

Sovražno in grobo besedišče v odzivnem Slovarju sopomenk sodobne slovenščine

Špela Arhar Holdt,* ‡ Polona Gantar,* Iztok Kosem,* Eva Pori,*
Nataša Logar,** Vojko Gorjanc,* Simon Krek*

* Filozofska fakulteta, Univerza v Ljubljani
Aškerčeva 2, 1000 Ljubljana
apolonija.gantar@ff.uni-lj.si, iztok.kosem@ff.uni-lj.si, eva.pori@ff.uni-lj.si,
vojko.gorjanc@ff.uni-lj.si, simon.krek@ff.uni-lj.si

** Fakulteta za družbene vede, Univerza v Ljubljani
Kardeljeva ploščad 5, 1000 Ljubljana
natasalogar@fdv.uni-lj.si

‡ Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
Večna pot 113, 1000 Ljubljana
spela.arharholdt@fri.uni-lj.si

Povzetek

V prispevku predstavljamo rešitve za identifikacijo in označevanje sovražnega ter grobega 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 identifikacijo sovražnega in grobega besedišča ter pripis oznak oziroma opozorilnih ikon z daljšimi pojasnili. Oznake temeljijo na sporočanjem namenu oziroma učinku, pri čemer je njihovo bistvo informacija o možnih posledicah rabe. Pri označevanju tako kot pri izdelavi celotnega slovarja posvečamo veliko pozornost digitalnemu mediju in vizualizaciji rešitev v njem. Ker je odzivnost eden ključnih konceptov slovarja, se tudi pri rešitvah glede označevanja zavedamo pomembnosti sodelovanja z uporabniško skupnostjo, zato predlagamo še rešitve za sodelovanje s skupnostjo pri dodajanju oznak.

Extremely Offensive and Vulgar Vocabulary in the Responsive Thesaurus of Modern Slovene

In the paper we present the solutions for identification and annotation of extremely offensive and vulgar vocabulary, which can be found in the responsive Thesaurus of Modern Slovene. As this is the first project of its kind, the prepared solutions are to a great extent innovative, and have been devised considering the use of automatic methods in dictionary compilation, open access nature of dictionary data, and the inclusion of users into the compilation process. The paper describes the process of identification of extremely offensive and vulgar vocabulary, as well as the attribution of labels and warning icons containing longer explanations. The labels are based on their communicative purpose or effect, and are focussed on providing the information about potential consequences of word use. During the processes of labelling and dictionary compilation, considerable attention is paid to the digital medium and related visualisation solutions. As responsiveness is one of the key concepts of the dictionary, a part of preparing the labelling solutions was to design ways of including user community in labelling.

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: 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) potrditi ali zavrniti.¹

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 objavljane strojne in uporabniško dodane (nepregledane) 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 obravnavati besedišča, ki vrednostno poimenuje posamezne družbene skupine in njihove

pripadnike. Tako besedišče se trenutno v slovarju (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 bodo vključene v prihajajočo nadgradnjo SSSS. Med temi želimo posebej izpostaviti nove načine identifikacije in označevanja sovražnega, grobega ter drugače negativno vrednotenega besedišča, ki SSSS presegajo in so uporabne za različne sodobne jezikovne vire.

2. Sovražno, grobo, tabuizirano, zaničljivo ... 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: 13), v daljši in bolj povedni obliki pa kot (Petković in Kogovšek Šalomon, 2007: 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

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).

in poslabšanju družbenega položaja tistih, proti katerim je naperjen.

Motl in Bajt (2016: 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 – predvsem jezikoslovje, povezano s slovarji. Ameriško slovaropisje (Hughes, 2011: 3. pogl.) 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: 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: 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 vpludne 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: 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: 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: 89). Gre torej za to, da pri slovaropisnem delu končne rešitve preprosto ne morejo biti “samo jezikoslovne; neizogibno morajo biti tudi ideološke” (Moon, 2014: 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: 399), še posebej pri družbeno občutljivih elementih slovarja.

3. Problemi trenutnega SSSS

SSSS je pripravljen strojno in je trenutno na voljo v prvi, nepregledani različici, v kateri 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.

