

Rezultati projekta "Slovar sopomenk sodobne slovenščine: od skupnosti za skupnost"

Špela Arhar Holdt,* Jaka Čibej*

* Center za jezikovne vire in tehnologije, Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
Večna pot 113, 1000 Ljubljana
spela.arhar@cjvt.si, jaka.cibej@cjvt.si

Povzetek

V prispevku predstavljamo rezultate projekta "Slovar sopomenk sodobne slovenščine: od skupnosti za skupnost", ki je izvedel vrsto diseminacijskih aktivnosti z namenom ozaveščanja javnosti o izidu Slovarja sopomenk sodobne slovenščine in o novem konceptu odzivnega slovarja, ki je digitalno zasnovan in v razvoj vključuje tudi uporabniško skupnost. Poleg diseminacije se je projekt osredotočal tudi na razvoj metod za spremljanje eksplicitne in implicitne uporabniške povratne informacije, ki pomaga nadgrajevati odzivne jezikovne vire v smeri večje uporabniške prijaznosti, ter z motivacijo uporabnikov za sodelovanje pri odzivnih slovarjih.

"The Thesaurus of Modern Slovene: By the Community for the Community": Project Results

The paper presents the results of a project titled "The Thesaurus of Modern Slovene: By the Community for the Community". The project carried out a range of dissemination activities with the aim of informing the public about the publication of the Thesaurus of Modern Slovene and the new concept of a responsive dictionary, which is digital-born and includes the language user community in its development. Aside from dissemination, the project also focused on developing methods for collecting explicit and implicit user feedback to help upgrade responsive language resources, particularly in terms of user-friendliness, as well as the motivation of users to contribute to the development of responsive dictionaries.

1. Uvod

Prispevek opisuje namen in rezultate projekta "Slovar sopomenk sodobne slovenščine: od skupnosti za skupnost", ki ga je v letih 2018 in 2019 financiralo Ministrstvo za kulturo Republike Slovenije v sklopu Javnega razpisa za (so)financiranje projektov, namenjenih promociji in uveljavljanju slovenskega jezika na področju jezikovnih virov, tehnologij in digitalizacije (JR-Promocija SJ-2018–2019). Projekt v finančnem obsegu 40.000 EUR je prijavila Fakulteta za računalništvo in informatiko Univerze v Ljubljani, izvedba pa je potekala v okviru razvojnih nalog Centra za jezikovne vire in tehnologije UL.

Vsebinsko je bil projekt vezan na Slovar sopomenk sodobne slovenščine, ki je na voljo od pomladi 2018 (<https://viri.cjvt.si/sopomenke/>). Slovar je v slovaropisni prostor uvedel nov tip jezikovnega vira: odzivni slovar. Slednji je v izhodišču pripravljen z avtomatskimi metodami, njegov nadaljnji razvoj pa poteka stopenjsko, odprto in v sodelovanju z uporabniško skupnostjo.

Primarni namen projekta je bil obvestiti javnost o novem slovarju, prednostih in omejitvah odzivnih slovarjev ter možnostih za uporabniško sodelovanje pri njihovem razvoju, s posebno mislijo na uporabniške skupine, za katere so informacije o sopomenskosti najbolj dragocene. V širšem smislu smo želeli s projektom opozoriti na novosti, ki jih skupnosti prinaša digitalno slovaropisje, na pomen prosto dostopnih in podatkovno odprtih jezikovnih virov, kot tudi vrednost uporabniške povratne informacije za njihov nadaljnji razvoj – ter uporabnike za sodelovanje pri razvoju tudi usposobiti in motivirati.

Pomemben del projekta je bilo zbiranje povratne informacije o novorazvitem odzivnem konceptu. V

projektu so bile izvedene različne uporabniške študije: anketa o novostih odzivnega slovarja, uporabniška evalvacija slovarskega vmesnika, analiza avtentičnih jezikovnih težav, ki jih slovar sopomenk lahko naslovi. Razvita je bila metodologija za spremljanje uporabniških aktivnosti med slovarsko rabo in izdelan je bil sopomenski modul za mobilno aplikacijo *Igra besed*, s katero v model odzivnega slovarja vključujemo igrifikacijo.

Rezultati projekta, vključno s projektnimi poročili, so na voljo na spletni strani <https://www.cjvt.si/promocija-sopomenk/>, niso pa še bili celovito predstavljeni raziskovalni skupnosti. Slovaropisne rešitve, ki omogočajo dostop do odprtih jezikovnih podatkov v čim krajšem času, so ključne za vzpostavitev in vzdrževanje temeljne jezikovne infrastrukture. Odpiranje slovaropisnih delotokov za sodelovanje uporabniške skupnosti je skladno s trendi digitalnega slovaropisja in pričakovani slovarskih uporabnikov v digitalni dobi. V tem kontekstu je projekt relevanten zlasti zato, ker preizkuša, evalvira in nadgrajuje koncept odzivnega slovarja, ki je možni odgovor na izpostavljene izzive in bo v nadaljevanju apliciran tudi na druge temeljne jezikovne vire za slovenščino.¹

2. Slovar sopomenk sodobne slovenščine kot odzivni slovar

Slovar sopomenk sodobne slovenščine je prvi primer odzivnega slovarja, ki se od drugih jezikovnih virov v določenih vidikih razlikuje in na slovaropisno področje vnaša nekatere novosti. Kot prvo, slovar je povsem digitalen in zasnovan izključno za digitalni medij, kar pomeni dvoje: a) zasnovan ni le kot slovar, temveč tudi kot odprto dostopna (slovarska) baza podatkov, ki jo je mogoče uporabiti za razvoj drugih jezikovnih virov in orodij za

¹ V dveh letih od objave Slovarja sopomenke sodobne slovenščine sta po tem modelu že nastala Kolokacijski slovar sodobne slovenščine (opisano v Kosem et al., 2018, dostopno prek vmesnika na <https://viri.cjvt.si/kolokacije/slv/>), slovarska baza na CLARIN.SI: <http://hdl.handle.net/11356/1250> (Kosem

et al., 2019)) in oblikoslovni leksikon Sloleks 2.0 (dostopno prek vmesnika na <https://viri.cjvt.si/sloleks/slv/>), slovarska baza na repozitoriju CLARIN.SI: <http://hdl.handle.net/11356/1230> (Dobrovoljc et al., 2019)), v načrtu pa so tudi novi viri.

obdelavo naravnega jezika; in b) uporabniku je na voljo preko spletnega vmesnika, ki je bil razvit tako za računalnike kot za mobilne naprave.

