

Simon Krek, Špela Arhar

Slovenski besedilni korpusi: kako v razred?

Povzetek: Prispevek osrednjo pozornost namenja možnosti uporabe besedilnega korpusa oz. korpusnih podatkov pri pouku slovenskega jezika. Korpus v razredu ponuja v prvi vrsti dopolnitev obstoječemu jezikovnemu opisu, saj prinaša za jezikovni pouk nepogrešljive sodobne podatke o realni jezikovni rabi, pa tudi nove možnosti razmišljanja o jeziku: učencu omogoča razumevanje jezika kot sredstva komunikacije, ki se dinamično ter demokratično dogaja med vsemi govorci, samega sebe pa kot člana jezikovne skupnosti, ki ima ob jezikovnem vprašanju pogosto na voljo več možnih odgovorov, med katerimi lahko ustrezno usposobljen samostojno ter odgovorno izbira. Preden pa lahko besedilni korpus v resnici uspešno zaživi v osnovnošolskem ter srednješolskem izobraževanju, je potrebno zagotoviti usposabljanje uporabnikov korpusa (tako učiteljev kot učencev) ter obstoj relevantnih korpusov s pedagoški situaciji ustrezno prilagojeno programsko opremo. V zadnjem delu prispevka opisujeva zasnovo pedagoškega korpusnega vmesnika za slovenski jezik, ki poteka kot ena od aktivnosti projekta Sporazumevanje v slovenskem jeziku.

Ključne besede: besedilni korpus, slovenski jezik, poučevanje slovenščine, pedagoški korpusni vmesnik, sporazumevanje v slovenskem jeziku

UDK: 372.880.863

Izvirni znanstveni prispevek

Simon Krek, Amebis, d.o.o. (Bakovnik 3, 1241 Kamnik) in Institut Jožef Stefan (Jamova 39, 1000 Ljubljana), Slovenija; e-naslov: simon.krek@guest.arnes.si

Dr. Špela Arhar, Trojina, Zavod za uporabno slovenistiko, Partizanska cesta 5, 4220 Škofja Loka, Slovenija; e-naslov: spela.arhar@trojina.si

Uvod

Ena od pomembnih zmožnosti, ki jo mora pridobiti otrok v času šolanja, je t. i. besedilna zmožnost oz. zmožnost razumevanja in tvorjenja besedil. Učni načrti za vse stopnje šolanja med cilji predvidevajo tudi pridobivanje znanj o rabi jezikovnih priročnikov, pri čemer je pri slovenskem jeziku omenjena predvsem raba priročnikov v tiskani obliki oz. takih, ki so bili primarno izdelani za rabo v knjižni obliki in preneseni v elektronsko. V zadnjih dvajsetih letih je razvoj računalništva in informacijsko-komunikacijskih tehnologij tako v jezikoslovju kot tudi v drugih družboslovno-humanističnih znanostih izzval razmeroma obsežne spremembe, ki še niso našle poti v učne načrte in v vsakdanjo prakso jezikovnega pouka. Obsežne zbirke besedil oz. besedilni korpusi so – skupaj s sodobnimi orodji za njihovo raziskovanje – eden od novih jezikovnih priročnikov, ki v pouk jezika potencialno vnašajo demokratičen naboj, ki si ga predtem ni bilo mogoče zamisliti.

Besedilni korpus

Korpus v jezikoslovju

Besedilni korpus v sodobnem jezikoslovju pomeni elektronsko zbirko avtentičnih besedil (pisnih ali transkribiranih govornih) izbranega jezika oz. jezikov. Korpusna besedila so pred vključitvijo v zbirko računalniško obdelana na način, ki omogoča napredno izrabo zajetih podatkov: besedila so shranjena v enotnem formatu, ki opredeljuje njihovo notranjo strukturo, če obstaja možnost (avtomatskega ali ročnega) označevanja pa zapis vključuje tudi jezikoslovne oznake posameznih elementov besedil.

Če govorimo o besedilnem korpusu na splošno, imamo pred očmi *referenčni korpus*. V primerjavi z drugimi vrstami korpusov ga odlikujeta predvsem velik obseg ter natančno domišljena metoda gradnje, ki se odraža v uravnoteženih razmerjih med zajetimi besedili (Gorjanc 2005, str. 8). Referenčni korpus je namenjen preučevanju t. i. jezika vsakdanje rabe in se npr. v aplikativnem jezikoslovju

uporablja za pridobivanje podatkov, ki so osnova za sestavljanje slovarjev, slovnice ter podobnih, uporabnikom jezika namenjenih podatkovnih virov. Temu ustrezno so v referenčne korpusne večinoma zajeta (pisna) besedila časopisov, revij, najbolj branih knjig, internetnih strani itd., vedno pogosteje pa tudi uravnotežen nabor transkribiranih govornih besedil različnih vrst.¹

Besedilni korpus v jezikoslovju brez dvoma predstavlja metodološko revolucijo, ki z novimi možnostmi statistične obdelave velike količine jezikovnih podatkov spreminja pogled na jezik s premikom interesa z vprašanja, kaj je v jeziku *mogoče*, do vprašanja, kaj je v jeziku *tipično*. Pomemben doprinos prinaša korpusno jezikoslovje tudi na področju obravnave pomena. Problem besedne dvoumnosti, ki izvira iz dojemanja jezikovnega znaka kot nerazdružljive vsote oblike ter (potencialno več) pomenov, postane nerelevanten s širitvijo pogleda na besedilno okolico obravnavane besede v jezikovni rabi: beseda je obravnavana v sklopu vzorcev, v katerih se tipično pojavlja, v ospredje obravnave stopi sopojavljanje jedrne besede z drugimi besedami ter s skladenjskimi vzorci. Pomen se tvori v besedilu, komunikaciji, diskurzu, zato ga je v komunikaciji oz. diskurzu tudi potrebno raziskovati, pri čemer je korpusni pristop nepogrešljiv.

Korpus za širšo rabo

Korpusi niso uporabno orodje le za jezikoslovce, pač pa tudi za številne druge uporabnike, ki se pri delu ali v prostem času srečujejo z jezikovnimi vprašanji – ožje gledano npr. novinarje, prevajalce, lektorje, učitelje slovenskega jezika itd., širše gledano pa pravzaprav za vse govorce določenega jezika. Tako je tudi referenčni korpus za slovenščino, korpus FidaPLUS, na spletni strani www.fidaplus.net prosto dostopen za uporabo, skupaj s programom Konkordančnik ASP32, pripravljenim posebej za raziskovanje slovenščine. Zmogljivosti programskega orodja ter tehnike uporabe korpusa FidaPLUS so predstavljene v priročniku za učenje dela s korpusom, ki je prav tako prosto dostopen na omenjeni internetni strani (Arhar 2007).

