



SLOVAK NATIONAL CORPUS
ĽUDOVÍT ŠTÚR INSTITUTE OF LINGUISTICS
SLOVAK ACADEMY OF SCIENCES

13th International Conference

SLOVKO 2025
Natural Language Processing
and Corpus Linguistics

15–17 October 2025
Hotel Devín, Riečna 4, 811 02 Bratislava

PROGRAMME

Wednesday, 15 October

8:30 – 9:10 **Registration & Coffee**

9:10 – 9:30 **Opening Session**

9:30 – 10:30 **Plenary Talk**

Aleksandar Trklja (Austria): Mapping Recurrent Lexico-grammatical Patterns in English Through Subtree Fragments **(in English)**

10:30 – 11:00 **Ján Mačutek, Radek Čech and Michaela Koščová** (Czech Republic): Longer Words, Easier-to-pronounce Phonemes: A Pilot Study **(in English)**

11:00 – 11:30 **Miroslav Kubát, Michaela Nogolová, Xinying Chen and Žaneta Stiborská** (Czech Republic): Syntactic Complexity and Political Ideology: A Study of Czechoslovak and Czech Presidential Speeches **(in English)**

11:30 – 12:00 **Petr Pořízka and Polina Ivankova** (Czech Republic): Quantitative Corpus Analysis of Vladimir Putin's Speeches **(in English)**

12:00 – 13:00 **Lunch Break**

13:00 – 13:30 **Josip Mihaljević** (Croatia): Challenges in Developing a Glagolitic Reverse Dictionary of Croatian Church Slavonic **(in English)**

13:30 – 14:00 **Veronika Kolářová, Václava Kettnerová and Jiří Mírovský** (Czech Republic): Advanced Syntactic Phenomena in the NomVallex Lexicon **(in English)**

14:00 – 14:30 **Jakub Absolon** (Slovakia): Redefining Terminology Work: The Role of Glossaries and TMS in Customized AI-driven Machine Translation **(in English)**

14:30 – 15:00 **František Kovařík, Vojtěch Kovář and Marek Blahuš** (Czech Republic): Do Frequency Types Matter in Lexicography? **(in Czech)**

15:00 – 15:30 **Vladimír Petkevič** (Czech Republic): Syntactic Structures In Czech Multiword Units: Types and Tokens in the Totalitarian Era and Today **(in English)**

15:30 – 16:00 *Coffee Break*

16:00 – 16:30 **Jiří Mírovský, Tereza Novotná and Barbora Hladká** (Czech Republic): MasKIT – Anonymization and Pseudonymization of Czech Legal Texts **(in English)**

16:30 – 17:00 **Miroslav Zumrík** (Slovakia): A Quantitative Analysis of Slovak and Czech Supreme Court Decisions **(in English)**

17:00 – 17:30 **Silvie Cinková, Jana Šamánková, Barbora Kubíková, Tereza Novotná and Vítěk Eichler** (Czech Republic): Annotation of Rhetorical Roles and Syllogistic Relations in Czech Argumentative Texts **(in English)**

17:30 – 18:00 **Daniel Hládek, Kristián Sopkovič, Ján Staš, Zuzana Sokolová and Matúš Pleva** (Slovakia): Financial Question-answering Dataset for Slovak Language Model Evaluation **(in English)**

Thursday, 16 October 2025

9:00 – 10:00 **Plenary Talk**

Thomas Hoffmann (Germany): Constructional Creativity – The Role of Corpus Data

(in English)

10:00 – 10:30 **Coffee Break**

10:30 – 11:00 **Markéta Ziková, Radek Čech and Pavel Kosek** (Czech Republic): The Sonority Sequencing Principle in Historical Czech: A Corpus-based Study **(in English)**

11:00 – 11:30 **Katarína Gajdošová, Ivica Hajdučková and Peter Malčovský** (Slovakia): A Corpus-based Analysis of Color Imagery in the Poetry Collection The Gift by Maša Haľamová **(in Slovak)**

11:30 – 12:00 **Václav Cvrček, Martin Stluka and Klára Pivoňková** (Czech Republic): When Data Meet Tools: Using the Monitor Corpus for the Analysis of Language Development **(in English)**

