

Slovko (2001 – 2009)

Päť ročníkov medzinárodného podujatia

V rámci lingvistických podujatí na Slovensku nemáme veľa takých, ktoré by sa uskutočňovali pravidelne a boli zamerané na jednu oblasť. Takéto podujatia sú naozaj skôr výnimočné, v súčasnom slovenskom kontexte sú širšie známe asi tri: na rôznych miestach Slovenska sa konávajú onomastické konferencie, v Banskej Bystrici konferencie o komunikácii a celú lingvistiku, ale aj interdisciplinárne oblasti zahŕňa každoročné Kolokvium mladých jazykovedcov (v r. 2010 sa už chystá 20. ročník), na ktorom sa ako mladí kedysi prezentovali aj jedni z priekopníkov slovenskej počítačovej a korpusovej lingvistiky (E. Páleš) a prví účastníci Slovka 2001 (K. Furdík, J. Ivanecký). Podujatí s názvom Slovko, ktoré sú zamerané interdisciplinárne na oblasti počítačovej a korpusovej lingvistiky, bolo zatiaľ len päť, no ukazuje sa, že sa založila dobrá tradícia s medzinárodným presahom.

Keď sa v r. 2001 pripravovalo a 26. – 27. októbra aj uskutočnilo prvé Slovko (ešte pod názvom Slovenčina a čeština v počítačovom spracovaní a za účasti výlučne českých a slovenských prednášateľov i poslucháčov), išlo predovšetkým o podujatie organizované „so zámerom zlepšiť vzájomnú informovanosť ľudí zaoberajúcich sa na Slovensku problematikou počítačov vo vzťahu k jazyku a naopak, jazyka vo vzťahu k počítačom“ (A. Jarošová: Malá inventúra pred hľadaním spoločného jazyka. In: Slovenčina a čeština v počítačovom spracovaní. Bratislava: Veda 2001, s. 7)¹. Ako konštatovala autorka a hlavná organizátorka seminára za Jazykovedný ústav Ľ. Štúra SAV (druhým organizátorom bol V. Benko za Pedagogickú fakultu Univerzity Komenského), v tejto oblasti boli v tom čase na Slovensku značne izolované ostrovcy aktivít prebiehajúcich v rôznych vedných odboroch a teoreticko-aplikačných kontextoch, z ktorých sa ako relevantné ukazovali aj aktivity v oblasti umelej inteligencie a kognitívnej lingvistiky. Neboli však k dispozícii výsledky automatizovaného spracovania slovenských jazykových dát ani relevantné jazykové dáta v podobe elektronického korpusu textov slovenského jazyka. Išlo o diametrálne odlišnú situáciu v porovnaní s Českou republikou, kde sa počítačová lingvistika cieľavedome budovala ako samostatný odbor viac ako tridsať rokov a národný korpus českého jazyka bol v r. 2000 zverejnený na internete v rozsahu 100 miliónov tokenov. Českí prednášatelia preto mohli „poskytnúť slovenskej odbornej verejnosti a študentom z lingvistických aj nelingvistických odborov ucelenejší pohľad na výsledky práce v oblasti počítačového spracovania češtiny, ktorá patrí v tomto smere medzi európsku a vo viacerých parametroch aj medzi svetovú špičku“ (tamže). Seminár splnil aj ďalší zo svojich cieľov: „nadviazať na vedeckú, pedagogickú a organizačnú prácu Jána Horeckého, ktorý sa od začiatku 60. rokov 20. storočia usiloval uplatňovať princípy a metódy matematickej lingvistiky na materiáli slovenského jazyka“ (tamže) a „ktorý stál aj pri revitalizácii počítačovej lingvistiky v Jazykovednom ústave, keď v rokoch

¹ Dostupný aj z <http://korpus.juls.savba.sk/structure/dicts.sk.html>

1988 – 1989 pripravil projekt bázy dát slovenského jazyka, v rámci ktorej sa začalo uvažovať aj o budovaní korpusu“ (tamže, s. 9). Na seminári odznelo a v zborníku z neho bolo publikovaných 14 príspevkov, z toho 5 od českých autorov. Publikovaný zborník zároveň predstavoval jeden z výsledkov účasti JÚLŠ SAV a PdF UK v mnohonárodnom projekte Trans-European Language Resources Infrastructure II – TELRI II, ktorý sa ako súbor coordinated action uskutočnil v rámci programu Európskej komisie INCO-COPERNICUS v r. 1999 – 2001.