Kot drugo, slovarska baza je v začetku zgrajena avtomatsko – v primeru Slovarja sopomenk sodobne slovenščine je bila baza zgrajena (Krek et al., 2017) na osnovi baze Velikega angleško-slovenskega slovarja Oxford DZS® (Šorli et al., 2006) in korpusa Gigafida 1.0 (Logar et al., 2012). Gradnja je temeljila na vzpostavljanju povezav med prevodnimi ustreznici v slovenskem delu Velikega angleško-slovenskega slovarja Oxford DZS® in na računanju relevantnosti teh povezav z algoritmom Personal PageRank (Page et al., 1999). Tako izračunane povezave so bile podlaga za sopojavitvene grafe, ki so služili kot osnova za iztočnice in nabor njihovih sopomenk v Slovarju sopomenk, sopomenke pa so bile naknadno razdeljene še na bolj relevantne (*jedrne sopomenke*, ang. *core synonyms*) in manj relevantne (*bližnje sopomenke*, ang. *near synonyms*) – v vmesniku sta kategoriji prikazani v ločenih razdelkih. Povezave med besedami so bile dodatno uravnotežene s korpusom Gigafida 1.0 in preverjene s Slovarjem slovenskega knjižnega jezika.

Kot tretje, odzivni slovar je objavljen relativno hitro po avtomatski izgradnji. Ustreznost rezultatov avtomatske gradnje z evalvacijo najprej potrjuje jezikoslovci, slovar pa je nato objavljen, čeprav baza kot avtomatsko generiran vir vsebuje določeno mero šuma. Iz tega razloga – in v skladu s sodobnimi trendi v slovaropisju, ki vse bolj vključuje množičenje v gradnjo jezikovnih virov (Čibej et al., 2015; Fišer in Čibej, 2015) – je v razvoj odzivnega slovarja vključena tudi uporabniška skupnost. Pri Slovarju sopomenk lahko uporabniki k izboljšavi slovarja pripomorejo na dva načina: a) s pomočjo razdelka v vmesniku lahko k iztočnici dodajajo manjkajoče sopomenke; b) kot ustrezne ali neustrezne lahko ocenjujejo obstoječe sopomenske kandidate v slovarju (tako uporabniško dodane kot tiste, ki so bile v slovar vključene z avtomatskim postopkom). Uporabniške sopomenke in glasovi se takoj pojavijo v vmesniku, tako pridobljeni podatki pa so upoštevani pri nadgradnjah slovarja. Vse uporabniške aktivnosti v slovarju so tudi sistematično in dosledno beležene v dnevnikih iskanja (Arhar Holdt et al., 2018: 407–408), kar omogoča tudi nadzor nad nekonstruktivnimi vnosi.

Odzivni slovar je odprto dostopen, se redno posodablja in je objavljen v različicah. V vmesniku Slovarja sopomenk je datum zadnje posodobitve izpisan ob iztočnici kot časovni žig, kar omogoča sledljivost. Vse različice slovarja so (in bodo) dostopne pod licenco Creative Commons 4.0 BY-SA in arhivirane v repozitoriju CLARIN.SI (npr. Krek et al., 2018 za Sopomenke 1.0), zato ima skupnost vedno možnost dostopati tudi do podatkov iz prejšnjih različic slovarja. Razvoj odzivnega slovarja tako nikoli ni končan – baza se postopoma veča oz. prečiščuje, slovar pa se z upoštevanjem uporabniškega doprinosa odziva tako na mnenja uporabniške skupnosti kot na spremembe v jeziku in se spreminja v skladu z jezikovno rabo – od tod tudi poimenovanje odzivni slovar.

3. Diseminacija slovarja v širši javnosti

Ena od aktivnosti v začetnem delu projekta je bilo obveščanje širše javnosti o izidu Slovarja sopomenk sodobne slovenščine, o novostih, ki jih prinaša odzivni slovar kot nov tip slovarja, posebej pa so bili uporabniki

pozvani tudi k sodelovanju pri nadaljnjem razvoju slovarja. Slovar je digitalen in predvideva sodelovanje skupnosti v digitalnem okolju, zato je velik del diseminacije potekal v digitalnem okolju. Ob tem smo poskrbeli tudi za diseminacijo med potencialnimi uporabniki, ki ostajajo zunaj digitalnega promocijskega dometa. Veliko pozornosti je bilo usmerjene v vzpostavljanje dialoga z uporabniki in zbiranje njihove povratne informacije. Diseminacija je potekala v štirih glavnih sklopih: obveščanje po e-pošti in po družbenih omrežjih, priprava promocijskih tiskovin, izvedba diseminacijskih dogodkov in anketa o novostih odzivnega slovarja.

3.1. Obveščanje po e-pošti in družbenih omrežjih

Za obveščanje po e-pošti je bila pripravljena obsežna adrema, ki je upoštevala tipologijo uporabniških skupin na področju slovarske rabe (Arhar Holdt, 2015) – vsebovala je vse najrelevantnejše skupine uporabnikov, v okviru katerih smo nato ciljno iskali naslovnike, za katere smo ocenili, da bi jih lahko zaradi poklicnih nalog, izobraževalnih potreb ali priložnostnih interesov Slovar sopomenk še posebej zanimal, npr. prevajalce, lektorje ter zaposlene v razvojnem sektorju, kulturi, izobraževanju in jezikovnih agencijah. Osredotočili smo se tudi na osnovne in srednje šole, fakultete in študentske klube ter različne skupine, kulturna in literarna društva, ki se z jezikom ukvarjajo ljubiteljsko. Slovar je bil diseminiran tudi s pomočjo strokovnih e-poštnih seznamov in skupin na družbenih omrežjih, kjer se skupnost organizira okrog jezikovnih vprašanj.

