

SLOVKO 2013
POČÍTAČOVÉ SPRACOVANIE PRIRODZENÉHO JAZYKA,
KORPUSOVÁ LINGVISTIKA, E-LEARNING

V dňoch 13. – 15. novembra 2013 sa v Bratislave uskutočnil 7. ročník medzinárodnej bienálnej konferencie SLOVKO 2013 s podtitulom *Počítačové spracovanie prirodzeného jazyka, korpusová lingvistika, e-learning*, ktorú zorganizovalo oddelenie Slovenského národného korpusu Jazykovedného ústavu Ľ. Štúra Slovenskej akadémie vied v Bratislave (ďalej SNK JÚLEŠ SAV). Konferenciu otvorili slávnostným príhovorom Zuzana Komárová, generálna riaditeľka Sekcie umenia a štátneho jazyka Ministerstva kultúry SR, a Nicol Janočková, zástupkyňa riaditeľa Jazykovedného ústavu Ľ. Štúra SAV v Bratislave.

Na konferencii odznelo 27 príspevkov (z toho 3 plenárne prednášky) od viac ako 50 vedcov zo Slovenska, Česka, Maďarska, Slovinska, Nemecka, Poľska a Grécka. Väčšina prezentácií bola prednesená v angličtine, niektoré odzneli v češtine a slovenčine. Po každom vystúpení bol vytvorený dostatočný priestor na diskusiu, ktorý využili účastníci konferencie na podnetné výmeny názorov a skúseností. Zborník tlačených príspevkov *Natural Language Processing, Corpus Linguistics, E-learning* (Eds. K. Gajdošová – A. Žáková. Lüdenschied: RAM-Verlag 2013. 303 s. ISBN 978-3-942303-18-7), ktorý mali účastníci k dispozícii už v priebehu konferencie, obsahuje 26 príspevkov v anglickom jazyku.¹ Kvalitu uverejnených príspevkov zaručuje ich posúdenie a schválenie členmi vedeckého výboru.

Príspevky sa zameriavajú na prehľad o súčasnom stave počítačovej lingvistiky, na možnosti zdokonaľovania a vývoja korpusových zdrojov a nástrojov, na gramatický výskum založený na korpusových dátach a využitie korpusov a databáz na tvorbu moderných vzdelávacích portálov a vyučovanie cudzích jazykov, ako aj na interdisciplinárne témy. V správe zachytávame všetky príspevky zo zborníka, ktoré boli zároveň prezentované v rámci konferencie SLOVKO 2013, ako aj 2 prednášky pracovníkov SNK JÚLEŠ SAV (M. Šimková, M. Debnár), ktoré odzneli v rámci programu vedeckého podujatia, pričom ich radíme podľa tematických okruhov.

Plenárne prednášky predniesli odborníci na súčasnú korpusovú lingvistiku a jazykové technológie Tamás Váradi, Tomáš Erjavec a expertka na online vyučovanie Katarína Pišútová. O výsledkoch projektu CESAR (Central and South Slavic Resources) informoval jeho hlavný koordinátor Tamás Váradi (Jazykovedný ústav Maďarskej akadémie vied v Budapešti) v štúdiu *Veni, Vidi, Vici: The Language Technology Infrastructure Landscape after CESAR*. Tento významný projekt financovaný Európskou úniou bol zameraný na zhromaždenie základných jazykových zdrojov pre šesť slovanských jazykov strednej a juhovýchodnej Európy. V nadväznosti na širokú škálu existujúcich zdrojov a predchádzajúcich národných aj medzinárodných aktivít sa vďaka realizácii projektu naplnila a sprevádzkovala komplexná platforma jazykových zdrojov umožňujúca a podporujúca rozsiahle multilingválne a intralingválne produkty a služby. Všetci partneri projektu vyvinuli značné úsilie o modernizáciu a skvalitnenie ponúkaných zdrojov v príslušných šiestich jazykoch. T. Váradi osobitne ocenil prácu SNK JÚLEŠ SAV, ktorý poskytol korpus slovenského jazyka, ako aj mnoho ďalších kvalitne anotovaných korpusov a nástrojov, ktoré fungujú v súlade s platnou legislatívou.

¹ V správe uvádzame prepis mien autorov príspevkov z iných ortografických systémov do latinky tak, ako ich prepísali autori a ako sú tieto mená uvedené v zborníku z konferencie.