Kljub navdušenju korpusnih jezikoslovcev nad uporabnostjo novih orodij pa kaže, da je interes širše javnosti za vključitev besedilnih korpusov med pogosto uporabljane jezikovne vire manjši, kot bi pričakovali. Razloge gre iskati predvsem v zasnovi konkordančnih orodij, ki so zaradi svoje specializiranosti uporabnikom neintuitivni ter posledično neprijazni (Kosem 2008). V tem prispevku se ukvarjava z vprašanjem možnosti uporabe korpusnih orodij (oz. splošneje korpusnih podatkov)² pri poučevanju slovenskega jezika. Kar se tiče priprave orodij, zahteva pedagoško delo povsem posebno obravnavo: tako s stališča gradnje korpusa kot tudi razvoja korpusnega vmesnika z vsemi ustrezajočimi programskimi funkcijami. Zasnova

¹ Besedilna struktura najnovejšega slovenskega referenčnega korpusa, ki je trenutno v fazi zbiranja besedil, je predstavljena v Logar Berginc in Šuster (2009).

² V prispevku ločujeva (I) uporabo korpusa od (II) uporabe korpusnih podatkov v razredu. Če prvo predvideva konkretno delo učenca (oz. učitelja) s konkordančnikom v času pouka, drugo zajema vrsto drugih možnosti, kako se lahko učenci s korpusnimi podatki pri pouku srečujejo, denimo na osnovi učnih gradiv, ki jih učitelj doma pripravi s pomočjo korpusa ter jih učencem predstavi npr. v obliki učnih listov, vaj, testov itd.

pedagoškega vmesnika, namenjenega specifično šolajoči se populaciji, v našem prostoru že poteka, predstavlja jo v poglavju Pedagoški korpusni vmesnik.

Korpus v razredu

Predvsem v angleškem prostoru se že več kot desetletje piše o možnostih uporabe korpusa oz. korpusnih podatkov v razredu, primarno seveda pri poučevanju angleščine, pri čemer je sicer večina raziskav usmerjenih v poučevanje angleščine kot drugega oz. tujega jezika.³ Tako npr. Mukherjee (2006) izpostavi tri možne načine uporabe korpusnih podatkov za pedagoške namene: (I) uporaba korpusa pri pripravi gradiva za poučevanje jezika (priročnikov, učbenikov itd.), (II) uporaba korpusa pri samem jezikovnem pouku ter (III) gradnja in analiza korpusov besedil učeče se populacije za evidentiranje močnih ter šibkih mest njihove jezikovne produkcije (prav tam). Prispevek se, kot rečeno, osredotoča na drugo od naštetih točk, kratka predstavitev v našem prostoru trenutno potekajočih korpusnojezikoslovnih aktivnosti, ki so povezane s prvo ter tretjo od naštetih točk, pa sledi v poglavju Sporazumevanje v slovenskem jeziku.

Zagovorniki vključevanja uporabe besedilnega korpusa oz. korpusnih podatkov v jezikovni pouk navajajo številne didaktične prednosti takšne metode: korpus prinaša primere realne jezikovne rabe, skupaj z informacijo o tem, kaj je v jeziku pogosto oz. tipično – česar v učbenikih vsaj trenutno ni najti. Samostojno preučevanje ter interpretacija korpusnih podatkov ponujata učencu možnost induktivnega usvajanja znanja, spodbujata tako razvoj splošnega jezikovnega občutka kot uzaveščenje specifičnih jezikovnih značilnosti ter struktur. Delo s korpusom krepi učenčevo avtonomijo pri raziskovanju jezika, analiza primerov realnega jezika pa razvija sposobnosti kritičnega mišljenja v povezavi z jezikoslovnim znanjem. Večja samostojnost pri raziskovanju jezika pozitivno vpliva na motivacijo za delo ter posledično učno uspešnost. Vključitev korpusa v razred, nenazadnje, usposablja učence za uporabo novih tehnologij pri iskanju odgovorov na jezikovna vprašanja (Gut 2006, str. 69–70).

Preusmeritev od deduktivnih metod poučevanja jezika k induktivnim se odraža v spremenjeni vlogi tako učitelja, ki namesto podajalca pravil zavzame vlogo voditelja, usmerjevalca raziskovanja, kot učenca, ki ob aktivnem preučevanju jezikovnih podatkov razvija metakognicijo o usvajanju jezika, pa tudi v vlogi slovnice kot sistema, ki je svoji klasični obliki pogosto predstavljen preveč abstraktno, da bi v pedagoški situaciji lahko bil resnično učinkovit (Bernardini 2004, str. 17).

Ob vseh naštetih prednostih pa je mogoče ugotoviti, da ostaja vključevanje korpusa v učni proces – že predolgo – zgolj na ravni posameznih poskusov, čemur botrujejo omejene možnosti dejanske uporabe korpusnih podatkov v razredu. Breyer (2006) navaja, da bi morali biti za uspešno uporabo korpusa pri jezikovnem pouku izpolnjeni trije temeljni pogoji: (I) zagotovitev kvalitetnega in kontinuiranega usposabljanja uporabnikov korpusov (tako učiteljev kot učencev), (II) obstoj pedagoško relevantnih korpusov ter (III) pedagoški situaciji ustrezno prilagojena programska oprema za delo s korpusom.

³ Glej npr. Burnard in McEnery 2000; Aston 2001; Aston, Bernardini in Stewart 2004; Hidalgo, Quereda in Santana 2007; Braun, Kohn in Mukherjee 2006; Sinclair 2004; Aijmer 2009.

Tako kot angleška tudi sedanja slovenska situacija izkazuje vrzeli pri vseh naštetih točkah: organizirana usposabljanja učiteljev za delo s korpusi so redka, obstoječi korpusi so, skupaj z vmesniki, za pedagoško rabo preveč specializirani. Posledično je tudi raba korpusa za pedagoške potrebe na ravni osnovnih ter srednjih šol v našem prostoru prej izjema kot pravilo.

Ugotovitev potrjuje analiza uporabe referenčnega korpusa FidaPLUS, ki je potekala v sklopu zasnove novih vmesnikov, o čemer bo več govora v nadaljevanju prispevka.⁴ 44 od 260 uporabnikov (16,9 %) je med vnaprej podanimi izbralo odgovor *korpus uporabljam za pripravo učnega gradiva*. Ker pa vprašanje ni nadalje razdelano, ni jasno, kolikšen delež znotraj navedenega vzorca predstavljajo osnovnošolski ter srednješolski učitelji, prav tako se ne ve, kako pogosta je omenjena aktivnost v primerjavi s potencialnimi drugimi nameni korpusne uporabe posameznega uporabnika. Nekoliko jasnejša je slika pri odgovoru *korpus uporabljam za reševanje domačih nalog*, ki ga je sicer izbralo 37 uporabnikov (14,2 %), vendar se v celotnem vzorcu vprašanih nihče ni opredelil za osnovnošolca oz. srednješolca, iz česar lahko sklepamo, da gre v tem primeru za izključno študentsko populacijo. Iz podatkov je torej mogoče zaključiti, da se referenčni korpus pri poučevanju uporablja predvsem na stopnji študija, čeprav ni izključeno, da ga uporabljajo tudi nekateri učitelji osnovnih oz. srednjih šol za pripravo gradiva.