E-poštna sporočila o novem slovarju sopomenk so bila poslana skupno na 1.545 e-poštnih naslovov, od tega 661 osnovnim in srednjim šolam, 223 visokošolskim in univerzitetnim profesorjem, 58 študentskim klubom, 70 centrom, zavodom in agencijam, 67 skupinam Slovencev v tujini, 252 na različne naslove v povezavi s slovensko prevajalsko skupnostjo, 64 kulturnim in literarnim društvom in 150 knjižnicam.

3.2. Promocijske tiskovine

Za obveščanje v nedigitalnem svetu so bile pripravljene promocijske tiskovine v sodelovanju z Znanstveno založbo Filozofske fakultete Univerze v Ljubljani za sejem akademske knjige Liber.Ac 2018. Vsa nastala gradiva (plakat, zloženka, knjižna kazala) so dostopna na projektni spletni strani in jih je mogoče ponovno uporabiti za promocijske in izobraževalne namene. V sklopu projekta smo natisnili 30 plakatov, 500 zloženek in 18.500 kazal. Tiskovine smo uporabnikom večinoma poslali po pošti, manjši del pa smo jih razdelili tudi na promocijskih dogodkih. Poslanih je bilo 150 pošilk splošnim knjižnicam in večini šolskih knjižnic slovenskih osnovnih ter srednjih šol, vsaka pošiljka je vsebovala 100 kazal in informativno zloženko ter poziv prejemnikom, naj gradivo prosto ponudijo obiskovalcem knjižnice.

3.3. Diseminacijski dogodki

Pri izvedbi diseminacijskih dogodkov smo skušali pokriti čimbolj raznolike uporabniške skupine (od dijakov do učiteljev, lektorjev in raziskovalcev), pri čemer so imeli prednost tisti uporabniki, ki so sami izkazali interes za sodelovanje, ter tisti, kjer je bilo mogoče slovar predstaviti tudi v obliki delavnic. Izvedenih je bilo 16 dogodkov, med drugim predstavitev Slovarja sopomenk za lektorje

slovenščine v tujini, delavnice za dijake 3. in 4. letnikov, predstavitev na Lingvističnem krožku na Filozofski fakulteti UL, sopomenski turnir na sejmu Liber.ac 2019 in predstavitev dela na Slovarju sopomenk na konferenci eLex 2019. Celoten seznam dogodkov je opisan na spletni strani projekta. Predstavitvena gradiva so na voljo v spletni knjižnici in jih lahko kdorkoli uporabi za nadaljnje (samo)izobraževanje. Štiri diseminacijske dogodke smo posneli in podnaslovili, da so na voljo za ogled tudi po koncu projekta (gl. razdelek 4.2).

3.4. Anketa o novostih odzivnega slovarja

Za strukturirano zbiranje povratne informacije uporabnikov smo oblikovali spletno anketo, ki sprašuje o rabi slovarja, njegovi uporabnosti ter ponuja uporabnikom možnost, da opredelijo odnos do novosti odzivnega slovarja. Izvorno so anketo izpolnjevali udeleženci na dogodkih, nato pa smo jo diseminirali tudi širše. Anketo je v celoti izpolnilo 671 uporabnikov, kar je v kontekstu evropskih uporabniških raziskav relativno visoka številka.² Rezultate ankete smo tudi statistično analizirali in predstavili na konferenci eReL 2019 v Zagrebu. Anketa, ki je natančneje predstavljena v (Arhar Holdt, 2020), kaže, da so uporabniki naklonjeni do večine novosti, ki jih prinaša odzivni slovar, npr. stalno posodabljanje, digitalni format, kolokacijski podatki, povezave na korpus, uporabniško vključevanje. Kot izboljšavo med drugim predlagajo dodatne iztočnice (tudi nestandardne), dodatek kvalifikatorjev k zaznamovanim iztočnicam in sopomenkam, nekatere izboljšave vmesnika in večjo povezanost z drugimi jezikovnimi viri.

4. Multimedijska izobraževalna gradiva

Multimedijska gradiva služijo za (samo)izobraževanje ciljnih uporabniških skupin tudi po koncu projekta in širijo domet informacij na uporabnike, ki imajo lahko korist od uporabe podnapisov (v slovenščini ali angleščini). V projektu so bila pripravljena video navodila, kako uporabljati slovar, posnetki predavanj o izdelavi slovarja in značilnostih odzivnega koncepta za različne vrste javnosti, in informativno-promocijske poddaje (podcasti), v katerih so avtorji slovarja, raziskovalci, predavatelji in predstavniki različnih uporabniških skupin Slovar sopomenk umestili v širši kontekst svojega poklicnega dela.

4.1. Poglavje na Portalu jezikovnih virov

Portal jezikovnih virov (<https://viri.trojina.si/>) je spletna knjižnica izobraževalno-promocijskih videoposnetkov o načinih uporabe prosto dostopnih spletnih jezikovnih virov in tehnologij.³ Po uveljavljeni metodologiji (Arhar Holdt et al., 2016) je bilo za portal pripravljenih šest posnetkov, ki razlagajo: a) iskanje po slovarju; b) filtriranje in sortiranje podatkov; c) ocenjevanje sopomenk; d) dodajanje sopomenk; e) dostop do kolokacij in zgledov ter e) izvažanje podatkov. Poleg tega je na voljo tudi posnetek *Pet minut za Sopomenke*, ki predstavlja koncept slovarja,

njegovo zasnovano in namen. Vsi posnetki vsebujejo slovenske in angleške podnapise, ki so bodisi integrirani v sam posnetek ali dodani z orodji platforme YouTube.

4.2. Posnetki predavanj in prosojnice

V spletni knjižnici projekta so zbrane povezave na posnetke štirih predavanj za različne tipe javnosti: a) za širšo javnost (Iztok Kosem: *Kam se obrniti, ko zmanjka besed?*); b) za šolskega naslovnika (Andreja Pavle Jurman: *Slovar sopomenk sodobne slovenščine in njegova uporabnost pri pouku slovenščine*); c) za študente slovenistike in prevajalstva (Špela Arhar Holdt in Jaka Čibej: *Digitalno slovaropisje: Kdo je slovar?* in č) za znanstveno javnost (Špela Arhar Holdt: *Thesaurus of Modern Slovene: By the Community for the Community*).⁴

Ob posnetih predavanjih so na voljo tudi prosojnice, ki so jih uporabljali predavatelji na posameznem od dogodkov. Na tej osnovi smo kot dodatno izobraževalno gradivo pripravili dopolnjene in odprto dostopne prosojnice o nastanku in sestavi Slovarja sopomenk sodobne slovenščine, ki jih učitelji in predavatelji lahko prenesejo na svoj računalnik in prosto prilagodijo za svoje delo v razredu oz. predavalnici.