Navedene značilnosti imajo več posledic. 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 zavajajoča (ne)zastopanost sopomenskega gradiva, npr. vse sopomenke, ki jih najdemo pri iztočnici *zmaj* – *ksantipa*, *veščca*, *strupenjača*, *babura*, *coprnica*, *pošast*, *kričava ženska* – so vezane na ženski spol in imajo izrazito negativno konotacijo, čeprav se beseda rabi tudi za moške in (v drugem pomenu) tudi s pozitivno konotacijo.

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*.

Pri uporabniško predlaganih sopomenkah 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*. Uredniška naloga je presoditi, kateri predlogi so relevantni za vključitev v slovarsko bazo (in s katerimi oznakami), že uporabnikom pa omogočiti, da problematično besedišče označijo kot tako, da se torej oznaka v vmesniku prikaže istočasno kot dodana sopomenka.

Besedišče, ki je problematično na ravni same leme, je mogoče označiti že v obstoječi različici slovarja. Primeri, pri katerih je oznaka vezana na posamezen pomen besede ali specifičen kontekst rabe, pa zahtevajo predhodno pomensko členitev ter z njo povezan slovaropisni pregled kolokacij in zgledov rabe.

V projektu Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL bomo uresničili dva cilja: (a) identificirali besedišče, ki je problematično na ravni leme in ga označiti po celotnem slovarju SSSS ter (b) dodali v slovarski vmesnik možnost, da uporabniki sami označijo svoje predloge. V nadaljevanju natančneje pojasnujemo, kako.

4. Identifikacija problematičnega besedišča

4.1. Slovaropisna izhodišča in sistem oznak

Prepoznavanje potencialnega, z vidika družbene občutljivosti problematičnega besedišča temelji na slovaropisnih izhodiščih, ki smo jih pred nekaj leti pripravili za slovarske vire na CJVT UL, prvič pa začeli uporabljati 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 negativnen odnos*). Omenjene oznake sodijo v širši okvir t. i. sporočanjških oznak,² ki opredeljujejo izraze ali pomene z vidika njihove rabe v sporočanjškem procesu in v situacijah, v katerih sporočanje poteka. V predlaganem slovaropisnem opisu so sporočanjške oznake namenjene označevanju izrazov, z izbiro katerih govorci dosežemo ali želimo doseči določen učinek pri naslovniku. Ta učinek je lahko povzročen s pozitivnim ali negativnim

² Celotni sistem označevanja, ki ga razvijamo v okviru virov CJVT UL, poleg sporočanjških 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.

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 sporočanjaškega namena oz. 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 govorca 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 pokazati tudi na okoliščine, v katerih je tako rabo mogoče prepoznati.

V slovarskem sistemu oznak označujemo z oznako *sovražno* izraze in pomene, ki so diskriminatorni, ksenofobični, rasistični in homofobični, ki so uperjeni proti predstavnikom skupin ali manjšin na podlagi njihove narodnosti, rase ali etničnega porekla, verskega prepričanja, spola, zdravstvenega stanja, spolne usmerjenosti, invalidnosti, gmotnega stanja, izobrazbe, družbenega položaja 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 *grob* 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.

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

4.2. Ročni pregled gradiva

Potencialno problematično besedišče v SSSS smo identificirali z ročnim pregledom iztočnic in sopomenk v slovarju. Na projektu smo se omejili 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 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 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 različne dolžine: 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; razprostreti; razviti; napeti; zavlčevati z; razpeti; prolongirati* in v nizu *pokavsati; nategniti; povaljati; porivati; pofukati; pojahati*.

Iz množice 65.615 nizov smo najprej umaknili 24.945 nizov (38,0 %), pri katerih sopomenke vsebujejo področne oznake, npr. *odbojnik, deflektor, ločilnik, membrana, opna, odbojna pregrada, zvočna stena z oznako elektrika* (ker so ti podatki terminološke narave, smo predvidevali zanemarljivo nizko vsebnost problematičnega besedišča in smo jih pustili za hiter pregled ob koncu naloge); ostalo je 496 nizov (0,8 %), ki vsebujejo lastnoimenske samostalnice, npr. *Antarktika, antarktično območje, južno polarno območje*, in 40.176 (61,2 %) občnoimenskih nizov, vsi relevantni za 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.