Kvalifikované informácie o korpusových zdrojoch, technickej podpore a výskume v oblasti počítačovej lingvistiky v Slovinsku podal Tomaž Erjavec (Inštitút Jožefa Stefana v Ľubl'ane) v príspevku *Slovene Corpora for Corpus Linguistics and Language Technologies*. Referenčné slovníky slovinčiny, špecializované a paralelné korpusy, ako aj korpusy cudzích jazykov sú dostupné zo serveru Inštitútu Jožefa Stefana v Ľubl'ane na adrese nl.ijs.si. Niektoré z korpusov, predovšetkým ručne anotované, sú prístupné aj na stiahnutie pod licenciou Creative Commons. Autor sa špeciálne venoval otázke dostupnosti korpusov, technickým aspektom kódovania, zachovaniu konzistentnosti a udržateľnosti korpusov. Pozornosť tiež upriamil na právne hľadisko a na ochranu autorských a osobných údajov. Odborné pracovisko plánuje naďalej presadzovať využívanie otvorených štandardov a odporúčaní, najmä usmernenia Text Encoding Initiative Guidelines, využitie open-source softvérových riešení, ako sú TreeTagger, CUWI a NoSketch Engine, a bezplatnú dostupnosť vytvorených korpusov.

Mária Šimková (Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied v Bratislave) v prezentácii *Cudzojazyčné texty v Slovenskom národnom korpuse* načrtla výhody a nevýhody prítomnosti cudzojazyčných textov v Slovenskom národnom korpuse (ďalej SNK), v Slovenskom hovorenom korpuse a v ďalších korpusových zdrojoch a databázach. Poukázala na to, že v paralelných korpusoch je originál nevyhnutne prítomný, ale prítomnosť cudzojazyčných textov v jednojazyčnom korpuse je nežiaduca. V slovenskej publicistike, hlavne po r. 1993, sú v periodikách často prítomné celé súvislé články v češtine, ale aj v angličtine a iných jazykoch. V odbornej literatúre sú problémom inojazyčné útvary, akými sú napr. resumé, abstrakty, citáty a bibliografia. Osobitnú kategóriu predstavuje umelecká literatúra s prekladovou zložkou a výpožičkami. Autorka svoj prehľad uzavrela konštatovaním, že existencii cudzojazyčných textov v korpuse sa nedá vyhnúť a ich selekcia je nevyhnutná, avšak je potrebné, aby táto selekcia prebiehala v rozumnej miere.

Radovan Garabík a Indrė Pileckytė (Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied v Bratislave; *From Multilingual Dictionary to Lithuanian WordNet*) opísali pilotnú verziu litovského WordNetu, ktorú vytvorili z dvojazyčného slovensko-litovského slovníka. Litovský WordNet je prepojený s Princeton WordNetom a so slovenským WordNetom; v súčasnosti obsahuje 7 874 substantívnych, 2 099 adjektívnych a 682 adverbiálnych synsetov (množiny slov s rovnakým významom), ktorým prislúchajú slovenské a anglické ekvivalenty, pričom všetky synsety prešli manuálnou opravou. V najbližšom čase plánujú tvorcovia litovského WordNetu skorigovať slovesné synsety, ktorých je zatiaľ 533, a kvantitatívne rozšíriť existujúcu databázu.

Vývoj korpusových nástrojov, ktoré zohrávajú významnú úlohu pri skvalitňovaní korpusov a zefektívňujú lexikografickú prácu, ako aj aplikácia jazykových technológií v praxi boli predmetom záujmu autorov aj autorských kolektívov z viacerých inštitúcií.

Nové vylepšené vyhľadávacie rozhranie NoSketch Engine (ďalej NoSkE) vhodné pre rôzne druhy korpusov predstavili jeho tvorcovia Tomáš Machálek a Michal Křen (Filozofická fakulta Karlovej univerzity v Prahe; *Query Interface for Diverse Corpus Types*). Medzi nové najdôležitejšie funkcie softvéru NoSkE patria hierarchické usporiadanie dostupných korpusov vrátane zobrazenia charakteristiky každého z nich, ponuka interaktívneho výberu morfológických kategórií pre CQL, podpora pre hovorené korpusy (možnosť prehrávať zvukové segmenty). Softvér NoSkE má v Českom národnom korpuse (ďalej ČNK) po ďalších vylepšeniach funkčnosťou prekonať a nahradiť vyhľadávacie rozhrania Park a Bonito. Na tomto mieste

treba pripomenúť, že nový vyhľadávací systém vďaka spolupráci s ČNK môžu využívať na vyhľadávanie aj používatelia SNK. Oddelenie SNK JÚLŠ SAV sa prostredníctvom testovania, návrhov a pripomienok spolupodieľa na zdokonaľovaní nového rozhrania NoSkE.

