

Slovko (2001–2009) Five Editions of the International Conference

Slovakia cannot boast of many linguistic events that have been organised on regular basis and focusing on one specific field. This kind of symposiums is actually rather unique, in the present-day Slovak context only three of them are known nationally: onomastic conferences are held in different parts of Slovakia, Banská Bystrica hosts conferences on communication and annual Young Linguists' Symposium covers all linguistic disciplines as well as interdisciplinary areas (in 2010 its 20th edition will be held). Moreover, this seminar once saw the early presentations of some of the pioneers of Slovak computational and corpus linguistics (Emil Páleš) and also hosted the first participants of Slovko 2001 (Karol Furdík, Jozef Ivanecký). Although there has been only 5 events named Slovko, all of them of interdisciplinary nature dealing with areas of computational and corpus linguistics, the conference gained the international character and its tradition seems to be well rooted.

2001 was the year of organization of the first Slovko conference, which was held on October 26–27 (at that time still called *Computer Processing of Slovak and Czech* and attended exclusively by Czech and Slovak lecturers and audience). This symposium represented in the first place an event organised “with the aim to improve mutual awareness and knowledge of people in Slovakia involved in the issues of computers related to the language and vice versa: language related to computers” (A. Jarošová: *Malá inventúra pred hľadáním spoločného jazyka*. In: *Slovenčina a čeština v počítačovom spracovaní*. Bratislava: Veda 2001, p. 7)¹. Quoting the author and the main organiser on behalf of the Ludovít Štúr Institute of Linguistics, Slovak Academy of Sciences (the second organiser being Vladmír Benko on behalf of the Faculty of Pedagogy, Comenius University), contemporary Slovakia featured only fairly isolated islets of respective activities in different scientific disciplines and theoretically-applied contexts, out of which also efforts in the area of artificial intelligence and cognitive linguistics revealed to be relevant. However, neither results of the automatised processing of Slovak language data nor relevant language data in terms of electronic corpus of texts in Slovak language were available at that time. In comparison with the Czech Republic it was a diametrically different situation since computational linguistics there had been intentionally developed as an autonomous field for more than 30 years and the 100 million token National Corpus of the Czech language has been made available on the internet in 2000. Czech lecturers could therefore “offer the Slovak professional public as well as students of linguistic and non-linguistic disciplines rather comprehensive overview of the work results in the area of Czech computer processing; this language belongs in this respect among the European and in terms of different parameters also among world's elite” (ibid). The seminar met also another of intended aims: “carry on with the scientific, pedagogic and organisational work of Ján Horecký who had been doing his best since the beginning of the 1960s to apply the

¹ Available from <http://korpus.juls.savba.sk/structure/dicts.sk.html>

principles and methods of mathematical linguistics on the material of the Slovak language” (ibid) and “who witnessed the revival of the computational linguistics in the Institute of Linguistics when he initiated in 1988–1989 a project of data base of the Slovak language, within which an idea arose to start to build a corpus” (ibid, p. 9). 14 papers altogether were presented, 5 of which by Czech authors, and consequently published in the proceedings. These proceedings were – at the same time – partial result of the L. Štúr Institute of Linguistics and Faculty of Pedagogy involvement in the multinational project Trans-European Language Resources Infrastructure II – TELRI II, carried out as a coordinated action within the European Commission programme INCO-COPERNICUS in 1999–2001.

The question whether the need of mutual informing of people involved in the field of computational linguistics in Slovakia and the exchange of experience with foreign specialists was only a one-time, temporal issue or whether it had a broader context, was answered in 2003 when Vladimír Benko (as a chief organiser on behalf of the L. Štúr Institute of Linguistics and Faculty of Pedagogy) held an international scientific seminar on 24–25 October, this time already called *Slovko*, subtitled as *Slavic Languages in Computer Processing*. The event was attended by 54 people interested in this area, 20 out of which were from Slovakia and the rest from abroad. This edition of *Slovko* saw the first presentation of the Slovak National Corpus as an actually existing electronic database of written texts covering the period 1955–2003 (in August 2003 it was made available for the public with its 26 million of tokens, in December 2003 the second version of the Corpus was released, containing 166 million of tokens) as well as the team of researchers of new department of the L. Štúr Institute of Linguistics, created in mid-2002. Fledgeling corpus linguists lead by Mária Šimková presented first of all the whole of Corpus Project and then partial research solutions in segmentation, lemmatisation and morphological annotation of texts of the Slovak language. The first Slovak morphological tagset presented at *Slovko* was shortly afterwards publicly examined and reviewed by Slovak and foreign specialists and at the beginning of 2004 started a manual annotation of selected texts of the Slovak National Corpus.

