

Končno poročilo o spremljanju uporabniških odzivov, mnenj in načinov uporabniškega vključevanja

Projekt Slovar sopomenk sodobne slovenščine: Od skupnosti za skupnost

* Poročilo vsebuje tudi *Specifikacije z opisom metodologije za napredno spremljanje spletnih uporabniških odzivov na jezikovne vire* (Delovni sklop 3).

Jaka Čibej (2019): Končno poročilo o spremljanju uporabniških odzivov, mnenj in načinov uporabniškega vključevanja. Projekt Slovar sopomenk sodobne slovenščine: Od skupnosti za skupnost. Ljubljana: Center za jezikovne vire in tehnologije.

Kazalo

1	Odzivni slovar in odzivni jezikovni viri	1
2	Uporabniške povratne informacije	2
2.1	Eksplicitna uporabniška povratna informacija	3
2.2	Implicitna uporabniška povratna informacija	8
2.3	Neposredni uporabniški doprinos k podatkovni bazi	13

1 Odzivni slovar in odzivni jezikovni viri

Slovar sopomenk sodobne slovenščine (dostopen na: viri.cjvt.si/sopomenke) vsebuje več kot 105.000 iztočnic in 300.000 sopomenk, s čimer je najboljšežnejša prosto dostopna avtomatsko generirana spletna zbirka sopomenk za slovenščino.

Slovar je v celoti zasnovan za uporabo v digitalnem okolju: uporabnik v iskalno okence vpiše iztočnico, za katero mu program ponudi sopomenske kandidate. Ker je pri razumevanju sopomenskosti nujno upoštevati besedilni kontekst, je v slovarju veliko pozornosti posvečene tudi povezavi s podatki o jezikovni rabi. Povsem nova v slovenskem prostoru je možnost, da uporabnik s pomočjo kolokacij (tipičnih sopojavitev besed) enostavno primerja rabo različnih sopomenk za posamezno iztočnico. Poleg tega so z rabo strojnih metod, ki omogočajo avtomatsko prepoznavo dobrih slovarskih zgledov (Kosem idr. 2011), v slovar uvoženi zgledi rabe, dodane pa so še povezave v korpus Gigafida, ki omogoča nadaljnje raziskovanje sodobne slovenščine.

Med glavnimi željami oz. potrebami slovarskih uporabnikov v digitalni dobi sta dostop do kvalitetnih, ažurnih jezikovnih podatkov, skupaj z možnostjo za izražanje lastnega mnenja in predlogov za izboljšave (Müller-Spitzer 2014, Arhar Holdt idr. 2016a). Skladno z uporabniškimi potrebami in novimi tehnološkimi možnostmi uvaja Slovar sopomenk sodobne slovenščine med slovenske jezikovne vire nov tip slovarja, ki ga imenujemo **odzivni slovar**. Za ta tip slovarja je značilno, da njegova gradnja poteka odprto javnosti in v sodelovanju z zainteresirano uporabniško skupnostjo. Pomemben del odzivnega slovarja je prilagojeni slovarski vmesnik, ki jasno prikazuje, na kateri stopnji jezikoslovne izdelanosti je posamezno slovarsko geslo, obenem pa je za vsak podatek razvidno, ali izhaja iz jezikovnih virov ali je dodan uporabniško. Poimenovanje 'odzivni slovar' izhaja iz dejstva, da se jezikovni podatki po predvideni metodologiji dinamično odzivajo: na eni strani na spreminjajočo se jezikovno realnost – sodobne metode podatkovne obdelave omogočajo, da se spremembe v jezikovni rabi lahko zelo hitro identificirajo in vključijo v slovar, s čimer je zagotovljena njegova ažurnost – na drugi strani pa je slovar odziven na stališča sodelujoče jezikovne skupnosti, saj omogoča napredno vključevanje znanja in mnenj različnih uporabniških skupin.

Med aktivnostmi za vključevanje uporabnikov, ki jih podpira platforma Slovarja sopomenk sodobne slovenščine, so ocenjevanje sopomenskega gradiva, neposredno dodajanje lastnih predlogov, javna razprava o obstoječih rešitvah in avtomatski uvoz morebitnega obstoječega uporabniškega gradiva v zbirko. Del navedenih možnosti je bil za slovenščino že preizkušen, vendar bolj ali manj sporadično, pogosto v smislu popestritve slovarske rabe. Nikoli do sedaj pa uporabniki še niso bili integrirani v slovarski koncept na način, ki bi omogočal zgoraj predstavljeni razvojni krog. Ko govorimo o promociji Slovarja sopomenk sodobne slovenščine, zato govorimo tudi o promociji novih načinov uporabniškega sodelovanja in vključevanja, pa tudi o zagotavljanju uporabniške motivacije za tovrstno sodelovanje.