Sporazumevanje v slovenskem jeziku

Projekt Sporazumevanje v slovenskem jeziku (v nadaljevanju: SSJ) poteka pod okriljem Ministrstva za šolstvo in šport⁵ in kot tak prinaša številne aktivnosti, osredotočene na gradnjo kvalitetnih, sodobnih ter prosto dostopnih jezikovnih virov za slovenščino: med cilji projekta sta denimo (I) pedagoška slovnica ter (II) slogovni priročnik za pomoč pri reševanju tipičnih zadreg pri pisanju, v postopku priprave pa je tudi (III) leksikalna baza, ki predstavlja osnovo za sestavljanje sodobnega slovarja. Vsi našteti jezikovni viri bodo pripravljene z najsodobnejšimi metodami ter dosledno na osnovi korpusnih, torej realnojezikovnih podatkov.

Osrednji produkt projekta bo novi referenčni korpus, ki bo prvič za slovenščino prinašal tudi uravnoteženi nabor govornih besedil. V zvezi s tem poteka v sklopu projekta tudi izboljšava avtomatskega označevanja slovenskih besedil ter priprava novega konkordančnika za delo s korpusom. Slednji bo na voljo v dveh oblikah: (I) t. i. splošni vmesnik bo namesto Konkordančnika ASP32 na internetu na voljo vsem zainteresiranim uporabnikom, (II) pedagoški vmesnik pa bo različica, namenjena šolajoči se populaciji.⁶

⁴ Več o zasnovi vprašalnika ter vmesnih rezultatih anketiranja je na voljo v Arhar 2009. V priču-jočem prispevku predstavljeni rezultati so z dne 4. 1. 2010.

⁵ »Operacijo delno financira Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za šolstvo in šport. Operacija se izvaja v okviru Operativnega programa razvoja človeških virov za obdobje 2007–2013, razvojne prioritete: razvoj človeških virov in vseživljenjskega učenja; prednostne usmeritve: izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja 2007–2013.« (Dostopno na: www.slovenscina.eu)

⁶ Posamezne aktivnosti projekta so natančneje predstavljene v prispevkih, ki so izšli v tematski številki revije Jezik in slovstvo (Gantar 2009).

Od projektnih aktivnosti velja na tem mestu omeniti še pripravo korpusa besedil šolajoče se populacije, delovno imenovanega KUS (korpus usvajanja slovenščine). V korpus bodo vključeni pisni izdelki osnovnošolcev zadnje triade ter srednješolcev, v prvi vrsti besedila, ki nastajajo pri samem pouku (npr. spisi, eseji, odgovori na vprašanja odprtega tipa pri različnih predmetih itd.). Korpus bo v sklopu projekta analiziran z namenom identifikacije problemov, ki jih imajo učenci s pisno produkcijo, pridobljeni podatki pa se bodo uporabili pri pripravi zgoraj naštetih jezikovnih virov.

Korpus in opismenjevanje

Izzivi opismenjevanja v slovenščini

Eden glavnih izzivov opismenjevanja v slovenščini izvira iz dejstva, da je razvoj komunikacijske zmožnosti v slovenščini pogojen tako z usvajanjem zakonitosti standardnega jezika v primerjavi z najprej usvojenim »maternim narečjem«, kot tudi z razvijanjem sposobnosti za odločanje, v katerih komunikacijskih situacijah uporabiti katero od omenjenih dveh jezikovnih izbir. Kot posledica se pri govoricah pojavlja diskrepanca med dojemanjem govora okolice kot poznanega, bližnjega, nerestriktivnega, obenem pa neprimernega za številne komunikacijske situacije, na drugi strani pa se standardni jezik oz. njegovo obvladanje dojema kot vrednota, ki pa se zdi – tudi zaradi mesta, ki ga v slovenski kulturi zaseda jezikovni predpis – v marsikaterem segmentu nedosegljiva.

Izkušnje kažejo, da se skrb pišočih ob tvorjenju besedila osredotoča na vprašanje, kaj je v jeziku dovoljeno oz. prepovedano (»kako se to napiše?«, »ali smem uporabiti to besedo?«, »je to sploh beseda?« itd.), pri čemer na vprašanja v obstoječih jezikovnih priročnikih, če po njih sploh posežejo, pogosto ne najdejo odgovorov – ali pa si podatke zaradi zahtevnega načina predstavitve neustrezno interpretirajo (Gantar in Krek 2009). V slovenskem prostoru so trenutno uporabnikom na voljo jezikovni priročniki, ki opredeljujejo *prav* ter *narobe*, malo prostora pa je v njih namenjenega vprašanju *kako* in *zakaj*. Opismenjevanje kot eden izmed učnih ciljev pouka slovenščine je še toliko težja naloga, ker se vrzel med sliko jezika, ki jo prikazujejo priročniki, ter sodobnim jezikom, v katerem živimo, z vsako novo generacijo šolajočih povečuje.

Uporaba besedilnega korpusa v razredu ponuja dopolnitev obstoječemu jezikovnemu opisu, čeprav slovarjev ter slovníc kot takih seveda ne more (in ne poskuša) nadomestiti. Poleg sodobnosti jezikovnih podatkov⁷ – ti so za uspešno opismenjevanje vsekakor nepogrešljivi – prinaša delo s korpusom predvsem *nove možnosti razmišljanja* o jeziku. S korpusom je mogoče preučevati tako primere rabe, ki sledijo jezikovnemu predpisu, kot tudi tiste, ki od predpisa odstopajo, obenem pa opazovati vzroke ter posledice takšnih odstopov. Veliko število na enem mestu zbranih primerov jezikovne rabe ponuja vpogled v odločitve številnih tvorcev

⁷ FidaPLUS prinaša denimo besedila, nastala med leti 1990 in 2006 (Arhar 2007), novi referenčni korpus bo zajemal besedila nastala med 1995 in 2010 (Logar Berginc in Šuster 2009.)

besedil in s tem omogoča razumevanje jezika kot sistema, ki se dinamično dogaja med vsemi govorci ter kot tak ne more biti uravnavan na osnovi introspekcijskih zmožnosti enega ali dveh posameznikov (Bernardini 2004, str. 30) – naj gre za avtorja jezikovnega priročnika, učitelja jezika, lektorja ali kako drugo avtoriteto na področju jezikovne norme.

Namen uporabe korpusa v razredu seveda ni – kot bi bilo mogoče iz zapisa-nega napačno sklepati – zmanjševati pomen pravil oz. usvajanja pravil uporabe jezika na poti razvijanja komunikacijske zmožnosti, pač pa predvsem spremeniti odnos učenca do lastne jezikovne produkcije. Ključno pri tem je spoznanje, da pri jezikovnih odločitvah večinoma ne gre za premike po črno-belih poljih pravilnega ter napačnega, pač pa da slednje prinašajo spekter sivin – da je torej posamezna izbira lahko bolj/manj v skladu z normo, obenem pa tudi bolj/manj tipična za določen tip diskurza. Prepoznava teh sivin omogoča učencu, da se ob srečanju z jezikovnim vprašanjem odloči za rešitev, ki se mu zdi za specifično situacijo najbolj ustrezna.