Kljub ogromni količini podatkov je bila tako oblikovana naloga 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.

4.3. Rezultati ročnega pregleda

Študentske odločitve smo pretvorili v končne odločitve 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.

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 %

Tabela 1: Številčna zastopanost in delež nizov glede na končno odločitev glede potencialne problematičnosti.

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. Tabela obenem ponazarja gradivo, ki bo deležno celovite in natančnejše obravnave v zaključnem delu projekta Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL (odločitev 1), podatke, ki so na tak ali drugačen način relevantni za nadaljnje delo (odločitev 2), in gradivo, ki ga z vidika negativne zaznamovanosti ne bomo nadalje obravnavali (odločitev 3).

Niz sopomenk	Študentske in končna 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; usmrtiti	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

Tabela 2: Primeri nizov s študentskimi odločitvami in končno odločitvijo o nadaljnji obravnavi.

V sodelovanju s študenti bomo v 1.810 nizih z odločitvijo (1) določili besede in zveze, ki so relevantne za slovarsko označevanje. Slednje bo potekalo ob upoštevanju pojavljanja oz. rabe identificiranega besedišča v raznovrstnih kontekstih, s čimer bomo željo po pohitritvi postopka prve selekcije ustrezno nadzorovali in obranili pred črno-belimi presojanjem primernosti. Za ponazoritev navajamo nekaj primerov, ki so na seznamu za presojo:

- **sovražno:** *črnuh, črnuhinja, zamorklja, hlapčevski črnc, rdečuh, rdečuhinja, beli prasec, bela prasica, lezba, lezbača, peder, buzerant, pička, prasica, kripelj;*
- **grobo:** *podjebavati, v kurcu, zdrkati, nabrisati, pokavsati, nategniti v rit, pofafati ga, sranje, poscan, fentati, crkniti, razpizden, sfukan, kurbarija, joški.*

V SSSS želimo poleg sovražnega in grobega označiti tudi besedišče, ki izraža negativen odnos. To se najde predvsem v nizih z odločitvijo (2), mestoma pa tudi v (1). Kot problematične so študentje prepoznavali tako izraze (npr. *budala, avša, bedast*) kot potencialne problematične pomene besed (npr. *nataknjen, zabit, nasekati*). Prve je mogoče označiti že v trenutni različici slovarja, saj je njihova problematičnost vezana na lemo ne glede na morebitno večpomenskost. V drugem primeru bi oznaka morala biti pripisana pomenu, zato bo označevanje možno šele, ko bo slovar vseboval pomenske členitve. Primeri besedišča, ki ga je mogoče označiti na ravni izraza:

- **izraža negativen odnos:** *trapa, bebav, počasne pameti, lolek, kozlarija, zarukan, špeglarca, luftar, blefer, snobovski, drhal, težakinja, mlatenje prazne slame, avša, otročaj.*

V "drugače negativno" so raznorodni primeri, saj so poleg zaznamovanih izrazov in pomenov študentje označevali tudi besedišče, ki poimenuje **negativne vsebine 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 besed ne bomo označevali, so sezname 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.

5. Vizualizacija v vmesniku SSSS 2.0

V slovarskem vmesniku SSSS 2.0 bomo na besedišče, o katerem razpravljamo tu, opozorili s kombinacijo opozorilne ikone in daljšega pojasnila, ki se bo izpisalo ob kliku nanjo. Trenutno rešitev, ki jo bomo po potrebi še nadgradili, kaže Tabela 3. Namenoma smo se odrekli pripisovanju (eno-)besednih oznak, saj bi te pri označevanju (mestoma tudi homonimnih) lem lahko vodile v napačno interpretacijo podatkov. Pri pomensko členjenih geslih bodo oznake seveda pripisane posameznim pomenom, pri pomensko nečlenjenih geslih pa bo kombinacija ikone in pojasnila omogočila, da je problematično besedišče na prvi pregled zelo opazno, pojasnilo pa je lahko daljše in vsebuje informacije o možnem učinku na naslovnik oz. možnih posledicah rabe označene besede.