V projektu smo razvili postopke za spremljanje uporabniških mnenj in odzivov na spletu [C3.1], kakor tudi metode za spremljanje različnih načinov uporabniškega vključevanja v slovarski razvoj oz. njihove slovarske rabe splošno, tj. napredno beleženje uporabniških aktivnosti v slovarskem vmesniku [C3.2]. Ključni del metodologije je sinteza in analiza pridobljenih informacij ter njihovo upoštevanje pri nadaljnjem razvoju promoviranega jezikovnega vira. Spremljanje uporabniških odzivov in načinov vključevanja zagotavlja, da je doprinos javnosti beležen in upoštevan že od samega začetka slovarskega razvoja, s čimer je zagotovljena odzivnost slovarja, ki je njegova temeljna značilnost.

V tem dokumentu poročamo o uresničitvi ciljev pri spremljanju uporabniških mnenj in opišemo načine upoštevanja uporabniškega vključevanja v nadaljnji razvoj Slovarja sopomenk sodobne slovenščine in drugih jezikovnih virov.

2 Uporabniške povratne informacije

V tem razdelku opisujemo različne vrste uporabniških povratnih informacij, ki smo jih zbirali in spremljali znotraj projekta. Delimo jih na eksplicitne uporabniške povratne informacije (tiste, ki se neposredno nanašajo na Slovar sopomenk sodobne slovenščine in jasno izražajo predloge in želje in po konkretnih izboljšavah vmesnika, baze ipd.), implicitne uporabniške informacije (tiste, iz katerih lahko posredno sklepamo o potencialnih izboljšavah vmesnika v smeri uporabniške prijaznosti) in na neposredni uporabniški doprinos k podatkovni bazi slovarja (torej doprinos, ki ga uporabniki v slovar dodajo s pomočjo mehanizmov, predvidenih v odzivnem slovarju).

2.1 Eksplicitna uporabniška povratna informacija

Pod eksplicitne uporabniške povratne informacije štejemo vse povratne informacije, v katerih so uporabniki in uporabnice podali neposredno mnenje o Slovarju sopomenk sodobne slovenščine.

Prvi vir eksplicitnih uporabniških povratnih informacij so bili komentarji uporabnikov ob promocijskih dogodkih in predstavitev Slovarja sopomenk sodobne slovenščine, ki so potekali v okviru projekta (popis dogodkov je na voljo na spletni strani projekta na naslovu <https://www.cjvt.si/promocija-sopomenk/dogodki/>). Dogodki so bili izvedeni za različne uporabniške skupine (npr. domača in tuja strokovna javnost, študenti in študentke, učitelji in učiteljice slovenščine in drugih jezikovnih predmetov na srednjih šolah, učitelji in učiteljice slovenščine kot drugega ali tujega jezika, lektorji in lektorice, dijaki in dijakinje, člani jezikoslovnih društev, govorniki slovenščine kot drugega jezika). Uporabniki so po predstavitev dogodkih (zlasti po tistih, ki so bili organizirani v obliki delavnic, na katerih so lahko udeleženci tudi v živo preizkusili delovanje Slovarja sopomenk sodobne slovenščine) imeli možnost podati komentarje oz. predloge za izboljšave.

Drugi vir eksplicitnih uporabniških povratnih informacij je bila uporabniška anketa, ki je bila izvedena v prvem letu projekta (glej poročilo z naslovom *Promocijski dogodki in uporabniške povratne informacije v letu 2018* v spletni knjižnici projekta na naslovu <https://www.cjvt.si/promocija-sopomenk/knjiznica/>, razdelek Projektna poročila). Namen ankete je bil preveriti odnos uporabnikov do novosti, ki jih v jezikovnih virih prinaša koncept odzivnega slovarja, in zbrati predloge za izboljšave (z dodatnim, neobveznim vprašanjem odprtega tipa: *Če bi lahko slovar izboljšali, kako bi ga spremenili?*). Vseh v celoti izpolnjenih anket je bilo 671. Zbrane odgovore smo nato razporedili v kategorije, ki jih skupaj s predlogi potencialnih rešitev predstavljamo v nadaljevanju. (Dodatne eksplicitne povratne uporabniške informacije, ki so bile zbrane s pomočjo neposredne evalvacije vmesnika s protokolom glasnega razmišljanja, so opisane v *Specifikacijah za adaptacije slovarskega vmesnika*, ki so na voljo s projektni spletni knjižnici pod razdelkom Poročila).

Obogatitev obstoječih podatkov v slovarju

V to kategorijo smo uvrstili 66 odgovorov, ki so se nanašali na obogatitev obstoječih podatkov v slovarju. Najpogosteje (43 odgovorov) so uporabniki izrazili željo po dodatnih kvalifikatorjih, ki bi sopomenke in njihovo rabo natančneje opredelili z vidika registra in slogovne zaznamovanosti. Nekoliko redkeje so se pojavile želje po več področnih oznakah (4 odgovori).

"Slovar je zelo dobro oblikovan in prijazen uporabniku, čeprav ni po vašem konceptu, bi besedam vseeno dodal uradne kvalifikatorje, saj bi to pohitrilo proces, o njihovi primernosti pa bi se s podrobnejšim iskanjem, ko bi bilo to potrebno, pozanimal sam."

"Mogoče edino uporaba kvalifikatorjev (narečno, pogovorno ipd.), sicer je slovar izredno kvaliteten in koristen. Pohvale!"

"Predvsem s kvalifikatorji, ki so za določen del praktikov izredno pomembni. Ocenjevanje uporabnikov je zanimivo in dobro, vendar poznavajoč množično napačno tolmačenje kvalifikatorskih napotil v splošnih in normativnem slovarju, veliko jezikoslovcev praktikov ne razume oziroma nima pravih orodij za ustrezno vrednotenje sopomenskosti. Kvalifikatorji nujni: izvrstna orodjarna s povezavami do korpusnega materiala in kolokacijske mreže ne bo tega nadomestila pri vseh uporabnikih."

"Če bi bilo mogoče na nek način vseeno vključiti kvalifikatorje, bi bilo dobro. Morda bi bilo to mogoče na osnovi virov, od koder v večini prihajajo primeri za posamezne besede. Koristne bi bile informacije o registru (vsaj groba razdelitev) in besedilnih tipih, kjer se besede pogosteje pojavljajo (npr. časopisi, internet, literatura ipd.)."

Redkeje se pojavljajo še predlogi o dodajanju naglasnih znamenj (2 odgovora) oz. podatkov o izgovoru (1 odgovor), o podrobnejši kategorizaciji sopomenk (6 odgovorov), o dodajanju normativnih razlag (1 odgovor), o izpisovanju celotnih oz. pravih slovničnih oblik (npr. *upati si* namesto *upati*, 2 odgovora), o oštevilčenju vnosov glede na pomen (1 odgovor), o dodajanju podatka o pogostosti v korpusu (1 odgovor), o slikovnem gradivu (3 odgovori) in o dodajanju več podatkov nasploh (2 odgovora).

Potencialne rešitve: V nadaljnjih različicah Slovarja sopomenk sodobne slovenščine bodo postopoma naslovljeni pomisleki glede manjkajočih kvalifikatorjev, začeni s kvalifikatorji, ki bodo v prihajajoči različici (2.0) dodani najbolj občutljivemu gradivu,

tj. tistim sopomenkam, ki so (v določenih kontekstih) lahko interpretirane kot žaljive (npr. *ženska – baba*).

Razširitev nabora podatkov v slovarju

V to kategorijo smo uvrstili 36 odgovorov, ki na več načinov zajemajo razširitev nabora podatkov v slovarju. Najpogosteje so uporabniki izrazili željo po razširitvi nabora iztočnic v slovarju nasploh (14 odgovorov). Odgovori nakazujejo, da želijo v slovar vključiti tudi starinsko besedišče (7 odgovorov) ter nestandardne/narečne/pogovorne izraze (6 odgovorov) oziroma lastnoimenske sopomenke (1 odgovor) ter blizu- in protipomenke (2 odgovora). Preostali odgovori zadevajo dodajanje več kolokacij (1 odgovor) in dodatne primere rabe (5 odgovorov) pri vnosih, ki so v trenutni različici brez.

"S kvalifikatorji in razširitvijo virov, da bi vseboval tudi starejše, zastarele in manj uporabljene izraze."

"Dopolnila bi ga s sopomenkami iz starejših (starih) književnih in strokovnih besedil."

"Vključila bi tudi sopomenke zastarelih izrazov oz izrazov, k niso več tako pogosti in se počasi opuščajo (namen je skrbeti za nadaljni razvoj slovenskega jezika brez okrnitve/pozabljanja lepih slovenskih izrazov, kar se danes rado dogaja (vpliv amerikanizacije, pomanjkanje branja "pravih" slovenskih knjig - prevladujejo prevodi, itd.))"

"Vem, da je to slovar sopomenk, ampak dobra informacija k sopomenki je tudi protipomenka. Če moraš obrniti stavek, na primer ..."

Potencialne rešitve: Želje uporabnikov po razširitvi nabora iztočnic bomo upoštevali pri nadgradnji, in sicer tako, da bomo pregledali dnevnik iskanj, s katerimi spremljamo uporabniške dejavnosti v Slovarju sopomenk sodobne slovenščine (glej tudi poglavje 2.2), in na ta način pridobili nabor iztočnic, ki so jih uporabniki najpogosteje iskali, a v slovarju še niso zabeležene. Iztočnice bomo dodali v slovar, njihove potencialne sopomenke pa bomo dodali s pomočjo avtomatskih postopkov, kot je npr. detekcija sopomenskosti s pomočjo vektorskih vložitev.