Vyhľadávaním a odstraňovaním duplikácií v dátach SNK sa zaoberal Vladimír Benko (Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied v Bratislave; *Data Deduplication in Slovak Corpora*). Detekcia presných duplikácií je podľa autora jednoduchá, náročnejšie je vyhľadávanie blízkych duplikácií a určenie primeranej miery ich eliminácie. Na detekovanie blízkych duplikácií použil kombináciu dvoch nástrojov: jednoduchú metódu fingerprint a nástroj Onion („One Instance Only“), ktorý vytvoril Jan Pomikálek z Masarykovej univerzity v Brne. V. Benko plánuje v budúcnosti skvalitniť použité metódy a aplikovať svoj výskum na všetky podkorpora Slovenského národného korpusu, ako aj na novovytvorený webový korpus SNK.

Kamil Barbierik, Martina Holcová Habrová, Pavla Kochová, Tomáš Liška, Zdeňka Opavská a Miroslav Virius (Ústav pro jazyk český Akadémie vied Českej republiky v Prahe) predstavili vo svojom príspevku *A New Path to a Modern Monolingual Dictionary of Contemporary Czech: the Structure of Data in the New Dictionary Writing System* nový špeciálny software Dictionary Writing System, ktorý sa využíva pri tvorbe výkladového slovníka súčasnej češtiny. Heslá slovníka je vygenerované na základe 3 vyvážených korpusov SYN2000, SYN2005, SYN2010. Vývoj softvéru a lexikografickú prácu podporuje grant z Ministerstva kultúry ČR v rámci programu aplikovaného výskumu a vývoja Národnej a kultúrnej identity.

Michal Laclavík, Štefan Dlugolinský (Ústav informatiky Slovenskej akadémie vied v Bratislave) a Michal Blanárík (Fakulta informatiky a informačných technológií Slovenskej technickej univerzity v Bratislave) diskutovali v článku *Experimenting with Slovak Wikipedia as a Source for Language Technologies* o možnostiach využitia slovenskej Wikipédie na spracovanie prirodzeného jazyka. Na jej dátach overili niektoré metódy a algoritmy na rozpoznávanie názvoslovných entít. Ich cieľom bolo poukázať na to, že slovenská Wikipédia má široké využitie v štatistickom spracovaní prirodzeného jazyka a môže slúžiť ako relevantný a dostupný zdroj textových dát.

V príspevku *Event Extractor: Email Events Detection and Calendar Integration* prezentovali Filip Ogurčák (Fakulta informatiky a informačných technológií Slovenskej technickej univerzity v Bratislave) a Michal Laclavík (Ústav informatiky Slovenskej akadémie vied v Bratislave) svoj prístup k automatickému detekovaniu udalostí (dátum, čas, miesto, názov) z malého korpusu slovenských emailových správ a ich následné editovanie a uloženie do kalendára pomocou extrakčnej metódy. Opísali plugin pre Thunderbird, ktorý predstavuje klient-sku časť aplikácie a grafické rozhranie na pridávanie udalostí do Google kalendára, ako aj serverovú časť aplikácie, ktorá zabezpečuje prepojenie detekovanej udalosti s pluginom pre Thunderbird a jednotlivými extrakčnými metódami.

Korpusové nástroje, programy a aplikácie, ktoré sú témou aj nasledujúcich príspevkov, vznikajú a zdokonaľujú sa s cieľom efektívneho vyhľadávania viacslovných pomenovaní, frazém a iných ľubovoľne zvolených slov a syntagiem, poskytujú tiež mnoho ďalších podnetov na lingvistické analýzy a zlepšujú možnosti práce s korpusmi. Vďaka tomu ich môžu efektívne využívať novinári, prekladatelia, pedagógovia a široká verejnosť.

