

Gramatika a korpus 2012: 4. mezinárodní konference

28.–30. listopadu 2012

Kateřina Najbrtová

Filozofická fakulta Masarykovy univerzity, Brno

145034@mail.muni.cz

Čtvrtá mezinárodní konference Gramatika a korpus se konala u příležitosti stého výročí narození Miloše Dokulila, význačné osobnosti české lingvistiky. Po předchozích setkáních, která se uskutečnila v Praze (2005), v Liblicích (2007) a v Mannheimu (2009), se konference vrátila zpět do Prahy, kde proběhla ve dnech 28.–30. listopadu 2012. Záštitu nad konferencí převzal Ústav pro jazyk český AV ČR, avšak mezi členy organizačního výboru jsme mohli nelézt mnoho zahraničních lingvistů.

Každý ze tří konferenčních dnů byl zahájen plenárním referátem, s výjimkou prvního dne setkání. Zahájení konference bylo věnováno Miloši Dokulilovi – připomenutí jeho vybraných myšlenek se ujal František Štícha.

V konferenčních příspěvcích byla zastoupena nejrůznější témata, více či méně spojená s korpusy. Kromě referátů zabývajících se všemi tradičně vymezovanými rovinami popisu jazyka byly prezentovány i příspěvky s tématem budování korpusů, pragmatických aspektů komunikace nebo příspěvky založené na jiných než českých korpusech. Zcela v duchu setkání početně dominovaly prezentace věnované výzkumu slovtvorby a také morfologie a syntaxe.

Plenární referáty prvního dne přednesli A. Polikarpov (*Regularities of Word-Formation in a System Theory of Language Evolution*) a P. Štichauer, který představil novou potenciální klasifikaci kompozit (*Is a new classification of Czech compounds possible?*). V dalším plenárním referátu (*Measuring Standard German? Frequency and Variation as Indicators for Acceptability*) M. Konopka a U. H. Waßner poukázali na to, že míra frekvence může být dobrým pomocníkem při diskuzi o přijatelnosti jevů, ale ne vždy se jedná o přímou závislost. V závěrečný den konference se A. Rosen věnoval vztahu empirie a teorie (v saussurovské terminologii parole a langue) a potíží při vytváření závislostních stromů, jež

by v ideálním případě měly zachycovat oboje (*Grammar based treebank – a happy marriage of empirism and theory?*).

V příspěvku nazvaném příhodně *Srovnávání nesrovnatelného* předvedl J. Králík, že společně s výrazným zvýšením objemu dat musí být upraveny i zažité metody zacházení s nimi – extrémní růst sledovaných textů nevede k extrémnímu růstu přesnosti získaných údajů.

Na korpus jako zdroj ověřování dat, nikoli jejich získávání se zaměřili P. Karlík, M. Ziková a studenti Masarykovy univerzity (*Osobní zájmena: ověřování hypotéz daty (intuice>KORPUS>dotazník)*).

Naopak to, že v určitých případech nemusí být ani rozsáhlé korpusy plně dostačující, prokázala K. Osolobě (*Korpusy a internet jako zdroje dat pro výzkum produktivity periferního slovtvorného typu: adjektiva typu hrůzoucí (hrůza) v korpusech a na internetu*).

Využití korpusu DIAKORP pro diachronní bádání demonstrovali A. Černá (*Člověk v množném čísle (v historické i v současné češtině)*) či F. Martínek (*Slovtvorné typy českých deverbativ a deadjektiv (diachronní pohled)*).

V příspěvku *Automatické generování slov ve slovanských jazycích na základě jejich společných historických kořenů* upozornili P. Pognan a J. Panevová na česko-francouzský projekt počítačového programu, který na základě vložených údajů utváří současné podoby staroslovanských kořenů v různých slovanských jazycích.

Deriváty adjektiv z hlediska jejich vztahu k základovým adjektivům a následného zachycení v anotaci PDT se zabývala M. Ševčíková (*Deadjektivní deriváty v češtině jako deriváty syntaktické vs. lexikální*). Oblast tvorby korpusů a využívání korpusových nástrojů dále zastoupily příspěvky H. Palátové a M. Gráce (*Segmentace textu na věty*) či M. Hnátkové (*Automatická slovnědruhová desambiguace slova „to“ v ustálených větných výrazech*).

Výzkum synchronních slovtvorných vztahů mezi zadanými vzory slovtvorných paradigmat umožňuje nový nástroj MORFIO, o němž informovali V. Cvrček a P. Vondříčka (*Nástroj pro slovtvornou analýzu jazykového korpusu*).

