

Natural Language Processing, Corpus Linguistics, Lexicography

Eighth International Conference
Bratislava, Slovakia, 21–22 October 2015
Proceedings

Editors
Katarína Gajdošová
Adriána Žáková

RAM-Verlag
2015

The articles have been reviewed by members of the Program Committee.

The articles can be used under the
Creative Commons Attribution-ShareAlike 3.0 Unported Licence.



Slovak National Corpus
E. Štúr Institute of Linguistics
Slovak Academy of Sciences
Bratislava, Slovakia 2015
<http://korpus.sk/slovko.html>

© by respective authors, 2015
Editors © Katarína Gajdošová and Adriána Žáková, 2015
Typography © Vladimír Radik and Radoslav Brída, 2015
Cover © Vladimír Benko, 2015

This edition © RAM-Verlag, 2015

Table of Contents

Foreword <i>Mária Šimková</i>	7
Úvod <i>Mária Šimková</i>	8
Fran: The Next Generation Slovenian Dictionary Portal <i>Kozma Ahačič, Nina Ledinek, and Andrej Perdih</i>	9
Recent Status of Development of the ALEXIS Dictionary Writing System <i>Kamil Barbierik, Martin Bodlák, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Josef Nový, and Miroslav Virius</i>	17
Corpus of Dialects of the Slovak National Corpus <i>Katarína Gajdošová, Radovan Garabík, and Mária Šimková</i>	26
A New Dialect Corpus: DIALEKT <i>Hana Goláňová</i>	36
The Automatic Identification of Type of Adverbials in Syntactically Annotated Texts <i>Milena Hnátková</i>	45
The Use of Acronyms in Different Communication Modes (a Corpus-Based Study) <i>Zuzana Komrsková</i>	63
Evaluating Automatic Idiom Annotation in Spoken Corpora: the Case of Somatic Idioms <i>Marie Kopřivová</i>	72
SK ANTIPLAG: Five Years After <i>Július Kravjar</i>	77
New Tools for Working with the ORAL Series Corpora of Spoken Czech: AchSynku and MluvKonk <i>David Lukeš</i>	90
Traitements automatiques en lexicographie de langues « non dotées » <i>Patrice Pognan and Miloud Taiïfi</i>	102
Identifying Corpus-specific Collocations: The Case of Spoken Slovene <i>Senja Pollak and Špela Arhar Holdt</i>	117
Myslím, já myslím, myslím, že... Forms and Functions of the Verb <i>myslet</i> in Spoken Czech <i>Petra Poukarová</i>	126

Czech Deadverbial Prepositions Used in Postposition to a Noun (Exemplified in Corpus SYN 2010)	
<i>Aksana Schillova</i>	133
On Building the Slovak Automatic Semantic Role Labeling System	
<i>Ján Staš, Daniel Hládek, Stanislav Ondáš, and Jozef Juhár</i>	141
Slavic Languages in Universal Dependencies	
<i>Daniel Zeman</i>	151

Foreword

Slovko 2015 – this year's edition entitled *NLP, Corpus Linguistics, Lexicography* – represents a follow up of previous autumn meetings in Bratislava. Organisers, both from the Slovak National Corpus of the L. Štúr Institute of Linguistics, Slovak Academy of Sciences, and from the Slovak Centre of Scientific and Technical Information, are honoured to welcome participants from five countries: Austria, Czech Republic, France, Slovakia and Slovenia.

Two conference days offer 18 presentations, including two plenary talks. Not all papers registered for presentation were also published – current programme comprises also two presentations that cannot be found in the proceedings. Members of the programme committee carefully reviewed every paper sent with the registration (two reviewers for each text) and thus contributed to the overall quality of the scientific event and of this publication, for which we would like to express our sincere gratitude.

The 8th edition of the biannual conference *Slovko 2015* experiences the increase of papers dealing with corpus linguistics including lexicography. On the other hand, computationally oriented papers are in a minority. There is a significant shift from presenting new written corpora and their analyses to the issues concerning the building and research of spoken, even dialect corpora. We believe that this focus of papers will also become a source of inspiration both for conference participants and readers of the proceedings in their further work in the area of NLP, corpus linguistics and related research in Slovakia and neighbouring countries.

We wish all the participants of *Slovko 2015* an enjoyable stay in the Slovak Centre of Scientific and Technical Information and in Bratislava in particular to those who came from abroad. We would also like to invite you to *Slovko 2017* that will be focusing, besides NLP and corpus linguistics, on computational terminology and terminography.

Mária Šimková
Translated by Jana Levická

Úvod

Slovko 2015 – tentoraz s podtitulom *počítačové spracovanie prirodzeného jazyka, korpusová lingvistika, lexikografia* – nadväzuje na predchádzajúce jesenné stretnutia v Bratislave. Organizátori zo Slovenského národného korpusu Jazykovedného ústavu Ľudovíta Štúra Slovenskej akadémie vied a Centra vedecko-technických informácií Slovenskej republiky vítajú na tomto podujatí účastníkov z piatich krajín: Česká republika, Francúzsko, Rakúsko, Slovensko a Slovinsko.

Počas dvoch dní rokovania odznie celkovo 18 príspevkov, z toho dve plenárne prednášky. Nie všetky príspevky prihlásené na prezentáciu sú aj publikované – v aktuálnom programe odznejú aj tri príspevky, ktoré sa v tomto zborníku nenachádzajú. Členovia vedeckého výboru starostlivo posúdili každý prihlásený článok (jeden text dvaja posudzovatelia) a prispeli tak ku kvalite celého podujatia a tejto publikácie, za čo im patrí naše poďakovanie.

Na 8. ročníku bienálnej konferencie *Slovko 2015* sa zvýšil počet príspevkov z oblasti korpusovej lingvistiky vrátane lexikografie. Technicky orientovaných príspevkov je tentoraz menej. Výrazný posun nastal od prezentácií budovania a analýz písaných korpusov smerom k riešeniu otázok tvorby a výskumov hovorených, dokonca aj nárečových korpusov. Veríme, že aj takto zamerané príspevky budú pre poslucháčov a diskutujúcich na konferencii, ako aj pre čitateľov publikovaných štúdií prínosom a inšpiráciou do ďalších prác v oblasti počítačového spracovania prirodzeného jazyka, korpusovej lingvistiky a súvisiacich výskumov na Slovensku i v okolitých krajinách.

Všetkým účastníkom konferencie *Slovko 2015* želáme príjemný pobyt v Centre vedecko-technických informácií, mimobratislavským účastníkom aj v Bratislave vôbec. Zároveň vás už teraz pozývame na *Slovko 2017*, ktoré bude okrem počítačového spracovania prirodzeného jazyka a korpusovej lingvistiky primárne venované počítačovej terminológii a terminografii.

Mária Šimková