

# Negativno zaznamovano besedišče v Slovarju sopomenk sodobne slovenščine 2.0

*Špela ARHAR HOLDT*

Filozofska fakulteta, Univerza v Ljubljani; Fakulteta za računalništvo in informatiko,  
Univerza v Ljubljani

*Iztok KOSEM*

Filozofska fakulteta, Univerza v Ljubljani; Institut Jožef Stefan;  
Fakulteta za računalništvo in informatiko, Univerza v Ljubljani

*Eva PORI*

Filozofska fakulteta, Univerza v Ljubljani

*Vojko GORJANC*

Filozofska fakulteta, Univerza v Ljubljani

*Simon KREK*

Filozofska fakulteta, Univerza v Ljubljani

*Polona GANTAR*

Filozofska fakulteta, Univerza v Ljubljani

V prispevku predstavljamo rešitve za prepoznavanje in označevanje zaznamovanega besedišča v okviru koncepta odzivnega Slovarja sopomenk sodobne slovenščine. Ker gre za prvi tovrstni projekt, so pripravljene rešitve v veliki meri inovativne, umeščene pa v okvir problematike avtomatske strojne izdelave slovarja, njegove odprtosti in vključenosti uporabniške skupnosti. Prispevek prikazuje postopek prepoznavanja sovražnega in grobega besedišča ter pripis

---

*Arhar Holdt, Š., Kosem, I., Pori, E., Gorjanc, V., Krek, S., Gantar, P.: Negativno zaznamovano besedišče v Slovarju sopomenk sodobne slovenščine 2.0. Slovenščina 2.0, 11(1): 8–32.*

*1.01 Izvirni znanstveni članek / Original Scientific Article*

DOI: <https://doi.org/10.4312/slo2.0.2023.1.8-32>

<https://creativecommons.org/licenses/by-sa/4.0/>



oznak, opozorilnih ikon in daljših pojasnil. Ukvarjamo se tako s tehničnimi kot vsebinskimi vprašanji označevanja. Vsebinsko oznake temeljijo na sporočanjsem namenu in učinku, pri čemer je njihovo bistvo informacija o možnih posledicah rabe, pri tehničnih rešitvah pa veliko pozornost posvečamo digitalnemu mediju in vizualizaciji rešitev v njem. Ker je odzivnost eden ključnih konceptov slovarja, se pri rešitvah glede označevanja zavedamo pomembnosti sodelovanja z uporabniško skupnostjo, zato tudi pri dodajanju oznak predlagamo rešitve za sodelovanje s skupnostjo. Izhodiščni konferenčni prispevek je bil razširjen v vseh poglavjih, dodano pa je povsem novo poglavje o obdelavi večpomenskih iztočnic, njihovi pomenski členitvi in pomenskem opisovanju z zgledi pomenov z negativno zaznamovanostjo.

**Ključne besede:** slovar sopomenk, odzivni slovar, slovarske oznake, sporočanski namen, uporabniška skupnost

## 1 Uvod

Slovar sopomenk sodobne slovenščine (SSSS) je oblikovan po modelu odzivnega slovarja: v prvem koraku je bil pripravljen strojno, nadaljnje urejanje podatkov pa poteka po korakih in v sodelovanju jezikoslovcev ter širše zainteresirane skupnosti (Arhar Holdt et al., 2018, str. 404). V SSSS lahko slovarski uporabniki ob strojno pripravljeno sopomensko gradivo dodajo lastne predloge sopomenk, za vse sopomenke v slovarju pa je mogoče tudi glasovati in gradivo na tak način (pomagati) urejati oz. selekcionirati.<sup>1</sup>

Vključevanje strojnih postopkov in predlogov uporabniške skupnosti v slovaropisne delotoke odgovarja na potrebe sodobnega časa, kot sta potreba skupnosti po odprto dostopnih jezikovnih podatkih in želja slovarskih uporabnikov po demokratičnem sodelovanju pri razvoju temeljne jezikovne infrastrukture. Na drugi strani pa ima neposredno objavlanje strojnega in uporabniško dodanega (nepregledanega) gradiva lahko tudi neželene posledice, ki jih je treba pri razvoju odzivnega modela predvideti in ustrezno obravnavati. Med prioriteta za razvoj SSSS je tako brez dvoma obravnavo besedišča, ki vrednostno poimenuje posamezne družbene skupine in njihove pripadnike. Tako besedišče

1 Slovar v vmesniku je na <https://viri.cjvt.si/sopomenke/slv/>, kot slovarska baza pa na repozitoriju CLARIN.SI (Krek et al., 2018). Strojno pripravo slovarja opisujejo Krek et al. (2017), koncept odzivnega slovarja pa Arhar Holdt et al. (2018).

se v nepregledani različici slovarskih gesel (lahko) pojavlja na različnih mestih in na različne načine.

Namen prispevka je predstaviti obseg problematike, ki se pri odzivnem slovarju pomembno razlikuje od tradicionalnih slovaropisnih projektov, in opisati rešitve, ki so vključene v nadgradnjo SSSS iz različice 1.0 v 2.0, kot tudi naloge za nadaljnje delo. V prispevku želimo posebej izpostaviti nove načine prepoznavanja in označevanja sovražnega, grobega ter drugače negativno vrednotenega besedišča, ki so uporabne za različne sodobne jezikovne vire, ne le SSSS.

Prispevek je nadgradnja dela (Arhar Holdt et al., 2022), ki je bilo strokovni javnosti predstavljeno na znanstveni konferenci Jezikovne tehnologije in digitalna humanistika 2022. Prispevek je posodobljen v vseh poglavjih, povsem nova pa so poglavja, ki opisujejo slovaropisni pregled gradiva in končne odločitve, implementirane v SSSS 2.0.

## **2 Negativno zaznamovano besedišče v družbi, jeziku in slovarju**

Na kratko je mogoče sovražni govor opredeliti kot “aktivno javno spodbujanje antipatije do določene, ponavadi šibke, družbene skupine” (Rebolj, 2008, str. 13), v daljši in bolj povedni obliki pa kot (Petković in Kogovšek Šalamon, 2007, str. 23):

ustno ali pisno izražanje diskriminatornih stališč. Z njim širimo, spodbujamo, promoviramo ali opravičujemo rasno sovraštvo, ksenofobijo, homofobijo, antisemitizem, seksizem in druge oblike sovraštva, ki temeljijo na nestrpnosti. Mednje sodi tudi nestrpnost, ki se izraža z agresivnim nacionalizmom in etnocentrizmom, z diskriminacijo in sovražnostjo zoper manjšine, migrante in migrantke. Žrtve sovražnega govora praviloma niso posamezniki, pač pa ranljive družbene skupine. V osrčju sovražnega govora je prepričanje, da so nekateri ljudje manj vredni, zato je cilj sovražnega govora v razčlovečenju, ponižanju, ustrahovanju in poslabšanju družbenega položaja tistih, proti katerim je naperjen.

Motl in Bajt (2016, str. 7) ugotavljata, da je sovražni govor deležen precejšnje pozornosti v različnih vedah, od prava, sociologije in

komunikologije do psihiatrije in informatike, pridružimo pa jim lahko tudi jezikoslovje.

## 2.1 Kritično slovaropisje

Ameriško slovaropisje (Hughes, 2009, str. 87–105) je že pred desetletji v svoje vire načrtno vgradilo tudi občutljivost do ranljivih družbenih skupin, pri čemer ni zanemarilo nobenega od delov geselskega članka: razlag, oznak in zgledov rabe (Logar et al., 2020, str. 104). V manjši meri in pozneje, a vendarle so se opozorila o nujni tovrstni družbeni občutljivosti ter odgovornosti pojavila tudi v slovenskem prostoru (npr. Gorjanc, 2005; Kern, 2015; Logar et al., 2020, str. 91, 104), a jih kljub temu do sedaj ni polno upošteval še noben slovarski projekt.

