

KONFERENCE O SLOVANSKÝCH HISTORICKÝCH KORPUSECH¹



Ve dnech 21. a 22. dubna 2015 se na Norské arktické univerzitě (The Arctic University of Norway) v Tromsø uskutečnila konference s názvem *Slavic Corpus Linguistics: The Historical Dimension*, kterou uspořádali místní diachronní slavisté Tore Nessel, Hanne Eckhoffová a Aleksandrs Berdičevskis pod dohledem Laury A. Jandy. Uvedení lingvisté pracují v projektu *Birds and Beasts* především na treebanku TOROT, sestávajícím ze staroslověnských, staro- a středněruských textů o celkovém rozsahu 220 000 tokenů, opatřených morfologickou a syntaktickou anotací. Vedle toho se zabývají statistickým zpracováním korpusových dat a vytvářejí typologii tzv. slovesných klasifikátorů (*verb classifiers*). Vyhledávání v korpusu TOROT a jeho funkce byly na konferenci představeny (zajistila to H. Eckhoffová) a řada dalších referátů z jeho dat vycházela.

Několik příspěvků se věnovalo samotné přípravě textů pro historické korpusy. Z nich nejdůležitější bylo metodologické vystoupení Davida J. Birnbauma (Pittsburgh). Tento americký profesor ukazoval výsledky svého edičního úsilí, zaměřeného především na středověké slovanské rukopisy, a apeloval přitom na mnohoúčelovost elektronických edic, na jejich přístupnost odborníkům z různých oborů. Editologickými otázkami, konkrétně mírou zachovávání původní interpunkce v textech vydávaných a zařazovaných do jazykových korpusů, se zabývala také Barbara Sonnenhauerová (Curych). Dokládala, že interpunkce, podmíněná v rukopisech „foneticky“, má pro zkoumání historického jazyka značný význam. Tím se dotkla obecnější otázky, jak vyřešit vztah mezi spolehlivostí dat a potřebnou jednotností jejich zpracování v zájmu snadného přístupu ke korpusovým datům.

U některých dalších příspěvků spočíval jejich primární účel v představení příslušného — zpravidla historického — korpusu či podobného projektu.² Leonid Iomdin (Moskva) referoval o hloubkově anotovaném korpusu současné ruštiny SynTagRus, který v současné době obsahuje cca 60 tisíc vět, což je 900 tisíc slov. Joanna Bilińska, Witold Kieras a Magdalena Derwojedowa (Varšava) představili probíhající projekt automatické morfologické analýzy polských textů z let 1830–1918 a poukázali např. na komplikace, které způsobuje nejednotný pravopis. Włodzimierz Gruszczyński (Krakov) představil korpus polských barokních textů (KorBa) a zmínil se o slovníku polštiny 17. a 18. století. Tento slovník, časově navazující na *Słownik polszczyzny XVI wieku*, se zpracovává tak, že hotová i podstatně rozpracovaná hesla — zahrnující pochopitelně i množství dokladů, v ideálním případě přímo s odkazy do korpusu — jejich autoři rovnou publikují na webu. (Že má v analogické problematice česká historická lexikografie co dohánět, snad ani není třeba psát.) Marija Lazarová (Lipsko) představila německou a (slovakizovanou) českou verzi *Saského práva* (tou druhou pochopi-

1 Tento text vznikl v souvislosti s řešením grantového projektu č. 406/12/0691 Korespondence Karla Havlíčka uděleného Grantovou agenturou České republiky.

2 Pro zjednodušení mluvím o korpusech historických, popř. historického jazyka, bez ohledu na to, zda zachycují jazyk krátkého, nebo dlouhého období, a lze k nim tedy přistoupit jako k reprezentaci jednoho stavu jazyka nebo jazykového vývoje. Pojmenování „diachronní“ by totiž v některých případech mohlo být zavádějící.



telně myslím tzv. *Žilinskou knihu*) a pomocí korpusového nástroje, četnosti n-gramů, zkoumala ustálení lexika (jejími slovy jeho standardizaci).

