



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

12th International Conference

SLOVKO 2023
Natural Language Processing and Corpus Linguistics

October 18–20, 2023
Hotel Devín, Riečna 4, 811 02 Bratislava

PROGRAMME

Wednesday, October 18

8.30 – 9.10 **Registration & Coffee**

9.10 – 9.30 **Opening Session**

09.30 – 12.00 *Session Chair: Jana Levická*

9.30 – 10.30 **Plenary Talk**

Łukasz Grabowski (University of Opole, Opole, Poland): From Solitary Corpus Analysis to Collective Insight: A Glimpse into Translation and Lexicography **(in English)**

10.30 – 11.00 **Filip Kalaš** (Slovakia): The Competition of German Adjectival Suffixes **(in English)**

11.00 – 11.30 **Renata Bronikowska** (Poland): Verbification of Feminine Forms of Adjectives *można* ‘possible’, *niemożna* ‘impossible’ and *niepodobna* ‘impossible’ – Corpus-based Approach **(in English)**

11.30 – 12.00 **Jaroslav David, Tereza Klemensová and Michal Místecký** (Czech Republic): Appellativization of Proper Names – In the Perspective of Corpus Analysis **(in English)**

12.00 – 13.00 **Lunch Break**

13.00 – 15.00 *Session Chair: Martina Ivanová*

13.00 – 13.30 **Jakob Horsch** (Germany): Corroborating Corpus Data with Elicited Introspection Data: A Case Study **(in English)**

13.30 – 14.00 **Barbora Štěpánková, Jana Šindlerová and Lucie Poláková** (Czech Republic): The Epistemic Marker *určitě* in the Light of Corpus Data **(in Czech)**

14.00 – 14.30 **Aksana Schillová** (Czech Republic): Adverbs Derived from Adjectival Present Participles in Polish, Slovak and Czech: A Comparative Corpus-based Study **(in Czech)**

14.30 – 15.00 **Marie Kopřivová and Kateřina Šichová** (Czech Republic & Germany): Proverbs in Contemporary Czech. Corpus Probe into Written Texts **(in Czech)**

15.00 – 15.30 **Coffee Break**

- 15.30 – 16.00 **Magdalena Majdak** (Poland): Keywords in Religious Literature of 17th and 18th Centuries in Light of the Data from the Electronic Corpus of 17th- and 18th-century Polish Texts
(in English)
- 16.00 – 16.30 **Marie Mikulová** (Czech Republic): Expressing Measure in Czech (A Corpus-based Study)
(in Czech)
- 16.30 – 17.00 **Edyta Jurkiewicz-Rohrbacher** (Germany): Dative Ambiguity in Russian: A Corpus Induced Study (in English)
- 17.00 – 17.30 **Alexandr Rosen** (Czech Republic): The *InterCorp* Parallel Corpus with a Uniform Annotation for All Languages (in English)
- 17.30 – 18.00 **ONLINE** **Martina Waclawičová** (Czech Republic): Differences in Spoken Language Processing in General Corpora (ORAL, ORTOFON) and in a Specialized Corpus (DIALEKT) and Their Reflection in the Mapka Application
(in Czech)

Thursday, October 19

09.00 – 10.00 *Session Chair: Ivor Uhliarik*

9.00 – 10.00 **Plenary Talk**

Daniel Kluvanec (European Commission, Luxembourg, Luxembourg): eTranslation as a EU Flagship Use Case of Natural Language Processing **(in English)**

10.00 – 10.30 **Coffee Break**

10.30 – 12.30 *Session Chair: Jana Levická*

10.30 – 11.00 **Nataliia Časnochová Zozuk** (Slovakia): Lexical Diversity and Language Impairment **(in Slovak)**

11.00 – 11.30 **Ondřej Pekáček and Irene Elmerot** (Czech Republic & Sweden): When Is a Crisis Really a Crisis? Using NLP and Corpus Linguistic Methods to Reveal Differences in Migration Discourse across Czech Media **(in English)**

11.30 – 12.00 **Martin Braxatoris and Anita Braxatorisová** (Slovakia): Use of Computer and Corpus Tools in the Research of a 19th Century German-language Manuscript Book of Notes and Extracts **(in Slovak)**

12.00 – 12.30 **Martin Diweg-Pukanec** (Slovakia): The Economy of Czech Exchange in the Slovak Marketplace of Austria after the Fall of Hungary **(in English)**

12.30 – 13.30 **Lunch Break**

13.30 – 15.30 *Session Chair: Alexandr Rosen*

13.30 – 14.00 **Dávid Držík and Kirsten Šteflovíč** (Slovakia): Text Vectorization Techniques Based on Wordnet **(in Slovak)**