Izboljšave vmesnika in njegovih funkcionalnosti

Skupno 24 odgovorov se je uvrstilo v kategorijo izboljšav uporabniškega vmesnika in njegovih funkcionalnosti. Poleg 4 odgovorov, ki so sicer izkazovali željo po

izboljšanem vmesniku, a konkretnih predlogov niso podali, so se pojavili še predlogi o večji preglednosti (2 odgovora), hitrejšem dostopu do kolokacij (1 odgovor), možnosti prikaza (uporabniških) sopomenk na enem mestu (1 odgovor), večjem številu primerov na prvi strani (1 odgovor), preprostejši različici vmesnika za osnovnošolce (1 odgovor) in večji interaktivnosti (3 odgovori).

"Da je čim bolj interaktiven, npr. če se z miško pomakneš na besedo, ki ni glavno geslo, ampak del razlage gesla, ta vodi do lastnega gesla v slovarju (i.e. te popelje naprej na svoje geslo, ne da sam iščeš to v iskalni vrstici slovarja)."

"Stremet k temu, da je čim bolj verodostojno gradivo. Morda več primerov in kvalifikatorjev. Stremeti k preprosti uporabi in izdelavi različice za osnovnošolce."

Kot dodatne funkcije v vmesniku bi uporabniki želeli tudi možnost popravljanja ali brisanja dodanih uporabniških sopomenk (1 odgovor), sistem za podrobnejše označevanje bolj ali manj ustreznih sopomenk (1 odgovor) in možnost javljanja napak (1 odgovor). Nekateri uporabniki menijo, da bi bilo koristno omejiti dodajanje sopomenk (3 odgovori), npr. z uporabo gesla (2 odgovora) oz. z odstranitvijo možnosti dodajanja uporabniških sopomenk (1 odgovor).

"Če želiš v slovar dodati sopomenko, moraš uporabiti izbrano uporabniško ime. Ni bi všeč, da je tako svobodno, saj se npr. lahko zgodi, da nekdo pod mojim pravim imenom vpisuje v slovar "bedarije" ..."

1 uporabnik je opozoril, da je vmesnik včasih nekoliko počasen, 1 uporabnik pa si želi navodila za uporabo.

Potencialne rešitve: Počasnost vmesnika (in obenem hitrejši dostop do kolokacijskih podatkov) smo vmes že naslovili s selitvijo Slovarja sopomenk sodobne slovenščine na zmogljivejši strežnik, navodila za uporabo pa smo v obliki videoposnetkov prav tako pripravili v okviru projekta in bodo vključena v slovarski vmesnik v prihajajoči različici.

Večja zanesljivost podatkov

Skupno 35 odgovorov se je nanašalo na izboljšanje natančnosti oz. zanesljivosti obstoječega gradiva v slovarju. Najpogosteje so uporabniki predlagali, da uporabniško dodane sopomenke sproti pregledujejo strokovnjaki (23 odgovorov).

"Redno preverjanje podatkov - novih besed, vpisanih s strani uporabnikov."

"redno preverjanje podatkov in dodajanje novih besed"

"Urejal bi ga strokovnjak za slovaropisje."

"Nekdo bi moral kritično strokovno spremljati zapisane vnose kot pri Wikipediji in posegati v netočne zapise."

1 odgovor je izrecno omenil odstranjevanje šuma, 6 odgovorov večjo natančnost in zanesljivost. 1 odgovor je predlagal ločevanje iztočnic glede na pomen (problem pri enakopisnicah tipa mrk – kot samostalnik in pridevnik). Preostali odgovori omenjajo boljše kriterije za ocenjevanje sopomenk (1), več funkcij (1) in stik z dejanskim jezikom (1).

Potencialne rešitve: Pomisleke uporabnikov glede omejitve dodajanja sopomenk v slovarski vmesnik je smiselno vzeti kot željo po večjem nadzoru nad uporabniško dodanim gradivom. V ta namen bodo pripravljene smernice za jasnejše ločevanje koristnih in nekoristnih uporabniških predlogov na podlagi dodanega gradiva (glej tudi razdelek 2.3), uredniški odbor Slovarja sopomenk sodobne slovenščine pa bo smernice integriral v svoj delotok in o spremembah obveščal tudi uporabnike.

Večja povezljivost z drugimi viri in mediji

Skupno 22 odgovorov je zajemalo večjo povezljivost z drugimi viri oz. drugimi mediji. Največkrat so uporabniki predlagali povezovanje z drugimi jezikovnimi viri, zlasti slovarji in korpusi (12 odgovorov).

"z zagotovitvijo baze, iz katere črpa - posodobitev GigaFide in redno dodajanje besedil v GigaFido"

"Povezal bi ga s sorodnimi slovarji na spletu."

"Dodajte SSSJ - Sinonimni slovar slovenskega jezika."

"mogoče povezava na sskj"

Predlagali so tudi integracijo v programsko opremo za urejanje besedil (npr. Microsoft Word, 1 odgovor) ter razvoj slovarja kot mobilne aplikacije (2 odgovora).