Další příspěvky se týkaly zpravidla obecnějších otázek jazykové změny. Aleksanders Berdičevskis (Tromsø) a Alexander Piperski (Moskva) dokládali na vývoji ruských sloves předpoklad, že s postupujícími morfologickými změnami směřuje jazyk k větší pravidelnosti. Rafał L. Górski (Krakov) představil možnosti statistických metod kvantitativní lingvistiky pro studium nedávných jazykových změn v polštině. Pracoval tedy „korpusově řízeným přístupem“ a zodpovídal otázku, zda je obraz jazykových změn, jak je podává korpus, v souladu s hranicemi epoch, jak si je stanovujeme podle společenských, politických aj. změn. Podle jeho výsledků lze změnu v jazyce zaznamenat s určitým zpožděním oproti změně vnějších událostí, např. nejvíce jazykových změn v 10. a 20. letech 20. století nebylo registrováno v roce 1918, ale o sedm let později. Změnám pro změnu lexikálním v souvislosti se společenským vývojem se pak věnovala Julija Kuzněcova (Chapel Hill), když zkoumala klíčová slova odrážející historické události v korpusu rusky psaných knih z 20. století, které jsou dostupné v Google Books. V podstatě metodologickým příspěvkem s ilustrací na vybrané hláskové změně bylo vystoupení Sergeje Saje (Petrohrad). Přednášející ukázal průběh ruské inovace *-oju > -oj* v instr. sg. fem., jež se nejrychleji prosazovala v 19. století (na počátku tohoto století ukázala korpusová data 26 % forem s *-oj* a na jeho konci už 96 %), a diskutoval, kam na škále od změny čistě generační ke změně skupinové lze tuto inovaci zařadit. Detailní analýza jazyka více než 50 spisovatelů po jednotlivých dekadách jejich života ukázala, že někteří změnu postupně přijímají, u jiných naopak vzrůstá konzervativnost a archaičnost.

Nepřekvapuje, že se pozornost několika příspěvků obracela k slovesnému vidu. Jekatěrina Mišina (Moskva) zkoumala vid ve staroslověnštině a staré ruštině, a to na třech typech dokladů: na formách, které by se „učebnicově“ neměly tvořit (imperfekta od dokonavých sloves, aoristu od nedokonavých apod.), na spojení infinitivu „dokonavého“ slovesa s *načati* a na užití prezentního tvaru s budoucím významem (což je kritérium nejméně spolehlivé, protože sémantické). Podle svých slov nenašla žádné doklady typu *načati* + pf. ani prezentní formu s budoucím významem utvořenou od ipf. Valerij Solovjov (Kazaň) se ve svém příspěvku věnoval výzkumu potenciálních ruských vidových párů, a to jak lexémů, tak vybraných významů příslušných sloves (resp. jejich výskytů v typických, frekventovaných spojeních). Podle jeho slov se osvědčilo sledovat průběh křivek výskytu potenciálních členů vidového páru v grafech získaných pomocí nástroje Google Books Ngram Viewer (viz <<https://books.google.com/ngrams>>), kde lze zobrazit data z let 1800–2008: pokud si průběh křivek obou sloves odpovídá, je velmi pravděpodobné, že se jedná o členy vidového páru.

Z příspěvků morfologických a syntaktických se nyní zmíním o těch, které pojednávaly o syntaktických funkcích zájmen. Iliana Krapova (Benátky) a Cvetana Dimitrova (Sofie) představily diachronní výzkum posese vyjádřené zájmenným dativem ve staroslověnských památkách. Popsaly změny postavení klitik ve větě (v souvislosti s — mými slovy — aktuálním členěním větným neboli řazením větných členů podle informační závažnosti) a synkretismus genitivu a dativu, proběhnuvší pod vlivem řečtiny (v řečtině stál na místě stsl. posesivního dativu genitiv). Mirjam Zumsteinová (Cambridge) se zamýšlela nad klasifikací staroruského *iže* jako vztažného zájmena nebo



jako tzv. tematizátoru (*topic marker*; částice uvozující téma věty). V příspěvku představila čtyři odlišné typy vedlejších vět a problematizovala anotaci dat v korpusu TOROT. Distribuci vztažných zájmen v historické perspektivě a jejich přechodem ve „vztažné částice“ se zabýval také Roland Meyer (Berlín): studoval rozšíření zájmen *jenž, který* a *co* ve staroslověněštině a v celém vývoji ruštiny, polštiny a češtiny a všiml si např. role jazykového kontaktu při zavedení určitého prostředku do některého z daných jazyků. Silvia Luraghiová a Erica Pinelliová (Pavia) pak sledovaly nárůst potřeby vyjádřit podmět pomocí zájmena, který probíhal v ruštině mezi 16. a 18. stoletím. A Łukasz Jędrzejowski (Postupim) srovnával vnořené otázky s infinitivem (*embedded infinitival questions*) a modální konstrukce se vztažnými výrazy (*modal existential wh-constructions*) v současné i starší polštině, které se (i mezijazykově) liší ve dvou ohledech — množstvím sloves, která je mohou uvozovat, a množstvím vztažných výrazů, které v nich jsou přijatelné. Autor zde mj. využíval (a polskými příklady i zpřesňoval) zjištění, která ve své srovnávací studii shromáždil Radek Šimík.³