12:00 – 12:30 **Petr Pořízka and Ján Mačutek** (Czech Republic): Corpus of Dramas by the Čapek Brothers from a Quantitative Perspective **(in English)**

12:30 – 13:30 **Lunch Break**

13:30 – 14:00 **Jakub Machura, Hana Žižková, Patrik Stano, Tereza Vrabcová, Dana Hlaváčková and Ondřej Trnovec** (Czech Republic): Tailored Fine-Tuning for the Comma Insertion in Czech **(in English)**

14:00 – 14:30 **Barbora Štěpánková, Lucie Poláková and Jana Šindlerová** (Czech Republic): No doubt, but... On Connective Functions of Epistemic Markers **(in Czech)**

14:30 – 15:00 **Hana Žižková, Zuzana Nevěřilová, Jakub Machura, Aleš Horák, Dana Hlaváčková and Patrik Stano** (Czech Republic): From Rule-based Proofreader Beta Opravidlo to AI-powered Opravidlo 2.0 **(in English)**



15:00 – 15:30 **Jakub Machura, Hana Žižková and Vojtěch Kovář** (Czech Republic): General Distribution of Commas in Czech Texts: Genre Variation and Error Analysis **(in English)**

15:30 – 16:00 ***Coffee Break***

16:00 – 16:30 **Jakub Sláma** (Czech Republic): A Mountain of Evidence? A Corpus Study of the Argument Structure of Transferred Senses of Nouns in English **(in English)**

16:30 – 17:00 **Tomáš Jelínek** (Czech Republic): Annotating Mood, Tense and Voice in Czech Corpora **(in English)**

17:00 – 17:30 **Kristína Piatková and Mária Stanková** (Slovakia): Adverbs and Particles: Part-of-speech Homonymy in Corpus Data and Media Discourse **(in Slovak)**

17:30 – 18:00 **Václava Kettnerová, Jiří Mírovský and Michal Olbrich** (Czech Republic): Modeling Valency Frames Using Inheritance: The Case of Czech Adjectives -telný ‘able’ **(in English)**

19:00 ***Social Event (Restaurant AuCafe)***

Friday, 17 October 2025

9:00 – 9:30 **Zuzana Sokolová, Maroš Harahus, Daniel Hládek and Ján Staš** (Slovakia): Annotated Slovak Datasets for Toxicity, Hate Speech and Sentiment Analysis **(in Slovak)**

9:30 – 10:00 **Ján Staš, Stanislav Ondáš, Matúš Pleva, Matej Horváth, Richard Ševc and Patrik Michalanský** (Slovakia): Development of a Database and Models for Children's Speech in the Slovak Language for Speech-oriented Applications **(in Slovak)**

10:00 – 10:30 **Filip Kalaš and Pavol Lipták** (Slovakia): AI and the Translation of Idioms – Challenges, Success, and a Corpus Perspective **(in English)**

10:30 – 11:00 *Coffee Break*

11:00 – 11:30 **Iryna Karamysheva** (Ukraine): A Corpus-based Approach to the Study of Permissive Constructions in English and Ukrainian **(in English)**

11:30 – 12:00 **Anna Vysloužilová and Dominika Kovářiková** (Czech Republic): Idioms in Disguise: How Grammatical Profiling Reveals Phraseological Patterns **(in English)**

12:00 – 12:30 **Hana Skoumalová, Přemysl Vítovec and Milena Hnátková** (Czech Republic): Phrasemes and Collocations in the Corpus – How to Find Unknown Variants **(in English)**

12:30 – 13:00 **Kamila Mrázková** (Czech Republic): The Words are Falling: The Syntagmatic Combination of the Lexemes slovo ‘word’ and padnout/padat ‘to (be) fall(ing)’ as a Building Block of Other Stereotyped Expressions and Phrasemes **(in Czech)**

13:00 – 13:30 **Katarína Gajdošová and Lucia Jasinská** (Slovakia): As Hungry as a Wolf, as Silent as a Lamb: Corpus Analysis of Adjectival Similes With an Animal Component **(in Slovak)**

13:30 *Closing Session & Coffee*