Ján Staš, Daniel Hládek, Jozef Juhár (Katedra elektroniky a multimediálnych telekomunikácií Technickej univerzity v Košiciach) a Martin Ološtiak (Filozofická fakulta

Prešovskej univerzity v Prešove) v príspevku *Automatic Extraction of Multiword Units from Slovak Text Corpora* predstavili vlastný spôsob automatickej extrakcie viacslavných pomenovaní z rozsiahlych textových korpusov. Pomocou dostupných textov na internete vytvorili 4 podkorpora zamerané na špecifické oblasti (novinové články, súdne texty, beletristické texty a web) a pre každú z nich navrhli morfológicky motivovaný a štatistický prístup na extrakciu relevantných viacslavných pomenovaní, ktorá môže byť užitočná pri klasifikácii a kategorizácii textových dokumentov, pri automatickej korekcii textu a porovnávacom lingvistickom výskume.

Ralitsa Dutsova a Ludmila Dimitrova (Inštitút matematiky a informatiky Bulharskej akadémie vied v Sofii) v príspevku *Software System for Processing Bulgarian Digital Resources: Parallel Corpora and Bilingual Dictionaries* informujú o systéme spracovania a webovej prezentácii bulharských digitálnych zdrojov. Modul Search je softvérový nástroj na extrakciu dát z bilingválnych slovníkov, ktorý má implementované rozličné parametre vyhľadávania z oblasti morfosyntaxe, derivatológie a metriky. Táto webová aplikácia je v súčasnosti experimentálnym nástrojom s flexibilnou štruktúrou a možnosťou rozšírenia o ďalšie funkcie. Podľa autoriek ide o užitočný nástroj so širokými možnosťami uplatnenia vo výskume, pri tvorbe lexikálnych databáz, v translológii a pri výučbe jazykov.

Rozpoznávaním združených pomenovaní v gréčtine sa zaoberali Vasiliki Foufi, Kyriaki Ioannidou a Olympia Tsaknaki (Filozofická fakulta Aristotelovej univerzity v Solúne; *Obstacles and Solution to Recognizing Compound Nouns in Greek: A Corpus Study*). Sústredili sa na vyhľadávanie syntagiem zložených z adjektív a substantív, ktoré sa v gréčtine často vyskytujú ako nekompaktné štruktúry prerušené v texte inými vloženými slovami. Na vyhľadávanie použili špeciálnu FST metódu (konečný stavový transducer, t. j. automat s 2 páskami), ktorú aplikovali na dáta korpusu gréckych publicistických textov s rozsahom približne 6 miliónov textových jednotiek. Výsledky hľadania pomocou FST metódy sú uspokojivé, nástroj sa však môže ďalej zdokonaľovať tak, aby sa pomocou neho dali vyhľadávať aj ďalšie typy združených pomenovaní.

V príspevku *Identification of Idioms in Spoken Corpora* predstavili Milena Hnátková a Marie Kopřivová (Filozofická fakulta Karlovej univerzity v Prahe) program FRANTA, ktorý bol vyvinutý na automatickú identifikáciu frazém v hovorených korpusoch ČNK (Pražský mluvený korpus, ďalej PMK, ORAL2006, ORAL2008). Doteraz sa frazémy v PMK vyhľadávali a identifikovali manuálne, preto sa teraz môže porovnávať efektivita automatickej a manuálnej identifikácie frazém. Autorky opísali výhody a nevýhody oboch postupov a načrtli ďalšie možnosti, ako spresniť vyhľadávanie idiómov v prepisoch hovorených prejavov.

Kateřina Veselovská (Matematicko-fyzikálna fakulta Karlovej univerzity v Prahe) predstavila v príspevku *Czech Subjectivity Lexicon: A Lexical Resource for Czech Polarity Classification* český lexikón subjektivity určený na analýzu nálady v českých textoch. Lexikón tvorí 4 947 hodnotiacich položiek zatriedených podľa slovných druhov s označenou pozitívnou alebo negatívnou polaritou. Jadro českého lexikónu subjektivity vzniklo automatickým prekladom voľne dostupného anglického slovníka subjektivity, ktorý sa následne využil na testovanie na dátach získaných z českých spravodajských a ďalších internetových portálov. Slovník je potrebné ďalej vylepšovať, už jeho prvá verzia však poskytuje účelný lexikálny zdroj na sémantické analýzy a výskum hodnotenia v jazyku.

