

# Mgr. Rudolf Rosa

Svatojánská 155/4

266 01 Beroun

Česko

☎ +420 776 577 617

✉ [rosa@ufal.mff.cuni.cz](mailto:rosa@ufal.mff.cuni.cz)

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

## Vzdělání

**Univerzita Karlova, Praha, Česko.**

Matematicko-fyzikální fakulta, Ústav formální a aplikované lingvistiky

2013–nyní **doktorské studium**, *matematická lingvistika*.

Odhalování struktury vět přirozeného jazyka pomocí částečně řízených metod

2007–2013 **bakalářské a magisterské studium**, *informatika, matematická lingvistika*.

práce “Automatic Post-editing of Phrase-based Machine Translation Outputs” [1]

## Programátorské dovednosti

plynně Perl, Bash, Subversion, vim

dobře Python, C++, Java, Linux, Git, SGE, LaTeX, SQL, webové technologie

základy C#, Prolog, XML

## Výzkum

2015–nyní **SloNLP**, *Slovenskočeský NLP Workshop*, hlavní organizátor a předseda PC.

2014–nyní **croSSSynt**, *Modelování závislostní syntaxe napříč jazyky*, projekt GAUK, hlavní řešitel.

2013–nyní **Universal Dependencies**, *Anotace závislostních korpusů konzistentní napříč jazyky*, dříve projekt HamleDT [5].

2013–2016 **TectoMT**, *Kvalitní překlad pomocí metod hloubkového jazykového inženýrství*, součást projektu QTLeap financovaného EU.

2011–2015 **Deepfix**, *Automatická post-editace výstupů statistického strojového překladu* [2], součást projektů FAUST, QTLeap a HimL financovaných EU, hlavní vývojář.

2011–2012 **FilmTit**, *Webová aplikace pro počítačem podporovaný překlad filmových titulů*, studentský softwarový projekt, vedoucí týmu.

## Ocenění

**ÚFAL best paper award**, *jeden z nejlepších článků ústavu*.

2017 “Slavic Forest, Norwegian Wood” [6]

2014 “HamleDT 2.0: Thirty Dependency Treebanks Stanfordized” [5]

2013 “Deepfix: Statistical Post-editing of Statistical Machine Translation Using Deep Syntactic Analysis” [3]

---

## Práce

- 2011–nyní **výzkumník a asistent**, *Univerzita Karlova*, Česko.  
syntaktická analýza; syntaktický strojový překlad; kurz Technologie pro NLP
- 2015–2016 **softwarový inženýr (šestiměsíční stáž)**, *Google*, Curych, Švýcarsko.  
automatické rozřešení anafor v textu
- 2004–2014 **vývojář**, *Jazyková škola Rosa*, Kladno, Česko.  
vývoj a údržba firemního administračního systému; webové stránky
- 2004–2014 **učitel angličtiny**, *Jazyková škola Rosa*, Kladno, Česko.  
výuka angličtiny pro děti a dospělé

---

## Jazykové dovednosti a osvědčení

- angličtina Certificate of Proficiency in English (C2)  
němčina na úrovni B2  
francouzština Diplôme d'étude de langue française (B1)

---

## Vybrané publikace

shrnutí **30 recenzovaných publikací**, 329 citací, h-index 10 (podle Google Scholar).

- [1] Rudolf Rosa. “Automatic Post-editing of Phrase-based Machine Translation Outputs”. Diplomová práce. Praha, Česko: Univerzita Karlova, Matematicko-fyzikální fakulta, 2013.
- [2] Rudolf Rosa, David Mareček, and Ondřej Dušek. “DEPFIIX: A System for Automatic Correction of Czech MT Outputs”. In: *Proc of WMT*. Montreal, Kanada: ACL, 2012, pp. 362–368.
- [3] 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*. Sofie, Bulharsko: ACL, 2013, pp. 172–179.
- [4] 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.
- [5] 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.
- [6] 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.