Due to the fact that no Slovak university or college offered either an autonomous course of study or a specific seminar of computational and corpus linguistics, the Slovak National Corpus department as a chief investigator of the *State Research Programme: Integrated Computational Processing of the Slovak Language for Linguistic Research Purposes* organised regular lectures and seminars focused on these two disciplines. A part of it was published in the proceedings entitled *Insight into the Slovak and Czech Corpus Linguistics*. Ed. M. Šimková. Bratislava: Veda 2006. 208 p.² As a natural conclusion of these and also of some other activities was the organisation of *Slovko* on 10–12 November in 2005 and receiving the Slovak Academy of Sciences Prize for Building Infrastructure in Science on 11 November 2005. The third international seminar *Slovko* had already a programme and organizing committee and thanks to aroused interest it was also extended to computer treatment of Slavic and East European languages. It was attended by more than 60 interested researchers, which has represented the highest number of participants of one edition of *Slovko*. Altogether 29 papers were presented in more-or-less homo-

² Available from <http://korpus.juls.savba.sk/publications/>

genous sessions: spoken corpora, speech analysis and synthesis, computer lexicography, parallel and historical corpora, terminology, e-learning, which were later on published in the proceedings *Computer Treatment of Slavic and East European Languages* (Ed. R. Garabík. Bratislava: Veda 2005. 246 p.).

Slovko in 2007 (explicitly held on October 25–27) was again organised by the Slovak National Corpus team (namely Radovan Garabík, Jana Levická, and Mária Šimková), this time as a solid fourth edition of a biannual international conference focused on NLP and computational lexicography and terminology. More than 50 participants had an opportunity to attend 37 lectures and presentations; the proceedings entitled *Computer Treatment of Slavic and East European Languages 2007* (Ed. J. Levická, R. Garabík. Bratislava: Tribun 2007. 318 p.) were available upon arrival at the conference. Beside the topics that have been fundamental for every Slovko (corpus development including spoken corpora: data collection, annotation and processing), the foreground was reserved for theoretical issues of computational lexicography and terminography, bilingual lexicography and terminography, co-occurrence analysis and pertinent collocations of lexicographical and terminographical relevance, new methods in data extraction and terminology mining from corpora, terminology databases and terminology management systems and so forth. Contributions of Slovak authors covering basic as well as applied research pointed to revolutionary changes that Slovak computational and corpus linguistics in particular had seen in the recent period. For the sake of the example, we can mention several results of the Slovak National Corpus team work: on January 1st 2007 began the second phase of the realisation of the State Research and Development Programme: *Building the Slovak National Corpus and Integrated Computational Processing of the Slovak Language for the Linguistic Research Purposes*, the 350 million token fully lemmatised and morphologically annotated general corpus was made available for public. Beside other products (database of digitised lexicographic and other linguistic works, parallel corpora), highest priority has been assigned to a partial project of the Slovak National Corpus, the Slovak terminology database, which was presented by its chief investigator Jana Levická, as well as the Slovak Spoken Corpus whose technical features were shown by (in cooperation with Milan Rusko) R. Garabík.

We decided to hold Slovko 2009 in the Congress Centre Smolenice of the Slovak Academy of Sciences to celebrate sort of a jubilee edition since it is the fifth in a row (all the previous editions took place in Bratislava but in different venues). Originally, the conference was to have been focused on parallel corpora, however, since the Institute of the Czech National Corpus, Charles University, Prague, organised a conference with the same topic on 17–19 September 2009 within their InterCorp Project, we sought to come up with a different key issue that could be of immediate interest in the Central-European context. Since the Institute of the Czech Language, Academy of Sciences of the Czech Republic ceased to organise regular conferences entitled *Grammar & Corpora* (hosted by František Štícha), the main conference topic was in fact right at hand: corpus linguistics, more precisely corpus based linguistic research. In this respect this year's Slovko has once again gathered more Slovak and Czech participants, which can be viewed also as a reminder of the beginnings of Slovko. Presented contributions point to other topics and, at the same time,

give evidence of the progress in this field: majority of them do not deal with a development of standard corpora of written or spoken texts but rather with the tools of text analysis, building, and usage of language resources, linguistic components of information systems, computer-aided translation systems, localisation and lexicography, computer-aided language learning etc. Similarly as the organisers of the first Slovko participated in international projects, the Slovak National Corpus team and Slovak computational and corpus linguistics have been involved in several international cooperation projects, especially in the 7th framework programme (EU FP7 INF 211983 Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and Their Digital Resources).

Five editions of Slovko reflect the corpus and computational linguistics in Slovakia, Czech Republic, as well as in neighbouring and more distant countries, namely Austria, Belgium, Bulgaria, France, Croatia, Hungary, Germany, Norway, Netherlands, New Zealand, Poland, Russia, Slovenia, Spain, Ukraine have had their participants so far. Approximately 130 participants altogether from 17 countries have attended Slovko conference (however, some of them have participated repeatedly, on regular basis). Overall, Slovko conferences hosted almost 300 participants that could listen to and/or read nearly 150 papers. We would like to express our gratitude for excellent cooperation to all members of programme committees and abstract peer reviewers. We also thank to all participants for creating fruitful working and social atmosphere, to Czech colleagues in particular for they have been showing us support from the beginning and attended regularly all Slovko's editions. We appreciate very much the work of all members of each organising committee and hope that the flag will be handed on.

*Mária Šimková
Jana Levická*