Uliana Yazhinova (Oddelenie slovanských štúdií Humboldtovej univerzity v Berlíne, Oddelenie slovanských štúdií Univerzity v Regensburgu; *A Corpus-based Analysis of the Functionality and the Meaning of Infinitive „Frustrative Construction” in Czech and Slovak*) skúmala infinitívne „konštrukcie frustrácie” v slovenčine a v češtine. Konštrukcie n(i)e a n(i)e + infinitív predstavujú jedinečné sémantické štruktúry, ktoré vyjadrujú subjektívnu, často negatívnu reakciu zo strany aktora alebo zúčastneného pozorovateľa na neočakávaný alebo nežiaduci výsledok akcie. Autorka na základe analýzy dát z korpusov ČNK, SNK a doplnkových dát z internetu podrobne opísala funkčné vlastnosti tejto kompaktnej konštrukcie a porovnala výsledky pre češtinu a slovenčinu.

Daša Munková (Pedagogická fakulta Univerzity Konštantína Filozofa v Nitre), Michal Munk a Martin Vozár (Fakulta prírodných vied Univerzity Konštantína Filozofa v Nitre) sa v príspevku *The Effect of Stop Words Elimination on Sequence Patterns Extraction in Comparable Corpora* zamerali na stop slová (hlavne predložky a spojky), ktoré skúmali na dátach dvoch porovnateľných korpusov reklamných textov, pričom každý z nich obsahoval 600 reklamných textov. Výsledky analýzy ukázali, že iba jazyk má signifikantný vplyv na množstvo a kvalitu extrahovaných pravidiel v porovnateľných korpusoch anglických a slovenských textov a odstraňovanie stop slov na základe vopred vytvoreného zoznamu je pri analýze takýchto krátkych textov neefektívne.

Korpusová lingvistika má svoje miesto v modernej filológii aj interdisciplinárnom výskume. Rozličné korpusy poskytujú veľké množstvo dostupných, prehľadných a utriedených dát, vďaka čomu sa mnohé hypotézy a teórie spoľahlivejšie a efektívnejšie overujú a samotný lingvistický výskum sa tak zdokonaľuje. Vlastný výskum vybraných javov z morfológie, syntaxe, valencie a sémantiky zachytilo vo svojich štúdiách viacero nižšie uvedených autorov z Česka a zo Slovenska.

K najfrekvencovanejším slovným tvarom v češtine patria reflexíva *sa* a *si*. Na báze korpusu súčasnej češtiny SYN2010 s rozsahom 100 miliónov slovných jednotiek ich detailne skúmal Vladimír Petkevič (Filozofická fakulta Karlovej univerzity v Prahe; *Formal (Morpho)Syntax Properties of Reflexive Particles se, si as Free Morphemes in Contemporary Czech*). V štúdiu sa zaoberá predovšetkým vzťahom reflexív k ich nadradeným slovám, ktorými môžu byť slovesá, deverbatívne adjektíva alebo deverbatívne substantíva. Špecifikácia vlastností reflexív môže prispieť k zlepšeniu automatickej morfologickej dezambiguácie korpusových dát, ako aj k spresneniu syntaktickej analýzy textov súčasnej češtiny.

Jedným z najnáročnejších problémov pri anotácii ručne značkovaneho korpusu slovenčiny r-mak (verzie 1.0 – 4.0) bola delimitácia participií. Podrobnému opisu anotácie participií a riešeniu hraničných prípadov sa venovala Agáta Karčová (Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied v Bratislave) v príspevku *Delimitation of Participles in the Manual Morphological Annotation*. Autorka objasňuje, prečo sa participiá v SNK vyčleňujú ako samostatná slovná trieda, venuje sa prípadom homonymie, analogicky vytvoreným tvarom a určovaniu lexém, ktoré stoja na rozhraní tried participiá – adjektíva a participiá – substantíva. Obhaja formálny prístup pri anotácii participií, ktorý sa ukázal ako logický a spoľahlivý.

Veronika Kolářová (Matematicko-fyzikálna fakulta Karlovej univerzity v Prahe) vo svojom príspevku *Agents Expressed by Prepositionless Instrumental Modifying Czech Nouns Derived from Intransitive Verbs* opísala české deverbatívne substantíva, ktoré umožňujú zmenu agensa vyjadreného bezpredložkovým inštrumentálom. Schopnosť substantíva byť rozvi-

tým konateľom v bezpredložkovom inštrumentáli sa tradične považuje za jedno z kritérií pri stanovovaní dejového významu substantíva. Hlavnou otázkou, na ktorú sa autorka zamerala, bolo, či je konateľ v bezpredložkovom inštrumentáli gramaticky správny aj pri substantívach odvodených z netranzitivných slovies.

