

Slovar sodobnega slovenskega jezika

Skupina avtorjev

Zadnje leto je bilo za tiste, ki se ukvarjamo s slovarji, dokaj pestro. Spomladi 2013 je izšel *Slovar novejšega besedja slovenskega jezika*, ki po besedah njegovih avtorjev dopolnjuje *Slovar slovenskega knjižnega jezika*. Kmalu zatem je bil objavljen še predlog za izdelavo *Slovarja sodobnega slovenskega jezika*, čigar izhodišče je v nasprotju s prej omenjenim slovarjem nov opis sodobnega jezika na način, ki ponuja sodobnejše leksikografske rešitve in išče evropske vzporednice. Po objavi se je sprožila javna debata o področju, ki navadno ni deležno pozornosti javnosti, kakor so komentirali novinarji; a ta vtis je pravzaprav napačen. Gre bolj za to, da je zadnja takšna javna debata potekala v sredini 60-ih let prejšnjega stoletja, ko je bil leta 1964 objavljen poskusni snopič *Slovarja slovenskega knjižnega jezika*, in se je spomnijo le najstarejše generacije. Za mlajše generacije je SSKJ enostavno danost. Srednja generacija se sicer še lahko spomni čakanja na izid posameznih knjig SSKJ, a takrat ni več šlo za spraševanje o ustreznosti ali neustreznosti slovarske zasnove.

Od leta 1964 do danes je svet doživel precej sprememb in slovarji pri tem niso nobena izjema. Kljub veliki časovni oddaljenosti od nastanka konceptualne zasnove SSKJ pa je pomembno predvsem dejstvo, da se je evropska in svetovna leksikografija začela spreminjati z množično uporabo računalnikov, še bolj radikalno pa z razcvetom svetovnega spleta in splošnim prehodom jezikov v digitalno okolje. Slovarji danes niso več debele knjige niti CD-ROM-i z v sebi zaključenim nespremenljivim avtorskim delom, temveč je njihova vsebina dojeta v smislu dinamične leksikalne informacije, prilagojene mediju, različnim uporabnikom in namenom rabe. Debata, ki je potekala v preteklem letu, v tem smislu ni brezplodna, kajti radikalnejše konceptualne spremembe se morajo brusiti z argumenti in protiargumenti.

V nadaljevanju prispevka je zbranih nekaj izhodišč avtorjev, ki so svoje poglede na sodobno leksikografijo predstavili na Posvetu o novem slovarju sodobnega slovenskega jezika, ki ga je

februarja letos organiziralo Ministrstvo za kulturo.

Kakšno je temeljno gradivo za slovar?

V zadnjih tridesetih letih so pomembne premike v raziskovanju jezikov prinesli korpusi. Korpusi kot elektronske zbirke avtentičnih besedil, nastale po vnaprej določenih merilih in z določenim namenom ter z ustrezno označenostjo, so vir podatkov za boljše opise jezikovne zgradbe in rabe, njihov računalniško obvladljiv format pa je v te opise med drugim prinesel natančnost meritev zelo različnih lastnosti jezika. Če se omejimo le na slovaropisje: prejšnje ročno podčrtovanje potencialnih slovarskih enot in njihovega besedilnega okolja, izpisovanje ter nato listanje kartotečnih listkov so zamenjale računalniške zbirke, ki skupaj z metodologijo, ki jih spremlja, in orodji, ki lajšajo obvladovanje velike količine podatkov, močno izboljšujejo vse vrste slovaropisnih izdelkov – in to predvsem v smislu verodostojnosti prikaza realnega jezikovnega stanja.