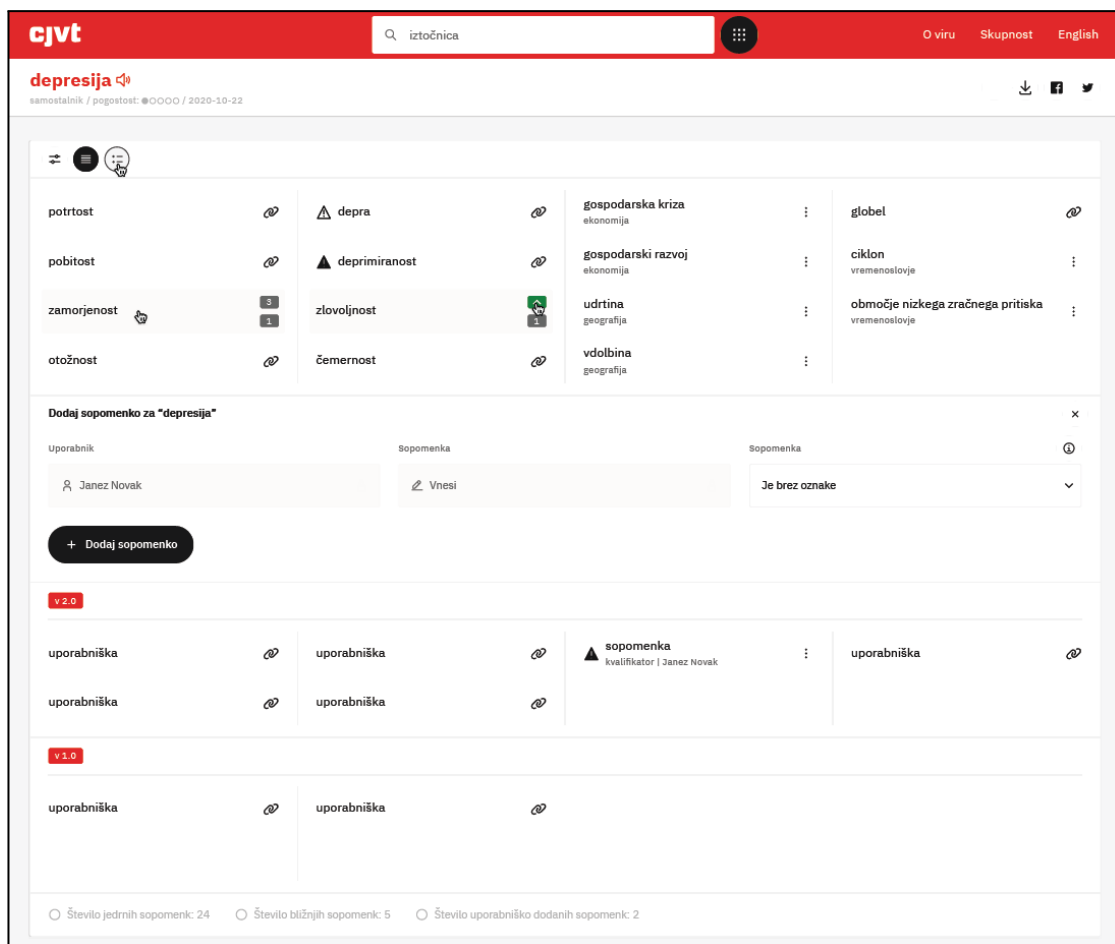
Oznaka	Ikona	Pojasnilo
Sovražno		Z uporabo besede lahko izražamo sovražni, nestrpni odnos do posameznika ali družbene skupine.
Grobo		Zaradi družbenih in moralnih norm se marsikateremu uporabniku jezika beseda lahko zdi groba ali neprimerna. Uporaba lahko povzroči nelagodje, razburi ali užali.
Izraža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

Tabela 3: Predvidene ikone in izhodiščna različica pojasnil za označevanje besedišča v SSSS.