Otázka, či potreba vzájomnej informovanosti pracovníkov v oblasti počítačovej lingvistiky na Slovensku a výmena skúseností s odborníkmi zo zahraničia bola iba jednorazová, dočasná, alebo má širší kontext, bola zodpovedaná v r. 2003, keď V. Benko (ako hlavný organizátor za JÚLŠ SAV i PdF UK) uskutočnil 24. – 25. októbra medzinárodný odborný seminár už pod názvom Slovko, v podtitule špecifikovaný na Slovánske jazyky v počítačovom spracovaní. Podujatia sa zúčastnilo 54 záujemcov o túto oblasť, z toho 20 zo Slovenska a 34 zo zahraničia. Na tomto Slovku sa prvýkrát prezentoval Slovenský národný korpus – ako už reálne existujúca elektronická databáza písaných textov z r. 1955 – 2003 (od augusta 2003 verejne prístupná v rozsahu 26 miliónov tokenov, v decembri 2003 bola zverejnená druhá verzia v rozsahu 166 miliónov tokenov) i ako kolektív pracovníkov nového oddelenia JÚLŠ SAV, založeného v podstate v polovici r. 2002. Začínajúci korpusoví lingvisti pod vedením M. Šimkovej prezentovali jednak celý korpusový projekt, jednak čiastkové riešenia pri segmentácii, lematizácii a morfolologickej anotácii textov slovenského jazyka. Prvá podoba slovenského morfologického tagsetu predložená na Slovku bola zakrátko osobitne verejne oponovaná domácimi aj zahraničnými odborníkmi a začiatkom r. 2004 sa rozbehla ručná morfologická anotácia vybraných textov SNK.

Vzhľadom na to, že na žiadnej vysokej škole na Slovensku neexistoval samostatný odbor a ani sa nevyučoval samostatný predmet počítačovej a korpusovej lingvistiky, oddelenie SNK JÚLŠ SAV ako riešiteľ štátnej úlohy Integrated Computational Processing of the Slovak Language for Linguistic Research Purposes organizovalo pravidelné prednášky a semináre z tejto oblasti. Časť z nich bola publikovaná v zborníku *Insight into the Slovak and Czech Corpus Linguistics*. Ed. M. Šimková. Bratislava: Veda 2006. 208 s.² Prirodzeným vyústením týchto a ďalších aktivít bolo zorganizovanie Slovka 10. – 12. novembra 2005 a prevzatie Ceny SAV za budovanie infraštruktúry pre vedu 11. novembra 2005. Tretí medzinárodný seminár Slovko už bol obohatený o vedecký výbor a organizačný výbor a vzhľadom na prejavovaný záujem bol rozšírený na Slovánske a východoeurópske jazyky v počítačovom spracovaní. Zúčastnilo sa ho vyše 60 záujemcov, čo bol a stále je najvyšší počet účastníkov jedného Slovka. Celkovo odznelo 29 príspevkov vo viac-menej homogénnych blokoch: hovorené korpusy, analýza a syntéza reči, počítačová lexikografia, paralelné korpusy, historické korpusy, terminológia, e-learning a pod. Tieto vyšli v zborníku *Computer Treatment of Slavic and East European Languages* (Ed. R. Garabík. Bratislava: Veda 2005. 246 s.).