Prvi pisni korpus, ki kaže celovito vzorčno podobo jezika, je za slovenščino nastal leta 2000, in sicer s povezavo strokovnjakov s Filozofske fakultete, Instituta Jožef Stefan, založbe DZS in podjetja Amebis. Temu korpusu – z imenom FIDA – je sledila nadgradnja, ki je prvotni obseg 100 milijonov besed povečala na 620 milijonov in je bila zaključena leta 2006 kot FidaPLUS. Sledila je še razširitev z več kot 500 milijoni besed, ki je bila kot korpus Gigafida zaključena leta 2012. Gigafido dopolnjuje iz nje vzorčeni 100-milijonski uravnoteženi korpus Kres. Gigafida in Kres sta ta hip za slovenščino najbolj primerna podlaga za uresničitev uveljavljenega sodobnega slovaropisnega izhodišča, ki pravi, da je o jeziku najprej treba vedeti čim več, da bi potem z analizo izluščili, kaj je osrednje in obrobno, standardno ali nestandardno, regionalno omejeno, stilno opredeljeno, dovolj stabilno za vključitev, dovolj marginalno za izključitev ipd. Seveda pa bo treba oba korpusa na osnovi slovaropisne uporabe in drugih povratnih informacij še nadalje dopolnjevati ali kako drugače spremeniti.

Kako pristopiti k izdelavi slovarja?

Da so slovarji danes veliko več kot le opis pomenov besed ali zvez, ki jim je dodana še oznaka za besedno vrsto, izgovor in zgledi rabe, se lahko zahvalimo predvsem novim medijem in

V zadnjih tridesetih letih so pomembne premike v raziskovanje jezikov prinesli korpusi. Korpusi kot elektronske zbirke avtentičnih besedil, nastale po vnaprej določenih merilih in z določenim namenom ter z ustrezno označenostjo, so vir podatkov za boljše opise jezikovne zgradbe in rabe, njihov računalniški format pa je med drugim prinesel natančnost meritev zelo različnih lastnosti jezika.

tehnologijam, zlasti spletu, pametnim telefonom, tablicam itd., ki omogočajo prikaz tako rekoč neomejenih količin podatkov na zelo različne načine. V računalniško berljivih leksikalnih bazah so danes strukturirani različni jezikovni podatki: pomenski opis besed, stalnih zvez in frazeoloških enot, slovnični podatki o pregibanju, skladijski podatki, npr. o vezljivostnih vzorcih, podatki o tipičnem besedilnem okolju besed, sinonimih, pravopisnih posebnostih, o pogostnosti besed glede na tip besedila, časovno obdobje, podatki o etimologiji – vsi tisti jezikovni podatki torej, ki smo jih bili še do nedavnega vajeni iskati po različnih tiskanih jezikovnih priročnikih. Bistvena pridobitev sodobnega časa je, da uporabniku, ki se znajde pred določeno jezikovno zadrego, ni treba razmišljati o tem, ali je njegova zadrega pravopisne, slovarske ali slovnične narave, niti mu ni treba kombinirati različnih priročnikov, saj je rešitev, če zna svoje vprašanje kolikor toliko razumljivo ubesediti, le klik ali dva stran.

Lahko bi torej rekli, da so elektronske leksikalne podatkovne baze, kot jih evropski prostor pozna že od 80. in 90. let dalje, osnova za izdelavo vseh vrst jezikovnih priročnikov. V slovenskem okolju je bila izdelavi take jezikovne podatkovne baze namenjena ena od aktivnosti projekta Sporazumevanje v slovenskem jeziku, ki sta ga v letih od 2008 do 2013 financirala Evropski socialni sklad in Ministrstvo za izobraževanje, znanost in šport RS. Pri tem projektu smo sodelavci z Univerze v Ljubljani, Inštituta Jožef Stefan, Zavoda za uporabno slovenistiko Trojina, Inštituta za slovenski jezik Frana Ramovša ZRC SAZU in iz podjetja Amebis zasnovali leksikalno podatkovno bazo za slovenščino, katere namen je dvojen: omogočiti izdelavo za slovenščino najpomembnejših eno- in dvojezičnih slovarjev ter specializiranih jezikovnih priročnikov ter zadostiti potrebam računalniške obdelave naravnega jezika oz. uporabi v jezikovnotehnoloških aplikacijah za slovenščino.