Ni pa zgolj sovražni govor tisti, ki ga je treba v slovarjih obravnavati posebej pozorno. Kritično slovaropisje opozarja, da je treba pri slovarskih opisih izrecne (in nove) rešitve iskati pri vseh elementih, ki prinašajo vpljudne in nevljudne vidike jezika, tabuiziranost, so usmerjeni v vrednotenje, konotacijo, kulturne aluzije ipd., še posebej pa je treba biti pozoren na nestabilna in spreminjajoča se poimenovanja vseh oblik drugosti (Moon, 2014, str. 85). Pri tem se sodobno slovaropisje ne more sklicevati na tradicionalne modele jezikovnega opisovanja in delovanja. Nikakor pri tem ni sprejemljivo tradicionalno razmišljanje, da “je slovar metajezikovni odsev dejanske hierarhizirane konceptualizacije sveta” (Vidovič Muha, 2013, str. 7), kar vodi v razpravljanje o resnicah v okviru slovaropisnega dela – prav nasprotno: slovaropisje mora jasno naslavljalati vprašanja, ki so v svojem bistvu ideološka, saj gre za “uravnoteževanje opisa tega, kar prinašajo podatki glede pomena, s tem, na kakšen način ‘naj bi bil’ v postmoderni vključujoči družbi določen koncept obravnavan in predstavljen” (Moon, 2014, str. 89). Gre torej za to, da pri slovaropisnem delu končne rešitve preprosto ne morejo biti “samo jezikoslovne; neizogibno morajo biti tudi ideološke” (Moon, 2014, str. 94). Pomembno je, da se ideološkosti pri slovarskih opisih zavedamo, da odkrito in jasno povemo, da je slovaropisno delo težavno prav zato, ker je tudi ideološko (Gantar, 2015, str. 399), še posebej pri družbeno občutljivih elementih slovarja.

## 2.2 Od oznak k daljšim pojasnilom

Oznake in pojasnila o rabi pri slovarskih opisih naj bi bile vključene kot informacija za govorce določenega jezika predvsem kot pomoč pri odločanju o ustrezni rabi besed v določenem kontekstu (Namatende Sakwa, 2011, str. 305). Pri tem se uporablja različne sisteme oznak in pojasnil, ki so si bili v času tiskanih slovarjev v veliki meri podobni. V slovenskem okolju se pri označevanju besedišča tudi pri nastajajočih novih slovarskih opisih v veliki meri prevzema uveljavljenega s Slovarjem slovenskega knjižnega jezika, ki tako izhaja iz obdobja tiskanega slovarja, kjer se je besedišče označevalo z oznakami tipično pred pomenskim delom iztočnice, v vsebinskem smislu pa se umešča v razumevanje slovarja kot informativno-normativnega, torej predvsem v razmerju do razumevanja koncepta knjižnega jezika in označevanje besedišča, ki izhaja iz njega (Kalin Golob in Gantar, 2015, str. 452). To pri oznakah pomeni osredinjenje na pojasnila o omejitvah pri jezikovni rabi glede na rabo v knjižnem jeziku.

V sodobnem slovaropisju je tako zaradi velike količine korpusnih podatkov kot tudi novega medija slovarjev možno prikazati informacije, ki so bile prej podane s klasičnim sistemom slovarskih oznak, na različne nove načine (Kosem, 2015, str. 483), pri tem čemer je smiseln razmislek tako o tehnični izvedbi označevanja v novem slovarskem mediju kot vsebinski. V primeru negativno zaznamovanega besedišča v SSSS je pri vsebinskem označevanju potreben razmislek, kako opozoriti na okoliščine rabe in podati informacijo o pragmatičnem pomenu (Šorli, 2014, str. 480; Šorli, 2015, str. 480), in sicer na način, ki bo slovarskemu uporabniku dal jasno informacijo ne le o zaznamovanosti, ampak na podlagi analize korpusnih podatkov tudi o nameri govorcev, ko je ta ključna sestavina pomena, ko je del pragmatičnega pomena namera govorca, da npr. izraža sovraštvo ali užali. Poleg jasne kratke oznake o zaznamovanosti želimo podati tudi pojasnilo, ki uporabniku daje jasno informacijo o tem, kaj njegova jezikovna izbira povzroča.

## 3 Problemi SSSS 1.0

SSSS 1.0 je pripravljen strojno in povsem ročno nepregledan. V tej različici slovarja so kot iztočnice in sopomenke navedene leme (brez

besednih vrst), pomensko členitev in opis začasno nadomeščajo strojno pripravljene pomenske gruče, slovar pa tudi ne vsebuje slovarskih oznak, razen področnih.

Raziskava o odnosu uporabniške skupnosti do SSSS 1.0, v kateri je sodelovalo 671 anketirancev, je pokazala naklonjenost do večine novosti, ki jih prinaša (odzivni) slovar, npr. stalno posodabljanje, strojni postopki, digitalni format, kolokacijski podatki, povezave na korpus, uporabniško vključevanje (Arhar Holdt, 2020, str. 470). Med problematičnimi značilnostmi sta bili izpostavljeni nezanesljivost strojno pridobljenih podatkov in primanjkljaj slovarskih oznak tako pri jedrnih in bližnjih sopomenkah kot pri uporabniško dodanih. To, da ni oznak, je motilo 37 % sodelujočih (ibid., str. 472).

Pomanjkljivosti SSSS 1.0, ki jih je izpostavila uporabniška skupnost, so pri negativno zaznamovanem besedišču še posebej pereče. Na eni strani se strojno pripravljene iztočnice in sopomenski kandidati pojavljajo brez oznak ali opozoril tudi pri izrazito problematičnih primerih, kot je npr. iztočnica *buzi* s sopomenkami *peder*, *buzerant*, *toplovodar*, *homič*, *poženščen moški*. Na drugi strani je problem potencialno zavarajoča (ne)zastopanost sopomenskega gradiva, npr. vse sopomenke, ki jih najdemo pri iztočnici *zmaj* – *ksantipa*, *veščča*, *strupenjača*, *babura*, *coprnica*, *pošast*, *kričava ženska* – so vezane na ženski spol in imajo izrazito negativno konotacijo, čeprav se beseda rabi tudi za moške in (npr. v pomenu članov športnega kluba) brez negativne konotacije.

Tudi kolokacije in zgledi, ki so namenjeni primerjavi rabe dveh sopomenk, so iz referenčnega korpusa izvoženi strojno in so v slovarju brez oznak. Posledica je lahko sopostavitev pomensko neustreznih podatkov, npr. pri primerjavi besed *ženska* – *kura* najdemo prekrivne kolokacije [*stara, prava, gola*] *ženska* in [*stara, prava, gola*] *kura* ali *ženska* [*brez glave, v postelji, na odru*] in *kura* [*brez glave, v postelji, na odru*]. Korpusni zgledi načeloma pomagajo razdvoumiti problematične primere, vendar niso na voljo za vse primerjane besede, zgledi, ki so na voljo, pa niso izbrani po vsebinskih kriterijih. To je zlasti problematično pri sovražnem besedišču, npr. kolokacije [*sovražiti, tepsti, ubiti*] *pedra* ali zgledi tipa *In reskiral sem celo, da bi me imel za pedra*.

Določene težave se pojavljajo tudi pri uporabniško predlaganih sopomenkah. Tu ločujemo na eni strani zlonamerne vnose, kot je npr.

uporabniški vpis *aljaz* pri iztočnici *gej*. Za takšne primere bi bilo treba določiti natančno uredniško politiko za sprotno obravnavo na ravni vmesnika. Na drugi strani uporabniki zaznamovano besedišče dodajajo kot dejanski sopomenski predlog, npr. pri iztočnici *južnjak*, kjer so uporabniki dodali dolg niz predlogov, mdr. *jugovič*, *južni brat*, *jugič*, *trenirkar*, *bosanec*, *z juga*. V SSSS 1.0 so nekateri uporabniki in uporabnice oznako ali kako drugo pojasnilo v oklepaju pripisali ob svoj sopomenski predlog, npr. *bojazljivec – pezde (vulg.)*, *Italijanka – makaronarka (slabš.)*, vendar je bilo takšno označevanje sporadično in nesistemizirano. Uredniška naloga je presoditi, kateri predlogi so relevantni za vključitev v slovarsko bazo (in s katerimi slovarskimi oznakami), že uporabnikom pa omogočiti, da problematično besedišče označijo kot tako, da se torej oznaka v vmesniku prikaže hkrati z dodano sopomenko.