V 5 odgovorih se je pojavil predlog o nadaljnji promociji slovarja, le v 2 odgovorih pa so si uporabniki želeli, da bi bil slovar na voljo tudi v tiskani obliki.

"tako da bi več ljudi vedelo zanj"

"Vec denarja bi vam dala za marketing in PR :)"

"Čim večja prepoznavnost, da bi ga ljudje redno uporabljali in tako dopolnjevali ter izpopolnjevali."

Potencialne rešitve: Povezljivost z drugimi viri smo v času projekta do določene mere že uvedli, saj Slovar sopomenk sodobne slovenščine omogoča tudi preklap na Kolokacijski slovar sodobne slovenščine 1.0, Slovenski oblikoslovni leksikon Sloleks 2.0 in Gigafido 2.0. Ob izidu novih jezikovnih virov CJVT bodo dodane tudi nove možnosti preklopa, pri čemer bo uporabnik zelo enostavno lahko preklopil med enakimi iztočnicami v različnih virih. Tudi promocija Sopomenk bo potekala še naprej v okviru promocije drugih odzivnih jezikovnih virov CJVT.

2.2 Implicitna uporabniška povratna informacija

Kot implicitne uporabniške povratne informacije obravnavamo vse informacije, ki jih od uporabnikov pridobimo posredno in iz katerih lahko posredno sklepamo, katere spremembe v slovarskem vmesniku so potrebne, da bi uporabniškim potrebam bolj zadostili.

Prvi vir implicitnih uporabniških povratnih informacij so dnevnik iskanja, ki jih v Slovarju sopomenk sodobne slovenščine sistematično zbiramo vse od objave. Uporabniško dejavnost beležimo v obliki dejanj s časovnim žigom, ki jih nato agregirano izvozimo. Izvoz dejanj novembra 2019 je podan v spodnji tabeli.

Iskanje iztočnice	1.363.282
Klik na sopomenko	24.380
Klik na kolokacijo	6.217
Filtriranje sopomenk po pogostosti	115.762
Premik s sopomenke na korpus Gigafida	8.219
Premik z zglada na korpus Gigafida	734
Glasovanje za sopomenko	39.719
Dodajanje uporabniške sopomenke	30.410

Nalaganje podatkov na osebni računalnik	264
Klik na povezavo na zunanjo stran	884
Obisk osnovne strani	282.997
Iskanje iztočnice, ki ni v slovarju	47.975
Deljenje vsebine na družbenih omrežjih	305
Preklop jezika vmesnika	372

Tabela 1. Beleženje v dnevnikih iskanj za Slovar sopomenk sodobne slovenščine (november 2019).

Iz podatkov je med drugim razvidno, da uporabniki pogosto iščejo tudi iztočnice, ki jih v Slovarju sopomenk sodobne slovenščine še ni (skoraj 48.000 iskanj in približno 17.000 nenajdenih nizov). Vidimo tudi, da je premikov neposredno s sopomenke na Gigafido več kot z zgledov na Gigafido, kar nakazuje, da uporabniki ob ogledu zgledov, ki so že vključeni v Slovar sopomenk, redkeje potrebujejo dodaten kontekst. Pogosto uporabljajo tudi funkcijo filtriranja sopomenk po pogostosti (skupno približno 115.000 dejanj).

Iz dnevnikov lahko razberemo tudi, s katero vrsto naprave uporabniki najpogosteje dostopajo do Slovarja sopomenk. Rezultate prikazuje spodnja tabela. Še vedno prevladujejo osebni računalniki, zanimivo pa je, da telefonski dostop zajema skoraj 14 odstotkov.

Nedoločeno	93.835
Osebni računalnik	1.866.130
Telefon	321.594
Tablica	19.817

Tabela 2. Dostop do Slovarja sopomenk sodobne slovenščine z različnimi napravami (november 2019).

Sopomenke so uporabniki razvrščali skupno 6.456-krat, zanimivo pa je, da je najpogostejši način razvrščanja prav razvrščanje po uporabniških ocenah, kar

nakazuje, da se uporabniki pri izbiranju prave sopomenke pogosto zanašajo na doprinos drugih uporabnikov.

Razvrščanje sopomenk – skupaj	6.456
Uporabniške ocene	2.201
Abecedni vrstni red	1.469
Razvrščanje po dolžini	1.386
Razvrščanje po relevantnosti	1.400

Tabela 3. Načini razvrščanja sopomenk v Slovarju sopomenk sodobne slovenščine po pogostosti zabeleženih dejanj v dnevnikih iskanja (november 2019).