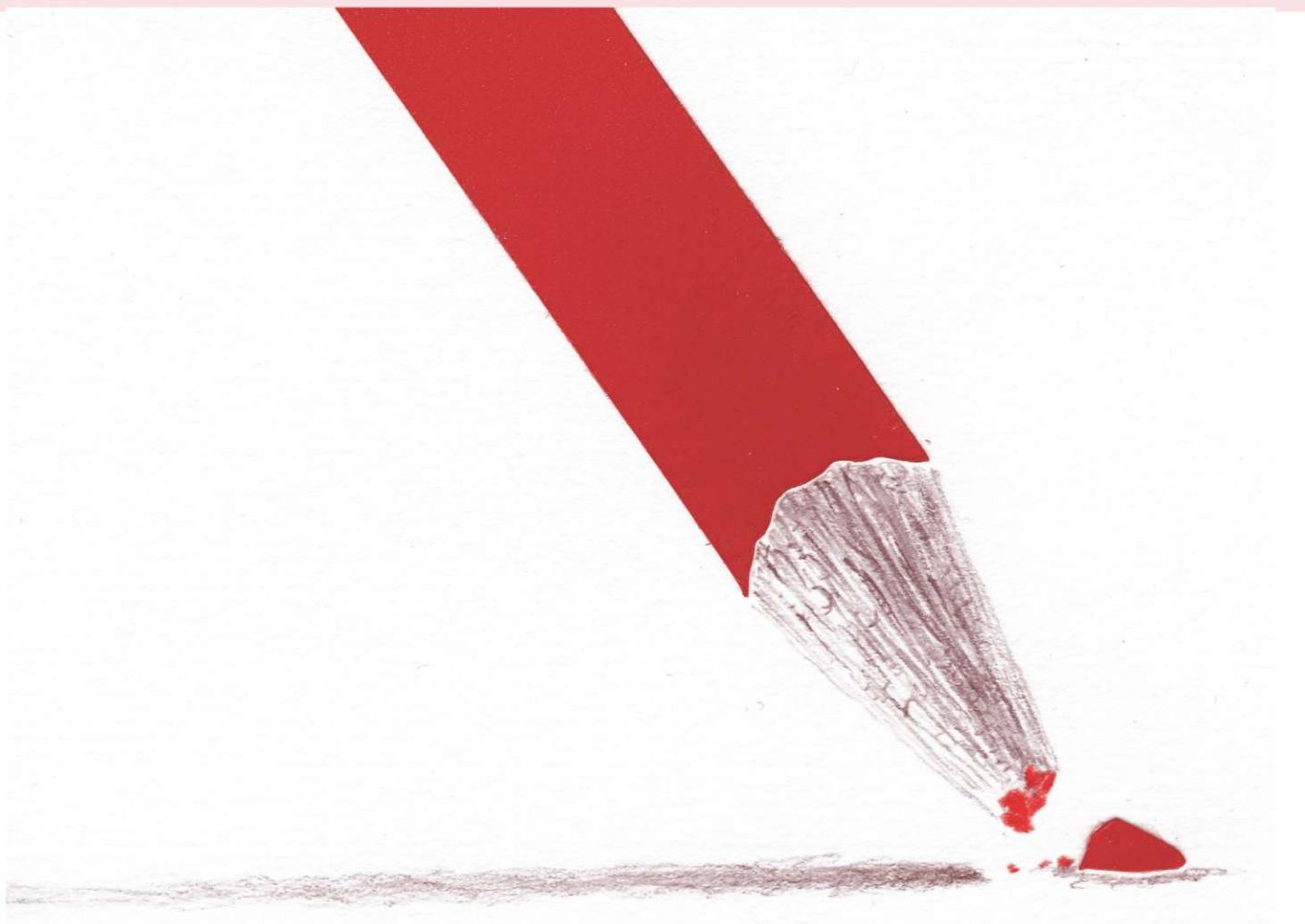
Leksikalno bazo smo zasnovali po vzoru dobrih evropskih praks, kot je denimo baza *Nizozemskega splošnega slovarja*, angleška podatkovna baza Dante, ki je namenjena izdelavi sodobnega irsko-angleškega slovarja, baza, na kateri temelji naj-

