

Mgr. Rudolf Rosa, Ph.D.

+420 776 577 617

rosa@ufal.mff.cuni.cz

<http://ufal.mff.cuni.cz/rudolf-rosa>

Selected research projects

- 2019–now **THEaiTRE**, *automatic generation of theatre play scripts*, PI.
- 2018–2020 **LSD**, *analyzing linguistic structures in deep neural networks*.
- 2013–now **HamleDT/Universal Dependencies**, *multilingual dependency treebanks*.
- 2015–2019 **SloNLP**, *Slovakoczech NLP Workshop*, main organizer and PC chair.
- 2014–2018 **croSSSynt**, *modelling dependency syntax across languages* [6].
- 2013–2016 **TectoMT**, *deep syntactic machine translation*.
- 2011–2015 **Depfix**, *automatic post-editing of statistical machine translation* [4].

Teaching

- 2014–now **Technology for NLP**, *NPFL092*, Bash, Git, Python, XML, JSON.
- 2017–now **NLP on computational cluster**, *NPFL118*, using our network and SGE.
- 2018–now **Multilingual NLP**, *NPFL120*, cross-lingual tagging and parsing, translation.
- 2019–now **Programming I.**, *NPRG030*, introduction to programming, Python.
- 2018–now **Supervising students**, *10× Bachelor, 1× Masters, 1× pre-PhD intern*.

Awards

- 2018 **GA UK kulatý stůl**, *best projects of the grant agency*, croSSSynt project.
- 3× **ÚFAL best paper award**, *institute's best papers*, 2013 [5], 2014 [7], 2017 [8].

Work

- 2011–now **Researcher, Teaching Assistant**, *Charles University, Czechia*.
Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
- 2015–2016 **Software Engineer (6 month internship)**, *Google, Zurich, Switzerland*.
automatic anaphora resolution for voice search
- 2004–2014 **Developer; Teacher of English**, *Jazyková škola Rosa, Kladno, Czechia*.
development of an administration system and web; teaching children and adults

Education

Charles University, Prague, Czechia.

Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics

- 2013–2018 **Doctoral Studies**, *computational linguistics*.
Discovering the structure of natural language sentences by semi-supervised methods [3]
- 2007–2013 **Bachelor and Master Studies**, *computer science, computational linguistics*.
Automatic post-editing of phrase-based machine translation outputs [2]

Programming Skills

Fluent Python, Bash, Subversion, vim, L^AT_EX
Operational C++, Java, Perl, Linux, Git, SGE

Language Skills and Qualifications

C2 level English (CPE), Czech (native speaker)
B1 level French (DELF), Slovak (passive knowledge), German
A2 level Spanish, Italian
A1 level Czech sign language

Selected Publications

Summary **40 peer-reviewed publications**, 756 citations, h-index 15 (Google Scholar).

- [1] David Mareček and Rudolf Rosa. “From Balustrades to Pierre Vinken: Looking for Syntax in Transformer Self-Attentions”. In: *Proc of BlackboxNLP*. Stroudsburg, PA, USA: ACL, 2019, pp. 263–275.
- [2] Rudolf Rosa. “Automatic Post-editing of Phrase-based Machine Translation Outputs”. Master’s Thesis. Prague, Czechia: Charles University, Fac. of Math. and Physics, 2013.
- [3] Rudolf Rosa. “Discovering the Structure of Natural Language Sentences by Semi-supervised Methods”. PhD thesis. Praha, Czechia: Charles University, Faculty of Mathematics and Physics, 2018.
- [4] Rudolf Rosa, David Mareček, and Ondřej Dušek. “DEPFI_X: A System for Automatic Correction of Czech MT Outputs”. In: *Proc of WMT*. Montréal, Canada: ACL, 2012, pp. 362–368.
- [5] Rudolf Rosa, David Mareček, and Aleš Tamchyna. “Deepfix: Statistical Post-editing of Statistical Machine Translation Using Deep Syntactic Analysis”. In: *Proc of ACL Student Research Workshop*. Sofia, Bulgaria: ACL, 2013, pp. 172–179.
- [6] Rudolf Rosa and Zdeněk Žabokrtský. “ KL_{cpos^3} – a Language Similarity Measure for Delexicalized Parser Transfer”. In: *Proc of ACL-IJCNLP, Short Papers*. Stroudsburg, PA, USA: ACL, 2015, pp. 243–249.
- [7] Rudolf Rosa et al. “HamleDT 2.0: Thirty Dependency Treebanks Stanfordized”. In: *Proc of LREC*. Ed. by Nicoletta Calzolari et al. Reykjavík, Iceland: ELRA, 2014, pp. 2334–2341.
- [8] Rudolf Rosa et al. “Slavic Forest, Norwegian Wood”. In: *Proc of VarDial4*. Ed. by Preslav Nakov et al. Stroudsburg, PA, USA: ACL, 2017, pp. 210–219.