4.3. Serija oddaj Sopomenki

V projektu smo pripravili serijo desetih poddaj z imenom *Sopomenki*. Z avtorji slovarja in drugimi predstavniki raziskovalne skupnosti smo obravnavali naslednje teme: a) interdisciplinarno sodelovanje računalničarjev in jezikoslovcev (gosta dr. Kaja Dobrovoljc in dr. Marko Robnik Šikonja); b) jezikoslovni pristopi k sopomenskosti (gosta dr. Polona Gantar in dr. Vojko Gorjanc); c) slovaropisje v preteklosti in prihodnosti (gosta dr. Simon Krek in dr. Marko Stabej); d) značilnosti odzivnih slovarjev (gostji dr. Špela Arhar Holdt in Barbara Gorišek); e) uporabniško vključevanje (gosta dr. Iztok Kosem in Jaka Čibej) in f) šibke točke odzivnih slovarjev (gostji dr. Nataša Logar in Eva Pori). V oddajah so sodelovali tudi predstavniki uporabniške skupnosti: g) predavatelj dr. Damjan Popič in študentka prevajalstva Sara Sever; h) učiteljica slovenščine Andreja Pavle Jurman in dijakinja Gimnazije Ledina Neja Šubic; i) prevajalca Andreja Golob in Josh Rocchio ter j) dr. Mojca Stritar Kučuk in Tanja Jerman, ki učita slovenščino kot drugi/tuji jezik. Posnetke je mogoče prek spletne povezave poslušati na različnih napravah ali jih prenesti na lastno napravo za poslušanje brez povezave.

5. Spremljanje uporabniške povratne informacije

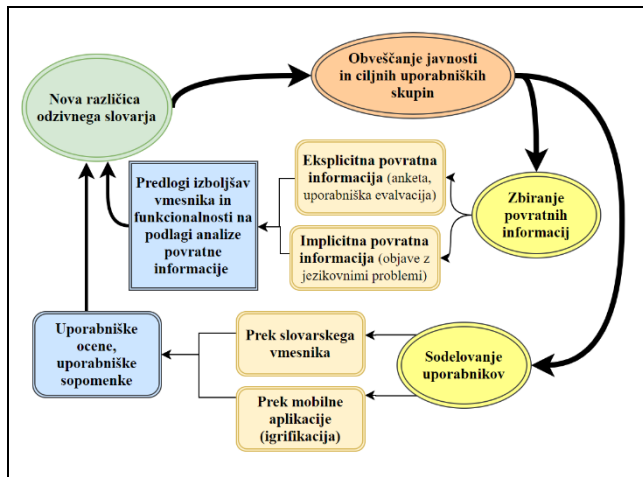
Projekt je zajemal razvoj in implementacijo postopkov za spremljanje uporabniških mnenj in odzivov na spletu ter metod za spremljanje različnih načinov uporabniškega vključevanja v slovarski razvoj. Uporabniško vključevanje pri razvoju odzivnega slovarja poteka po modelu, ki ga predstavlja Slika 1.

² Temeljna raziskava s področja slovarskih uporabnikov v digitalnem okolju, ki sta jo izvedla Koplenig in Müller Spitzer (2014: 83), je denimo vključevala 684 udeležencev.

³ Portal je bil zasnovan v letih 2014–15 s podporo Ministrstva za kulturo: "Izdelava spletne strani z opisi jezikovnih virov in orodij za slovenščino ter osnovnimi (video)navodili za njihovo

uporabo" in "Nadgradnja in popularizacija predstavitvenega portala spletnih jezikovnih virov za slovenščino".

⁴ Zadnje predavanje je bilo posneto v okviru mednarodnega kongresa Euralex 2018 in je dostopno na portalu VideoLectures (<http://videolectures.net/>). Ostale tri posnetke smo zagotovili v sklopu projekta, skupaj s slovenskimi in angleškimi podnapisi.

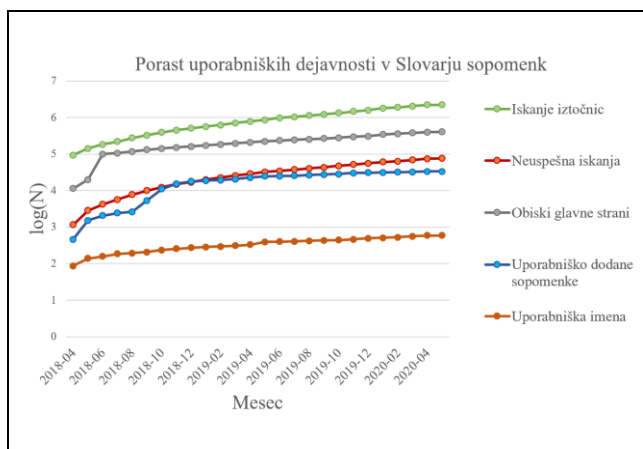


Slika 1. Uporabniško vključevanje v razvoj odzivnega slovarja.

Poleg *eksplicitnih povratnih informacij* uporabnikov, ki smo jih preverili z anketo o novostih odzivnega slovarja (gl. razdelek 3.4), smo v projektu spremljali tudi njihove *implicitne povratne informacije*, in sicer na dva načina: s spremljanjem dnevnikov iskanja in z analizo jezikovnih objav na družbenem omrežju Facebook.

