

UVODNIK

Digitalizirani jezikovni viri, procesiranje naravnega jezika, korpusne analize slovničnih in drugih jezikovnih pojavov, rudarjenje besedil, označevalniki, luščilniki, leksikografska orodja, sinteza govora, strojno prevajanje, avatarski sogovorci, pametne hiše ... Skupna točka: jezik.

Raziskovanje, ki ob jeziku združuje informatike in strokovnjake z drugih področij ter jezikoslovce, je v zadnjih dveh desetletjih prineslo rezultate, ki v svoji aplikativni obliki postajajo vsak dan bolj običajen del življenja slehernika, na spoznavni ravni pa znova potrjujejo inovativnost in znanstveno relevantnost interdisciplinarnih sodelovanj. Slovensko jezikoslovje – če pogled zožimo le na domače raziskave z jezikoslovjem kot osrednjo stroko – se je temu razvoju priključilo konec 90. let 20. stoletja. Poslej je del jezikoslovcev, spodbujen zlasti z empiričnimi podatki, ki jih je prinašal prvi korpus slovenskega jezika FIDA, z raziskavami slovenščine in primerjalno ali povezano tudi drugih jezikov sledil živahnemu dogajanju na področju elektronskih jezikovnih virov ter sodobnih jezikovnih tehnologij in prinašal nova ali deloma drugačna, obenem pa vendarle tudi temeljna spoznanja o mnogih segmentih slovenščine. Vzporedno s širjenjem tovrstnega raziskovanja je nastajala vrzel – rezultati opisanih dejavnosti niso imeli svojega osrednjega prostora, v katerem bi bili povezano in kontinuirano predstavljeni ter utemeljevani, zato je marsikatera empirična raziskava slovenščine, ki bi jo bilo mogoče na ravni jezikovnotehnoloških pristopov prenesti tudi na druge jezike, ostala spregledana. Ker smo prepričani, da se bodo omenjene vsebine še naprej nadgrajevale in sledile svetovnim trendom raziskovanja sodobnih jezikov, jim namenjamo novo znanstveno jezikoslovno revijo: *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*.

Prva številka *Slovenščine 2.0* prinaša devet prispevkov dvanajstih avtorjev. Razpravni del začenja besedilo, ki korpusno jezikoslovje umešča v jezikoslovnoteoretski kontekst 20. stoletja: od pogleda na jezik, ki ga začenja de Saussurjev strukturalizem, prek zanikanja predpostavke o inherentni in nedeljivi povezavi pomena ter izraza pri Chomskem do (ponovne) Sinclair-

Hallidayeve obravnave jezika oz. govornice kot celote, v kateri je kakršnokoli ločevanje med leksiko in slovnico dejansko lažno. Prispevek o referenčnih, specializiranih in vzporednih korpusih, do katerih je mogoče dostopati prek konkordančnikov na strežniku nl.ijs.si, pregledno in v bistvenem predstavlja slovenske korpusne, ki so nastali v preteklih več kot 15 letih – korpusne, do večine katerih (javni) dostop brez angažiranja avtorja prispevka ne bi bil (več) mogoč – ter konkordančnika CUWI in noSketchEngine, prek katerih poteka iskanje in prikaz korpusnih rezultatov. Sledi prispevek o korpusnem pristopu k luščenju prevodnih ustreznice ter lažnih prijateljev v slovenskem in hrvaškem jeziku. Predstavljeni pristop je jezikovno neodvisen, zato je njegova metodologija širše uporabna, in to tako v leksikografiji kot pri poučevanju tujih jezikov. Četrto besedilo prikazuje primerjavo dveh korpusov: korpusa Gigafida in spletnega korpusa slovenščine slWaC. Avtorja sta na oba korpusa aplicirala metodo tematskega modeliranja in ugotovila, da so med Gigafido in slWaCom dokajšnje tematske razlike, kar bi veljalo upoštevati pri prihodnjih gradnjah referenčnih korpusov slovenščine. Sledi prispevek o avtomatski evalvaciji strojnega prevajanja. Avtorici ugotavljata, da bolj kot izbrana metrika (NIST, BLEU, TER) na avtomatsko evalvacijo prevajalnika vpliva evalvacijsko gradivo, kar je bilo še posebej izrazito vidno pri gradivu, ki sta ga uporabili v razpravni raziskavi – pri podnapisih. V naslednji razpravi avtorica na podlagi obsežnih korpusnih podatkov ugotavlja tendence pri razvrščanju končnice *-je* in končnice *-i* v imenovalniku množine samostalnikov prve moške sklanjatve tipa *študentje* : *študenti*. Raziskava med drugim prinaša spoznanje, da končnico *-je* v veliki večini ohranjajo samostalniki, ki poimenujejo človeško, in samostalniki, ki se končajo na *-ob*, *-of* in *-d*. Razpravni del zaključuje korpusna študija o rabi kratkega in dolgega nedoločnika, ki je potrdila, da se kratki nedoločnik v govornem jeziku redkeje pojavlja v javnih, pogostejše pa v nejavnih govornih položajih, ter da je kratki nedoločnik prevladujoča oblika povsod po Sloveniji razen v Prekmurju. Razpravnemu delu sledi še ocena knjige *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba* ter poročilo o jezikovnotehnološkem izpopolnjevanju učiteljev, ki je potekalo jeseni 2012 v organizaciji Slovenskega društva za jezikovne tehnologije in v sodelovanju s

Filozofsko fakulteto v Ljubljani ter Trojino, zavodom za uporabno slovenistiko. Želimo vam navdihujoče branje.

Urednici:

Polona Gantar, Nataša Logar Berginc

Gantar, P., Logar Berginc, N. (2013): Uvodnik. Slovenščina 2.0, 1 (1): 1–3.

URL: http://www.trojina.org/slovenscina2.0/arhiv/2013/1/Slo2.0_2013_1_01.pdf.