ITAT (2016, 2017, 2018), BlackBoxNLP (2018, 2019)

Reviewing journal papers “Special Issue of Prague Bulletin of Mathematical Linguistics on Machine Translation” (2012, 2013, 2014, 2015, 2016), “Prague Bulletin of Mathematical Linguistics” (2016), “ACM Computing Surveys” (2013, 2014)

Supervision

defended 2 master students, 1 bachelor student

current 1 doctoral student, 2 master students

Selected publications

- 2018 David Mareček, Rudolf Rosa: Extracting Syntactic Trees from Transformer Encoder Self-Attentions. In: Proceedings of the First Workshop on Analyzing and Interpreting Neural Networks for NLP, pp. 347-349, Association of Computational Linguistics, Brussels, Belgium, ISBN 978-1-948087-71-1
- 2017 Bedřich Pišl, David Mareček: Communication with Robots using Multilayer Recurrent Networks. In: Proceedings of the First Workshop on Language Grounding for Robotics, pp. 44-48, Association for Computational Linguistics, Vancouver, Canada, ISBN 978-1-945626-64-7
- 2016 David Mareček: Delexicalized and Minimally Supervised Parsing on Universal Dependencies. In: Statistical Language and Speech Processing, pp. 30-42, Springer International Publishing, Cham, Switzerland, ISBN 978-3-319-45924-0
- 2015 David Mareček: Multilingual Unsupervised Dependency Parsing with Unsupervised POS tags. In: MICAI 2015: Advances in Artificial Intelligence and Soft Computing, Part I, pp. 72-82, Springer, Berlin / Heidelberg, ISBN 978-3-319-27059-3
- 2014 Pavel Pecina, Ondřej Dušek, Lorraine Goeriot, Jan Hajič, Jaroslava Hlaváčová, Gareth J.F. Jones, Liadh Kelly, Johannes Leveling, David Mareček, Michal Novák, Martin Popel, Rudolf Rosa, Aleš Tamchyna, Zdeňka Urešová: Adaptation of machine translation for multilingual information retrieval in medical domain. In: Artificial Intelligence in Medicine, ISSN 0933-3657, vol. 61, no. 3, pp. 165-185
- 2014 Daniel Zeman, Ondřej Dušek, David Mareček, Martin Popel, Loganathan Ramasamy, Jan Štěpánek, Zdeněk Žabokrtský, Jan Hajič: HamleDT: Harmonized Multi-Language Dependency Treebank. In: Language Resources and Evaluation, ISSN 1574-020X, vol. 48, no. 4, pp. 601-637
- 2013 David Mareček, Milan Straka: Stop-probability estimates computed on a large corpus improve Unsupervised Dependency Parsing. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 281-290, Association for Computational Linguistics, Sofia, Bulgaria, ISBN 978-1-937284-50-3
- 2012 Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, Zdeněk Žabokrtský: Coordination Structures in Dependency Treebanks. In: Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 517-527, Association for Computational Linguistics, Sofia, Bulgaria, ISBN 978-1-937284-50-3
- 2012 David Mareček, Zdeněk Žabokrtský: Exploiting Reducibility in Unsupervised Dependency Parsing. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 297-307, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-937284-43-5

Citations

Google Scholar 51 papers, 787 citations, h-index: 16

Scopus 21 papers, 191 citations, h-index: 8

Web of Science 12 papers, 43 citations, h-index: 4