5.1. Dnevnik iskanja

V projektu smo razvili metodologijo sistematičnega beleženja dnevnikov iskanja in spremljanja uporabniških aktivnosti med uporabo vmesnika Slovarja sopomenk. Beležijo se vsi podatki, ki jih je mogoče legitimno spremljati, od uporabniškega sodelovanja (ocenjene in dodane sopomenke) do iskanj (uspešnih in neuspešnih), premikov znotraj vmesnika (npr. razvrščanje, filtriranje podatkov, kliki na prikaz kolokatorjev in povezave na korpus Gigafida), izvoza podatkov, deljenja na družbenih omrežjih itd. (za natančnejši pregled gl. Arhar Holdt et al., 2018). Dnevnik izvažamo in analiziramo mesečno. Slika 2 prikazuje rast nekaterih uporabniških dejanj v slovarju od objave do 6. maja 2020.



Slika 2: Porast uporabniških dejavnosti v Slovarju sopomenk sodobne slovenščine.

Iz podatkov je med drugim razvidno, da uporabniki pogosto iščejo tudi iztočnice, ki jih v Slovarju sopomenk

sodobne slovenščine še ni. Do 6. maja 2020 je bilo zabeleženih 2.251.093 iskanj, 75.672 neuspešnih iskanj, 406.454 obiskov glavne strani, 597 uporabniških imen, 26.683 uporabniško dodanih sopomenk in 23.730 nenajdenih nizov. Seznam najpogostejših nenajdenih nizov razkrije, katere iztočnice je treba prioritarno dodati pri nadgradnji oz. kako izboljšati iskalno vrstico (npr. avtomatsko dodajanje prostega morfema se, če uporabnik vpiše glagol, ki brez morfema ne obstaja, npr. *smejati*).

Opazen je tudi stalen porast uporabniškega sodelovanja, kar gre neposredno pripisati tudi aktivnostim promocijskega projekta. Visok odziv je presegel pričakovanja in priča o naklonjenosti jezikovne skupnosti do slovarske vključevalnosti, obenem pa potrjuje vrednost diseminacijskih projektov za seznanjanje javnosti o možnosti in omejitvah, ki jih prinašajo odzivni slovarji.

Prve analize uporabniško dodanih sopomenk (Čibej, 2019: 13–16) kažejo, da so predlogi v veliki večini dobronamerni in konstruktivni: na uravnoveženem vzorcu 1.662 sopomenk (vzorec zajema vseh 467 uporabnikov, ki so do decembra 2019 v slovar dodali vsaj en vnos) je bilo več kot 70 % sopomenk ustreznih in slogovno nevtralnih, več kot 20 % pa ustreznih in slogovno zaznamovanih (regionalizmi, termini ipd.). Le dobrih 6 % gradiva je neustreznega (lapsusi, zlonamerni vnosi). Opombe v oklepajih (npr. v *športu*; *slabšalno*), ki jih uporabniki pripišejo dodanim sopomenkam, kažejo, da želijo uporabniki dodano besedišče ob vnosu tudi ustrezno kvalificirati. Vsebnost kvalifikatorjev, zlasti možnost opozarjanja na potencialno žaljivo ali celo sovražno besedišče, je bila tudi na drugih mestih projekta izpostavljena kot ključna točka za nadaljnji razvoj slovarja. V prvi slovarski nadgradnji bomo zato prioritarno dodali tovrstne kvalifikatorje, omogočili pa bomo tudi, da svoje vnose s temi oznakami kvalificirajo že sami uporabniki.

Veliko število uporabniških predlogov priča o tem, da lahko jezikovna skupnost pomembno izboljša uporabnost slovarja. Ključna pri odzivnem konceptu pa je tudi kontrola kakovosti uporabniških vnosov. Ti se v vmesniku najprej prikazujejo v ločenem, spodnjem delu slovarskega gesla, ki sugerira jezikoslovno nepregledanost, vmesnik pa uporabnikom omogoča tudi, da na predloge reagirajo (pozitivno ali negativno). Po jezikoslovnem pregledu se uporabniški predlogi lahko umestijo v slovarsko bazo in premestijo v gornji del gesla, v primeru, da so zlonamerni, pa se lahko tudi izbrišejo. Uredniški posegi, ki bodo zagotovili kontinuirano izboljšavo kakovosti slovarskega gradiva, se bodo opravljali ob vsaki slovarski nadgradnji.

5.2. Objave na družbenem omrežju Facebook

Izdelali smo skripto, s katero lahko iz vnaprej določenih jezikovnih uporabniških skupin na Facebooku izluščimo objave in komentarje nanje ter nekatere metapodatke (npr. čas objave, število komentarjev na objavo, število všečkov in ostalih odzivov), skladno z zakonodajo pa skripta poskrbi za varovanje osebnih podatkov, tako da anonimizira uporabniška imena. Skripta je jezikovno neodvisna in odprto dostopna na voljo drugim zainteresiranim raziskovalcem (gl. Čibej in Arhar Holdt, 2018).

S programom smo izluščili podatke iz Facebookovih skupin *Za vsaj približno pravilno rabo slovenščine* ter *Društvo ljubiteljskih pravopisarjev in slovničarjev*. Vzorec 500 najnovejših objav smo ročno uredili in tako dobili 501

jezikovno vprašanje. Vprašanja smo kategorizirali, da bi ugotovili najpogostejše uporabniške jezikovne zadrege, ki jih Slovar sopomenk (skupaj z ostalimi odzivnimi slovarji) lahko naslovi (Čibej, 2019: 10–13).

Največ vprašanj (158 oz. slabih 31 % vzorca), ki so relevantna za Slovar sopomenk, spada v scenarij, pri katerem uporabnik sprašuje za mnenje drugih uporabnikov pri odločanju med dvema različicama besed ali besednih zvez. Na ravni sopomenskih besed in razlike med njimi je uporabnikom v tem primeru slovar lahko v pomoč, saj omogoča primerjavo sopomenk v kontekstu. Ugotovitev nakazuje, da je treba pri nadgradnji še nekoliko več pozornosti nameniti primerjavi sopomenk s kolokatorji. V trenutni različici so kolokatorji, s katerimi je mogoče primerjati sopomenke, izluščeni avtomatsko, navedeni pa so samo najbolj tipični po besednih vrstah. Izboljšati bi jih bilo mogoče z evalvacijo, pri kateri bi uporabniki npr. izrazili, katere besedne vrste kolokatorjev se jim zdijo najbolj intuitivne za primerjavo (npr. kombinacije prislov + pridevnik, pridevnik + samostalnik) in podobno.