Václava Kettnerová, Markéta Lopatková, Eduard Bejček, Anna Vernerová (Matematicko-fyzikálna fakulta Karlovej univerzity v Prahe) a Marie Podobová (Slovanský ústav Akadémie vied Českej republiky v Prahe) v príspevku *Corpus Based Identification of Czech Light Verbs* zhrnuli možnosti identifikácie českých kategoriálnych slovies. Na odlíšenie použitia slovies v ich kategoriálnom význame od ich plnovýznamového použitia aplikovali test redukcie, koreferenčný test a doplnkové kritérium na anotáciu pomocných slovies. Prijaté kritériá overovali paralelnou anotáciou veľkého množstva dát získaných z ČNK. Výsledkom anotácie sú tzv. „zlaté dáta“, ktoré pozostávajú z 893 kombinácií kategoriálnych slovies a predikatívnych podstatných mien, ktoré sa využijú pri tvorbe valenčného slovníka českých slovies VALLEX.

Autorky príspevku *Verb Valency and Argument Non-correspondence in a Bilingual Treebank* Jana Šindlerová, Zdeňka Urešová a Eva Fučíková (Matematicko-fyzikálna fakulta Karlovej univerzity v Prahe) prezentovali na materiáli česko-anglického závislostného korpusu rozličné realizácie valenčných rámcov slovies v rámci rovnakej sémantickej triedy, zamerali sa na slovesá vyjadrujúce súd (v širokom zmysle slova, napr. obviňovať, prisúdiť, usvedčiť, vytknúť a iné). S ohľadom na túto triedu slovies analyzovali príklady a zvažovali prirodzenosť „tretieho argumentu“ vo valenčnej štruktúre. Navrhli konkrétny spôsob unifikovania anotačnej praxe pre danú triedu slovies a možné spôsoby na upravenie valenčnej teórie pre potreby viacjazyčných dát.

Kateřina Najbrtová (Filozofická fakulta Masarykovej univerzity v Brne) v príspevku *Valency of Selected Primary Adjectives in the SYN2010 Corpus* opísala valenčné rámce primárnych adjektív analyzovaných na textoch korpusu SYN2010. Autorka vychádzala zo starších štúdií H. Prouzovej a M. Kopřivovej, porovnávala ich valenčné rámce adjektív a vybrala z nich konkrétne adjektíva, ktoré testovala na materiáli korpusu SYN2010. Vytvorila rozsiahly zoznam primárnych adjektív s valenciou a frekvenčnou štatistikou a podrobne opísala valenciu najčastejších z nich. Zistila tiež, že ak existuje možnosť valenčných doplnení primárnych adjektív infinitívom, potom je toto doplnenie vždy početnejšie ako doplnenie vedľajšou vetou.

Zdeňka Hladká (Filozofická fakulta Masarykovej univerzity v Brne) príspevkom *Corpora of Private Correspondence as a Source of Material Focused on a Research of Diminutives* upriamuje pozornosť na dôležitý zdroj lingvistických dát – korpus súkromnej korešpondencie. Zo špecializovaného korpusu KSKdopisy, ktorý vznikol na Masarykovej univerzite a obsahuje 2 000 súkromných listov, Z. Hladká vytvorila podkorpus mladých autorov (mladších ako 30 rokov) obsahujúci 300 listov, ktorý slúžil na skúmanie deminutív. Súčasťou príspevku je zoznam vyexcerpovaných deminutív, údaje o ich frekvencii a základné informácie týkajúce sa ich lexikografickej reflexie.

Nasledujúce príspevky sa venujú využívaniu korpusov vo vyučovaní, vytváraní špeciálnych korpusov a cvičení na didaktické ciele a tiež prehľadu o online vzdelávaní a jeho špecifikách v porovnaní s klasickým spôsobom vyučovania.

Jedným z cieľov príspevku Kataríny Pišútovej (Centrum informačných technológií Univerzity Komenského v Bratislave; *Introduction to Online Learning*) bolo poskytnúť kľú-

čové informácie o online vzdelávaní. Autorka sa venovala základným historickým faktom o vzniku online vzdelávania, na základe ktorých sa pokúsila o výstižnú definíciu tohto moderného spôsobu vzdelávania. Vymedzila aj rozdiely medzi tradičným a online vyučovaním. Druhá časť príspevku bola venovaná rozvoju online vzdelávania na Slovensku a problémom, ktoré sú špecifické pre našu situáciu (technické problémy, novosť tejto formy vzdelávania, nedostatky v štruktúre kurzov a i.).

