

# Workshop on Deep Learning and Neural Approaches for Linguistic Data

Radovan Garabík and Dagmar Gromann

Deep learning and neural network approaches are indispensable in modern Natural Language Processing and generally in all kinds of linguistic data analysis approaches. Latest advances suggest that hitherto intractable problems can be solved in near future, especially as demonstrated by ongoing revolution in NLP influenced by the “unreasonable effectiveness” of Transformer architectures.

Those huge and complex language models require a lot of training data. Fortunately, at least for official EU languages, there is substantial text data collected in various repositories through many years of corpus linguistic research (and we can often resort to web corpora too). However, such resources are highly non-uniform, often difficult to find, and in various states of usability. While projects to comprehensively collect and describe language resources or even provide a unified cloud-based access to the tools exist (e.g. Common Language Resources and Technology Infrastructure – [CLARIN](#), [META-SHARE](#), [European Language Grid](#), [European Language Equality](#)), the situation is still very fragmented and access to data is difficult, especially in automated modes. The Linguistic Linked Open Data ([LLOD](#)) initiative attempts to improve the situation by designing a database-like access method based on existing standard web technologies, to provide a uniform mode of accessing the data, regardless of physical location or the natural language in question. Several of these initiatives, e.g. the European Language Grid, also include language technologies within their repository.

The COST Action CA18209 on European network for Web-centred linguistic data science ([NexusLinguarum](#)) organizes within the context of its Working Group 3 *Support for linguistic data science* a workshop focused on deep learning and neural approaches in connection with LLOD.

The goal of the workshop is connecting researchers working on all aspects of deep learning in relation with linguistic data and the effective use of deep learning in understanding the specificities of linguistic data, to be better exploited and combined with linked data mechanisms.

Researchers from all areas of NLP, corpus and computational



**Radovan Garabík** is leader of T3.2 in NexusLinguarum. He works at the Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences, and is the principal architect of the Slovak National Corpus and author of the main Slovak dictionary portal. His main domains of interests are corpus linguistics, NLP and computer lexicography, especially applied to the Slovak language.

linguistics or any other fields working with linguistic data are invited to submit a short description of their research, results, tools and applications. There is also a place to submit not only finished novel results, but also ongoing research, existing projects, including work in progress and future plans, as well as notable already published results.

The workshop is focused on, but not restricted to, these topics:

- Language models for the Multilingual Semantic Web
- Enhancement of language models with structured linguistic data
- Neural Machine Translation for LLOD interlinking
- Structured linguistic data to improve Neural Machine Translation
- Use cases combining language models and structured linguistic data

The workshop will be collocated with the next NexusLinguarum plenary meeting, to be held in Skopje, North Macedonia on 30 September 2021. It will be held in hybrid mode – both online and physical. The deadline for submitting proposals is 15 July 2021.

For further details please visit the [workshop webpage](#).



**Dagmar Gromann** is leader of Working Group 3 for deep learning and linguistic data in NexusLinguarum. She is Assistant Professor at the University of Vienna, and her work focuses on computational linguistics, with a particular interest in knowledge and information extraction for linguistic data.

<http://dagmargromann.com/>

ORCID:

0000-0003-0929-6103