6. Motiviranje uporabnikov za sodelovanje

Za uporabniško sodelovanje, ki je pomemben del koncepta odzivnega slovarja (Slika 1), je treba skupnost ustrezno motivirati. Vmesnik Slovarja sopomenk sodobne slovenščine vsebuje lestvico uporabnikov glede na število vnesenih sopomenk, v zavijku Skupnost pa je vloga sodelovanja pri nadaljnjem razvoju slovarja tudi eksplicitno izpostavljena. Če motivacijo za sodelovanje delimo na družbeno, psihološko in ekonomsko (Lew, 2013), se vmesnik dotika predvsem družbene ravni: naslavlja uporabniško željo sodelovati v skupnosti in v njej doseči določen status. V okviru različnih možnosti, ki jih ponuja množičenje na področju slovaropisja (gl. tudi Čibej et al., 2015; Fišer in Čibej, 2015), je trenutna rešitev precej osnovna. Pomemben del projekta je bil zato identifikacija dodatnih motivacijskih možnosti, ki bi raziskovanje sopomenskosti in sodelovanje pri razvoju slovarja naredili uporabnikom privlačnejše in dolgoročneje zanimivo. Zelene prilagoditve smo na eni strani iskali s pomočjo uporabniških evalvacij slovarskega vmesnika, na drugi strani pa v smeri igrifikacije, ki bi v šolskem okolju lahko služila tudi didaktičnim namenom.

6.1. Uporabniška evalvacija slovarskega vmesnika

V sklopu projekta smo izvedli uporabniško evalvacijo slovarskega vmesnika, pri kateri je sodelovalo dvajset študentov jezikoslovnih smeri: deset s področja slovenistike in deset s področja prevajalstva. Evalvacije so potekale po metodi glasnega razmišljanja (ang. *think aloud protocol*), ki ji je sledil vodeni intervju z uporabniki (Štingl et al., 2018). Teme intervjuja so pokrile konceptualna vprašanja (npr. odnos do neprečiščenosti avtomatsko pridobljenih podatkov, nedokončanosti in spremenljivosti slovarja) kot tudi specifične lastnosti in funkcionalnosti vmesnika (npr. postavitev in poimenovanje gumbov, navigacija in preklapljanje med različnimi vrstami podatkov, opaznost in razumljivost vmesniških možnosti). Fokus je bil na vprašanjih, kaj uporabniki v zvezi s

slovarjem dojemajo kot moteče in kaj bi jih za sodelovanje pri razvoju slovarja posebej motiviralo.

Izsledke predstavlja poročilo, ki je dostopno v spletni knjižnici (Arhar Holdt, 2019). Glavni pomisleki evalvatorjev so bili vezani na nejasno kontrolo uporabniških vnosov, tako nad tem, kdo predloge v slovar (lahko) dodaja, kot nad tem, kaj v slovarju po uredniškem pregledu in potrditvi (lahko) ostane. Ker uredniški delotoki v trenutnem vmesniku niso jasno predstavljeni, si sodelujoči niso znali razlagati, kako bo nadgradnja slovarja potekala. Na drugi strani so izpostavili pomanjkljivosti mehanike uporabniškega sodelovanja: moteče je, da je pri ocenjevanju avtomatsko pridobljenega gradiva sopomenki mogoče obenem pripisati tako pozitiven kot negativen glas; kot drugo, vmesnik uporabniku ne omogoča, da izbriše lasten predlog, zato v slovarju na uredniški poseg čakajo tudi lapsusi, ki bi jih lahko odstranili uporabniki sami. Našteti in še drugi predlogi evalvatorjev so strnjeni v seznam nalog za vmesniško nadgradnjo, ki vsebuje tudi določene izboljšave slovarske baze, npr. opredelitev besedne vrste pri slovarskih iztočnicah in s tem pripravo ločenih slovarskih vnosov za enakopisne iztočnice.

Evalvacija je pokazala, da elementi tekmovalnosti, ki jih v vmesnik vnaša tekmovalna lestvica, na sodelujoče nimajo posebej spodbudnega vpliva. Motivacija, kot so jo opredelili sodelujoči evalvatorji, je predvsem psihološka: pomagati drugim, pomagati odpraviti specifične probleme slovarske baze. V družbenem smislu bi jih motiviralo, če bi sodelovanje pri razvoju slovarja lahko uporabili za referenco v okviru študijskih nalog ali kot vnos v življenjepis.

6.2. Sopomenski modul za Igro besed

Dobro izhodišče za motivacijo skupnosti ponuja igrifikacija. Tudi na področju razvoja jezikovne infrastrukture so že bile preizkušene t. i. igre z namenom (ang. *Games with a Purpose*, GWAPs), npr. za označevanje jezikovnih podatkov Phrase Detectives (Poesio et al., 2013), JeuxDeMots (Lafourcade, 2007) ali ZombiLingo (Guillaume, 2016). Ideja tovrstnih iger je, da motivirajo z igralskostjo, igranje pa je zasnovano tako, da obenem prispeva k določenemu uporabnemu cilju.

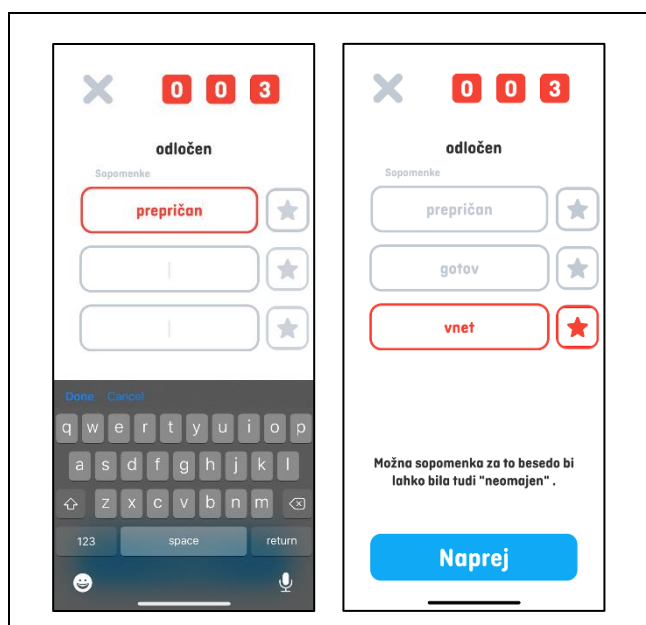
Igra z namenom, ki dopolnjuje model odzivnega slovarja, je mobilna aplikacija "Igra besed: kolokacije". Aplikacija, ki je nastala v projektu "Promocija jezikovne igralne aplikacije za mobilne naprave" (2018–2019),⁵ se osredotoča na prepoznavo in dopolnjevanje slovenskih kolokacij. Vpisovanje, razvrščanje in selekcioniranje kolokacijskega gradiva, ki ga v obliki izzivov ponuja igra, posredno pomaga urejati bazo Kolokacijskega slovarja sodobne slovenščine (Kosem et al., 2018).