Grzegorz Krynicki (Fakulta angličtiny Univerzity Adama Mickiewicza v Poznani) prezentoval vo svojom príspevku *Corpus-based Online Word Formation Exercises for Advanced Learners of English – Challenges and Solutions* návrh a prevádzku online platformy určenej na vyučovanie angličtiny ako cudzieho jazyka. Vytvoril cvičenia zamerané na tvorbu slov, doplnené o ukážkové vety, ktoré sa jednoduchou metódou automaticky extrahujú z Britského národného korpusu. Takéto cvičenia obohatené o autentický vetný kontext rozvíjajú schopnosti žiaka učiť sa gramatiku a uľahčujú pochopenie významov, vzťahov a kolokácií slov cudzieho jazyka.

Špecializovaný korpus môže slúžiť ako cenný zdroj údajov, z ktorých sa dajú zistiť časté chyby vyskytujúce sa v procese osvojovania si cudzieho jazyka. K tomuto záveru dospela na základe výskumu Andrea Hudousková (Filozofická fakulta Karlovej univerzity v Prahe; *The Corpus CzeSL in the Service of Teaching Czech for Foreigners – Errors in the Use of the Pronoun který*). Autorka pracovala s dátami nereferenčného korpusu CzeSL, ktorý bol sprístupnený v roku 2012 a je zostavený z písomných prác študentov učiacich sa češtinu ako cudzí jazyk s rozsahom približne 2 mil. slov. Autorka sa zamerala na vyhľadávanie chýb pri používaní vzťažného zámena *který* a konfrontovala rozsah zistených chýb s opismi vzťažnej vety v učebniciach a gramatikách češtiny.

Špecifický prienik sémantiky, antropológie, filozofie a korpusovej lingvistiky badať vo výskume T. Timoshchenkovej a M. Debnára.

Tatiana Timoshchenko (Filozofická fakulta Karlovej univerzity v Prahe) vo svojom príspevku *Determination of Czech BCT Prototypes on the Basis of Corpus Data* sémanticky analyzovala prototypy základnej farebnej terminológie v češtine (Basic Color Terms, skr. BCT) skúmanej na korpusových dátach. Stanovenie jadra farebných kategórií spočívalo v hľadaní charakteristických prírodných entít ako kognitívnych referenčných bodov. V štúdiu vyčleňuje 3 typy lingvistických konštrukcií: prirovnávacie (*černý jako uhel*), príslovkové (*křídově bílý*) a inštrumentálne (*šedý prachem*). V prirovnaniach sa BCT prototypy odhaľujú rovnomerne, adverbiálna konštrukčná analýza je efektívnejšia na opis sémantického poľa BCT a inštrumentálna konštrukcia sa podľa autorky osvedčila ako zdroj koncepcnej metonymie.

Marek Debnár (Jazykovedný ústav Ľ. Štúra Slovenskej akadémie vied v Bratislave) predstavil vo svojej prezentácii *Možnosti spracovania sémantických polí v Slovenskom národnom korpuse* niektoré prieniky filozofie a jazyka. Načrtol možnosti využitia korpusov pre širšie humanitné bádanie, najmä literárno-teoretické, semiotické a filozofické. Takisto navrhol zaviesť do štýlovo-žánrovej anotácie textov v Slovenskom národnom korpuse kľúč „narrator“, ktorý by zaznamenával perspektívu rozprávača v umeleckých textoch, čo by umožnilo štatistické skúmanie textov rôznych umeleckých období a oblastí z hľadiska subjektívizácie výpovede.

V závere medzinárodnej konferencie SLOVKO 2013 sa Mária Šimková, vedúca oddelenia SNK JÚLŠ SAV, poďakovala všetkým zúčastneným aj prednášajúcim za aktívnu účasť na

konferencii, príjemnú a priateľskú atmosféru počas celého podujatia, recenzentom zborníka za ich kvalitnú prácu a zaželala všetkým pozitívne prínosy z rokovaní pre ďalšie projekty v oblasti počítačového spracovania prirodzeného jazyka. Zároveň vyjadrila želanie, aby sa v duchu profesionality a spolupráce odohrala aj plánovaná konferencia SLOVKO 2015.

Agáta Karčová – Beáta Kmeťová
Jazykovedný ústav Ľ. Štúra SAV, Bratislava