Opismenjevanje v slovenščini s pomočjo besedilnega korpusa naj bi torej vodilo k naslednjim ciljem: (I) ob srečanju z jezikovnim vprašanjem je učenec usposobljen poiskati jezikovne informacije različnih vrst, (II) učenec je usposobljen jezikovne podatke, ki so mu na voljo, interpretirati v skladu z ustrojem ter namenom jezikovnega vira, (III) učenec je usposobljen na osnovi podatkov pridobiti odgovor na svoje vprašanje, pri čemer se zaveda, *kako* in *zakaj* se je odločil za določeno izbiro, in je posledično sposoben (IV) svojo odločitev ustrezno argumentirati. Šele našteje usposobljenosti lahko vzpostavijo pogoje za razvoj samostojnih ter odgovornih članov jezikovne skupnosti, o kakršnih piše Stabej (2009): »Če so [govorci in govorke] (še posebej v javnem in poklicnem sporazumevanju) samostojni in odgovorni do sebe, svojih besedil in sporočil, sogovorcev, bralcev in še do česa, potem bodo slovarske [oz. jeziko(slo)vne nasplošno, op. Arhar in Krek] informacije še kako znali in hoteli uporabljati. V jezikovni skupnosti, kjer obstaja močno jezikovno skrbništvo v obliki lektorstva in drugih jezikovnih intervencij, kjer je pojmovanje jezikovne kulture povezano pretežno s (pre)občutljivostjo na površinske (pravopisne, oblikoslovne) in ideološke (domačnost/tujost) prvine diskurza, je pot do samostojnega in odgovornega govorca (oziroma pisca) še dolga.« (Prav tam, str. 124)

Povzeti je torej mogoče: učenje jezika je namenjeno pridobivanju komunikacijske zmožnosti, tj. védenja o tem, kako jezik ustrezno uporabljati v različnih komunikacijskih situacijah. Korpus prinaša primere realne rabe jezika (dejanske rabe jezika v realnih situacijah), zato je smiselna vključitev uporabe teh podatkov v povezavi s čim več aspekti razvijanja učenčeve komunikacijske zmožnosti (Braun 2006, str. 26). Z drugimi besedami: »Ker korpusi zagotavljajo dostop do avtentičnih jezikovnih interakcij (tako pisnih kot govornih, monoloških ali dialoških), so idealno orodje za opazovanje in usvajanje družbeno vzpostavljenih parov oblika/pomen. Z drugimi besedami, učencem omogočijo opazovanje tega, *kaj* tipično povemo v danih okoliščinah in *kako* to povemo, ter povezovanje obojega.« (Bernardini 2004, str. 17–18)

Korpus in učitelj

Ena glavnih ugotovitev korpusnega jezikoslovja je, da statistična obdelava velikega števila primerov dejanske jezikovne rabe pogosto daje rezultate, ki niso v skladu z posameznikovo jezikovno intuicijo.⁸ Razlog za to gre deloma iskati v dejstvu, da si ob recepciji jezika govorniki vtisnemo v spomin pojavnosti, ki se v specifični situaciji zdijo nepričakovane, ne ustavljamo pa se zavestno pri tistem, kar je vsakdanje: takšen način obravnave jezikovnih informacij bistveno pospeši njihovo procesiranje, kar omogoča uspešno komunikacijo, vendar pa na drugi strani vpliva na sposobnost posameznika, da realno oceni, kakšno je razmerje med jezikovnimi pojavi v rabi: »[Jezikovni opisi, temelječi na introspekciji] opisujejo, kaj ljudje o jeziku vedo oz. kako ga *dojemajo*, ne pa tega, kako jezik *uporablja*o.« (Tsui 2004, str. 39)

Omenjeno dejstvo je potrebno imeti pred očmi predvsem v primerih, ko je govorec jezika (po službeni ali kaki drugi dolžnosti) dolžan s podatki o jezikovni rabi razpolagati in jih uporabljati – kar za učitelja jezika vsekakor drži. Učitelj slovenščine projicira svoj občutek o jezikovni rabi na učence z izbiro slovnične tematike, ki ji posveča več pozornosti, jezikovnih virov, h katerim napotuje učence ob jezikovnih vprašanjih, s posameznimi besedilnimi gradivi, ki jih uporablja, vse do specifičnih primerov jezikovnih pojavov, ki jih ob obravnavi izpostavlja v obliki zgledov.

S tega stališča je zanimiva npr. problematika gradiv, ki se pri jezikovnem pouku uporabljajo veliko pogosteje oz. neposredneje kot jezikovni priročniki, o katerih je bilo govora v prejšnjem poglavju, tj. učbenikov, učnih listov, e-gradiv itd. V razpravo o slednjih se v pričujočem prispevku ne podajava, navajava pa v razmislek: analize učbeniškega gradiva za pouk jezika konsistentno izkazujejo, da se med temami, ki sovpadajo s korpusno sliko o tem, kaj je v jeziku pogosto, povsem enakovredno pojavljajo tudi teme, ki preučujejo obrobne, v jeziku redke pojave (za angleški prostor jih povzema denimo Tsui 2004; Conrad 2004). V okolju, kjer je čas dragocen, se vsakršno osredotočanje na manj relevantno zdi vsekakor nesmiselno. Kako doseči spremembo zornega kota pri tem segmentu izobraževanja, pa je že novo vprašanje.

Omeniti je potrebno še: avtorji poročajo, da so od vseh lastnosti jezika učiteljem jezikovnega pouka med najbolj zoprnimi jezikovne variante (Sinclair 2004; Conrad 2004). Od učitelja se pričakuje, da bo posredoval učencem jasne smernice glede jezikovne rabe in jim hkrati na logičen način predstavil izjeme k pravilom, pri tem pa naj bi se spretno gibal med jezikovnim predpisom, ki zaradi potrebe po generalizaciji prinaša za konkretno situacijo pogosto pomanjkljive informacije, ter jezikovno realnostjo, ki lahko s svojimi številnimi *kaj pa če* v učencu zbudi občutek neobvladljivosti ter posledično nemoči. Korpusni podatki se v tovrstnih situacijah izkažejo za več kot dobrodošle, saj na enem mestu ponujajo dvoje: (I) nabor možnosti različnih izbir pri določenem jezikovnem vprašanju ponuja relativizacijo jezikovnega predpisa, ki pogosto (eksplicitno ali implicitno) navaja kot

⁸ Ob čemer je potrebno poudariti, da na drugi strani korpusni pristop mnoge jezikoslovne ugotovitve, ki izvirajo iz časa pred možnostjo uporabe korpusov, potrjuje.

ustrezno eno samo ter (II) podatki o viru, sobesedilu ter pogostnosti pojavitve teh izbir, ki pričajo o tem, da izbira med variantami ni arbitrarna.