V projektu "Slovar sopomenk sodobne slovenščine: od skupnosti za skupnost" je bil razvit nov modul igre, ki se osredotoča na sopomenske podatke – modul "Sopomenke". Pri izbiri in pripravi nalog je bil ključen razmislek o didaktični vrednosti uporabniškega sodelovanja za spoznavanje koncepta sopomenskosti, usvajanje besedišča in razvijanje jezikovnega metaaparata. Igra Sopomenke posledično združuje množičenje in jezikovno didaktiko ter je kot taka primerna tudi za uporabo v razredu.

⁵ Igra besed je bila prvotno razvita kot spletna aplikacija, dostopna na <https://www.igra-besed.si/>. Njeno pripravo je sofinanciralo Ministrstvo za kulturo RS v letu 2013.

Igra ponudi iztočnico, na katero igralka ali igralec vpiše do tri sopomenke. Vnosi se primerjajo z bazo Slovarja sopomenk in ustrezno točkujejo. Didaktično vrednost ima prikaz dodatne sopomenke iz baze (Slika 3), pa tudi progresivno stopnjevanje težavnosti igre po stopnjah, ki na začetku vsebujejo iztočnice z bogatim naborom sopomenk (npr. *hud*, *živahen*, *odličen*), kasneje pa iztočnice z redkejšimi sopomenkami (npr. *šumeč*, *utrujen*, *zvezda*). Učitelj ima možnost, da zahtevnost določene stopnje preveri vnaprej in za delo pri določeni šolski uri izbere težavnost, ki je ustrezna določeni stopnji ali skupini učencev.

Slika 3 prikazuje dva mobilna ekrana. Igralec se nahaja na stopnji 003 (skupno igra ponuja 500 stopenj s po 10 sopomenkami). Za iztočnico *odločen* igralec vnese sopomenke *prepričan*, *gotov* in *vnet*, od katerih je le zadnja v slovarski bazi, zato je vnos ocenjen z eno zvezdico. Kot "didaktično sopomenko" igra ponudi besedo *neomajen*. Z novimi sopomenkami se bo posodabljala tudi sama igra, s čimer bo izboljšana igralnost in uporabniška izkušnja z ocenjevanjem vnesenih besed.



Slika 3. Sopomenski modul v Igru besed.

7. Sklep in bodoče delo

Projekt "Slovar sopomenk sodobne slovenščine: od skupnosti za skupnost" prinaša rezultate, ki so uporabni za nadaljnje (samo)izobraževanje uporabniške skupnosti, kot tudi izsledke, kako skupnost razume in sprejema koncept odzivnega slovarja, katere novosti vidijo kot prednost in kaj se jim zdi v povezavi z novim konceptom problematično. Rezultati na drugi strani osvetljujejo, kakšne adaptacije bi bilo treba zagotoviti za izboljšavo uporabniške izkušnje z vmesnikom Slovarja sopomenk sodobne slovenščine. Predlogi uporabnikov so bili vključeni v načrt za nadgradnjo slovarja v različico 2.0, ki je že v pripravi.

Projektna izkušnja potrjuje, da je motivacija uporabnikov za sodelovanje pri nadaljnjem razvoju slovarja pomemben del odzivnega koncepta. Težave slovarja in širše odzivnega koncepta, ki jih uporabniki izpostavljajo kot demotivacijske, bodo naslovljene pri nadgradnji Slovarja sopomenk, kot tudi pri nadaljnjem

razvoju odzivnih virov. Na drugi strani vpeljava igrifikacije v odzivni model ponuja nove, še ne preverjene možnosti za sodelovanje skupnosti. V naslednjem koraku zato želimo oceniti učinkovitost tovrstnega sodelovanja in kvaliteto podatkovnega doprinosa, igralni modul pa bi radi tudi evalvirali s pomočjo uporabnikov in ga ustrezno izboljšali.

Koncept odzivnega slovarja je odgovor na potrebe novega časa, vendar bo pri njegovem razvoju tudi v bodoče nujno upoštevati uporabniško povratno informacijo. Model uporabniškega vključevanja, ki je bil razvit v projektu, je mogoče videti kot podstat za nadaljnji, kontinuirani razvoj v smer vedno boljšega rezultata, pri razvoju katerega – za razliko od preteklih praks – slovaropisje uporabnikov ne želi le "ciljati", ampak jih po mnenju, potrebah in vedenju aktivno sprašuje ter povratno informacijo v čim večji meri tudi upošteva.

8. Zahvala

Projekt, opisan v prispevku, je med letoma 2018 in 2019 financiralo Ministrstvo za kulturo Republike Slovenije. Raziskovalni program št. P6-0411 (Jezikovni viri in tehnologije za slovenščino) je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna. Raziskave množičenja in uporabniškega vključevanja v projektu so potekale v povezavi z vsebinami akcije COST CA160105 eNetCollect (European Network for Combining Language Learning with Crowdsourcing Techniques).

Zahvaljujemo se vsem, ki so prispevali k projektnim rezultatom: dr. Kaji Dobrovoljc, dr. Poloni Gantar, mag. Teji Goli, Andreji Golob, Barbari Gorišek, dr. Vojku Gorjancu, Tanji Jerman, Leonu Noetu Jovanu, dr. Iztoku Kosmu, dr. Simonu Kreku, dr. Cyprianu Laskowskemu, dr. Nataši Logar, Andreji Pavle Jurman, dr. Damjanu Popiču, Evi Pori, Juretu Preglau, Davidu Rajsterju, Joshu Rocchii, dr. Marku Robniku Šikonji, Sari Sever, dr. Marku Stabeju, dr. Mojci Stritar Kučuk, Mihi Šajini, Neji Šubic, Petri Štingl, Lenki Trdina in Gašperju Uršiču.