Nesnáze rozpoznávání dativu při anotaci se staly tématem příspěvku V. Petkeviče (*Automatické rozpoznávání substantivního dativu a jeho syntaktických funkcí v českých textech*). Zmíněný úkol je velmi náročný, jeho (byť ne zcela dokonalé) řešení spočívá především ve využití slovesné i substantivní valence.

Webovou aplikaci Corpus Architect umožňující pohodlné budování textových korpusů včetně případné tokenizace a lemmatizace představili J. Pomikálek a V. Suchomel (*Creating Corpora Using Corpus Architect*).

Zajímavý způsob analýzy jazyka využívající závislosti funkce substantiv ve větě na funkčním stylu nabídl T. Jelínek (*Porovnávání funkčních stylů v korpusu SYN2005 na základě frekvence syntaktických funkcí substantiv*).

Problematické vidu a možnostem a omezením jeho výzkumu v korpusech věnovali svá vystoupení F. Esvan (*K střídání vidu při vyprávění v češtině*), T. Berger (*Funkcionální vidové dvojice v ČNK*) a P. Nádeníček (*Dílčí druhy imperfektivnosti a perfektivnosti. Příspěvek k vidovému systému slovanských jazyků*).

Součástí bohatého programu byly i referáty zahraničních badatelů zaměřené např. na ruštinu – S. Gorokhova (*The frequency effect in the production of adjective-noun agreement: A corpus based study of Russian speech errors*) či I. B. Itkin a kol. (*Research on stressed prepositions in Russian: a corpus based approach*), německý jazyk regionů Rakouska, Německa a Švýcarska – S. Ueberwasser (*Pluri-Regional German Grammar: A Corpus based Approach*) či slovo tvorbu litovského jazyka – J. Pakerys (*The co-occurrence of denominal and deadjectival verbs with their base words in Modern Lithuanian (a corpus based study)*).

Z. Hladká předvedla, jakým způsobem je možné na příkladech hypokoristik z KSK odhalit slovo tvorné postupy (*Využití korpusů korespondence pro účely (nejen) slovo tvorných analýz*), deminucí – ať už z pohledu češtiny či češtiny v komparaci s jinými jazyky – se zabývali J. Bílková (*Pseudodeminutiva v češtině*) a T. Káňa (*Česká substantivní deminutiva ve světle korpusových dat*), primární deminutiva odvozená sufixy *-ek* a *-ík* analyzovala z fonologického hlediska M. Ziková (*Maskulinní deminutiva v současné češtině*). Jak interpretovat nejrůznější zvláštnosti a inovace slovo tvorných procesů v češtině ukázal J. Šimandl (*Nárůst nepravidelného tvoření slov: o čem svědčí?*). Kompozicí jako slovo tvorným postupem se zabývala I. Bozděchová (*Kompozice a její potenciál v současné slovní zásobě češtiny*), vybrané sufixy a jejich funkce spolu s možnostmi spojování byly předmětem zájmu D. Hlaváčkové a K. Paly (*Některé vybrané substantivní sufixy v češtině (-ák, -ec, -ík, -ník) – případová studie*).

Ve vystoupeních tematicky orientovaných na syntax využili jejich autoři korpus jak ke zkoumání spíše okrajových jevů (A. Karčová: *Postponovaný přívlastek v slovenštině a češtině*, M. Šimková: *Předložky kvůli/kvůli a pre/pro v systéme a v použití (na materiálu slovenského a českého jazyka)*), tak jevů ovlivňujících široké úrovně jazyka (M. Dočekal, H. Strachonová: *Negace a skopus kvantifikátorů*, M. Mikulová a další: *Liší se mluvené a psané texty ve valenci?*), o diachronní náhled se postaral P. Kosek (*Vývoj slovosledných vlastností auxiliáru préterita ve starší češtině*).

Konference byla nejen důstojnou poctou M. Dokulilovi, ale především podnětným setkáním jazykovědců různých věkových skupin a národností, vzájemně spojených tématem korpusového zkoumání jazyka. Autoři referátů obsáhli široké spektrum oblastí výzkumu, přičemž důraz byl kladen na témata slovo tvorná a syntaktická. Přednesené příspěvky jsou vynikajícím dokladem toho,

jak veliké (de facto neomezené) využití nabízí počítačově zpracovaný jazyk – ještě nikdy v historii neměli lingvisté v rukou zdroje dat v takovém rozsahu. Nutno dodat, že k příjemnému průběhu setkání nemalou měrou přispěla i společná slavnostní večeře, která umožnila neformální setkání (a často i diskuzi) všech zúčastněných.

Všechny příspěvky jsou vydány na CD nakladatelstvím Gaudeamus Hradec Králové.