Različico SSSS 2.0 smo pripravili pod okriljem projekta Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL, ki ga je med leti 2021–22 financiralo Ministrstvo za kulturo Republike Slovenije. V slovarsko nadgradnjo smo vključili več ciljev, ki naslavlja zgoraj našteje probleme: (a) razviti in preizkusiti sistem označevanja negativno zaznamovanega besedišča znotraj koncepta odzivnega slovarja; (b) identificirati besedišče, ki je negativno zaznamovano v vseh pomenih besede in ga je zato mogoče označiti na ravni leme, in ga označiti po celotnem slovarju SSSS; (c) na izbranem naboru gradiva preizkusiti označevanje negativne zaznamovanosti pri pomensko členjenih slovarskih geslih ter (d) dodati v slovarski vmesnik možnost, da uporabniki sami označijo svoje predloge. V nadaljevanju predstavimo izhodišča, metodologijo in rezultate za vsakega od ciljev, prikažemo primere implementacije in vizualizacije v prenovljenem slovarskem vmesniku, prispevek pa sklenemo z opredelitvijo načrtov in prioritet za nadaljnji razvoj slovarja.

## 4 Identifikacija in označevanje problematičnega besedišča

### 4.1 Slovaropisna izhodišča in sistem oznak

Prepoznavanje potencialnega, z vidika družbene občutljivosti problematičnega besedišča temelji na slovaropisnih izhodiščih, ki jih pripravljamo

za slovarske vire na CJVT UL, prvič pa smo jih začeli uveljavljati pri izdelavi Velikega slovensko-madžarskega slovarja (Kosem et al., 2018a). V izhodišča je vključeno prepoznavanje elementov sovražnega govora (oznaka *sovražno*), elementov nevljudnosti, žaljivosti (*grobo*) ter elementov negativnega vrednotenja ali konotacije (*izraža negativen odnos*). Omenjene vrednotenjske oznake sodijo v širši okvir t. i. sporočanj-skih oznak,<sup>2</sup> ki opredeljujejo izraze ali pomene z vidika njihove rabe v sporočanjem procesu in v situacijah, v katerih sporočanje poteka. V predlaganem slovaropisnem opisu so sporočanje oznake namenjene označevanju izrazov, z izbiro katerih govorniki dosežemo ali želimo doseči določen učinek pri naslovniku. Ta učinek je lahko povzročen s pozitivnim ali negativnim vrednotenjem, z uporabo v določenem govornem položaju (npr. javnem, nejavnem) ali z namenom izraziti odnos do predmetnosti ali vsebine, ki temelji na določenih družbenih normah, pričakovanjih in odstopanjih od njih. Ta sistem se od tradicionalnega označevanja besed na podlagi odnosa do knjižne norme, kot ga pozna SSKJ (t. i. stilno-zvrstni in ekspresivni kvalifikatorji), ločuje v kvalificiranju besedišča na podlagi sporočanje namena in učinka, pri čemer izhodišče kvalificiranja ni v opozarjanju na odstop od knjižne norme, pač pa v informiranju glede možnih posledic rabe. S takim sistemom se želimo izogniti morebitnemu kvalificiranju govornika samega, hkrati pa opozoriti na kontekst potencialno problematične rabe v informativnem smislu. To pomeni, da ne želimo uporabnikov slovarja obveščati samo o možnih učinkih rabe grobega in sovražnega besedišča, pač pa posredno pokazati tudi na okoliščine, v katerih je tako rabo mogoče prepoznati.

V slovarskem sistemu vrednotenjskih oznak s t. i. negativnega pola<sup>3</sup> označujemo z oznako *sovražno* izraze in pomene, ki so diskriminatorni, ksenofobični, rasistični in homofobični, ki so uperjeni proti predstavniki skupin ali manjšin na podlagi njihove narodnosti, rase ali etničnega porekla, verskega prepričanja, spola, zdravstvenega stanja, spolne

2 Celotni sistem označevanja, ki ga razvijamo v okviru virov CJVT UL, poleg sporočanje oznak, ki jih notranje členimo na vrednotenjske, registrske in stilne, zajema še nabor pragmatičnih, kontekstualnih, področnih, slovničnih, časovnih in trendovskih oznak ter nabor oznak, vezanih na tuja poimenovanja in prevodne ustreznice.

3 Poleg oznak negativnega vrednotenjskega pola uporabljamo tudi pozitivni vrednotenjski oznaki "izraža pozitiven odnos" in "lahko izraža pozitiven odnos" (npr. pri *hribovka*, *blazen*). Oznako "lahko izraža negativen odnos" uporabljamo za omilitev negativne konotacije (npr. pri *češplja*).



usmerjenosti, invalidnosti ter drugih lastnosti in prepričanj. Z oznako *sovražno* se torej opredeljujemo do vseh izrazov, ki spodbujajo sovražstvo, predsodke ali nestrpnost in s tem lahko predstavljajo – kot je bilo opredeljeno že v razdelku 2 – elemente sovražnega govora.

Na drugi strani z oznako *grobo* označujemo izraze ali pomene, ki so za naslovnika lahko žaljivi, z vidika družbenih in moralnih norm pa neprimerni. Tipično se nanašajo na človeško ali živalsko telo, spolnost, prehranjevanje in izločanje – zlasti torej na tabuizirano predmetnost – in so rabljeni v neformalnih govornih situacijah.

Tretji sklop predstavlja besedišče, ki izraža neodobravanje, nenaklonjenost, posmehljivost ali kritiko do lastnosti posameznikov, predmetov ali dejanj. Z oznako *izraža negativen odnos* želimo tako opozoriti na izraze z negativno konotacijo ali vrednotenjem, ki so lahko za naslovnika žaljivi ali neprijetni.

## 4.2 Študentska identifikacija potencialno problematičnega besedišča

Potencialno problematično besedišče v SSSS smo identificirali z ročnim pregledom iztočnic in sopomenk v slovarju. Omejili smo se na slovarske (jedrne in bližnje) sopomenke, saj pregled uporabniških predlogov zahteva dodatne uredniške premisleke in bo zato opravljen kasneje s prilagojeno metodologijo. Zaradi obilja gradiva smo delo organizirali v dva koraka: širši pregled, v katerem smo v grobem ločili potencialno problematično in neproblematično gradivo, nato pa natančnejši pregled problematičnih primerov.

Najprej smo s specializirano programsko skripto iz slovarske baze izvozili nize sopomenk, urejenih na podlagi pomenskih gruč (Krek et al., 2017), npr. *speljati se; izginiti; pobrati se; skidati se; spokati se; spizditi*, pri čemer smo strojno odstranili nize, ki so se glede nabora sopomenk podvajali, in tiste, ki so bili podmnožica kakega drugega niza. Na tak način smo pripravili 65.615 nizov različnih obsegov: od posameznih sopomenskih parov do zelo dolgih nizov, ki pa so redki: več kot 30 sopomenk vsebuje le 156 nizov, povprečje je 5 sopomenk na niz.

Čeprav strojno pomensko gručenje ni povsem natančno in se razlikuje od slovaropisne pomenske členitve, tovrstna organizacija podatkov

dobro naslovi dva pomembna problema: (a) tak pristop bistveno pohitri pregledovanje, kot bo razvidno v nadaljevanju; (b) presojanje je lahko bolj natančno, saj problematičnost posamezne leme nakazujejo ostale besede v nizu, prim. npr. *nategniti* v nizu *raztegniti; dilatirati; iztegniti; nategniti; pogrniti; razgrniti; razmakniti; razpreti; razprostrti; razviti; napeti; zavlačevati z; razpeti; prolongirati* in v nizu *pokavsati; nategniti; povaljati; porivati; pofukati; pojahati*.

Iz množice 65.615 nizov smo nato odstranili 24.945 nizov (38,0 %), pri katerih sopomenke vsebujejo področne oznake, npr. *odbojnik, deflektor, ločilnik, membrana, opna, odbojna pregrada, zvočna stena z oznako električna*. Ti podatki so terminološke narave, zato smo predvidevali zanemarljivo nizko vsebnost problematičnega besedišča. Predpostavko smo preverili s pregledom 200 naključno izbranih nizov, v katerih se problemi skladno s pričakovanji niso pojavljali. Preostalega terminološkega gradiva nismo ročno pregledovali. Ostalo je 496 nizov (0,8 %), ki vsebujejo lastnoimenske samostalnike, npr. *Antarktika, antarktično območje, južno polarno območje*, in 40.176 (61,2 %) občnoimenskih nizov, ki so bili vsi vključeni v ročni pregled.

Podatke so pregledovali študentke in študenti jezikoslovnih smeri, in sicer po trije vzporedno. Pregledovanje je potekalo v okolju *Google Sheets*. Sopomenske nize smo organizirali v vrstice tabele, kjer jim je bilo mogoče pripisati eno od naslednjih odločitev: (1) niz vsebuje sovražno ali grobo besedišče; (2) niz vsebuje besedišče, ki je drugače negativno ali (v določenem pomenu, kontekstu) izraža negativen odnos; (3) z vidika sovražnosti, grobosti, negativnosti je niz neproblematičen. Če so pregledovalci želeli, so lahko opredelili tudi, da je (4) v nizu kako drugače zaznamovano besedišče, da (5) ne razumejo vseh besed v nizu, lahko pa so vpisali tudi dodaten komentar na svoje odločitve ali podatke.

Odločitve so študentke in študenti sprejemali na podlagi kratkih navodil za prepoznavanje potencialno negativne zaznamovanosti besed, kot so elementi sovražnega govora (rasna, etična diskriminacija, diskriminacija na podlagi spola, spolne usmerjenosti, hendikepa) negativnega vrednotenja (glede na družbeni status, gmotni položaj, obnašanje in značaj, izgled ipd.) in grobosti (glede na tabuizirano predmetnost, npr. spolnost, telesno izločanje, nasilje, in tipično neformalnost

govornega položaja). Pri mejnih primerih so izbrali tisto izmed odločitev, ki nakazuje višjo stopnjo negativnosti (npr. pri mejnih primerih med kategorijo 1 in 2 so izbirali 1).

Študentski pregled je zajemal 40.672 sopomenskih nizov, ki jih je pregledalo 6 sodelujočih, kar pomeni, da je vsak v povprečju pregledal 20.336 nizov. Priprave na označevanje (priprava podatkov in smernic, testno označevanje in uvodni sestanki za reševanje izhodiščnih dilem) so potekale v januarju in februarju, pregledovanje pa od marca do maja 2022, pri čemer je treba upoštevati, da zaradi drugih obveznosti študenti na projektih sodelujejo z omejenim naborom tedenskih ur. V povprečju so v navedenih treh mesecih porabili 72 ur na osebo, kar pomeni 12,7 sekund na sopomenski niz. Kljub veliki količini podatkov je bila naloga torej izvedljiva v relativno kratkem času, saj so študentje lahko odločitev podali takoj, ko so v nizu našli eno samo problematično besedo, natančnejše razmisleke o vrsti zaznamovanosti oz. označevanja posameznih besed pa so prepustili za drugi korak dela s podatki.

Vsak sopomenski niz so pregledali po trije študenti in študentke. Pri nalogi nas ni zanimalo ne/ujemanje med njihovimi odločitvami (čeprav posredno nakazuje težja mesta za označevanje), ampak je trojni pregled služil za zagotavljanje večje natančnosti pri ločevanju neproblematičnega gradiva od tistega, ki potrebuje nadaljnji pregled. Posamezne študentske odločitve smo pretvorili v skupne po naslednjem ključu: (1) **sovražno/grobo**: če je vsaj eden od študentov presodil, da se v nizu pojavlja sovražno ali grobo besedišče; (2) **drugače negativno**: kombinacije odločitev “druga negativnost” in “neproblematično” ali (3) **neproblematično**: če so vsi študenti presodili, da je z vidika sovražnosti, grobosti, negativnosti niz neproblematičen. Rezultate prikazuje Tabela 1.

**Tabela 1:** Številčna zastopanost in delež nizov glede na skupno študentsko odločitev potencialne problematičnosti

Kategorija končne odločitve	Število nizov v kategoriji	Delež glede na vse pregledano
Sovražno/grobo	1.810	4,5 %
Drugače negativno	12.730	31,3 %
Neproblematično	26.132	64,3 %
Skupaj	40.672	100,0 %

V Tabeli 2 navajamo nekaj nizov s po tremi sopomenkami, ki so jim študentke in študenti pripisali skladne ali različne odločitve. Kot je razvidno, lahko posamezen niz vsebuje raznoliko zaznamovano besedišče, kot tudi nezaznamovano besedišče. Sopomenske nize z najvišjo stopnjo negativne zaznamovanosti (odločitev 1) smo na projektu natančneje obravnavali, kot opisujemo v nadaljevanju. Z ročnim pregledom pa smo identificirali tudi gradivo, ki je na tak ali drugačen način relevantno za nadaljnje delo (odločitev 2), in gradivo, ki ga z vidika negativne zaznamovanosti ne bomo nadalje obravnavali (odločitev 3).

**Tabela 2:** Primeri nizov s študentskimi odločitvami o nadaljnji obravnavi

Niz sopomenk	Posamezne in skupna odločitev
fukati; porivati; natepavati	111 -> 1
skozlati; izbruhati; zbruhati	111 -> 1
pedrski; buzerantski; toplovodarski	111 -> 1
črnuhinja; zamorka; zamorklja	111 -> 1
pofukanka; prasica; zajebanka	111 -> 1
debilen; bebast; duševno zaostal	121 -> 1
kripelj; pohabljenec; pohabljenka	211 -> 1
kurnik; pajzelj; temačna luknja	222 -> 2
bedastoča; glupost; nesmisel	222 -> 2
eliminirati; likvidirati; usmrtili	222 -> 2
izmozgano; izčrpano; mršavo	223 -> 2
imenski; nazivni; nominalni	333 -> 3
kopirni papir; indigo; karbon	333 -> 3
zaustaviti se; izklopiti se; izključiti se	333 -> 3

V “drugače negativno” so raznorodni primeri, saj so poleg zaznamovanih izrazov in pomenov (npr. *budala*, *avša*, *bedast*) študentje označevali tudi vrednotenjsko nevtralno besedišče, ki poimenuje negativne vsebine, dejanja in predmetnost. Gre zlasti za poimenovanja agresivnega obnašanja: *uničiti*, *dotolči*, nekaterih osebnih lastnosti: *pokvarjen*, *hudoben*, *ničvreden*, *grozljiv*, *grd*, *apatičnost*, *pokvarjenost*; videza, stanja: *neurejenost*, *razdejanje*, *zanikrnost* itd. V slovarju večina teh besed ne potrebuje oznake. Čeprav teh besed ne bomo označevali, so seznanili tovrstnega besedišča pomemben rezultat ročnega pregleda, saj so koristni za različne druge namene na področju slovaropisja in strojne obdelave jezika, npr. za filtriranje gradiva z negativnim pomenom iz jezikovnih iger ali učnih gradiv, strojno pripisovanje sentimenta ipd.

### 4.3 Slovaropisni pregled in izbira slovarskih oznak

V 1.810 nizih z odločitvijo (1) smo določili besede in zveze, ki so relevantne za slovarsko označevanje z oznako *sovražno* ali *grobo*. Prvi izpis problematičnega gradiva iz sopomenskih nizov so pripravili študenti in študentke, nato pa je odločitve pregledala, sprejela in mestoma spremenila skupina treh jezikoslovcev. Odločanje je potekalo ob upoštevanju pojavljanja oz. rabe identificiranega besedišča v korpusih Gigafida 2.0 in Janes. Kvalitativna empirična analiza je dopolnila željo po pohitritvi prve selekcije, s čimer smo se obranili pred intuitivnim ali črno-belim presojanjem primernosti. V analizi smo identificirali tako primere, ki jih je mogoče označiti na ravni leme (npr. *črnuh*, *razpizditi*), kot primere, pri katerih bi bilo oznako *sovražno* ali *grobo* mogoče pripisati enemu ali več (ne pa vsem) pomenom besede (npr. *zamorec*, *debilen*; *nategniti*, *batina*).

Jezikoslovna analiza je razkrila tudi določen delež gradiva, ki je v slovar SSSS prišlo pomotoma oz. zaradi specifik v metodologiji njegove priprave iz baze Velikega angleško-slovenskega slovarja Oxford-DZS. Iz SSSS smo tako odstranili identificirane napačno izluščene primere (npr. *zamoreka*, *poneediti se*), množinske leme, ki se v bazi sicer pojavljajo tudi v ednini (npr. *babe*, *švabi*), in zveze, ki so v izvornem slovarju nastopale kot prevodne ustreznice razlagalnega tipa, v SSSS pa kot iztočnice nimajo pravega smisla (npr. *neuglajena podeželanka*, *bogat vulgarnež*). Rezultate kaže Tabela 3.

**Tabela 3:** Številčna zastopanost in primeri iztočnic (za pripis oznak ali za pomensko členitev)

	Primeri	Število besed oz. zvez
Oznaka <i>sovražno</i> na ravni leme	<i>črnuh</i> , <i>cigo</i> , <i>čifut</i> , <i>rdečuhinja</i> , <i>beli prasec</i> , <i>bela sodrga</i> , <i>lezba</i> , <i>lezbača</i> , <i>peder</i> , <i>buzerant</i>	27
Oznaka <i>grobo</i> na ravni leme	<i>podjebavati</i> , <i>v kurcu</i> , <i>zdrkati</i> , <i>zrajcan</i> , <i>pofafati ga</i> , <i>sranje</i> , <i>fentati</i> , <i>razpizden</i> , <i>sfukan</i> , <i>kurbarija</i>	140
Odstranitev iz SSSS	<i>zamoreka</i> , <i>poneediti se</i> , <i>babe</i> , <i>tipi</i> , <i>švabi</i> , <i>neuglajena podeželanka</i> , <i>bogat vulgarnež</i> , <i>črnski yuppie</i> , <i>bela golazen</i>	86
Besedišče za pomensko členjenje in (morebitno) označevanje na ravni pomenov	<i>baba</i> , <i>batina</i> , <i>blazen</i> , <i>češplja</i> , <i>coprnica</i> , <i>črv</i> , <i>debil</i> , <i>kmetica</i> , <i>kripelj</i> , <i>nabrisati</i>	234

Primere, kjer je problematičnost vezana na lemo ne glede na morebitno večpomenskost, smo v SSSS 2.0 označili in vizualizirali, kot predstavlja poglavje 6. Kjer se oznaka nanaša na posamezni pomen, pa smo iztočnice pomensko razčlenili in jim pripisali pomensko informacijo v obliki t. i. pomenskega indikatorja, kot predstavljamo v poglavju 5. Omeniti velja tudi, da smo pri jezikoslovni analizi veliko gradiva, ki so ga študenti umestili v kategorijo (1), premestili v kategorijo (2), kar pomeni, da ga bomo pri nadaljnjem razvoju slovarja predvidoma označili z oznako *izraža negativen odnos*, npr. *trapa*, *kozlarija*, *špeglarca*, *luftar*.

## 5 Pomenska obdelava iztočnic

Iztočnice, ki so bile s seznama relevantnih za slovarsko označevanje določene za pomensko analizo, so morale biti večpomenske, pri čemer so ob enem ali več vrednotenjsko nezaznamovanih pomenov morale nakazovati vsaj en, lahko pa tudi več negativno vrednotenih pomenov, npr. *jalov*, *batina*, *črv*, *gnoj*, *prasica* itd. Takih iztočnic je bilo na seznamu 234. Pomenska analiza je vključevala pomensko razčlenitev, ki je obsegala določitev števila in zaporedja pomenov, ter pomenski opis s pomočjo t. i. pomenskih indikatorjev.

### 5.1 Pomenska členitev

Znotraj pomenske obdelave iztočnic je proces pomenske členitve potekal na podlagi analize rabe besede v referenčnem korpusu standardnega jezika Gigafida 2.0 ter v korpusu Janes, za katerega velja, da se jezik v njem v marsičem razlikuje od pisnega standarda (Fišer et al., 2016, str. 68). Čeprav je pomenska členitev vključevala prepoznavanje vseh pomenov analizirane besede, torej tudi vrednotenjsko nezaznamovanih, smo bili pri pomenski analizi pozorni zlasti na pomene, ki izražajo katero od negativnih vrednotenj, še zlasti, če je bilo teh pomenov več. Večpomenskost je bilo mogoče prepoznati pri več samostojnih pomenih z negativno konotacijo, (npr. *bastard*, *degeneriran*) in pri stopnjevanju negativne konotacije (*cigan*, *debil*, *češplja*):

bastard

1. *izraža negativen odnos* nezakonski otrok
2. *izraža negativen odnos* izprijena, nasilna ali uporniška oseba

cigan

2. *izraža negativen odnos* podlež; lopov
4. *sovražno* pripadnik etnične skupine

debil

1. *izraža negativen odnos, lahko ironično nepremišljena, nespametna oseba*
2. *sovražno* oseba z motnjo v razvoju

češplja

2. *lahko izraža negativen odnos, neformalno ženska ali dekleta*
3. *grobo, neformalno* ženski spolni organ

## 5.2 Pomenski opis

Za pomenski opis pomensko razčlenjenih besed smo uporabili t. i. pomenske indikatorje (Gantar, 2015, str. 164), ki so eden od treh segmentov pomenske informacije v slovarskih virih CJVT. Poleg indikatorjev (v zgornjih primerih na drugem mestu) pomenski opis tvorijo še oznaka (na prvem mestu) in razlaga, ki je v zaenkrat SSSS ne predvidevamo.

Osnovna vloga pomenskega indikatorja je na kratko in prepoznavno opredeliti pomen besede glede na njene druge pomene. Pomenski indikatorji so za to v prvi vrsti namenjeni oblikovanju t. i. pomenskega menija, ki ga poznamo iz tujejezičnih slovarjev za tujce, uvajamo pa ga tudi v slovarske vire CJVT (prim. Kolokacije 1.0, Veliki slovensko-madžarski slovar 1.0). Specifičnost posameznega vira, v našem primeru slovarja sopomenk, narekuje tudi način oblikovanja indikatorjev. Ker gre za pomene z negativno konotacijo, je bilo treba premisliti ubesedenje indikatorjev z vidika opisovanja družbeno občutljivih vsebin na eni strani ter zagotavljanja pomenske obvestilnosti in konsistentnosti na drugi. Pri oblikovanju indikatorjev smo si zato prizadevali ohranяти čim večjo nevtralnost oz. splošnost ubeseditve, zlasti pri večpomenskih besedah, ki poimenujejo posameznike, npr. "oseba" (*debil, gnoj, kača, prasica*). V primeru ženskih oblik, ki se tudi uporabljajo zgolj za ženske osebe, smo ohranili indikator "ženska" (*češplja, klošarka, debilka*). Razlog za to je odraz dejanske rabe, je pa v zvezi s temi primeri smiselno razmisliti tudi o vlogi definicije

kot tretje pomenske informacije, ki lahko prevzame pojasnjevanje specifičnih pomenskih lastnosti, indikator pa bi zato lahko ostal bolj splošen tudi pri ženskih oblikah.

Pri besedah, ki označujejo telesni ali duševni hendikep, smo skušali čim bolj konsistentno slediti uporabi nevtralnih in družbeno korektnih indikatorjev, npr. 'oseba z motnjo v razvoju' (sovražno: *debil, imbecil*), 'o motnji v razvoju' (izraža negativen odnos: *debilnost*), 'o osebi z motnjo v razvoju' (izraža negativen odnos: *defekten, bebast*); 'o telesni okvari' (izraža negativen odnos: *defekten, liliputanski*) 'o osebi s telesno okvaro' (izraža negativen odnos: *kruljav*), 'oseba s telesno okvaro' (sovražno: *kripelj, nakaza*).

Na drugi strani pa zlasti enopomenskih besed in besed z več izrazito vrednotenjskimi pomeni ni mogoče pomensko opisati s povsem nevtralnimi indikatorji. V takih primerih smo indikatorje ubesedili tako, da ustrezno ponazarjajo pomensko specifiko besede in se tako na nek način približujejo razlagam, npr.

blazen

3. izraža poudarek, lahko izraža pozitiven odnos o nenavadnem ali neverjetnem

psiho

1. izraža negativen odnos, neformalno čudaška ali nevarna oseba

Tudi pri t. i. nenevtralnih indikatorjih smo skušali slediti konsistentni ubeseditvi, če je bilo z njo mogoče zajeti sorodne pomenske lastnosti besede, npr. 'ničvredna, izprijena oseba' (izraža negativen odnos: *degeneriranec, podgana, pes*), 'ničvredna oseba' (izraža negativen odnos: *rit, ušivec*), 'nepreudarna, nespametna ženska' (izraža negativen odnos: *gos, goska, kura*), 'neuglajena, nespametna oseba' (izraža negativen odnos: *govedo, kmet*).

S problemom nevtralne ubeseditve indikatorjev smo se tokrat prvič bolj poglobljeno ukvarjali prav zaradi kvalificiranja in pomenskega opisovanja družbeno občutljivih vsebin. Menimo, da bi bilo iskanje rešitev za ustrezno ubeseditve pomenske informacije pri družbeno občutljivih vsebinah v prihodnje smiselno nadgraditi tako z jezikoslovnimi analizami besedilnega okolja kot s sociolingvističnimi in žanrskimi analizami






okoliščin rabe. Na ta način bi bilo mogoče prepoznati tudi druge, ne zgolj leksikalne možnosti vrednotenja in pridobiti vpogled v celotno pomensko sliko. Spoznanja bodo koristna tudi pri oblikovanju strategij za vključitev tovrstne informacije v celotni pomenski opis leksike v Digitalni slovarski bazi in posledično v jezikovne vire CJVT.

## 6 Rešitve v SSSS 2.0

V slovarskem vmesniku SSSS 2.0<sup>4</sup> na sovražno in grobo besedišče opozarjamo s kombinacijo opozorilne ikone in daljšega pojasnila, ki se izpiše ob kliku nanjo. V različici 2.0 slovarske oznake *izraža negativen odnos* še nismo pripisovali, bo pa na voljo uporabnikom, zato smo pripravili pojasnila za vse tri vrednotenjske oznake hkrati (Tabela 4). Pri vizualizaciji smo se namenoma odrekli pripisovanju (eno-)besednih oznak, saj bi te pri označevanju (mestoma tudi homonimnih) lem lahko vodile v napačno interpretacijo podatkov. Pri pomensko členjenih geslih so oznake pripisane posameznim pomenom, pri pomensko nečlenjenih geslih pa kombinacija ikone in pojasnila omogoči, da je problematično besedišče na prvi pregled zelo opazno, pojasnilo pa je lahko daljše in vsebuje informacije o možnem učinku na naslovnika oz. možnih posledicah rabe označene besede.

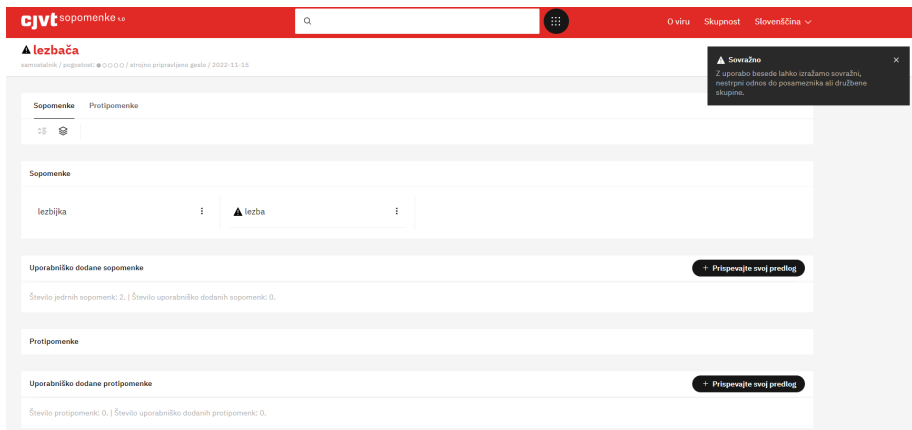
**Tabela 4:** Ikone in njihova pojasnila v SSSS 2.0

Oznaka	Ikona	Pojasnilo
Sovražno		Z uporabo besede lahko izražamo sovražni, nestrpni odnos do posameznika ali družbene skupine.
Grobo		Zaradi družbenih in moralnih norm se marsikateremu uporabniku jezika beseda lahko zdi groba ali neprimerna. Uporaba lahko povzroči nelagodje, razburi ali užali.
Izraža negativen odnos		Beseda lahko ni nevtralna. Z uporabo besede se lahko posmehujemo, izražamo neodobranje ali kritiko do nekaterih lastnosti posameznikov, predmetov ali dejanj.

Slika 1 kaže posnetek zaslona vmesnika SSSS 2.0 Slika ponazarja, kakšna bo vizualizacija z ikonami, ki lahko stojijo pri iztočnici (*lezbača*) ali pri sopomenki (*lezba*). Klik na ikono odpre pojasnilo desno zgoraj – v

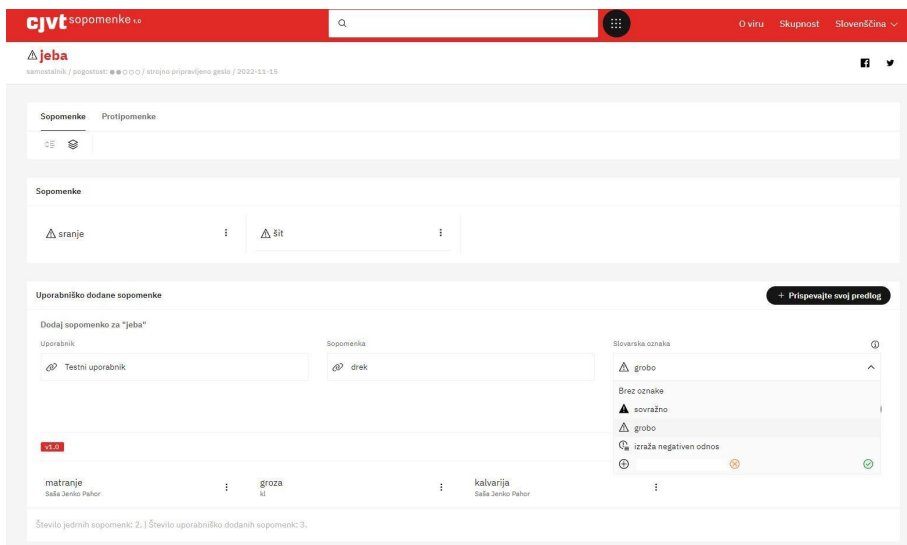
4 V času priprave prispevka je razvojno različico slovarja mogoče videti na naslovu: <https://viri.cjvt.si/sopomenke-beta/slv/>.

obliki pojavnega okenca, ki po nekaj sekundah samo izgine z ekrana. Razvidne so tudi nekatere druge novosti slovarja, ki se jim v tem prispevku ne posvečamo, npr. prenovljena oblikovna podoba vmesnika, informacija o besedni vrsti in pogostnosti iztočnice ter možnost dodajanja protipomenk. Ker vse težave SSSS niso enostavno in hitro rešljive, smo želeli slovarske uporabnike bolj opozoriti na trenutne omejitve. Čeprav je bila metodologija priprave SSSS 1.0 pojasnjena v razdelku *O viru*, pri samih iztočnicah ni bilo izrecnih opozoril, da je slovar pripravljen strojno, in to na vseh ravneh: sopomenke, kolokacije, korpusni zgledi, kar je lahko vodilo v napačne interpretacije slovarske vsebine. V SSSS 2.0 smo zato v zaglavje vključili tudi status izdelanosti gesla (*strojno pripravljeno geslo* na Sliki 1).



**Slika 1:** Iztočnica *lezbača* v SSSS 2.0: klik na opozorilno ikono odpre pojasnilo.

Protokol dodajanja sopomenk (in skladno tudi protipomenk) smo nadgradili, da bodo predlagani besedi ali zvezi uporabnice in uporabniki lahko dodali tudi slovarsko oznako oz. oznake. Privzeta izbira je, da je uporabniški predlog “brez oznake”, ostale možnosti so na voljo v spustnem meniju (Slika 2). V različici SSSS 2.0 so na klik na voljo oznake *sovražno*, *grobo* in *izraža negativen odnos*, poleg tega pa okence, v katerega je mogoče vtipkati morebitno drugo oznako.