9. Literatura in viri

- Špela Arhar Holdt. 2015. Uporabniške raziskave za potrebe slovenskega slovaropisja: prvi koraki. V: Vojko Gorjanc, Polona Gantar, Iztok Kosem in Simon Krek, ur., *Slovar sodobne slovenščine: problemi in rešitve*, str. 136–149. Ljubljana: Znanstvena založba Filozofske fakultete.
- Špela Arhar Holdt, Kaja Dobrovoljc in Iztok Kosem. 2016. Predstavitveni portal spletnih jezikovnih virov za slovenščino. V: *Zbornik konference Jezikovne tehnologije in digitalna humanistika 2016*, str. 27–31.
- Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Polona Gantar, Vojko Gorjanc, Bojan Klemenc, Iztok Kosem, Simon Krek, Cyprian Laskowski in Marko Robnik Šikonja. 2018. V: *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts*, str. 401–410. Ljubljana: Ljubljana University Press.
- Špela Arhar Holdt. 2019. *Specifikacije za adaptacije slovarskega vmesnika. Poročilo projekta Slovar sopomenk sodobne slovenščine: Od skupnosti za skupnost*. Ljubljana: CJVT. <https://www.cjvt.si/promocija-sopomenk/knjiznica/>.
- Špela Arhar Holdt. 2020. How Users Responded to a Responsive Dictionary: The Case of the Thesaurus of

- Modern Slovene. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovje*, 46(2): v tisku.
- Jaka Čibej, Darja Fišer in Iztok Kosem. 2015. The role of crowdsourcing in lexicography. V: *Electronic lexicography in the 21st century: linking lexical data in the digital age: proceedings of eLex 2015 Conference*, str. 70–83. Ljubljana: Trojina, Institute for Applied Slovene Studies; Brighton: Lexical Computing.
- Jaka Čibej in Špela Arhar Holdt. 2018. Researching Dictionary Needs of Language Users Through Social Media: A Semi-Automatic Approach. V: *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts*, str. 67–76. Ljubljana: Ljubljana University Press.
- Jaka Čibej. 2019. *Končno poročilo o spremljanju uporabniških odzivov, mnenj in načinov uporabniškega vključevanja. Projekt Slovar sopomenk sodobne slovenščine: Od skupnosti za skupnost*. Ljubljana: CJVT. <https://www.cjvt.si/promocija-sopomenk/knjiznica/>
- Kaja Dobrovoljc et al. 2019. Morphological lexicon Sloleks 2.0, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1230>.
- Darja Fišer in Jaka Čibej. 2015. Potencial množičenja v sodobni leksikografiji. V: Vojko Gorjanc, Polona Gantar, Iztok Kosem in Simon Krek, ur., *Slovar sodobne slovenščine: problemi in rešitve*, str. 542–564. Ljubljana: Znanstvena založba Filozofske fakultete.
- Bruno Guillaume, Karén Fort in Nicolas Lefebvre. 2016. Crowdsourcing Complex Language Resources: Playing to Annotate Dependency Syntax. V: *Proceedings of the International Conference on Computational Linguistics (COLING)*. <https://hal.inria.fr/hal-01378980/>.
- Alexander Koplenig in Caroline Müller-Spitzer. 2014. The first two international studies on online dictionaries – background information. V: Caroline Müller-Spitzer, ur., *Using Online Dictionaries*. Berlin, Boston: De Gruyter Mouton.
- Iztok Kosem, Simon Krek, Polona Gantar, Špela Arhar Holdt, Jaka Čibej in Cyprian Laskowski. 2018. Collocations dictionary of modern Slovene. V: *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts*, str. 989–997. Ljubljana: Ljubljana University Press.
- Iztok Kosem et al. 2019. Collocations Dictionary of Modern Slovene KSSS 1.0, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1250>.
- Simon Krek, Cyprian Laskowski in Marko Robnik Šikonja. 2017. From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. V: *Proceedings of eLex 2017: Lexicography from Scratch*, str. 93–109. Leiden, Netherlands.
- Simon Krek et al. 2018. Thesaurus of Modern Slovene 1.0. Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1166>.
- Mathieu Lafourcade. 2007. Making people play for Lexical Acquisition with the JeuxDeMots prototype. V: *SNLP'07: 7th International Symposium on Natural Language Processing*, str. 7. Pattaya, Chonburi, Thailand.
- Robert Lew. 2013. User-generated content (UGC) in online English dictionaries. V: Andrea Abel, Annette Klosa, ur., *Ihr Beitrag bitte! – Der Nutzerbeitrag im Wörterbuchprozess (OPAL – Online publizierte Arbeiten zur Linguistik)*, str. 9–30. Mannheim: Institut für Deutsche Sprache.
- Nataša Logar, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek. 2012. *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- Lawrence Page, Sergey Brin, Rajeev Motwani in Terry Winograd. 1999. *The PageRank Citation Ranking: Bringing Order to the Web. Technical Report*. Stanford InfoLab.
- Massimo Poesio, Jon Chamberlain, Udo Kruschwitz, Livio Robaldo in Luca Ducceschi, 2013. Phrase Detectives: Utilizing Collective Intelligence for Internet-scale Language Resource Creation. V: *ACM Transactions on Interactive Intelligent Systems* 3, 3:1–3:44.
- Mojca Šorli, Katarina Grabnar, Simon Krek in Tomato Košir. 2006. Oxford-DZS comprehensive English-Slovenian dictionary. V: *Proceedings of the XII EURALEX International Congress*, str. 631–637. Edizioni dell'Orso: Università di Torino: Accademia della Crusca.
- Petra Štingl, Apolonija Gantar in Špela Arhar Holdt. 2018. Research on the User Experience of the Thesaurus of Modern Slovene. V: *Proceedings of MEi:CogSci Conference 2018, Bratislava*, str. 98.