14.00 – 14.30 **Daniel Hládek, Maroš Harahus, Ján Staš and Matúš Pleva** (Slovakia): Slovak Language Models for Basic Preprocessing Tasks in Python **(in English)**

14.30 – 15.00 **Lucie Benešová, Martin Stluka and Klára Pivoňková** (Czech Republic): Lemmatization of the DIA1900 Diachronic Corpus **(in Czech)**

15.00 – 15.30 **Ján Staš, Daniel Hládek and Tomáš Koctúr** (Slovakia): Slovak Question Answering Dataset Based on the Machine Translation of the SQuAD v2.0 **(in Slovak)**

15.30 – 16.00 *Coffee Break*

16.00 – 18.00 *Session Chair: Václav Cvrček*

16.00 – 16.30 **Markéta Ziková, Martin Březina, Radek Čech and Pavel Kosek** (Czech Republic): Syllabic Consonants in Historical Czech and How to Identify Them **(in English)**

16.30 – 17.00 **Jakub Machura, Hana Žižková, Adam Frémund and Jan Švec** (Czech Republic): Is It Possible to Re-educate RoBERTa? Expert-driven Machine Learning for Punctuation Correction **(in English)**

17.00 – 17.30 **Ilia Afanasev, Olga Lyashevskaya, Stefan Rebrikov, Yana Shishkina, Igor Trofimov and Natalia Vlasova** (independent researchers): The Effect of (Historical) Language Variation on the East Slavic Lects Lemmatisers Performance **(in English)**

17.30 – 18.00 **Katarína Gajdošová, Michaela Mošat'ová and Petra Švancarová** (Slovakia): Errors in the Congruent Attribute among Students Learning Slovak as a Foreign Language (Learner Corpus-based) **(in Slovak)**

19.00 *Social Event (Gala Dinner at the Restaurant Hradná)*

Friday, October 20

09.00 – 11.00 Session Chair: *Vladimír Petkevič*

- 9.00 – 9.30 **Cristina Fernández-Alcaina, Eva Fučíková, Jan Hajič and Zdeňka Urešová** (Czech Republic): Spanish Synonyms as Part of a Multilingual Event-Type Ontology **(in English)**
- 9.30 – 10.00 **Veronika Kolářová, Václava Kettnerová and Jiří Mírovský** (Czech Republic): Through Derivational Relations to Valency of Non-verbal Predicates in the NOMVALLEX Lexicon **(in English)**
- 10.00 – 10.30 **Michaela Nogolová, Michaela Hanušková, Miroslav Kubát and Radek Čech** (Czech Republic): Linear Dependency Segments in Foreign Language Acquisition: Syntactic Complexity Analysis in Czech Learners' Texts **(in English)**
- 10.30 – 11.00 **Daniel Zeman, Pavel Kosek, Martin Březina and Jiří Pergler** (Czech Republic): Morphosyntactic Annotation in Universal Dependencies for Old Czech **(in English)**

11.00 – 11.30 *Coffee Break*

11.30 – 13.00 Session Chair: *Natália Kolenčíková*

- 11.30 – 12.00 **Vladimír Petkevič and Hana Skoumalová** (Czech Republic): Annotation of Analytic Verb Forms in Czech – Complex Cases **(in English)**
- 12.00 – 12.30 **Michal Škrabal and Karel Piorecký** (Czech Republic): Corpus of Contemporary Poetry – A Gauntlet to Pick Up to Slovak Colleagues **(in Czech)**
- 12.30 – 13.00 **Václav Cvrček** (Czech Republic): Associations and Collocations in Corpora **(in English)**

13.00 – 14.00 *Lunch Break*

14.00 – 15.30 Session Chair: *Miroslav Zúmrík*

- 14.00 – 14.30 **Petr Pořízka** (Czech Republic): CapekDraCor: A New Contribution to the European Programmable Drama Corpora **(in English)**

14.30 – 15.00 **Dmitri Sitchinava** (Germany): Multiple Interpretation and Fragmented Texts within a Historical Corpus: The Case of Old East Slavic Vernacular Writing **(in English)**

15.00 – 15.30 **Richard Holaj and Petr Pořízka** (Czech Republic): ANOPHONE: An Annotation Tool for Phonemes and L2 Annotation Systems for Czech **(in English)**

15.30 – 16.00 **Adrian Jan Zasina** (Czech Republic): Can Low Proficiency Level Learners Produce Diverse Texts? A Multidimensional Approach to Czech as a Foreign Language **(in English)**

16.00 *Conference Close & Coffee*



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

WiFi of the conference room

NETWORK: Devin

PASSWORD: none – wifi of the conference room is without password

Signal Chat

For those who would like to be updated about conference last-minute changes and reminded about all important conference information live online – during the conference, Signal chat will be used. For joining the chat, please download the

Signal app at

<https://signal.org/download/>

and then

scan the following QR code:

