

Mojca Šorli

Trojina, zavod za uporabno slovenistiko

Sodobni sporazumevalni slovar slovenskega jezika: izhodišča, viri, izvedba

V prvem delu prispevka orišemo dve osnovni možni izbiri strokovnih izhodišč in se zavzamemo za komunikacijsko-pragmatični pristop, ki bo proizvedel sodobni sporazumevalni tip slovarja, namenjen zlasti mlajšemu in/ali jezikovno nespecializiranemu uporabniku slovenščine. Pri tem sta ključni korpusno izhodišče in dejstvo, da morata lociranje in opis pomena potekati znotraj frazeoloških kontekstov in da je zato potrebno pričeti opuščati “besede” kot osnovne slovarske enote s stabilnim semantičnim pomenom. V drugem delu se navežemo na Leksikalno bazo za slovenščino kot predvideni vir novega oz. novih slovarjev slovenščine. Kritično ovrednotimo rezultate projekta in podrobneje predstavimo pomenske opise oz. razlagalne strategije, ki predstavljajo pomembno prednost v sodobnem opisu jezika, saj omogočajo konsistenten in sporazumevalno usmerjen opis vseh tipov leksemskih denotatov. V zadnjem delu se opredelimo do nekaterih konkretnih rešitev, ki jih predvideva Predlog za novi slovar sodobnega slovenskega jezika, pri čemer izpostavimo predlagane metode dela (izraba moči množic) ter nekatere kadrovske in finančno-organizacijske vidike projekta.

A Contemporary Communicative Dictionary of Slovene: Sources, Design And Execution

In the first section, two possible lexicographical approaches are presented with regard to the design of a new Dictionary of Standard Slovene. Of these, the approach that would produce a communicatively and pragmatically oriented dictionary aimed at younger and/or non-specialised users of Slovene is deemed more appropriate in view of the current situation in the field of language resources. The need for corpus-based analysis of the lexicon is of crucial importance, as is recognition of the fact, ensuing from contemporary lexicography, that meaning descriptions must be located and compiled within phraseological contexts, and that it is therefore necessary to begin to phase out “words” as basic lexical units with a stable semantic meaning. In the second part, we bring into the argument the Slovene Lexical Database (2008-2012), which was designed from scratch as a source for new electronic monolingual dictionaries. We critically evaluate the results of the project and describe in

detail its definition strategies, which represent an important advantage in modern language description, enabling a consistent and communication-oriented description of all types of denotata. The last part of the article provides an assessment of some of the concrete solutions in the “Proposal for a New Dictionary of Modern Slovene”, in particular some of the proposed methods (crowdsourcing), as well as the financial and organisational aspects of the project.

Ključne besede: slovar slovenščine, korpusna analiza, sporazumevanje, pragmatika, pomenski opis, leksikografija, Leksikalna baza za slovenščino

Key words: dictionary of Slovene, corpus analysis, communication, pragmatics, meaning description, lexicography, Slovene Lexical Database

1 Uvod

Ob spremembah, ki jih narekujejo nove tehnologije in elektronski medij, je jasno, da je bilo treba modele slovnično-pomenskih opisov besedišča zastaviti na novo. Spletni slovarski in širše jezikovni viri bodo morali uporabnikom predvsem zagotavljati kvalitetne, tj. zanesljive podatke, ki ne bodo pridobljeni introspektivno, temveč **na podlagi empiričnih podatkov**, tj. velikih elektronskih zbirk besedil. Pri tem bo zelo pomembna vpetost podatkov v širši kontekst možnosti, ki jih ponujata informacijska in komunikacijska tehnologija. Eden glavnih izzivov sodobne leksikografije bo pravzaprav njena **sposobnost, da prepreči t. i. informacijski stres, ki izhaja iz preobilice podatkov**. Soočanje z velikimi količinami podatkov v leksikografskem procesu nas nauči, da je kljub številnim lastnostim jezika, ki jih je mogoče kategorizirati ali posplošiti, praktično vsaka beseda v nečem, bodisi oblikoslovno, skladenjsko, po tipu večpomenskosti, rabi itd. tudi posebna in kot taka zahteva specifične rešitve. Vedno več vemo o tem, katere postopke pri izdelavi slovarja je smiselno (pol)avtomatizirati (rutinski postopki) in katerih ne (prevladujoči del analitičnih in opisnih postopkov), ker jih lahko kvalitetno opravijo le leksikografi (gl. npr. Kosem idr. 2012; Kilgarriff in Rundell 2011). Strokovne smernice razvoja, ki jih lahko zagotovo povzamemo tudi za slovenski jezik in njegove uporabnike: **slovar bo vir dinamičnih podatkov, ki bodo ustrezali specifičnim potrebam specifičnih uporabnikov v njihovi specifični situaciji**.

Trenutno vidimo kot ključni razlog, da ni mogoče doseči konsenza o tem, kako naprej v slovenski leksikografiji, različne poglede na to, 1) ali lahko še naprej vzdržujemo stanje, v

katerem imamo vsi uporabniki slovenščine na voljo **en sam, “temeljni” slovar**, ki naj zadošča sporazumevalnim potrebam tako različnih uporabniških skupin, kot so osnovnošolci in dijaki, študenti, odrasli nezahtevni uporabniki, učitelji, jezikovni delavci (lektorji, prevajalci itd.), jezikoslovci in tujci, ki se učijo slovenščino, 2) ali je po jezikovnotehnoškem preboju in izsledkih sodobne leksikografske teorije in prakse še upravičeno vztrajati pri tradicionalnem **jezikovnosistemskem modelu slovarja**, kakršen je *Slovar slovenskega knjižnega jezika*, in ne nazadnje 3) kdo naj izdela novi slovar slovenskega jezika oz. kakšne vrste znanj zahteva danes optimalna slovarska ekipa. V nadaljevanju se podrobneje posvetimo zlasti točkama 1 in 2, v poglavju 3 pa spregovorimo o kadrovskih in organizacijskih vidikih izvedbe SSSJ. Mnenja in predlogi, navedeni v tem prispevku, izhajajo iz večletne leksikografske prakse, pa tudi iz s to prakso motiviranega znanstvenega raziskovanja avtorice.

2 Strokovna izhodišča sporazumevalnega slovarja slovenščine ali kakšen slovar sploh potrebujemo?

2.1 Korpusno izhodišče in implikacije

Delo J. Sinclairja je zlasti s projektom COBUILD (1987) in iz njega izhajajočimi enojezičnimi slovarji za tujce (prvi med njimi *The Collins COBUILD English Language Dictionary* iz leta 1987, dalje: CCELD) izzvalo tradicionalne modele enojezičnega in dvojezičnega slovarja kot glosarsko urejenega seznama besed in pomenov z novim modelom, po katerem prepoznavamo pomene preko in znotraj frazeologije (Moon 2008: 243). Prototip resnične leksikografske inovacije tako ključno določajo **korpusna analiza, lociranje in opis pomena znotraj frazeoloških kontekstov ter pričetek opuščanja “besed” kot semantičnih enot v osami** (ibid. 242), torej kot primarnih slovarskih enot s stabilnim semantičnim pomenom. Beseda vse bolj postaja zgolj točka dostopa do podatkov o njenih različnih (kontekstualnih) rabah. P. Hanks,¹ eden najvidnejših sodobnih leksikografov, govori o tem, da bi morali videti pomen zlasti kot dinamičen pojav oz. dogodek, ter da je treba pri leksikalni analizi ločevati med psihološkim in družbenim oz. javnim ali konvencionalnim pomenom, ki je slovarsko pravzaprav najbolj relevanten. Ta radikalno drugačni pogled na naravo in izvor pomena prinaša s seboj določene rešitve, ki jih tradicionalno usmerjeni leksikografi radi

¹ <http://newbooksinlanguage.com/2013/06/10/patrick-hanks-lexical-analysis-norms-and-expectations-mit-press-2013/>.

opredeljujejo preprosto kot “napake”. Za primer lahko navedemo obravnavo homonimnih iztočnic (npr. kos), ki se v klasičnih slovarjih obravnavajo kot samostojne iztočnice, prav tako legitimno pa jih je obravnavati tudi kot večpomenske iztočnice. Takšna odločitev temelji na novo definiranjem razmerju med “lemo”, ki je predkorpusni leksikografiji večinoma tuj formalni, oblikovni kriterij, ter “leksemom”, kjer je odločilen pomen (gl. npr. Moon 1998: 5). Po takšnem razumevanju “kos” pripada eni sami lemi ter dvema leksemoma (ptič in del celote).

2.2 Obseg in vrsta slovarja

Posledica empiričnega pristopa v leksikografski praksi je tudi **identificiranje in opisovanje tistega, kar je (proto)tipično**, torej pogosto, in ne vsega, kar v jeziku obstaja (ali obstaja celo hipotetično, kot besedotvorna možnost). Korpusi nam lahko ogromno povedo zlasti o tem, katere so tipične rabe besed (Hanks 2013), zato je odsotnost določenega pomena, pogosto opredeljena kot “napaka”, običajno posledica odločitve, da opisujemo dejansko jezikovno rabo tukaj in zdaj. Pri tem seveda ne gre odrekati legitimnosti historičnim ali kako drugače specializiranim slovarjem, ki jih prav tako potrebujemo. Toda namesto “zajeti vse” je danes v veljavi načelo, po katerem je treba **čim bolj ciljno izbrati obseg, vrsto in način predstavitve podatkov o besedišču**, v skladu s pričakovanimi potrebami uporabnikov, pri tem pa lahko različne vire v elektronski obliki povezujemo med seboj. Prednost elektronsko zasnovane slovarske baze je prav v tem, da omogoča naknadno in postopno dodajanje specifičnih vrst podatkov, to pomeni, da so lahko posamezni, glede na tip slovarja manj prioritetni segmenti slovarskega gesla (npr. izgovorjava, etimologija, povezave z zunanjimi viri, multimedijske vsebine itd.) predmet ločenih oz. odloženih projektov. Premišljen izbor leksike in tipa podatkov bo občutno zmanjšal čas in stroške izdelave novega slovarja. Nekateri segmenti leksike so, poenostavljeno rečeno, bolj potrebni nove (in sodobne) obravnave kot drugi. To je zlasti **jedrni del slovenskega besedišča**, ki ga je treba tako z jezikoslovnega kot kulturološkega vidika ustrezno posodobiti, pa tudi neologizmi in (pol)terminološke enote. Nujne so izboljšave pri vsebini in obliki pomenskih opisov nekaterih povsem vsakdanjih besed, pogosto na ravni jezikovne rabe.

2.3 Slovarski koncepti: SSKJ in Collins Cobuild English Language Dictionary (CCELD)

Tako rekoč edini razlagalni slovar slovenščine, Slovar slovenskega knjižnega jezika (1970–1991) (dalje: SSKJ), je izšel v enozvezkovni obliki leta 1994, od takrat pa ni bil več posodobljen. Leta 2012 smo dobili še Slovar novejšega besedja slovenskega jezika (dalje: SNB), ki “predstavlja osnovno novejše leksikalno dopolnilo Slovarju slovenskega knjižnega jezika,” od leta 1991 naprej,² in ki hkrati nakazuje, da bo tudi novi SSKJ temeljil na tradicionalnih slovaropisnih metodah. Ob primerjavi komunikacijskih in pragmatičnih prvin slovarjev SSKJ in CCELD A. Vidovič Muha (2013: 98 [2000]) identificira **dva povsem različna slovarska koncepta**: SSKJ je zasnovan kot “**predstavitev sistemiziranih poimenovalnih enot jezika**”, medtem ko CCELD “**v celoti temelji na sporočanjski vlogi jezika**”; “/p/ragmatična kompetenca, se pravi ‘sposobnost učinkovite jezikovne rabe za izpolnitev namere in cilja’, [je] bistvo sporočajske zasnove slovarja.” Avtorica tako ugotavlja, da CCELD **ne prinaša “(slovarske) vrednosti svoje leksike, ampak prototipski opis določenega denotata**, torej kar se da splošno predstavo (pojem) o njem, izoblikovano na podlagi mnogih, v določenem trenutku po največ možnih zapisov njegove rabe.” Poleg tega (ibid.: 98–99) pripoznava, da “[r]ačunalniške zbirke besedil omogočajo torej popolnoma nov tip slovarskih zasnov oz. slovarjev z ambicijo prek jezika **prepoznavati kulturne, civilizacijske, ideološke idr. okoliščine, ki omogočajo rabi biti učinkovita /.../**”. Kot zagovornica jezikovnosistemske oz. skladijske naravnosti slovarske zasnove si odgovarja na vprašanje, kakšne so možnosti za posredovanje podatkov o sporazumevalnem vidiku in pragmatičnih okoliščinah rabe v okvirih “klasičnega tipa slovarstva” (ibid.: 99), kamor sodi tudi SSKJ: “V vseh primerih, ko SSKJ pojasnjuje, kaj človek s tem ‘izraža’, in ne kaj ‘pomeni’, gre za izpostavitve sporočajskega vidika leksema in s tem njegove ‘pragmatične vloge’” (ibid.: 100–101). Tudi med slovarji, kot sta SSKJ in CCELD, bi torej lahko našli vzporednice, vendar ostaja prvi naravnan izrazito jezikovnosistemsko, drugi pa sporazumevalno-pragmatično. Pri tem pa vendarle moti dejstvo, da **osnovna spoznanja, ki jih o jezikovnem pomenu prinaša korpusno jezikoslovje, niso bila prevzeta** v slovensko leksikografsko (institucionalno) prakso. Naj na tej točki pojasnimo, da temelji vsebina naslednjih poglavij v celoti na predpostavki oz. trditvi, da potrebujemo uporabniki slovenščine v tem trenutku zlasti **nov slovar z opisom jezika izključno s stališča njegove sporazumevalne vloge**.

² <http://isjfr.zrc-sazu.si/sl/publikacije/slovar-novejšega-besedja-slovenskega-jezika-prvi-natis#v>.

2.4 Leksikalna baza za slovenščino

Leksikalna baza za slovenščino (dalje: LBS) je nastajala v letih 2008–2012, zasnovana pa je bila z namenom izgradnje novega slovenskega slovarja v digitalni obliki. Načrtovani slovar bo namenjen zlasti splošnim uporabnikom in šolajoči se mladini, leksikalna zbirka s svojim jezikoslovnim ustrojem in jezikovnotehnološkimi načeli gradnje pa se prilega tudi potrebam jezikoslovcev in jezikovnih tehnologov. V nadaljevanju podrobneje predstavimo segment pomenskih opisov, ki predstavlja eno pomembnejših inovacij klasičnega pristopa v slovenskem slovaropisju, razvili pa smo ga v okviru LBS prav z namenom približati se konceptu sporazumevalnega slovarja, ki je hkrati uporabniško prijazen in utemeljen v sodobnih leksikografskih teorijah. Zasnovo za tip “sporočanjsko-pragmatičnega slovarja”, kakršnega s SSKJ zgoraj vzporeja Vidovič Muha, tako že imamo.

2.4.1 Podatki o rabi besede – vrste definicij in razlagalne strategije

V nasprotju s klasično slovarsko definicijo opisi v LBS temeljijo na predpostavki, da mora razlaga hkrati s pomenom v ožjem smislu prikazati tudi obnašanje iztočnice v njenem naravnem skladijskem in besedilnem okolju. Taksonomija, kot je v nadaljevanju podana v kratki, pregledni obliki, ne izhaja iz posamezne besedne vrste, temveč temelji na razlikovalnih lastnostih, ki so praviloma v prvem delu razlage:

Tabela 1: Tipologija pomenskih opisov LBS in primerjava z opisi v SSKJ (oznaka * pomeni le delno ustreznost)

| | LBS: | SSKJ: |
|---|--|--|
| <i>TIP A: Stavčna razlaga brez če/ko-stavka</i> | | |
| <i>A1 NOTA</i> | nota je posamezen ton, ki ga proizvedemo z inštrumentom ali glasom | znak za ton |
| <i>A2 STISNITI 3</i> | (trdno, a z občutkom oprijeti) kdo stisne komu roko v pozdrav ali v znak hvaležnosti, spodbude: <i>Ko je B. Netanjahu stisnil roko Arafatu, je stisnil roko tudi vsem njegovim vojakom /.../ ko sta si na brdu stisnila roke G. Bush in V. Putin /.../</i> | I. trdno držeč, oprijemajoč z roko, rokami narediti, da na kaj deluje sila: <i>močno mu je stisnil roko / pri pozdravu stisniti roko v slovo</i> |
| <i>TIP B: Če/ko-stavek</i> | | |
| <i>B1 NEPRILAGOJEN</i> | če je ČLOVEK, RASTLINA ali PREDMET neprilagojen RAZMERAM v okolju, nima za bivanje ali delovanje v njem ustreznih lastnosti ali pogojev | ki ni prilagojen: neprilagojeno ravnanje, vedenje / neprilagojen človek; družbeno neprilagojen |

| | | |
|--|---|--|
| <p>B2 ENAČITI</p> | <p>če človek enači kaj s POJAVOM, POJMOM ali LASTNOSTJO, meni, da gre za enaki stvari, pri tem pa navadno spregleda bistvene razlike, bodisi zaradi nevednosti ali namerno: <i>Krščanskega verovanja ne kaže enačiti z zunanjo strukturo Cerkve, tako imenovanim klerikalnim vidikom. / Dodala je, da Slovenci poslikavo telesa žal še vedno enačimo s pustnimi maskami.</i></p> | <p>imeti, šteti za enako: <i>nepravilno je enačiti normo z akordom; pojma se v praksi večkrat enačita /.../ // po določeni značilnosti, lastnosti imeti kako stvar za enako z drugo: le kako bi enačili stroj in človeka; opise v povesti lahko enačimo s pisateljevim otroškim svetom /.../</i></p> |
| <p>B2 PLEZATI <i>čez/prek(o)/skozi kaj; iz česa</i></p> | <p>če ČLOVEK pleza preko OVIRE, skozi ODPRTINO ali iz neudobnega POLOŽAJA, se skuša s pomočjo celega telesa premakniti v želeno smer, navadno s trudom ali težavo: <i>Zaposleni v studiu morajo včasih plezati prek kupov daril, pleniv in kupov igrač. / Med snemanjem namreč ne moreš plezati s tribun. / Velikokrat so morali plezati skozi okna, iskati stranske izhode in pred dekleti pobegniti skozi kuhinjo.</i></p> | <p>2. premikati se, pomagajoč si z nogami in rokami, tako, da je telo blizu podlage:* <i>otroci radi plezajo; plezati čez plot, po klopek / ekspr. po vseh štirih plezati navkreber</i></p> |
| <p>TIP C: Projekcija</p> | | |
| <p>ČAS ZACELI (VSE) RANE, ČAS CELI RANE</p> | <p>če kdo izreče, da čas celi vse rane, se strinja s splošno razširjenim prepričanjem, da bolečina, zlasti zaradi izgube drage osebe, prej ali slej popusti: <i>»Boš videla ... vzemi si čas. Čas zaceli vse rane, čas je edini gospodar usode« / Čas ni zacelil rane, čas ni ozdravil bolečine, dragi Polde / Težka je izguba ljubljene sopotnika, čas celi vse rane.</i></p> | <p>čas celi rane; čas prinese svoje po daljšem trajanju se stvari omilijo</p> |
| <p>TIP D: Neimenovalniška oblika iztočnice</p> | | |
| <p>DEŽEVEN</p> | <p>v deževnem OBDOBJU ali KRAJU veliko dežuje: <i>Naziv deževno mesto pa si je prislužilo čisto po krivici, saj so podatki, da je to mesto z največ dežja v ZDA, čisto napačni. / Pa lep pozdrav izpod deževnega Pohorja!</i></p> | <p>nanašajoč se na dež, deževanje: <i>deževen dan; deževna doba, deževno vreme / deževna pokrajina / deževni oblak; deževne kaplje</i></p> |

Prednost opisanih razlagalnih strategij, ki predstavljajo izhodišče za oblikovanje slovarskih razlag in ki smo jih v Navodilih za avtorje LBS (Gantar idr. 2011) razdelali še glede na posamezne besedne vrste, je tudi v tem, da omogočajo **konsistenten in sporazumevalno usmerjen opis vseh tipov leksemskih denotatov**. Nekateri pomeni so konvencionalno povezani z družbenim neodobravanjem ipd., npr. pomen zveze “sosedova trava je bolj zelena”, kjer gre za konvencionalno negativno pojmovano naravnost posameznika, ki meni, da je življenje drugih boljše ali lažje (Šorli 2012: 110). Tega pomena ne moremo izpeljati s kompozicionalnimi postopki in tudi ne s sklepanjem, pač pa ga lahko ugotovljamo z analizo

velikega števila stavčnih vzorcev s to leksikalno zasedbo in širšega sobesedila. S korpusno analizo je namreč postalo jasno, da ilokucijsko moč posameznih izrazov prepoznavamo, ker jih sestavljajo **konvencionalne jezikovne oblike**, ki si jih delijo govorniki v jezikovni skupnosti (Stubbs 2007); ker so **produkt družbene konvencije**, in ne zato, ker bi bili razvidni iz semantične in skladijske zgradbe pomenskih enot oz. razmerij med njihovimi sestavinami. Poskus, da bi ločevali med nekakšnim stabilnim semantičnim pomenom in pragmatičnim, od konteksta odvisnim pomenom, se tako zdi vse bolj neutemeljen.

Na tem mestu smo izpostavili zgolj inovacijo LBS na področju pomenskih opisov, drugi pomembni vidiki leksiko-gramatičnega opisa LBS so izčrpno podani v ustreznih dokumentih (npr. Gantar idr. 2009, 2011). Morebitne slabosti LBS izhajajo iz dejstva, da je imela v njej osrednjo vlogo sintagmatika, torej skladijske strukture, stavčni vzorci in kolokacije, kar je omejilo možnosti za doslednejšo izpostavitve inherentnih slovničnih lastnosti besed in paradigmatičnih razmerij. Prav tako bi bilo mogoče nadgraditi obravnavo pragmatike, zlasti v smislu sistematičnega kodiranja in označevanja (govorna dejanja, aksiološki vidiki pomena itd.). Kljub temu menimo, da je leksikalna zbirka dokaj uspešno zasledovala izhodiščni cilj ustvariti model za celovit leksikalni opis slovenskega besedišča z osredotočanjem na leksikografsko pomembna dejstva. V izdelavo LBS so bila vložena štiri leta intenzivnega dela širše skupine sodelavk in sodelavcev, na podlagi tega in ob pomoči še nekaterih drugih rezultatov projekta *Sporazumevanje v slovenskem jeziku* ESS in RS (2008–2013) pa je nastal tudi Predlog za Slovar sodobnega slovenskega jezika (dalje: PSSSJ). V nadaljevanju se opredeljujemo do nekaterih rešitev, ki so v njem zapisane.

3 Pomisleki: viri, metode in organizacija dela

3.1 Korpusni viri za SSSJ

Predlagani osnovni vir je – poleg gradiva v LBS – milijardni referenčni korpus **Gigafida**, iz katerega je bil izpeljan še taksonomsko uravnoteženi, reprezentativni 100-milijonski podkorpus **KRES** (Logar Berginc idr. 2012).³ Uravnoveženost korpusa je za ustrezno sliko “sodobnega slovenskega jezika” pomembna vsaj toliko kot njegova velikost, zato predlagamo ponovno ovrednotenje vloge korpusov KRES, Gigafida in morebitnih dodatnih besedilnih virov pri izdelavi SSSJ. Nova slovarska podoba slovenščine bo namreč zgolj in samo odraz taksonomije besedil v izbranem korpusu; vključevala bo jezik, ki ga najdemo v najrazličnejših

³ Korpusa sta nastala v okviru projekta *Sporazumevanje v slovenskem jeziku*, www.slovenscina.eu. Več na: <http://www.slovenscina.eu/korpusi/gigafida> in <http://www.slovenscina.eu/korpusi/kres>.

zvrsteh in žanrih današnje pisane (in govorjene) slovenščine, kot taka pa bo tudi presegala tradicionalno delitev na knjižni in neknjižni jezik. Z odločitvijo o zasnovanju slovarske baze na izjemno obsežnem korpusnem gradivu in posledično o avtomatizaciji postopkov v prvi fazi izdelave je povezana tudi potreba po uvedbi “množičenja”.

3.2 Avtomatsko luščenje podatkov in “množičenje”

Sodobni slovarski portali ponujajo uporabnikom možnost aktivnega vključevanja, tipično tako, da ti prispevajo predloge za nova gesla (ang. UGC ali User-generated Content). Pozitivna plat podobnih projektov je zlasti v tem, da popularizirajo leksikografske vire in postopke, npr. **Macmillan English Dictionary Online**, ali nekoliko drugače ameriški **Wordnik**,⁴ ki po pričakovanjih kaže, da dajejo splošni uporabniki večji pomen partikularnemu kot tipičnemu in pogostemu. Ker tudi PSSSJ predvideva obsežen segment dela, ki bi ga v postopku izdelave novega slovarja opravile “množice”, je temu vredno posvetiti nekaj razmisleka. Zagotovo je del leksikografovega dela takšne narave, da ga je treba čim bolj avtomatizirati, ali, kot ugotavlja M. Rundell (2013), “smer potovanja je jasna: postopoma oblikujemo nabor robustnih aplikacij, ki bodo družno usmerjale proces sestavljanja in urejanja slovarskega besedila.” Vse to z namenom izdelave slovarjev, ki so bolj zanesljivi, notranje bolj konsistentni in manj odvisni od subjektivnih presoj avtorjev. T. i. množičenje (ang. crowdsourcing) ponuja možnost, da delo, ki je lahko dokaj rutinsko in ne zahteva nujno jezikoslovnega znanja, opravijo kar splošni uporabniki spleta, kar naj bi prihranilo čas in denar (npr. Fišer in Tavčar 2013: 125), saj se od udeležencev tipično pričakuje prostovoljna udeležba. Ugotovitve o tem, ali je razmerje med kvaliteto in vloženim delom oz. sredstvi dovolj ugodno, niso enoznačne, znano pa je, da je v projektih, ki izkoriščajo moč množic, težje in dražje nadzorovati kvaliteto (Ooi 2010). Koncept množičenja v kontekstu majhne slovenske govorne skupnosti je sam po sebi vprašljiv. Pilotski projekt na manjših segmentih gradiva, kot je poskus množičenja z orodjem sloWCrowd (2012–2013) (Fišer in Tavčar 2013), vzpostavljen za odpravljanje napak v avtomatsko zgrajenem semantičnem leksikonu za slovenščino sloWNet (Fišer 2009), je pokazal, da je na ta način mogoče pridobiti relativno majhno skupino dejavnih sodelavcev,⁵ ki delujejo od zunaj in “lahko v proces sodelovanja pri izdelavi slovarja poljubno vstopajo in izstopajo” (PSSSJ: 55). Kot je mogoče sklepati iz opisa v PSSSJ in Kosem idr. (2013), je narava vprašanj v

⁴ <http://www.wordnik.com/>.

⁵ V času enega leta se je k reševanju prijavilo 310 uporabnikov, pri čemer je bila distribucija števila odgovorov zelo neenakomerna (100 uporabnikov 10 nalog ali manj, število uporabnikov, ki so prispevali več kot 500 odgovorov, pa le 11).

načrtovanem projektu zahtevnejša, količina zahtevanih odločitev pa občutno večja (100.000 lem oz. 26.500.000 predvidenih odločitev). Ne nazadnje, če v zameno za 1 strokovno odločitev pridobimo (in plačamo) 5 odločitev laičnih uporabnikov, se zastavlja vprašanje, koliko z množičenjem v resnici privarčujemo. Poleg tega bi bilo treba upoštevati realne vsebinske zadrege, ki bi lahko ovirale udeležbo širše množice pri reševanju nalog, tj. o ravni strokovnosti jezika, ki bo uporabljen v programu za množičenje, in s tem povezana dostopnost podatkov, na podlagi katerih se bo uporabnik odločal med možnimi rešitvami. Ob dejstvu, da izraba moči množic uvaja v slovarski proces celo **ново polje delovnih nalog** (vzpostavljanje in vzdrževanje ustrezne tehnične podpore, sprotna evalvacija obdelanih podatkov, izdelava navodil za uporabnike spleta, mobilizacija, nadzor nad potekom itd.), ki niso časovno zanemarljive, in da predpostavlja **povsem novo delitev ključnih leksikografskih nalog in postopkov**, prinaša ta segment izdelave glede na nepreizkušeno metodo v vsakem primeru veliko neznank.

3.3 Kadrovski in finančni vidiki

Finančni razrez za segment množičenja, ki ne zahteva strokovne usposobljenosti, predvideva finančni delež (pribl. 600.000 €), le malo manjši od deleža (860.000 €), namenjenega strokovno usposobljeni leksikografski ekipi, ki naj bi delala honorarno (PSSSJ: 105), kar nas pripelje do vrednotenja leksikografskega dela. V predlogu za SSSJ je kadrovska struktura zastavljena piramidno, pri čemer je redna zaposlitev predvidena le za ožji krog članov ekipe. V situaciji, ko so sredstva za raziskave in nove projekte omejena, se tako perpetuirajo projektna praksa, ki temelji na prekernih oblikah zaposlitve večjega dela strokovne ekipe. Izražamo pobudo, da se bodoči javni leksikografski projekti načrtujejo tako, da ne bodo izrabljali že tako šibkega položaja sicer ozko specializirane skupine jezikoslovcev leksikografov, praviloma s statusom samozaposlenih v kulturi, saj bo morala glede na predvidevanja in izkušnje iz preteklih projektov ta skupina – če naj bo projekt izveden v 5 letih – prevzeti tudi večji del redatorskega dela v t. i. rumeni fazi in končnega pregleda gesel (PSSSJ: 105). Avtorska in redatorska faza izdelave slovarja, leksikalne zbirke itd. predstavljata jedro vsakega leksikografskega projekta. Ob predpostavki, da bodo za to delo izbrani strokovno usposobljeni sodelavci z izkušnjami, ni prav nobenega razloga, da bi bili za primerljivo delo in izobrazbo deležni manjšega plačila in socialne varnosti kot tisti s statusom raziskovalca. Velja poudariti, da je **ustrezna usposobljenost ekipe ključnega pomena**, saj čas, iz različnih razlogov vložen v usposabljanje vedno novih kadrov in razširjeno redakcijo njihovih gradiv,

občutno poviša stroške izdelave. Predvsem pa je utečenost oz. stalnost ekipe bistvenega pomena za **doseganje optimalnega razmerja med učinkovitostjo in kakovostjo leksikografskega dela.**

Poleg tega, da novi slovar slovenskega jezika ne more (p)ostati talec prestižnih bojev med posamezniki ali skupinami strokovnjakov, so prav vsi, ki trdijo, da si prizadevajo za kvaliteten novi slovar slovenščine, dolžni osebne interese in preference podrežati interesu javnosti. To med drugim pomeni, da bodo z vso odgovornostjo sestavili slovarsko ekipo izključno na podlagi strokovnih meril, z ustreznim vodenjem pa zagotovili optimalen izkoristek potenciala in znanja prav vsakega posameznega sodelavca pri projektu.

4 Zaključek

Izhodišče opisanih pogledov in pomislekov je mnenje, da mora predlog projekta za novi slovar slovenščine prinašati realno oceno možnosti izvedbe za obljubljeni obseg in kvaliteto v danih časovnih, finančnih in kadrovskih okvirih. V prvem poglavju smo orisali **dve osnovni možni izbiri strokovnih izhodišč** in se zavzeli za komunikacijsko-pragmatični pristop, ki bo proizvedel sodobni sporazumevalni tip slovarja, namenjen zlasti mlajšemu in/ali manj zahtevnemu uporabniku slovenščine, kot smo ga predvideli v Leksikalni bazi za slovenščino. V nadaljevanju smo razpravljali o **virih, izboru leksike in dejanskih potrebah** ter predstavili kritični pogled na nekatere **organizacijske in kadrovske vidike** PSSSJ, za zaključek pa navajamo še nekaj konkretnih predlogov:

Slovar v predlaganem obsegu, tj. 100.000 lem, je po našem mnenju zastavljen dokaj velikopotezno. Po naši oceni bi bila v predlaganem vsebinskem okviru realna izdelava na Leksikalni bazi temelječega **srednje velikega slovarja v obsegu 50.000–70.000 iztočnic**, ob upoštevanju delovnih postopkov, ki so bili vzpostavljeni na projektu SVSJ in v tem času ustrezno preizkušeni.

Moč množic bi bilo morda smiselno izrabiti zlasti v **segmentu sprotnega dopolnjevanja gradiva** z najsodobnejšimi rabami besed in novimi besedami, na primer v obliki jezikovnega foruma, segment prve faze pa prilagoditi tako, da bo predstavljal **minimalno tveganje za potek dela** v odločilni, “rumeni” fazi. Gotovo je treba upoštevati tudi, da se relativna uspešnost avtomatskih postopkov, preizkušenih v okviru LBS, nanaša na besede z enim ali največ dvema pomenoma, ki tako za računalnik kot človeške sestavljalce ne predstavljajo največjega izziva.

Kljub načelnemu stališču MZK, da se javna sredstva investirajo v en sam novi slovar slovenskega jezika, želimo opozoriti, da bi bil glede na večkrat izražena nekompatibilna strokovna izhodišča dveh glavnih skupin pobudnikov morda vendarle na mestu razmislek o dveh ločenih slovarskih projektih, ne zato, ker bi se na ta način želeli izogniti strokovnim nesoglasjem, temveč zato, ker **gre**, kot smo pokazali, **v resnici za dva povsem različna slovarska koncepta**; poleg tega en sam “temeljni” slovar (glede na obseg, izbrano besedišče, težavnost metajezika, tip zgledov itd.) **ne more zadovoljiti potreb vseh govorcev**. Tudi številni navedki v PSSSJ poudarjajo potrebo po diferenciaciji uporabnikov in vsebin (npr. 3.1.1), pri čemer so tisti, ki jih je za svojo ciljno skupino izbrala ekipa avtorjev ob načrtovanju LBS, predvsem mlajši in jezikovno nespecializirani uporabniki: “Določitev končnega uporabnika je za izdelavo slovarja ključnega pomena in mora biti izhodišče tako pri izbiri medija kot sestavnih delov slovarske vsebine” (PSSSJ: 20).

Predlagamo tudi **izvedbo preliminarne raziskave**, npr. z anketiranjem, ki bo ugotovila, kakšne so dejanske potrebe in pričakovanja različnih skupin slovarskih uporabnikov. V vsakem primeru bo lahko šele odziv uporabnikov na končni izdelek podal dokončno sodbo o uporabnosti, kvaliteti, zanesljivosti (kateregakoli) slovarja kot tudi upravičenosti naših pričakovanj in kritike.

Literatura

Slovarji in podatkovne zbirke

Collins COBUILD English Language Dictionary, (1. izdaja), Sinclair, J. (Ed.), 1987, London, HarperCollins (CCELD).

Leksikalna baza za slovenščino - <<http://www.slovenscina.eu/spletni-slovar>>.

Macmillan English Dictionary Online - <<http://www.macmillandictionary.com/>>.

Slovar slovenskega knjižnega jezika - <bos.zrc-sazu.si/sskj.html>.

Sporazumevanje v slovenskem jeziku - <www.slovenscina.eu>.

Wordnik - <<http://www.wordnik.com/>>.

Druga literatura

COOK, Paul, idr., 2013: A lexicographic appraisal of an automatic approach for detecting new word-senses. V: KOSEM, Iztok (ur.), idr.: *Electronic lexicography in the 21st century : thinking outside the paper : proceedings of eLex 2013 Conference*, 17.–19. oktober 2013, Tallinn, Estonia. Ljubljana: Trojina, zavod za uporabno slovenistiko. Tallinn: Eesti Keele Instituut, 2013. 49–65.

FIŠER, Darja, in TAVČAR, Andrej, 2013: Več glav več ve: uporaba množičenja za čiščenje sloWNeta. V: ŽELE, Andreja (ur.). *Družbena funkcijskost jezika : vidiki, merila, opredelitve*, Obdobja 32. 1. natis. Ljubljana: Znanstvena založba Filozofske fakultete, 2013. 125–132.

FIŠER, Darja, 2009: Pristopi za avtomatizirano gradnjo semantičnih zbirk. *Terminologija in sodobna terminografija*. Ljubljana: Založba ZRC, ZRC SAZU. 357–370.

GANTAR, Polona, Katja GRABNAR, Polonca KOCJANČIČ, Simon KREK, Olga POBIRK, Rok REJC, Mojca ŠORLI, Simon ŠUSTER in Petra ZARANŠEK, 2009: *Specifikacije za izdelavo leksikalne baze za slovenščino (Kazalnik 6): Projekt Sporazumevanje v slovenskem jeziku*: Kamnik. Slovenija.

http://www.slovenscina.eu/Media/Kazalniki/Kazalnik6/SSJ_Kazalnik_6_Specifikacije-leksikalna-baza_v1.pdf.

(Dostop 23. 1. 2014.)

GANTAR, Polona, KOSEM, Iztok, KREK, Simon, ŠORLI, Mojca, 2011: *Leksikalna baza za slovenščino. Navodila za avtorje*. Projekt Sporazumevanje v slovenskem jeziku (2008–2013).

HANKS, Patrick, 2013: *Lexical Analysis: Norms and Exploitations*. MIT Press.

KOSEM, Iztok, GANTAR, Polona, in KREK, Simon, 2013. Automation of lexicographic work: an opportunity for both lexicographers and crowd-sourcing. KOSEM, Iztok (ur.), idr.: *Electronic lexicography in the 21st century : thinking outside the paper : proceedings of eLex 2013 Conference*, 17.–19. oktober 2013, Tallinn, Estonia. Ljubljana: Trojina, zavod za uporabno slovenistiko. Tallinn: Eesti Keele Instituut, 2013. 32–48.

KREK, Simon, KOSEM, Iztok, GANTAR, Polona, 2013: Predlog za Slovar sodobnega slovenskega jezika (PSSSJ).

LOGAR BERGINC, Nataša, ARHAR HOLDT, Špela, GRČAR, Mitja, BRAKUS, M., KREK, Simon, 2012. Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

MOON, Rosamund, 1998: *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon.

MOON, Rosamund, 2008: Sinclair, Phraseology and Lexicography. *An International Journal of Lexicography* 21(3). 243–254.

OOI, Vincent B. Y. 2010: *English Internet Lexicography and Online Dictionaries*. *Lexicographica*. Volume 26. 143–154.

- RUNDELL, Michael, in KILGARRIFF, Adam, 2011: Automating the creation of dictionaries: where will it all end? V: Meunier, F., Cock, S. D., Gilquin, G., and Paquot, M. (ur.): *A Taste for Corpora. In honour of Sylviane Granger*. Amsterdam, Netherlands: John Benjamins. 257–282.
- SINCLAIR, John McH., 1987: *Looking Up. Account of the Cobuild Project in Lexical Computing* (Collins Cobuild dictionaries), London in Glasgow: Collins.
- Slovar novejšega besedja slovenskega jezika (SNB), 2012: Ljubljana: SAZU.
- Slovar slovenskega knjižnega jezika (SSKJ), 1994: Ljubljana: SAZU, DZS.
- STUBBS, Michael, 2007: On texts corpora and models of language. V: M. Hoey, M. Mahlberg, M. Stubbs in W. Teubert (ur.): *Text, Discourse and Corpora: Theory and Analysis*. London in New York: Continuum. 127–161.
- ŠORLI, Mojca, 2012: Semantična prozodija v teoriji in praksi – korpusni pristop k proučevanju pragmatičnega pomena : primer slovenščine in angleščine. V: ŠORLI, M. (ur.): *Dvojezična korpusna leksikografija : slovenščina v kontrastu: novi izzivi, novi obeti* (Zbirka Trojinski konj). 1. izd. Ljubljana: Trojina, zavod za uporabno slovenistiko, 2012. 90–116.
- VIDOVIČ MUHA, Ada, 2013a (1. izdaja 2000): *Slovensko leksikalno pomenoslovje. Govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.