Kakorkoli dolg je že niz argumentov za vpeljavo besedilnega korpusa v pouk slovenskega jezika, v vsakem primeru je predpogoj usposabljanje učiteljev za delo z novim jezikovnim virom. Prvi korak na tej poti bo opravljen, ko bo za slovenščino na voljo uporabniku prijazen ter intuitiven korpusni vmesnik, dela s katerim se bo mogoče hitro ter enostavno privaditi. Lahek dostop do korpusnih podatkov bo učiteljem omogočil primerjanje generaliziranih pravil z dejanskimi jezikovnimi pojavitvami, identifikacijo tipično uporabljanih struktur oz. vzorcev za izbrane jezikovne pojave, na voljo bo ogromno število primerov, orodja za organizacijo podatkov na različne za jezikovni pouk uporabne načine itd. (Granath 2009). Znanje uporabe korpusnih podatkov učiteljem torej ponuja možnost širitve lastnega pogleda na jezik, izpopolnjevanja jezikovnega občutka ter védenja o jeziku, obenem pa predstavlja korpus zmogljivo orodje za izdelavo kvalitetnih, sodobnih učnih gradiv za poučevanje jezika.

Kaj pa računalniško opismenjevanje?

V pričujočem poglavju se ne lotevava diskusije o različnih vrstah opismenjevanja, kakor bi bilo mogoče sklepati iz naslova, ampak na izbranem primeru predstavlja eno od možnosti uporabe korpusnih podatkov, tj. za dopolnitev informacij, ki so o jezikovni rabi na voljo v obstoječih jezikovnih priročnikih slovarskega tipa.

Iskanje besede *opismenjevanje* v Slovarju slovenskega knjižnega jezika (v nadaljevanju: SSKJ) ne ponudi nobenega rezultata,⁹ zato pa je kratka razlaga na voljo pri izvornem glagolu:

opismenjevatí -újem nedov. (á û) publ. *učiti pisati in brati*: opismenjevatí nepismene

Primer 1: Glagol »opismenjevatí« v SSKJ.

Podobna slika se kaže pri istem geslu v slovarju Slovenskega pravopisa (v nadaljevanju: SP), kjer je glagolnik – kakor je sicer navada pri tem slovarju – zabeležen znotraj glagolskega gesla¹⁰:

opismenjevatí -újem nedov. -ujóč, -eváje; -ál -ála, -át, -án -ána; opismenjevánje; (-át) (á û) publ. |učiti pisati in brati|: *koga ~ odrasle*

Primer 2: Glagol »opismenjevatí« v SP.

⁹ Izkušnje kažejo, da premnogi uporabniki slovarja ob tovrstni izkušnji žal še vedno sklepajo, da beseda posledično »ne obstaja« oz. je potemtakem »ni dovoljeno uporabiti«. Neobstoj v jezikovnem viru kot argument za neobstoj v jeziku se je prenesel tudi na korpuse, kar je – kljub njihovi obsežnosti ter reprezentativnosti – seveda nesmisel, saj je nemogoče pričakovati od kakršnegakoli vira, da bo v njem zajet oz. predstavljen jezik v celoti. Prav zato je ključno, da se govorniki jezika naučijo predvsem *interpretacije* jezikovnih podatkov, ki so jim na voljo, saj je, kot pravi Sinclair (2004) »iz dokazov enako lahko izpeljati domiselne kot nesmiselne zaključke.« (prav tam, str. 2)

¹⁰ V SP je najti tudi *opisméniti, opisménjen / opisménjenost ter opismenjeválen / opismenjeválni*.

Informacije, ki jih ponujata slovarja, lahko dopolnimo s korpusnimi podatki, npr. z naborom kolokatorjev obravnavane besede. Spodnja tabela prinaša besede, ki se v korpusu FidaPLUS statistično relevantno pojavljajo levo oz. desno od glagolnika *opismenjevanje* (ta se v korpusu pojavlja s 1331 zadetki). V obravnavo so zajete besede do treh mest levo oz. desno, uporabljena statistika pa je logaritem verjetnosti. Navedenih je prvih 15 primerov glede na statistično vrednost (ta je navedena v robnih stolpcih tabele poleg vsakega kolokatorja):

Statistična vrednost	Besede, ki se tipično pojavljajo LEVO od jedra	Jedro	Besede, ki se tipično pojavljajo DESNO od jedra	Statistična vrednost
4.249.649.143	računalniški		RO	252.343.713
1.437.605.883	računalniško		ro	168.552.924
612.251.616	začetni	opisme-	odrasla	167.489.544
551.080.655	plavalen	njevanje	otrok	105.775.344
534.197.150	program		šola	100.500.150
169.532.201	informatijski		učenec	99.463.985
163.298.124	projekt		branje	65.223.969
146.224.452	sistematičen		slaboviden	60.885.437
139.560.216	okvir		osnovnošolec	57.675.560
119.075.555	proces		slep	46.741.396
106.476.890	metoda		mlada	44.472.109
96.174.544	funkcionalen		prebujanje	43.192.368
94.657.652	medijski		Braillov	40.031.921
88.319.508	tečaj		Rom	39.267.837
76.527.285	informatijsko		učitelj	38.124.109

Primer 3: Nabor kolokatorjev za »opismenjevanje« v korpusu FidaPLUS.

Za boljšo preglednost je mogoče po natančnejšem pregledu korpusnih zadetkov¹¹ gornje primere razvrstiti v skupine glede na skladišne vzorce, v katerih se jedrna beseda ter kolokator pojavljata. Na ta način se nekoliko približamo zapisom, ki smo jih vajeni iz priročnikov slovarskega tipa:

¹¹ Slednji je na poti do ustrezne interpretacije podatkov v vsakem primeru nujen. V obravnavanem primeru npr. analiza podatkov denimo pokaže nekatere probleme na ravni avtomatskega pripisovanja osnovne oblike besedam – *računalniško*, *informatijsko* namesto *računalniški*, *informatijski* ter *odrasla*, *mlada* namesto *odrasel*, *mlad*. Poleg tega ugotovimo, da se beseda *prebujanje* med kolokatorji znajde zaradi t.i. korpusnega šuma, velike količine delno ponovljenih zadetkov iz istega vira. Prav tako je šele iz primerov mogoče identificirati morebitne širše skladišne vzorce, v katerih jedrna beseda nastopa – čeprav je v tabeli kolokatorjev navedena beseda *okvir*, se slednja redko pojavlja v zvezi *okvir opismenjevanja*, pač pa gre večinoma za *okvir programa / projekta / razvoja opismenjevanja*.

računalniško, začetno, plavalno, informacijsko, sistematično, funkcionalno, medijsko

opismenjevanje

program, projekt, proces, metoda, tečaj **opismenjevanja**

okvir (programa, projekta, razvoja) **opismenjevanja**

opismenjevanje odraslih, otrok, učencev, slepih in slabovidnih, osnovnošolcev, mladih, Romov, učiteljev

opismenjevanje v šolah, v Braillovi pisavi

opismenjevanje in branje

računalniško opismenjevanje (RO)

Primer 4: Tipične besedne zveze s samostalnikom »opismenjevanje«.