Drugi vir implicitnih uporabniških informacij so bile jezikovne objave na družbenem omrežju Facebook, ki niso neposredno zadevale Slovarja sopomenk sodobne slovenščine, temveč so obravnavale konkretne jezikovne probleme uporabnikov. Proučevali smo jih tako, da smo v okviru projekta najprej izdelali program, s katerim smo s pomočjo uradnega Facebookovega aplikacijskega programskega vmesnika (<https://developers.facebook.com>) iz vnaprej določenih jezikovnih uporabniških skupin na Facebooku izluščili vse objave in komentarje nanje. (Skladno z zakonodajo je bilo poskrbljeno za varovanje osebnih podatkov, saj so vsa uporabniška imena avtomatsko psevdonimizirana z naključnimi šiframi, s čimer smo prikrili občutljive osebne podatke.) Program je jezikovno neodvisen in ga je brez prilagoditev mogoče uporabiti tudi za druge jezike in druge vrste podatkov, kot prosto in odprto dostopna programska oprema pa je na voljo tudi drugim zainteresiranim raziskovalcem na repozitoriju GitHub (povezava je na voljo v projektni spletni knjižnic, razdelek Strokovna literatura).

Za vsako od izluščenih skupin (opazovali smo npr. skupine *Društvo ljubiteljskih pravopisarjev in slovničarjev*, *Za vsaj približno pravilno rabo slovenščine* in *Skropucalo.si*) smo analizirali po uporabnikih uravnotežen vzorec 500 objav, ki smo jih kategorizirali v skladu s tipologijo, razvito v okviru obstoječih uporabniških raziskav (Arhar Holdt idr. 2016a, Arhar Holdt idr. 2015, Čibej idr. 2015a). Kategorizirane objave (na voljo v spletni knjižnici pod razdelkom *Gradiva*) smo nato razdelili v sklope, ki ponazarjajo tipične scenarije uporabniških problemov in

ponujajo implicitno povratno informacijo o najbolj perečih vrzelih v jezikovnih virih za slovenščino. Rezultate prikazuje spodnja tabela.

Nejezikovno vprašanje	87
Iščem besedo	20
Ločila	50
Razlika med besedama	11
Skladnja	11
Želim nekaj zunanjega	34
Katera varianta je boljša?	148
Kako se pravilno črkuje?	14
Kaj gre s to besedo?	2
Mala ali velika začetnica?	20
Kako naj pregibam besedo/zvezo?	42
Ali je ta beseda/zveza v redu ali ne?	36
Ločila, Katera varianta je boljša?	1
Kako se pravilno izgovori?	5
Metajezikovno vprašanje	6
Kaj pomeni beseda/zveza?	4
Od kod izvira beseda/zveza?	4
Ali niso moji predlogi boljši?	3
Ali je ta beseda še aktualna?	1
Zakaj se tako črkuje?	2

Tabela 4. Kategorizacija uporabniških vprašanj.

Daleč največ vprašanj spada v scenarij *Katera varianta je boljša?*, 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 lahko v pomoč Slovar sopomenk sodobne slovenščine, ki omogoča primerjavo sopomenk v kontekstu.

"Bi v strokovnem članku, kjer se ve, o čem teče beseda, uporabili "tablični računalnik" ali "tablica"? Hvala."

"Čer, zanima me, kaj je pravilno oz. bolj pravilno: v razmiku 9 mesecev ali v razmaku 9 mesecev? :)"

"Danes sem prvič naletela na besedo nogometašinja. Po krajšem raziskovanju sem ugotovila, da je nogometašica sicer precej bolj pogosto uporabljena beseda, pravilni pa sta obe. Drži?"

Na ravni oblik in besednih zvez bi lahko uporabniku bili v pomoč različni viri, npr. oblikoslovni leksikon in slogovni priročnik ter kolokacijski slovar.

"črtomiru ali črtomirju?"

"Hladno stisnjeno ali hladno stiskano olje?"

Kolokacijski slovar sodobne slovenščine, ki je na Centru za jezikovne vire in tehnologije Univerze v Ljubljani izšel jeseni 2018 (v času projekta), neposredno odgovarja tudi na scenarij Kaj gre s to besedo?, pri katerem uporabnik sprašuje po besedi, ki se tipično veže z drugo besedo.

"Jutro! Pečico ogrejemo ali segrejemo? Mleko? Hvala. :)"

"Majhni stroški ali nizki stroški? Hvala :)"

"Manjša dilema - ali napravo vklopimo v električno omrežje ali priklopimo na električno omrežje? Oboje?"

"rheumatic heart disease - revmatska ali revmatična bolezen srca?"

Z vidika Slovarja sopomenk sodobne slovenščine je zanimiv tudi scenarij *Iščem besedo*, pri katerem uporabnik največkrat opisno izrazi, kaj želi sporočiti, od uporabnikov pa želi povratno informacijo v obliki besede, ki bi izpolnjevala opisane kriterije. V tem primeru lahko gre za narečno ali pogovorno besedo, za katero išče izraz v standardni slovenščini. Ker uporabniki v Slovar sopomenk pogosto dodajajo tudi narečne sopomenske kandidate (glej razdelek 2.3), bi bilo smiselno razmisliti o

tem, da iskalni niz v iskalni vrstici išče tudi po uporabniških sopomenkah, česar trenutna različica slovarja ne omogoča.