² Dostupný aj z <http://korpus.juls.savba.sk/publications/>

Slovko v r. 2007 (25. – 27. októbra) organizoval kolektív SNK (najmä R. Garabík, J. Levická, M. Šimková) už ako stabilizovanú 4. bienálnu medzinárodnú konferenciu so zameraním na NLP a počítačovú lexicografiu a terminológiu. Vyše 50 účastníkov si malo možnosť vypočuť 37 príspevkov a prezentácií, zborník *Computer Treatment of Slavic and East European Languages* (Ed. J. Levická, R. Garabík. Bratislava: Tribun 2007. 318 s.) mali všetci k dispozícii priamo na podujatí. Okrem tém, ktoré sú ako základné súčasťou každého Slovka (tvorba korpusov vrátane hovorených: zber dát, anotácia a spracovanie), sa do popredia dostali teoretické otázky komputačnej lexicografie a terminografie, bilingválna lexicografia a terminografia, kookurenčná analýza a lexicograficky alebo terminograficky relevantné kolokácie, nové metódy v extrahovaní dát a získavanie terminológie z korpusov, terminologické databázy a systémy terminologického manažmentu a pod. Príspevky slovenských autorov, pokrývajúce základný aj aplikovaný výskum, poukázali na prevratné zmeny, ktoré slovenská počítačová a najmä korpusová lingvistika zaznamenala za posledné obdobie. Napr. z produkcie SNK, ktoré 1. januára 2007 začalo 2. etapu riešenia štátnej úlohy *Budovanie Slovenského národného korpusu a elektronizácia jazykovedného výskumu na Slovensku*, bol verejnosti k dispozícii hlavný korpus v rozsahu 350 miliónov tokenov s plnou lematizáciou a morfológickou anotáciou na báze vlastného tagsetu. Okrem ďalších produktov (databáza elektronických lexicografických a iných lingvistických diel, paralelné korpusy) sa prioritným čiastkovým projektom SNK stala Slovenská terminologická databáza, ktorú na konferencii predstavila jej hlavná riešiteľka J. Levická, a budovať sa začal aj Slovenský hovorený korpus, ktorého technické parametre prezentoval (v spolupráci s M. Ruskom) R. Garabík.

Slovko 2009 sme ako polojubilejné, piate v poradí umiestnili do Kongresového centra SAV v Smoleniciach (všetky predchádzajúce boli v Bratislave, a to na viacerých miestach). Hlavným zameraním mali byť pôvodne paralelné korpusy, ale keďže Ústav Českého národného korpusu FF UK v Prahe organizoval takto profilovanú konferenciu v súlade s cieľmi svojho projektu InterCorp 17. – 19. septembra 2009, hľadali sme inú hlavnú tému, ktorá by mohla byť aktuálnou v stredo európskom priestore. A keďže Ústav pro jazyk český AV ČR prestal organizovať pravidelné konferencie Grammar & Corpora (pod vedením F. Štíchu), priam sa nám ponúkla ako hlavná téma korpusová lingvistika, resp. korpusovo zamerané lingvistické výskumy. V tejto súvislosti je na tohtoročnom Slovku opäť viac slovenských a českých účastníkov, čo je aj istá pripomienka východiskovej situácie Slovka 2001. Posun v tejto oblasti nielen v našich krajinách naznačujú aj ďalšie prezentované témy: už to nie sú otázky budovania štandardného korpusu písaných či hovorených textov, ale viac nástroje textovej analýzy, tvorba a využitie jazykových zdrojov, lingvistické zložky informačných systémov, preklad s počítačovou podporou, lokalizácia a lexicografia, didaktika vyučovania cudzích jazykov s počítačovou podporou a pod. A tak ako organizátori prvého Slovka boli členmi medzinárodných projektov, aj Slovenský národný korpus a slovenská počítačová a korpusová lingvistika sa v súčasnosti prezentuje vo viacerých medzinárodných spoluprákach, predovšetkým v 7. RP (EU FP7 INF 211983 s názvom *Conceptual Modelling of Networking of Centres for High-Quality Research in Slavic Lexicography and Their Digital Resources*).

Päť ročníkov Slovka je odrazom vývinu počítačovej a korpusovej lingvistiky na Slovensku, v Čechách, v okolitých i vzdialenejších krajinách, z ktorých sú doteraz zastúpené Belgicko, Bulharsko, Francúzsko, Gruzínsko, Holandsko, Chorvátsko, Maďarsko, Nemecko, Nórsko, Nový Zéland, Poľsko, Rakúsko, Rusko, Slovinsko, Španielsko, Ukrajina. Zahraničných účastníkov bolo na všetkých konferenciách Slovko spolu asi 130 (nie sú to unikátne výskyty, viacerí sú našimi pravidelnými hosťami) zo 17 krajín. Spolu so slovenskými účastníkmi sme privítali na týchto podujatiach takmer 300 záujemcov, ktorí si mali možnosť vypočuť a/alebo prečítať takmer 150 príspevkov. Za vynikajúcu spoluprácu ďakujeme všetkým členom vedeckého výboru a posudzovateľom abstraktov. Ďakujeme všetkým účastníkom za vytváranie tvorivej pracovnej i spoločenskej atmosféry, osobitne českým kolegom, ktorí nás od začiatku podporovali a pravidelne sa zúčastňovali aj Slovka. Ďakujeme všetkým členom doterajších organizačných výborov a dúfame, že štafeta pôjde ďalej.

*Mária Šimková
Jana Levická*