**Slika 2:** Iztočnica *jeba* v SSSS 2.0: uporabniško dodajanje nove sopomenke in slovarske oznake.

Pomen in raba oznak *sovražno*, *grobo* in *izraža negativen odnos* je razložena in ponazorjena s primeri, s čimer bo lahko dosežena določena stopnja enotnosti uporabniškega označevanja (informacije so na voljo na klik, gl. ikono (i) na Sliki 2). Predvideno pa je, da bodo uporabniki oznake mestoma interpretirali in uporabljali drugače, kot bi jih slovaropisci. Vse dodane oznake bodo (skupaj z dodanimi sopomenkami oz. protipomenkami) preverjene in uporabniški predlogi bodo dragoceno gradivo ne le za dopolnitev odprto dostopne slovarske baze, ampak tudi za analize širšega dojemanja označevalnega sistema ter dometa in meja oznak za negativno vrednotenje. Prav tako pomemben uvid bodo ponudile ročno vpisane oznake, ki jih bomo analizirali z vidika vsebine in pogostosti ter uporabili izsledke za nadaljnji razvoj slovarja.

## 7 Sklep in nadaljnje delo

Sodobno slovaropisno delo ima ob zavedanju ideološkosti, vključevanju novih pristopov, uporabi tehnologije, moči množic itd. danes veliko možnosti, da tudi vprašanja označevanja konotacije naslavlja na novo in zanj pripravlja inovativne rešitve (Gorjanc, 2017, str. 154).

V prispevku smo opisali, kako poteka obravnava sovražnega in grobega besedišča v SSSS in katere spremembe so na voljo v različici 2.0. Rešitve naslavljajo dve pomembni značilnosti SSSS: njegovo strojno izdelanost in odprtost, da pri razvoju slovarja sodeluje tudi uporabniška skupnost. V novi različici slovarja so sovražnemu in grobemu besedišču pripisane slovarske oznake oz. opozorilne ikone s pojasnili o možnih učinkih rabe in dodana je možnost, da uporabniki pripišejo oznako svojim predlogom sopomenk.

Prepoznano sovražno in grobo besedišče bo koristno tudi pri izdelavi drugih virov, kjer se za pomene izbirajo reprezentativne kolokacije in zgledi. Pri izdelavi novih gesel za Kolokacijski slovar sodobne slovenščine (Kosem et al., 2018b) npr. že zdaj pri pripravi podatkov (pred slovaropisno analizo) označujemo kolokacije, ki vsebujejo sovražno in grobo besedišče, pa tudi besedišče, ki izraža negativen odnos. Tako slovaropiske in slovaropisce opozorimo na potencialno problematične kolokacije in posledično pohitrimo delo oz. se izognemo vključevanju problematičnih vsebin. Sezname problematičnega besedišča, ki jih uporabljamo trenutno, so pripravljene *ad hoc* iz odprto dostopnih jezikovnih virov in precej krajši od seznamov, ki bodo (lahko) nastali na osnovi predstavljenega dela.

Izražanje negativnega odnosa je pogosto vezano na posamezen pomen besede, zato bo velik del naloge izvedljiv šele ob pripravi pomensko členjenih gesel. Pri pomenski členitvi in nadaljnjem označevanju gradiva SSSS bomo uporabljali metodologijo, ki jo razvijamo pri izdelavi Velikega slovensko-madžarskega slovarja (Kosem et al., 2018a) in podatke oz. informacije, ki so na voljo v obstoječih odprto dostopnih virih za slovenščino. Preizkus prenosa metodologije smo izvedli pod okriljem projekta Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL, kjer je bila med cilje vključena tudi nadgradnja SSSS z 2.000 pomensko členjenimi gesli, ki imajo slovaropisno pregledane in razvrščene sopomenke ter kolokacije.

Nadaljnji razvoj v tej smeri omogoča predvsem projekt Nadgradnja portala jezikovnih virov CJVT, ki poteka v letih 2022–2023, financira ga Ministrstvo za kulturo. V okviru projekta se izdeluje urejevalnik za Digitalno slovarsko bazo (DSB – ta ob drugih tipih podatkov vsebuje tudi kolokacijske podatke in podatke o sopomenskosti), kar omogoča lažje

pomensko členjenje večje količine iztočnic. V okviru istega projekta v DSB integrirana večja količina pomenskih podatkov iz slovarjev, enciklopedij in podobnih virov, ki so bili odkupljeni v okviru projekt Razvoj slovenščine v digitalnem okolju in so bili dani v odprt dostop. Ob upoštevanju korpusnih podatkov, ki so že vključeni v slovaropisni proces, je s tem omogočeno tudi strojno gručenje pomenov (ang. *word sense induction*) na ravni vseh iztočnic, ki so trenutno vključene v slovarsko bazo.

Zato bi bilo v nadaljnje premisleke glede sovražnega in grobege besedišča znotraj koncepta odzivnega slovarja smiselno celoviteje vključiti tudi analizo okoliščin rabe, kar je do neke mere mogoče izvesti tudi strojno. Zanimivo bi bilo obravnavati zaznavanje in presojanje sovražnosti, grobosti v različnih tipih besedil, npr. medijskih. Ob tem se odpira tudi vprašanje formalnosti in neformalnosti položajev, na katere se ta presoja nanaša: ali posega na vse ravni izražanja ali gre zgolj za formalne, javne položaje in ali je neodvisna od generacijske ali kake druge pripadnosti presojevalca. V tej luči bi bilo lahko zanimivo sodelovanje jezikoslovk in jezikoslovcev s strokovnjakinjami in strokovnjaki s področja sociologije, etnologije, antropologije, psihologije, ki obravnavajo družbene okoliščine, generacijske kontekste, formalne in neformalne pozicije itd. Ob vsem naštetem se je treba zavedati, da slovaropisje naslavlja zlasti pojavnosti sovražnosti in grobosti, ki so v besedilnem kontekstu leksikalno izražene. Tudi s tega vidika bi bilo delo smiselno povezati s področji žanrske analize, pragmatike, kritične analize diskurza, strojne identifikacije sovražnega govora in podobnih pristopov, ki naslavlajo prikrite, implicitne, posredne načine izražanja sovražnosti oz. negativnega vrednotenja nasploh.

## Zahvala

Projekt Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL je v letih 2021–2022 financiralo Ministrstvo za kulturo Republike Slovenije. Raziskovalne programe št. P6-0411 (Jezikovni viri in tehnologije za slovenski jezik), št. P6-0215 (Slovenski jezik – bazične, kontrastivne in aplikativne raziskave) ter št. P6-0436 (Digitalna humanistika: viri, orodja in metode) sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije iz državnega proračuna.

## Literatura

- Arhar Holdt, Š. (2020). How Users Responded to a Responsive Dictionary: The Case of the Thesaurus of Modern Slovene. *Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje*, 46(2), 465–482. doi: 10.31724/rihjj.46.2.1
- Arhar Holdt, Š., Gantar, P., Kosem, I., Pori, E., Logar, N., Gorjanc, V., & Krek, S. (2022). Sovražno in grobo besedišče v odzivnem Slovarju sopomenk sodobne slovenščine. V D. Fišer & T. Erjavec (ur.), *Jezikovne tehnologije in digitalna humanistika: zbornik konference*, 1. izdaja (str. 10–16), Ljubljana, Slovenija. Inštitut za novejšo zgodovino. Dostopno prek [https://nl.ijs.si/jtdh22/pdf/JTDH2022\\_Proceedings.pdf](https://nl.ijs.si/jtdh22/pdf/JTDH2022_Proceedings.pdf)
- Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, A., Gorjanc, V., Klemenc, B., Kosem, I., Krek, S., Laskowski, C., & Robnik Šikonja, M. (2018). Thesaurus of Modern Slovene: By the Community for the Community. V J. Čibej, V. Gorjanc, I. Kosem & S. Krek (ur.), *Proceedings of the 18th Euralex International Congress: Lexicography in Global Contexts* (str. 401–410). Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Dostopno prek <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1>
- Fišer, D., Erjavec, T., & Ljubešič, N. (2016). JANES v0.4: korpus slovenskih spletnih uporabniških vsebin. *Računalniško posredovana komunikacija*, 4(2), 67–99. Dostopno prek <https://journals.uni-lj.si/slovenscina2/article/view/7003/6694>
- Gantar, P. (2015). *Leksikografski opis slovenščine v digitalnem okolju*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Dostopno prek <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/download/62/138/2602-1?inline=1>
- Gorjanc, V. (2005). Neposredno in posredno žaljiv govor v jezikovnih priložnikih: diskurz slovarjev slovenskega jezika. *Družboslovne razprave*, 21(48), 197–209.
- Gorjanc, V. (2017). *Nije rečnik za seljaka*. Biblioteka XX vek, Beograd.
- Hughes, G. (2009). *Political Correctness: A History of Semantics and Culture*. 1st edition. Wiley-Blackwell, MA.
- Kalin Golob, M., & Gantar, P. (2015). Stilistika in enojezični slovar: Označevanje jezikovne variantnosti. V V. Gorjanc, P. Gantar, I. Kosem & S. Krek (ur.), *Slovar sodobne slovenščine: Problemi in rešitve* (str. 446–465). Dostopno prek <https://ebooks.uni-lj.si/zalozbau//catalog/download/15/47/530-1?inline=1>
- Kern, B. (2015). Politična korektnost v slovaropisju. V D. Zuljan Kumar & H. Dobrovoljc (ur.), *Zbornik prispevkov s simpozija 2013* (str. 144–154). Nova Gorica: Založba Univerze.