Alternativno informacijam, ki jih o rabi besede *opismenjevanje* SSKJ prinaša na osnovi podatkov iz časa svojega nastanka, kolokacije iz korpusa FidaPLUS realneje odražajo sodobno družbeno stvarnost. Za zveze, kot denimo *računalniško, plavalno, informacijsko, medijsko opismenjevanje*, razlaga »učiti brati in pisati« ne zadostuje, saj izkazujejo pomenski premik v smer »učiti nabora osnovnih veščin določene aktivnosti«. *Opismenjevanje* samo na sebi sicer tudi v korpusu prinaša pomen učenja branja ter pisanja, vendar pa je iz podatkov obenem razvidno, da beseda najpogosteje nastopa v besednih zvezah tipa *računalniško opismenjevanje*, tudi npr. pri primerih *program / projekt računalniškega opismenjevanja, računalniško opismenjevanje odraslih / v šoli* itd.

Pa še odgovor na v naslovu zastavljeno vprašanje: da je računalniško opismenjevanje v sodobnem času resnično vroča tema, priča tudi podatek, da se zveza v korpusu FidaPLUS pojavlja s 490 primeri (ko je govora o opismenjevanju, gre torej v 36 % za eksplicitno navedbo zveze *računalniško opismenjevanje*). *Opismenjevanje odraslih*, ki ga za primer podaja SP, se v korpusu pojavi z 28 primeri, od tega 8-krat kot *računalniško opismenjevanje odraslih*. Zveza *opismenjevanje nepismenih*, ki jo kot primer rabe podaja SSKJ, pa se – vsaj v primerljivi besednozvezni obliki – sploh ne pojavlja, najbrž zato, ker je pravzaprav pleonastična.

Pedagoški korpusni vmesnik

Zasnova vmesnika

Pričujoče poglavje se osredotoča na zasnovo pedagoškega korpusnega vmesnika, ki je, kot rečeno, ena od aktivnosti projekta Sporazumevanje v slovenskem jeziku. Pedagoški vmesnik je namenjen povsem specifični populaciji – ciljni uporabniki so osnovnošolci višjih razredov ter deloma srednješolci – in njegov razvoj poteka ločeno od razvoja korpusnega vmesnika za splošno rabo, vseeno pa sta aplikaciji v mnogih točkah sorodni, saj sta vezani na funkcionalnosti istega konkordančnika. Na tem mestu je potrebno opozorilo, da so grafične predstavitve posameznih elementov vmesnika v nadaljevanju poglavja namenjene zgolj ponazoritvi – vmesnika sta v času priprave prispevka v postopku razvoja, zato končne oblikovne rešitve še niso na voljo.

Prvo vodilo pri snovanju pedagoškega vmesnika je *zagotavljanje preglednosti*. O tem priča postavitvev informacij na internetni strani:¹² (I) vsi osnovni elementi vmesnika so prikazani na isti strani – iskalno okence, tabela s konkordančnim nizom, kolokacijska slika – tako da preklapljanja med stranmi s podatki pravzaprav ni; (II) vsa pomoč, ki je učencu pri delu na voljo, je dosegljiva s klikom z osnovne strani, pri čemer povezave nosijo povedna imena, npr. *kako lahko iščem, kaj lahko počnem s primeri* itd.; (III) tako pomoč kot vse ostale povezave se odprejo v obliki premičnega okenca, ki se pojavi na osnovni strani (učenec lahko okence po želji premika po zaslону oz. zapre, ko pomoči ne potrebuje več). Navigacija po straneh je sicer potrebna pri pregledovanju konkordančnega niza, vendar se pri tem (IV) spreminja le nabor korpusnih zadetkov, medtem ko preostala struktura osnovne strani ostaja enaka.

Druga pomembna točka razvoja vmesnika je *zagotavljanje razumljivosti*. S tega stališča je veliko pozornosti namenjene jeziku vmesnika, naj gre za naslove vmesniških elementov ter povezav na dodatne informacije, navodila za uporabo programa oz. vmesniško pomoč itd. Jezikoslovna terminologija, ki se sicer pri tovrstnih programih uporablja na vseh ravneh vmesnika, se pri pedagoškem nadomešču s popreprostenim opisom (npr. *oglej si, s katerimi besedami se tvoja beseda v jeziku družu namesto oglej si kolokatorje jedrne besede* ipd.).

Iskalni pogoj

Izdelava ustreznega iskalnega pogoja je predpogoj za dostop do korpusnih podatkov, zato je pri pedagoškem vmesniku še zlasti pomembno, da je iskanje zasnovano čim bolj logično ter preprosto. Vmesnik v osnovi ponuja dve iskalni možnosti, (I) iskanje posamezne besede ter (II) iskanje besedne zveze oz. besednega niza. Ob iskanju se vnesenim besedam avtomatsko pripiše osnovna oblika,¹³ v primeru dvoumnosti pa program učencu v drugem koraku ponudi nabor možnih iskalnih pogojev, med katerimi lahko izbira.¹⁴

Učenec ima pri izdelavi iskalnega pogoja na voljo tako pomoč v obliki kratkih navodil (s povednim naslovom povezave *Kako lahko iščem?*) kot tudi dveh »bližnjic« ob iskalnem okencu, ki omogočata dostop do začetnega primera korpusne rabe tudi učencem, ki terminov *beseda* ter *besedna zveza* ne poznajo (primera obenem implicitno razlagata pomen terminov):

Poizkusi poiskati:

- besedo (klikni primer *metulj*),
- besedno zvezo (klikni primer *nočni metulj*).

Primer 5: Bližnjice za iskanje pri pedagoškem vmesniku.

¹² Pedagoški korpus bo za uporabo prosto dostopen na internetu, zato je na ta medij vezana tudi zasnova vmesnika.

¹³ Na ta način je iskanje omogočeno tudi v primeru, da učenec ne pozna osnovne (slovarske) oblike za besedo, ki jo želi iskati. V primeru, da učenec želi točno tisto besedno obliko, ki jo je vnesel, lahko uporabi iskanje z narekovaji, npr. »*metulja*«.

¹⁴ V primeru iskanja besednega niza *metulj leti* npr. program izdela konkordančni niz z vsemi primeri, ki ustrezajo danemu pogoju v najširšem smislu, obenem pa ponudi učencu možnost omejitve nabora zadetkov s klikom na katero od navedenih možnosti: *metulj leteti, metulj leto, metulj let*.

Konkordančni niz

Osnovni namen konkordančnika je izdelava konkordančnega niza, tj. nabora besedilnih fragmentov z iskalnim nizom v jedru ter minimalnim sobesedilom levo ter desno (konkordanc). Konkordančni niz je grafično prikazan v obliki pregledne, večbarvne tabele, ki jo uvaja spremno navodilo učencu: *Oglej si primere, v katerih nastopa tvoja beseda ali besedna zveza*. Neposredno ob konkordančnem nizu je na voljo tudi povezava na pomoč (*Kaj lahko počnem s primeri?*), ki učenca nauči, kako lahko dostopa do širšega sobesedila posamezne konkordance, kako lahko primere prenese v urejevalnik besedil za nadaljnjo obravnavo oz. kako jih lahko natisne.