novejši *Veliki poljski slovar*, ter z upoštevanjem metodoloških pristopov in jezikoslovnih teorij, ki temeljijo na prepričanjih, da (1) besedni pomen kot sama na sebi zaključena enota ne obstaja, pač pa ima beseda le pomenske tendence, ki jih ureničuje v vsakokratnem konkretnem besedilnem (in zunajbesedilnem) kontekstu, (2) da je jezik mogoče opazovati kot statistično merljiv sistem preferenc ali z drugimi besedami, da so pomeni neposredno povezani z vzorci, v katerih se pojavljajo, (3) da namen slovarjev ni podajanje absolutnih definicij besed, pač pa pomagati uporabnikom razumeti pomen tako, da povežejo že znano z neznanim, (4) da se govorniki pomenov besed učimo iz kontekstov, v katerih jih tipično slišimo, zato je smiselno slovarski opis prilagoditi diskurzu naravnega pojasnjevanja besed, čemur najboljše ustrezajo stavčne definicije, (5) da so med pričakovanji slovarskih uporabnikov najvišje ovrednotene zanesljivost vsebine, jasnost informacij, hitra in brezplačna dostopnost, prilagodljivost vsebin, možnost sprotnega posodabljanja, večpredstavnost in povezave na druge jezikovne vire in enciklopedične podatke.

Gradnji slovarske baze se pri izdelavi sodobnih priročnikov ne bo mogoče izogniti, če želimo jezikovni opis slovenščine dvigniti na evropsko raven, če želimo slovenščini zagotoviti preživetje v digitalnem okolju, če želimo rešiti problem jezikovnih priročnikov za slovenščino na dolgi rok in nenazadnje, če želimo s finančnim vložkom ravnati gospodarno in z mislijo na rezultat, ki bo uporaben in vreden zaupanja.

Kako k izdelavi pripomorejo nove tehnologije?

Izdelava slovarjev vse od osemdesetih let prejšnjega stoletja postaja vse bolj računalniško usmerjena, saj je s pomočjo informacijskih in jezikovnih tehnologij slovaropisje učinkovitejše in bolj konsistentno, končni izdelek pa kvalitetnejši. Jezikovne tehnologije imajo pomembno vlogo že pri zbiranju in obdelavi ter jezikoslovnem označevanju korpusnega gradiva, kar omogoča iskanje in posploševanje identificiranih jezikovnih pojavov. S pomočjo jezikovnih tehnologij je prav tako mogoče izdelati utemeljen nabor ter prednostni seznam izdelave slovarskih



gesel, bodisi z iskanjem ključnih besed v korpusu ali z analizo zgodovine najpogostejših iskanj v spletnem slovarju. Tipično vedênje besed v sobesedilu, stalne besedne zveze in strokovno besedišče raziskujemo z orodji za luščenje kolokacij, ki z uporabo statističnih funkcij identificirajo tiste kombinacije besed, ki se skupaj pojavljajo pogosteje kot z ostalimi besedami v korpusu, na podoben način pa delujejo tudi algoritmi za avtomatsko luščenje zglédov. Med ambicioznejšimi jezikovnotehnološkimi postopki, v razvoj katerih bi bilo treba še vlagati, je avtomatsko prepoznavanje pomena večbesednih enot glede na sobesedilo in avtomatska pomenska členitev slovarskih iztočnic. Za tuje jezike prav tako že obstajajo orodja za avtomatsko iskanje semantičnih relacij med besedami, predvsem sopomenskosti ter nad- in podpomenskosti. Ena najzahtevnejših nalog pa ostaja avtomatsko luščenje razlag iz korpusov. Ker se jezik nenehno spreminja, je ob predpostavki, da imamo na voljo spremljevalni korpus, z jezikovnimi tehnologijami mogoče zagotoviti

tudi identifikacijo novih besed ter pomenskih premikov, kar omogoča avtomatsko posodabljanje slovarja.