- Kosem, I. (2015). Oznake: Vrednotenjski pomen in pragmatična funkcija v slovarju. Slovarska baza in slovar. V V. Gorjanc, P. Gantar, I. Kosem & S. Krek (ur.), *Slovar sodobne slovenščine: Problemi in rešitve* (str. 482–494). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Kosem, I., Čeh Bálint, J., Gorjanc, V., Kolláth, A., Kovács, A., Krek, S., Novak-Lukanovič, S., & Rudaš, J. (2018a). *Osnutek koncepta novega velikega slovensko-madžarskega slovarja*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani. Dostopno prek <https://www.cjvt.si/komass/wp-content/uploads/sites/17/2020/08/Osnutek-koncepta-VSMS-v1-1.pdf>
- Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J., & Laskowski, C. (2018b). Kolokacijski slovar sodobne slovenščine. V D. Fišer & A. Pančur (ur.), *Jeziškovne tehnologije in digitalna humanistika*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Dostopno prek [http://www.sdj.si/wp/wp-content/uploads/2018/09/JTDH-2018\\_Kosem-et-al\\_Kolokacijski-slovar-sodobne-slovenscine.pdf](http://www.sdj.si/wp/wp-content/uploads/2018/09/JTDH-2018_Kosem-et-al_Kolokacijski-slovar-sodobne-slovenscine.pdf)
- Krek, S., Laskowski, C., & Robnik Šikonja, M. (2017). From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. V I. Kosem idr. (ur.), *Proceedings of eLex 2017: Lexicography from Scratch* (str. 93–109), Leiden, Netherlands. Dostopno prek <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>
- Krek, S., Laskowski, C., Robnik Šikonja, M., Kosem, I., Arhar Holdt, Š., Gantar, P., Čibej, J., Gorjanc, V., Klemenc, B., & Dobrovoljc, K. (2018). Thesaurus of Modern Slovene 1.0, Slovenian language resource repository CLARIN. SI, <http://hdl.handle.net/11356/1166>
- Logar, N., Perger, N., Gorjanc, V., Kalin Golob, M., Kogovšek Šalomon, N., & Kosem, I. (2020). Raba slovarjev v slovenski sodni praksi. *Teorija in praksa*, 57, 89–108. Dostopno prek: [https://www.fdv.uni-lj.si/docs/default-source/tip/tip\\_pos\\_2020\\_logar\\_idr.pdf?sfvrsn=0](https://www.fdv.uni-lj.si/docs/default-source/tip/tip_pos_2020_logar_idr.pdf?sfvrsn=0)
- Moon, R. (2014). Meanings, Ideologies, and Learners' Dictionaries. V A. Abel idr. (ur.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* (str. 85–105), Bolzano/Bozen. Institute for Specialised Communication and Multilingualism. Dostopno prek [https://euralex.org/elx\\_proceedings/Euralex2014/euralex\\_2014\\_004\\_p\\_85.pdf](https://euralex.org/elx_proceedings/Euralex2014/euralex_2014_004_p_85.pdf)
- Motl, A., & Bajt, V. (2016). *Sovražni govor v Republiki Sloveniji: Pregled stanja*. Mirovni inštitut, Ljubljana. Dostopno prek <https://dlib.si/stream/URN:NBN:SI:DOC-F2YZP2RB/c117f4c6-8fe9-437d-8c64-5b7987a856b6/PDF>

- Petković, B., & Kogovšek Šalamon, N. (2007). *O diskriminaciji: Priročnik za novinarje in novinarke*. Ljubljana: Mirovni inštitut. Dostopno prek <https://www.mirovni-institut.si/wp-content/uploads/2014/08/Prirocnik-o-diskriminaciji-final-all.pdf>
- Rebolj, D. (2008). Uporabnejša opredelitev politične korektnosti. V S. Autor & R. Kuhar (ur.), *Politična (ne)korektnost* (str. 4–15). Ljubljana: Mirovni inštitut. Dostopno prek <https://www.mirovni-institut.si/wp-content/uploads/2014/08/nestrpnost-6.pdf>
- Sakwa, N. (2011). Problems of Usage Labelling in English Lexicography. *Lexikos*, 21, 305–315. Dostopno prek <https://lexikos.journals.ac.za/pub/article/view/47>
- SSKJ. (2014). *Slovar slovenskega knjižnega jezika: Uvod*. Druga, dopolnjena in deloma prenovljena izdaja. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. Dostopno prek <https://fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika>
- Šorli, M. (2014). Pragmatic Meaning in Lexicographical Description: Semantic Prosody on the Go. V A. Abel, C. Vettori & N. Ralli (ur.), *Proceedings of the XVI EURALEX International Congress: The User in Focus* (str. 477–491), Bolzano/Bozen. Institute for Specialised Communication and Multilingualism.
- Šorli, M. (2015). Vrednotenjski pomen in pragmatična funkcija v slovarju. V V. Gorjanc, P. Gantar, I. Kosem & S. Krek (ur.), *Slovar sodobne slovenščine: Problemi in rešitve* (str. 466–480). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Vidovič Muha, A. (2013). *Moč in nemoč knjižnega jezika*. Ljubljana: Znanstvena založba Filozofske fakultete.



## Negative Vocabulary in the *Thesaurus of Modern Slovene 2.0*

The paper describes an upgraded version of the *Thesaurus of Modern Slovene 1.0*, which is currently the largest open-access collection of Slovene synonyms generated automatically. The creation of the thesaurus has introduced a new type of dictionary, referred to as a responsive dictionary, which allows the data to respond continuously to the opinions of the contributing language community. The upgrade was motivated by the results of a survey of the user community's attitudes towards the *Thesaurus of Modern Slovene*, which revealed a lack of dictionary labels, particularly for non-neutral vocabulary. As a result, the updated version of the thesaurus focuses on developing solutions for identifying and annotating extremely offensive and vulgar vocabulary. To address this, the digital medium is utilized to display information about potentially problematic vocabulary in new ways. The updated version of the thesaurus incorporates a combination of warning icons and longer explanations to provide a clear visual tag as well as an explanation about the potential consequences of word use. The identification of potentially negative words was primarily conducted manually. Synonym sets were exported from the dictionary database, ordered in semantic clusters, and reviewed by students who were provided with brief instructions to identify potentially negative words, such as elements of hate speech (discrimination based on race, ethnicity, gender, sexual orientation, or disability), negative attitudes (related to social status, wealth, behaviour and character, appearance, etc.), and vulgarity (related to taboo topics, e.g., sexuality, bodily excretions, and violence, in the typical informal speech situation). The decisions made by the students were reviewed and modified by a team of linguists, based on corpus data. As responsiveness is a key concept of the thesaurus, involving the user community in future labelling procedures is an important part of the preparation of final labelling solutions.

**Keywords:** thesaurus, responsive dictionary, dictionary labels, communicative purpose, user community