Konkordance so izpisane v dovolj veliki, berljivi pisavi, predvideno je manjše število konkordanc na stran (okrog 15). Ob konkordančnem nizu je na voljo informacija o celotnem številu vseh zadetkov v korpusu. Posamezna konkordanca prinaša za boljšo preglednost še zaporedno številko ter osnovno informacijo o besedilnem viru primera:

1	časopis	Rdeč kožni izpuščaj na obrazu v obliki	metulja,	ki se še poveča ob izpostavljanju
2	revija	metuljčki in cvetijo rožice. Tudi Trije	metulji	bodo pod ošabnimi rožami iskali
3	časopis	očeta ter na brate in sestre. Oče je lovil	metulje	in se ukvarjal s fotografijo, to pa je
4	časopis	žuželke npr. kačji pastirji, nekateri	metulji,	čebele in čmrlji, mravlje rodu
5	revija	pa ima? Katere igre igra najrajši? Ali zbira	metulje?»	Hoteli bodo vedeti: »Koliko je star
6	revija	v nekaj sekundah naredite bližinski posnetek	metulja,	nato z zoomom približate predmet
7	časopis	plavalka, ki je blestela v kategorijah 100 metrov	metulj	in na 200 metrov prosto, se je

Primer 6: Del konkordančnega niza z jedrom »metulj«.

Klik na konkordanco odpre (na istem zaslonu) okence s širšim sobesedilom, poleg katerega so na enostaven način predstavljene osnovne informacije o besedilnem viru:

Zavarovane so vse vrste dvoživk (žabe, pupki, močeradi) in plazilcev (kače, kuščarice). Zavarovanih je tudi precej sladkovodnih vrst (npr. raki) in morske živali. Prav tako so zavarovane nekatere žuželke npr. kačji pastirji, nekateri **metulji**, čebele in čmrlji, mravlje rodu Formica.

Od kod je to besedilo?
časopis SAVINJSKE NOVICE

Kdaj je bilo to besedilo napisano? leta 1998

Kdo je napisal to besedilo?
neznani novinar oz. novinarka

Primer 7: Širše sobesedilo ter informacije o viru pri pedagoškem vmesniku.

Vmesnik učencu omogoča enostavno navigacijo po straneh konkordančnega niza, prav tako je na voljo zgodovina iskanj, s pomočjo katere učenec po želji preklaplja med konkordančnimi nizi, ki jih je izdelal.

Kolokacijska slika

V primeru, da učenec kot iskalni pogoj vnese (posamezno) polnopomensko besedo, torej samostalnik, glagol, pridevnik ali prislov, se na osnovni strani vmesnika pod konkordančnim nizom izriše kolokacijska slika za izbrano jedro. Sliko uvaja spremno navodilo *Oglej si, s katerimi besedami se tvoja beseda v jeziku druži*, poleg slike je na voljo tudi povezava na dodatne informacije (*Kaj lahko počnem z besedami?*).

Slika prikazuje kolokatorje, ki so privzeto urejeni po skupinah glede na besednovrstno oznako, omogočeno pa je tudi njihovo razvrščanje zgolj glede na statistično relevantnost. Pri izdelavi kolokacijske slike so uporabljeni vnaprej določeni parametri, ki jih učenec ne more spreminjati.¹⁵ Sama vizualna reprezentacija podatkov je lahko različna, v spodnjem primeru so kolokatorji, kot rečeno, razvrščeni v skupine glede na besedno vrsto, velikost pisave pa je vezana na statistično relevantnost kolokacije (višja kot je vrednost, večja je pisava).



Primer 8: Kolokacijska slika za jedro »metulj«.

S klikom na kolokator v kolokacijski sliki se avtomatsko izdelata konkordančni niz s primeri, ki vsebujejo izbrani kolokator skupaj z jedrno besedo. Na ta način je učencu omogočen hiter dostop do informacij, ki jih potrebuje za ustrezno interpretacijo kolokacijske slike:

¹⁵ Izbrani bodo parametri, ki bodo ob testiranju dali najboljše rezultate. V prispevku je seznam kolokatorjev izdelan s statistiko logaritma verjetnosti, za okno do vključno tri besede levo ter desno od jedra. Spreminjanje parametrov v pedagoškem vmesniku ne bo omogočeno, bo pa ta možnost na voljo v splošni različici vmesnika.

1	časopis	mraz težko pride do živega. Spomladi se iz GOSENIC razvijejo metulji , ki na novih posevkih
2	knjiga	strani poslikana z ogromnimi sonci, metulji in GOSENICAMI . Ustavil sem avto na ovinku, kjer se
3	revija	hranijo nekateri redki in ogroženi metulji oziroma njihove GOSENICE . Vlažni travniki in svetli
4	revija	To so pravzaprav GOSENICE majhnega, rjavega metulja . Če so razmere primerne in je na
5	revija	Govora je predvsem o GOSENICAH , odrasli metulji so precej bolj neodvisni.

Primer 9: Del konkordančnega niza z jedrom »metulj« ter kolokatorjem »gosenica«.

Sklep

Besedilni korpus kot jezikovni vir je za pouk jezika nepogrešljivo orodje, saj omogoča pridobivanje védenja o tem, kako se jezik v komunikacijskih situacijah različnih vrst dejansko uporablja. Preučevanje korpusnih podatkov spodbuja pri učencih razvoj splošnega jezikovnega občutka ter omogoča boljši vpogled v možnosti rabe določenih jezikovnih struktur, pri čemer je ključno spoznanje, da jezikovne odločitve niso zgolj pravilne oz. napačne glede na jezikovni predpis, pač pa da nanje pogosto vpliva širši spekter dejavnikov. Pod vodstvom učitelja (razbremenjenega vloge avtoritete, ki mora v vsakem trenutku znati presoditi, kaj je v jeziku prav in kaj narobe) se lahko učenec postavi v vlogo raziskovalca ter interpretorja realnih jezikovnih podatkov, s čimer mu je omogočen razvoj kritičnega mišljenja v povezavi z jezikoslovnim znanjem, pa tudi dojemanja samega sebe kot enakovrednega soustvarjalca diskurza, člana jezikovne skupnosti, ki ima kot tvorec besedil veliko svobode, pa tudi odgovornosti, ki jo slednja prinaša s seboj.