A jezikovne tehnologije v uspešnem slovarskem projektu ne smejo biti uporabljene le kot podpora ostalim delovnim fazam, temveč morajo biti sodobni leksikografski projekti jezikovnotehnološko že zasnovani, tehnologije pa morajo igrati ključno vlogo tudi pri vseh nadaljnjih korakih. Zgolj na tak način lahko namreč dosežemo objektivizacijo leksikografskega dela in posledično zadostno mero učinkovitosti in natančnosti ter zagotovimo sledljivost, reproduktibilnost, povezljivost z drugimi viri, kot so drugi slovarji, Wikipedija in korpusi, ter dolgoročno ter večnamensko uporabnost. Prav tako je že zdaj treba načrtovati hitro in postopno objavlanje rezultatov in redne posodobitve slovarja, poleg tega je treba predvideti tudi njegove prilagoditve za druge potrebe, kot so nadaljnji slovarski projekti, jezikovne tehnologije in aplikacije, s čimer bo enkratni finančni, časovni in strokovni vložek večstransko izkoriščen.

S temi izzivi se bo bodoča slovarska ekipa lahko uspešno spopadla le tako, da bodo z možnostmi in omejitvami jezikovnih tehnologij seznanjeni vsi, od vodje projekta do urednikov slovarja in leksikografov, in da bodo projektni sodelavci svoje interdisciplinarno znanje uporabili pri pripravi zasnove slovarja, gradnji slovarske baze ter razvoju leksikografskih orodij in uporabniških vmesnikov. Zasnova in izdelava slovarja brez jezikovnotehnoloških kompetenc z naknadno računalniško implementacijo bi namreč vodila v že videne neuspešne leksikografske poskuse pri nas.

Kdo so uporabniki slovarja?

Določitev zahtev različnih vrst uporabnikov je za izdelavo slovarja ključnega pomena in mora biti izhodišče tako pri izbiri medija kot sestavnih delov slovarske vsebine. Slovar mora slediti predvsem zahtevam šolskih in splošnih, tj. odraslih uporabnikov; gre za uporabnike, ki se bodisi v izobraževalnem procesu ali pri svojem poklicnem ali vsakdanjem delu ukvarjajo s tvorjenjem in razumevanjem besedil in pri tem potrebujejo ustrezno lahko dostopno podporo glede vseh vrst jezikovnih informacij. Ob tem je treba upoštevati, da sta se besedilna produkcija, dostopna širokemu krogu uporabnikov, in število piscev, ki pri tem sodelujejo, v času širitve spleta in družabnih omrežij potisočerala, s čimer se je spremenil tudi temeljni družbeni model publiciranja, ki je zamenjal prejšnjega, temelječega na tisku in založniški industriji. Slovar mora torej svojim ciljnim uporabnikom uspešno služiti v novem digitalnem in spletnem modelu komuniciranja.

Za splošnega uporabnika *Slovarja sodobnega slovenskega jezika* se predvideva, da doživlja jezik in z njim povezane zadrege skozi sporazumevalne situacije, kar pomeni, da je njegov cilj rešitev določene sporazumevalne zagate, manj pa raziskovanje jezika oziroma njegovih prvin. Informacija, ki jo pričakuje, je aktualna pomenska opredelitev besed in zvez, s poudarkom na tistih, ki so v slovenskem besedišču nove in še neuveljavljene, ter besed in zvez, ki so sicer vezane na določena specializirana področja, a pod vplivom družbenopolitičnih, gospodarskih in kulturnih okoliščin prehajajo v splošno rabo, zlasti prek informativnih medijev. Uporabniki od sodobnega slovarja pričakujejo, da je hitro in enostavno dostopen, npr. prek različnih digitalnih, tudi mobilnih naprav, ki so jih pri izobraževanju ter v svojem poklicnem ali vsakdanjem delu vajeni uporabljati, in sicer na preprost in intuitiven način. To pomeni, da posebno izobraževanje ne sme biti potrebno oz. da uporabniki ne bodo

potrebovali posebnih navodil za uporabo slovarja.