"Pozdravljeni, ena šiviljska: kako se pravilno reče 'cvikelnom' na srajčkah in podobnih zadevah?" (potrebujem besedo)

"Pomoč pri oblikovanju povedi: Ali veš, da izkušen lovec ... (fora je, da vidi sled/stopinjo in ve, kateri živali pripada). (potrebujem besedo)"

Na podoben način lahko na vprašanja iz scenarija *Kako se pravilno črkuje?* in na nekatera vprašanja iz scenarija *Katera varianta je boljša?* pomaga odgovoriti oblikoslovni leksikon, ki ponuja pregibne oblike besed.

"Pozdravljeni. Ali je pravilno Javierov, Javierev ali Javierjev?"

"Živalski vrt, množina: živalski vrti ali živalski vrtovi? Hvala :)"

"sukajo ali sučejo okoli nje?..(sukajo je glupo total)"

Različica Slovenskega oblikoslovnega leksikona Sloleks 2.0 je izšla oktobra 2019, glede na vsebino analiziranih uporabniških vprašanj pa bi jo bilo smiselno razširiti z dodatnimi lastnimi imeni (tako osebnimi kot krajevnimi).

Analiza kaže, da pri precejšnjem številu vprašanj jezikovni vir, ki bi pomagal odgovoriti na vprašanje, že obstaja (Slovar sopomenk sodobne slovenščine, Kolokacijski slovar sodobne slovenščine, Slovenski oblikoslovni leksikon), treba pa je zagotoviti, da so uporabniki z njim seznanjeni in ga znajo uporabljati. Prav tako bi bilo smiselno uporabniška vprašanja kategorizirati z mislijo na širši nabor virov in na povezljivost med njimi. Nekatera vprašanja je namreč mogoče rešiti na več načinov, koristnost virov pa uporabnikom lahko prikažemo tudi z dodatnimi multimedijskimi gradivi (npr. videoposnetki s pomočjo), ki naslavlajo konkretne probleme in najpogostejše scenarije, identificirane v raziskavi. Na ta način bi bilo mogoče tudi promovirati različne vire in različne načine, kako priti do rešitve.

2.3 Neposredni uporabniški doprinos k podatkovni bazi

V Slovarju sopomenk sodobne slovenščine neposredni uporabniški doprinos beležimo na dva načina. Kot prvo, uporabniki lahko iztočnicam dodajajo svoje predloge sopomenk, ki se beležijo v slovarski bazi in prikazujejo v ločenih razdelkih za uporabniško dodane sopomenke. Kot drugo, uporabniki lahko obstoječe

sopomenke ocenijo bodisi pozitivno (+1) bodisi negativno (-1), ocene pa se prav tako zabeležijo v slovarski bazi in prikažejo v vmesniku, kar omogoča razvrščanje sopomenk od najboljše do najslabše ocenjene.

Rezultati izvoza obeh vrst uporabniškega doprinosa novembra 2019 kažejo, da je bilo v Slovar sopomenk do sedaj dodanih 23.140 sopomenk (s čimer se je število sopomenk iz začetne različice povečalo za skoraj 13 odstotkov) in 34.210 pozitivnih ocen ter 5.509 negativnih ocen.

Sopomenke so v slovar uporabniki dodajali pod 445 različnimi uporabniškimi imeni. Analiza uravnoveženega vzorca 1.662 sopomenk (največ 10 na uporabnika) je pokazala, da uporabniki večinoma v slovar prispevajo konstruktivne predloge – podrobnejši pregled prikazuje spodnja tabela. Poleg standardnih, nezaznamovanih predlogov (1.175 vnosov) so v slovar dodajali tudi narečne in pogovorne sopomenke (218) in slengovske izraze (67). Le manjšina (110) vnosov je neustreznih, ne pa nujno tudi nekonstruktivnih. Sistem dodajanja uporabniških sopomenk je torej uspešen, zahteva pa še natančnejše smernice za obravnavo sopomenskosti, ki bodo razvite na podlagi zbranega gradiva. V skladu s smernicami bo nato določeno, katere uporabniško dodane sopomenke bodo vključene v različico Sopomenk 2.0.