Za učitelja so korpusni podatki dragoceni zato, ker ponujajo odgovor na vprašanje, kaj je v jeziku tipično, kaj pa se pojavlja redkeje in zgolj v specifičnih situacijah, na osnovi česar je mogoče usmeriti pouk k problemom, s katerimi se bodo učenci kot govorci jezika verjetneje srečevali. Korpusna orodja na drugi strani ponujajo številne možnosti organizacije jezikovnih podatkov na nov, zanimiv ter posledično privlačen način. Čeprav je za uspešno opismenjevanje potrebno v pouk vključiti tudi samostojno delo učencev s korpusom, lahko učitelj vrzel med podobo jezika, ki jo prinašajo trenutno obstoječi jezikovni viri, ter jezikom, v katerem živimo tukaj in zdaj, v veliki meri premosti tudi z uporabo na korpusnih podatkih temelječega učnega gradiva – učnih listov, vaj, domačih nalog, testov itd.

Tako učiteljem kot učencem morajo biti za delo na razpolago ustrezni korpusi s pedagoški situaciji prilagojenimi programskimi orodji. V sklopu projekta Spora-zumevanje v slovenskem jeziku poteka vzporedno z gradnjo novega referenčnega korpusa, ki bo prosto na voljo za uporabo na internetu, razvoj novega konkordančnika z dvema vmesniškima različicama. Prva bo namenjena širšemu krogu korpusnih uporabnikov, torej tudi učiteljem slovenskega jezika, druga pa bo razvita posebej

za potrebe šolajoče se populacije, predvsem osnovnošolcem višjih razredov. Zasnova pedagoškega vmesnika temelji na omogočanju preglednosti ter razumljivosti na vseh ravneh programa, kar se med drugim odraža v postavitvi vmesniških elementov na način, ki zahteva minimalno navigacijo med podatki, popolni poenostavitvi načinov iskanja po korpusu ter uporabi jezika, ki namesto jezikoslovne terminologije prinaša popreprosten opis – tako na ravni vmesniških elementov kot (enostavno dostopne) uporabniške pomoči.

Literatura in viri

- Aijmer, K. (ur.) (2009). *Corpora and language teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Arhar, Š. (2007). Kaj početi z referenčnim korpusom. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta. www.fidaplus.net (4. 1. 2010).
- Arhar, Š. (2009). Uporabniška evalvacija korpusa FidaPLUS: zasnova vprašalnika, prvi rezultati. V: Stabej, M. (ur.). *Infrastruktura slovenščine in slovenistike (Obdobja 28)*. Ljubljana: Znanstvena založba Filozofske fakultete, str. 19–26.
- Aston, G. (ur.) (2001). *Learning with Corpora*. Houston: Athelstan.
- Aston, G., Bernardini, S. in Stewart, D. (ur.) (2004). *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Bernardini, S. (2004). *Corpora in the classroom: An overview and some reflections on future developments*. V: Sinclair, J. (ur.). *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 15–36.
- Braun, S. (2006). ELISA: a pedagogically enriched corpus for language learning purposes. V: Braun, S., Kohn, K. in Mukherjee, J. (ur.). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt am Main: Peter Lang, str. 25–48.
- Braun, S., Kohn, K. in Mukherjee, J. (ur.) (2006). *Corpus technology and language pedagogy: new resources, new tools, new methods*. Frankfurt am Main: Peter Lang.
- Breyer, Y. (2006). *My concordancer: Tailor-made software for language learners and teachers*. V: Braun, S., Kohn, K. in Mukherjee, J. (ur.). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt am Main: Peter Lang, str. 157–176.
- Burnard, L. in McEnery, A. M. (ur.) (2000). *Rethinking language pedagogy from a corpus perspective*. Frankfurt am Main: Peter Lang.
- Conrad, S. (2004). *Corpus linguistics, language variation, and language teaching*. V: Sinclair, J. (ur.). *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 68–85.
- Gantar, P. (ur.) (2009). Slovarji, več kot le besede. *Jezik in slovstvo: tematska številka*, 54, št. 3–4.
- Gantar, P. in Krek, S. (2009). Drugačen pogled na slovarske definicije: opisati, pojasniti, razložiti?. V: Stabej, M. (ur.). *Infrastruktura slovenščine in slovenistike (Obdobja 28)*. Ljubljana: Znanstvena založba Filozofske fakultete, str. 151–159.
- Gorjanc, V. (2005). *Uvod v korpusno jezikoslovje*. Domžale: Izolit.

- Granath, S. (2009). Who benefits from learning how to use corpora?. V: Aijmer, K. (ur.). *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 47–65.
- Gut, U. (2006). Learner speech corpora in language teaching. V: Braun, S., Kohn, K. in Mukherjee, J. (ur.). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt am Main: Peter Lang, str. 69–86.
- Hidalgo, E., Quereda, L. in Santana, J. (ur.) (2004). Corpora in the foreign language classroom: Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6). Amsterdam: Rodopi.
- Kosem, I. (2008). User-friendly corpus tools for language teaching and learning. V: Garcia Garcia, A. (ur.). *Proceedings of the 8th Teaching and Language Corpora Conference*. Lizbona: Associação de Estudos e de Investigaç o Científica do ISLA-Lisboa, str. 183–192.
- Logar Berginc, N. in Šuster, S. (2009). Gradnja novega korpusa slovenščine. *Jezik in slovestvo*, 54, št. 3–4, str. 57–68.
- Mukherjee, J. (2006). Corpus linguistics and language pedagogy: The state of the art – and beyond. V: Braun, S., Kohn, K. in Mukherjee, J. (ur.). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt am Main: Peter Lang, str. 5–24.
- Sinclair, J. (ur.) (2004). *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Sinclair, J. (2004). New evidence, new priorities, new attitudes. V: Sinclair, J. (ur.). *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 271–299.
- Stabej, M. (2009). Slovarji in govornici: kot pes in mačka? *Jezik in slovestvo*, 54, št. 3–4, str. 115–138.
- Tsui, A. (2004). What teachers have always wanted to know – and how corpora can help. V: Sinclair, J. (ur.). *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 39–61.

Simon KREK (Amebis, d.o.o. and Jožef Stefan Institute, Slovenia),
Špela ARHAR, Ph.D. (Trojina – Institute for Applied Slovene Studies, Slovenia)

SLOVENE TEXT CORPORA: HOW TO GAIN ACCESS TO THE CLASSROOM?

Abstract: The paper focuses on the prospects of the use of text corpora or corpus data in teaching the Slovene language. Corpora in the classroom represent an addition to the existing language description by providing up-to-date facts about real language use, indispensable for any language teaching. Furthermore, corpus data open the possibility of a different kind of thinking about language in general: they enable the learner to understand the language as a means of communication operating dynamically and democratically in the entire speaking community. As a consequence, the learner recognizes himself/herself as a member of the same community that offers diverse possible answers to language dilemmas. However, a properly trained learner can make a qualified and autonomous choice in the end. Before this kind of language instruction can be introduced into primary and secondary schools, the existence of corpora relevant for classroom use and the instruction on corpus use must be provided both for teachers and for learners. In the last section the outline of the new pedagogic corpus interface is described, which is one of the activities of the »Communication in Slovene« project.

Keywords: text corpus, Slovene language, Slovene language teaching, pedagogic corpus interface, communication in Slovene