Uporabniki slovarja bodo seveda tudi jezikoslovci, vendar pa bo zanje najbrž bolj zanimiva slovarska baza, ki bi morala biti odprto dostopna v obliki podatkovne zbirke v formatu XML. Na podlagi teh podatkov bo namreč mogoče izvajati različne raziskave o pomenskih, skladdenjskih, vezljivostnih, kolokacijskih, normativnih in drugih vidikih sodobnega slovenskega jezika.

Na kakšen način naj bo slovar dostopen?

Najpomembnejša in večinoma tudi edina namembnost slovarja, kot jo vidijo uporabniki in pogosto tudi stroka, je ta, da je slovar referenčni priročnik, v katerega pogledajo, kadar jih zanima definicija, pomeni, frazeologija itd. določene besede. Vendar imajo slovarji, vsaj če so izvorno zapisani kot digitalna leksikalna baza, tudi širšo uporabo, saj lahko služijo kot zelo dobrodošel vir podatkov za jezikovne tehnologije, torej aplikacije, ki omogočajo računalniško obdelavo jezika in s tem pripomorejo k informatizaciji slovenščine. Takšne aplikacije so npr. sinteza govora, strojno prevajanje, poizvedovanje po informacijah in avtomatsko generiranje povzetkov, nenazadnje pa tudi izdelava označenih korpusov. Slovar namreč vsebuje obilico informacij o jeziku, ki so nepogrešljive za izdelavo takšnih aplikacij za določen jezik: od besednovrstnih informacij, izgovorjavi in pomenski členitvi do podatka o vrstah besedil, v katerih se uporabljajo.

Novi slovar slovenskega jezika tako lahko opravlja dvojno funkcijo – kot referenčni vir za najširši krog uporabnikov in kot podatkovna baza za uporabo v jezikovnih tehnologijah. Za uresničevanje prve vloge zadošča prosti dostop prek spletnega iskalnika, za polno uresničevanje druge vloge pa mora biti slovar oz. njegova baza odprto dostopna. Pod pojmom odprti dostop razumemo možnost in dovoljenje, da lahko celoten slovar prenesemo na svoj računalnik, in to ne samo v kateri od oblik, predvidenih za predstavitev, kot sta HTML ali PDF, temveč kot kopijo slovarske baze v izvornem zapisu XML.

Naša teza je, da bi morali biti vsi jezikovni viri slovenskega jezika, ki nastanejo z javnim financiranjem, javni in s tem maksimalno odprti, kolikor to pač dopuščajo predhodne avtorske pravice in varovanje pravice do zasebnosti, kar pa pri predvidenem slovarju ne bi smel biti problem. Zapiranje namreč koristi zgolj institucijam, ki si nad slovarjem lastijo avtorske pravice, slovenskim uporabnikom, ki so slovar posredno financirali, pa samo škoduje. Odprti dostop do jezikovnih virov predvideva tudi »Resolucija o nacionalnem programu za jezikovno politi-

co 2014–2018, pa tudi sicer se slovenska politika financiranja raziskav počasi, a vztrajno bliža zahtevi po obveznem odprtem dostopu do rezultatov javno financiranih raziskav – to je npr. obvezen pogoj tudi za vse projekte novega programa EU za raziskave in inovacije Obzorje 2020.

Za odprti dostop do podatkov obstaja večje število licenc, vendar so se od vseh najbolj uveljavile licence *Creative Commons* oz. Ustvarjalna gmajna. Licence CC so namenjene urejanju dostopa do (umetniških, znanstvenih) avtorskih del in uporabniku dajejo ustrezno informacijo, imajo pravno veljavo, so strojno čitljive in kot take povsem primerne tudi kot pravni okvir za odprti dostop do slovarjev. Licence CC dovoljujejo prevzem avtorskih del in njihovo nadaljnje razširjanje, obstajajo pa v več različicah, pri katerih posamezne omejitve odprtosti lahko medsebojno kombiniramo.

Za slovar se nam zdita najbolj ustrezni licenca CC BY (priznanje avtorstva) oz. CC BY-SA (priznanje avtorstva, deljenje pod enakimi pogoji). Vse nadaljnje omejitve, predvsem CC BY-ND (brez predelav) in CC BY-NC (nekomercialno) ali celo CC BY-ND-NC bi bistveno zmanjšale uporabnost in s tem uporabo slovarja. Omejitev ND na primer pomeni, da bi bilo treba izdelavo odprte, a predelane slovarske baze za specifične jezikovno-tehnološke namene financirati povsem na novo, kar bi vodilo v dvojno javno financiranje podobnih jezikovnih virov, ob tem pa dodatno upočasnilo razvoj jezikovnih tehnologij za slovenski jezik. Tudi omejitev NC ima negativne posledice za uporabo slovarja in s tem za informatizacijo slovenskega jezika. Slovar je del jezikovne infrastrukture, v katero država vlaga tudi zato, da omogoča razvoj in konkurenčnost gospodarstva. Dejstvo, da bi bilo uporabo slovarskih podatkov treba plačati, bi odgnalo večino že tako redkih podjetij, ki se trudijo razvijati informacijske produkte, vezane na uporabo slovenščine. Res bodo takšni produkti plačljivi in avtorji slovarja od teh prihodkov ne bodo dobili nič, zato pa vsi, ki so z davki financirali izdelavo slovarja, dobijo možnost, da kupijo izdelke za slovenski jezik, ki jih sicer ne bi bilo, obenem pa se s tem spodbuja konkurenčnost domačega gospodarstva.

Kdo lahko naredi slovar?

Za nastanek slovarja sodobnega slovenskega jezika potrebujemo sodelovanje različnih strokovnjakov – torej tudi morda na prvi pogled nepričakovanih in novih akterjev, ki so izkazali interes, reference in znanje – ter skupen koncept. Na fakultetah vseh slovenskih univerz nastajajo raziskave in gradivo, ki v

Uporabniki od sodobnega slovarja pričakujejo, da je hitro in enostavno dostopen prek različnih digitalnih, tudi mobilnih naprav, ki so jih pri izobraževanju ter v svojem poklicnem ali vsakdanjem delu vajeni uporabljati, in sicer na preprost in intuitiven način.

marsičem pripomorejo k temeljnemu raziskovanju in prinašajo tudi v slovaropisje nova znanja ter metodologijo. Zato bi bilo treba raziskovalno in strokovno delo za nastajanje temeljnih sodobnih priročnikov reorganizirati in povezati v okviru resne in finančno podprte jezikovne politike, ki bi v najkrajšem možnem času nadoknadila zamudništvo pri temeljnih jezikoslovnih delih.

Na slovenskih visokošolskih in drugih raziskovalnih inštitucijah so odlični jezikoslovci, specializirani za delo s korpusi, in jezikoslovci, ki obvladajo vse vidike, ki jih odpira priprava slovarja (zvrstnost, norma, (iz)govor, stilistika ...), pa tudi jezikovni tehnologi, računalničarji in drugi strokovnjaki, ki so si pridobili evropsko aktualno leksikografsko znanje in ga že prenesli na slovensko gradivo. V okviru konzorcija s projektno zadanim ciljem je mogoče vsa ta znanja povezati. Če je bila univerza v zgodovini umaknjena s temeljnih slovaropisnih del, je sedaj, ko prav na univerzah opozarjamo na zaostajanje pri sinhronem raziskovanju in digitalizaciji slovenščine, čas, da pobudo enakopravnega konzorcijskega interdisciplinarnega sodelovanja, ki je bil zapisan med sklepi posveta o novem slovarju, čim prej uresničimo. ●

*** Avtorji:**

dr. Tomaž Erjavec, Institut »Jožef Stefan«

dr. Darja Fišer, Filozofska fakulteta, Univerza v Ljubljani

dr. Polona Gantar, Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU

dr. Monika Kalin Golob, Fakulteta za družbene vede, Univerza v Ljubljani

dr. Iztok Kosem, Trojina, zavod za uporabno slovenistiko

dr. Simon Krek, Institut »Jožef Stefan«

dr. Nataša Logar, Fakulteta za družbene vede, Univerza v Ljubljani