V mém přehledu budou následovat příspěvky věnované slovesům, nejprve jednotlivým tvarům, poté slovesným doplněním. Vladimir Plungjan (Moskva) a Anna Urmančijeva (Petrohrad) se na základě dat z korpusu PROIEL věnovali sémantice staroslověnského opisného perfekta. Příslušný slovesný čas podle nich na rozdíl od jiných minulých časů nekalkuje řecké předlohy. Na rozdíl od současné bulharštiny ve staroslověněštině nevystupuje jako rezultatívum, tj. nevyjadřuje děj působící do přítomnosti, mající na ni vliv. Pavel Petruchin (Moskva) se ve svém příspěvku zabýval složenými slovesnými tvary tvořenými aoristem či imperfektem slovesa *byti* a činným přičestím ve staroslověnských a staroruských textech. Daná konstrukce má nejasný původ a řečník upozornil, že kromě průběhového významu, který se jí obvykle připisuje, má i význam habituální a stavový. Dále ukázal ústup těchto slovesných tvarů a vysvětlil jej mj. postupným narůstáním počtu slov mezi určitým a neurčitým slovesným tvarem, což vedlo opisovače rukopisů k neporozumění a reinterpetaci tvaru. Další přednášející Marija Ovsjannikova (Petrohrad) se v příspěvku s atraktivním názvem *Včerejší argumenty jsou dnešní okolnostní určení* věnovala změnám v ruské rekci od 18. do 20. stol., a to na základě dat z ruského národního korpusu. František Martínek (Praha) pak sledoval změny významu a syntaktického chování sloves *hodlat* a *uspět* v českých textech z 19. a 20. století.

Také další příspěvky byly věnovány aktuálním otázkám slovanské gramatiky. Ruprecht von Waldenfels (Krakov) sledoval pomocí synchronního paralelního korpusu slovanských jazyků (ParaSol) výraznější užití předložky *do* (*s* genitivem) na úkor *v* (*s* akuzativem) a zčásti také na úkor *k* v jazycích, které lze označit jako severoslovanské — v lužické srbštině, češtině, slovenštině, polštině, ukrajinštině a běloruštině. Ilja Seržant (Mohuč/Vilnius) se zabýval značením životnosti v diachronním pohledu, a to rovněž na materiálu staroruských památek. A konečně Anna Ptencova (Moskva) se věnovala slovnědruhovému platnosti výrazu *krivo* („zlý“, „zlo“) ve staroruských památkách, přičemž u některých dokladů zpochybňovala běžnou slovníkovou interpretaci adverbialní a navrhovala je klasifikovat jako substantivum.

3 Šimík, Radek (2011): *Modal Existential Wh-Constructions* [Ph.D. thesis]. Groningen: Rijksuniversiteit Groningen.



Jak je patrné, konference nabídla nejrůznější inspirace pro přípravu jazykového materiálu i jeho vytěžování a zpracovávání, především v gramatických výzkumech, ale také lexikografickém popisu. Jako nejstručnější výběr z daných inspirací uvádím následující webové odkazy.

Prezentace a postery z konference jsou přístupné na adrese: <<http://site.uit.no/slavhistcorp/>>.

TOROT: The Tromsø Old Russian and OCS Treebank je dostupný na adrese: <<https://nestor.uit.no/>>.

ParaSol: A Parallel Corpus of Slavic and Other Languages je k dispozici na adrese: <www.parasolcorpus.org>.

PROIEL: Pragmatic Resources in Old Indo-European Languages je dostupný na adrese: <<http://www.hf.uio.no/ifikk/english/research/projects/proiel/>>.

Ediční počiny Davida J. Birnbauma a jeho spolupracovníků jsou k dispozici na adrese: <<http://www.obdurodon.org/>>.

Korpus polských barokních textů KorBa je přístupný na adrese: <<http://wiki.nlp.ipi-pan.waw.pl/korba>>.

Elektroniczny słownik języka polskiego XVII i XVIII wieku je k dispozici na adrese: <http://xvii-wiek.ijp-pan.krakow.pl/pan_klient/>.

František Martínek