

## **RNDr. Jana Straková, Ph.D.**

Researcher at Institute of Formal and Applied Linguistics

E-MAIL: [strakova@ufal.mff.cuni.cz](mailto:strakova@ufal.mff.cuni.cz)

ORCID: [0000-0003-0075-2408](https://orcid.org/0000-0003-0075-2408)

URL: <https://ufal.mff.cuni.cz/jana-strakova>

### **Research Interests:**

- deep learning for NLP,
- automatic text analysis: POS tagging, lemmatization, NER, dependency parsing, semantic parsing.

### **Projects:**

- [UDPipe](#),
- [NameTag](#): Named Entity Recognizer,
- [MorphoDiTa](#): Morphologic Dictionary and Tagger.

### **Employment History:**

- since 2009: researcher at Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University, Prague

### **Education and Academic Degrees:**

- 2009 – 2017: graduated (Ph.D.) in Computer Science from Faculty of Mathematics and Physics, Charles University, Prague; with thesis “Neural Network Based Named Entity Recognition”,
- 2007 – 2009: graduated (master's degree) in Computer Science from Faculty of Mathematics and Physics, Charles University, Prague; with thesis “Syntax based Information Retrieval”,
- 2004 – 2007: graduated (bachelor's degree) in Computer Science from Faculty of Mathematics and Physics, Charles University, Prague; with thesis “Word Alignment using Minimum Weight Edge Cover”.

### **Grants:**

Team member of 3 national and international grant projects.

### **Stays Abroad:**

- October – November 2010: internship at Institute of Cognitive Science, University of Colorado at Boulder, Boulder, USA,
- January – March 2010: internship at Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany,
- July – September 2009: internship at Google London as software engineer intern, London, UK,
- June – September 2008: internship at Google Zurich as software engineer intern, Zurich, Switzerland.

### **Publications and Citations:**

peer-reviewed publications: 24

citations: 3614

h-index: 17 (Google Scholar)

### **Selected Bibliography:**

- Jana Straková, Milan Straka, Jan Hajič (2019): [Neural Architectures for Nested NER through](#)

- [Linearization](#). In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 5326-5331, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-950737-48-2,
- Milan Straka, Jana Straková (2017): [Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe](#). In: Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, pp. 88-99, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-945626-70-8,
  - Milan Straka, Jan Hajič, Jana Straková (2016): [UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing](#). In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), pp. 4290-4297, European Language Resources Association, Paris, France, ISBN 978-2-9517408-9-1,
  - Jana Straková, Milan Straka, Jan Hajič (2014): [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, pp. 13-18, Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-00-6,
  - Eneko Agirre, Enrique Alfonseca, Keith Brendan Hall, Jana Kravalová, Marius Pasca, Aitor Soroa (2009): [A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches](#). In: Proceedings of NAACL-HLT 09, pp. 19-27, Association for Computational Linguistics, Boulder, CO, USA, ISBN 978-1-932432-41-1.

Last update: March 20<sup>th</sup>, 2024