Opis kategorije in uredniškega delotoka	Primeri sopomenk	Frekvenca v vzorcu
Uporabnik je dodal sodobno, nezaznamovano sopomenko. Gradivo je dobronamerno in po večini ustreza kriterijem sopomenskosti. Potreben je uredniški (leksikografski) pregled in potrditev novega gradiva, sledi vpis v bazo slovarja.	alarm – opomnik ambiciozen – slavohlepen atribut – prednost avtomatsko – nezavedno arhivirati - shraniti	1.175
Uporabnik je dodal regionalno (narečno) ali pogovorno sopomenko, pri čemer je včasih pripisal tudi kvalifikator. Gradivo je dobronamerno in po večini ustreza kriterijem sopomenskosti. Potreben je uredniški (leksikografski) pregled in ustrezna kvalifikacija.	babica – nona (lokalno) nogomet – fuzbal (pogovorno) biser – perla gneča – dren oblič – flajšter obraz – fris	218
Uporabnik je dodal gradivo, ki je bodisi napačno (lapsus ali rezultat nerazumevanja sopomenskosti) ali ni	čast - biti v zborniku <3 uwu @ šemi	110

dobronamerno (šaljivi vnosi). Deloma je mogoče težave identificirati in odstraniti avtomatsko, po ročnem pregledu se te predloge odstrani iz slovarja.	cesta – cesta dobro – vredu nazaj – krokodil	
Uporabnik je dodal slengovsko, šaljivo sopomenko pri čemer je včasih pripisal tudi kvalifikator. Gradivo je dobronamerno in po večini ustreza kriterijem sopomenskosti. Potreben je uredniški (leksikografski) pregled in ustrezna kvalifikacija.	ideja – fleš ne priti – zahinaviti prijatelj – buraz popivanje – ulivanje razžalostiti - zamoriti	67
Uporabnik je dodal neologistično sopomenko, npr. besedo, ki predstavlja slovensko ustreznico prevzete besede. Gradivo je dobronamerno, vendar ne ustreza kriterijem pojavljanja v dejanski rabi. Potreben je pregled pogostnosti v korpusu Gigafida 2.0.	potrpežljivost – učakanost marmelada – sadjemaz analiza – pregleddek semafor – preklopno svetilo revija - objavnik	40
Uporabnik je dodal sopomenko, ki v zgodnji fazi prevzetosti iz tujega jezika, predvsem angleščine. Gradivo je dobronamerno in po kontekstih rabe slengovsko oz. žargonsko. Potreben je uredniški pregled in odločitev glede potencialne vključitve.	dogodek – event povezava – link predloga – templejt prijatelj – frend sledilec – follower ponarejen – fejk	28
Uporabnik je dodal termin oz. strokovno sopomenko, pri čemer je včasih pripisal tudi kvalifikator. Gradivo je dobronamerno in po večini ustreza kriterijem sopomenskosti. Potreben je uredniški (leksikografski) pregled in ustrezna kvalifikacija.	destilat – prekapina šipek – pasja roža sirast – kazeozen stalen – perenialen izkop - ekskavacija	24
Skupaj		1.662

Tabela 5. Kategorije uporabniško dodanih sopomenk.

V predhodnih korakih projekta je bilo identificirano, da uporabniško sodelovanje pri razvoju slovarja presega pričakovanja na ravni dodajanja novega sopomenskega gradiva (uporabniki so dodali visoko število novih sopomenk, ki so po večini uporabne za dopolnitev slovarske baze), redkejša od pričakovanega pa je sodelovanje pri čiščenju baze z glasovi za »slabe« in »dobre« sopomenke. Ti glasovi so tudi manj dosledno pripisani, kot smo pričakovali, poleg tega pa so razpršeni in prisotni le pri okrog 5 odstotkih celotnega slovarja, največkrat pri pogosto iskanih iztočnicah.

Problem zbiranja ocen je torej tudi neenakomerna razporeditev iztočnic, saj nimajo vse enake možnosti, da bi bile ocenjene, izkazalo pa se je, da ocenjevanje sopomenskosti z negativno ali pozitivno oceno pogosto zelo težavno (polje sopomenskosti je sivo, zato se uporabnik težko odloči v eno ali drugo smer). Na podlagi teh ugotovitev smo se odločili, da se namesto neposrednih izboljšav glasovanja v samem vmesniku za mobilne naprave raje osredotočimo na razvoj sopomenskega modula znotraj mobilne aplikacije Igra besed (glej *Specifikacije za adaptacijo slovarskega vmesnika*), ki s pomočjo igrifikacije omogoča zanesljivejše in po iztočnicah enakomernejše zbiranje ocen.

Kot dodatno dejavnost znotraj projekta smo izvedli tudi ciljno označevanje večbesednih sopomenk, ki so bile že pri izdaji prve različice Slovarja sopomenk sodobne slovenščine identificirane kot problematične (šum zaradi avtomatske gradnje slovarja). Pregledanih je bilo 18.365 večbesednih sopomenskih kandidatov, ki so bili označeni kot ustrezni ali neustrezni. 5.882 jih je bilo označenih kot pretežno do povsem neustreznih, 9.027 kot pretežno do povsem ustreznih, 3.456 pa je ostalo neodločenih. V naslednji različici Slovarja sopomenk sodobne slovenščine 2.0 bodo iz vmesnika odstranjene večbesedne sopomenke, ki so bile ocenjene negativno, ostale pa bodo dodatno pregledane v skladu s smernicami o določanju sopomenskosti, ki bodo razvite na podlagi uporabniško dodanega gradiva.