Slika 1 kaže oblikovalski predlog vmesnika SSSS 2.0, kakršen je na voljo v času priprave prispevka. Slovarske informacije na sliki so provizorične. Slika ponazarja, kakšna bo vizualizacija pri bližnjih in jedrnih (pri *depra* in *deprimiranost*) ter pri uporabniških sopomenkah (pri *sopomenka*). Razvidne so tudi nekatere druge novosti, npr. delitev uporabniških sopomenk glede na slovarsko verzijo, v kateri so bile predlagane, ter možnost dodajanja slovarskih oznak ob predlagane sopomenke.



Slika 1. Oblikovalski predlog vmesnika SSSS 2.0 (vsebina je provizorična).

6. Uporabniško dodajanje oznak

Nedavno izvedena raziskava o odnosu uporabniške skupnosti do SSSS, v kateri je sodelovalo 671 anketirancev, je pokazala naklonjenost do večine novosti, ki jih prinaša slovar, npr. stalno posodabljanje, strojni postopki, digitalni format, kolokacijski podatki, povezave na korpus, uporabniško vključevanje (Arhar Holdt, 2020: 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.: 472).

V trenutnem slovarskem vmesniku nekateri uporabniki in uporabnice težavo rešujejo tako, da oznako ali kako drugo pojasnilo v oklepaju pripišejo ob svoj sopomenski predlog, npr. *babica – nona (lokalno)*, *bojazljivec – pezde (vulg.)*, *Italijanka – makaronarka (slabš.)*. Kot omenjeno v poglavju 3, pa večina predlaganih sopomenk oznake nima.

Skladno z uporabniškimi željami in potrebami želimo nadgraditi protokol dodajanja sopomenk, da bodo predlagani besedi ali zvezi uporabnice in uporabniki lahko dodali tudi slovarsko oznako oz. oznake. Privzeta izbira bo, da je predlog "brez oznake", ostale možnosti bodo na voljo v spustnem meniju (Slika 1). V različici SSSS 2.0 bodo na klik na voljo oznake *sovražno*, *grobo* in *izraža negativen odnos*, poleg tega pa bomo ponudili okence, v katerega bo mogoče vtipkati morebitno drugo oznako.

Pomen in raba oznak *sovražno*, *grobo* in *izraža negativen odnos* bo razložena in ponazorjena s primeri, s

čimer bo lahko dosežena določena stopnja enotnosti uporabniškega označevanja (informacije bodo na voljo na klik, gl. ikono (i) na Sliki 1). Predvideno pa je, da bodo uporabniki oznake mestoma interpretirali in uporabljali drugače, kot bi jih slovaropisci. Vse dodane oznake bodo (skupaj z dodanimi sopomenkami) preverjene in označene sopomenke bodo dragoceno gradivo ne le za dopolnitev odprto dostopne slovarske baze sopomenk, ampak tudi za analize širšega dojemanja označevalnega sistema ter dometa in meja oznak. Prav tako pomemben vidik 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: 154).

V prispevku smo opisali, kako poteka obravnava sovražnega in grobega besedišča v SSSS in katere spremembe so v načrtu za različico 2.0, ki bo objavljena jeseni 2022. Rešitve naslavlajo dve pomembni značilnosti SSSS: njegovo strojno izdelanost in odprtost, da pri razvoju slovarja sodeluje tudi uporabniška skupnost. V novi različici slovarja bodo sovražnemu in grobemu besedišču pripisane slovarske oznake oz. opozorilne ikone s pojasnili o možnih posledicah rabe in dodana bo možnost, da uporabniki pripišejo oznako svojim predlogom sopomenk.

Ker vse težave trenutnega SSSS niso enostavno in hitro rešljive, želimo slovarske uporabnike bolj opozoriti na trenutne omejitve. Čeprav je metodologija priprave SSSS pojasnjena v razdelku *O viru*, pri samih iztočnicah ni izrecnih opozoril, da je slovar pripravljen strojno, in to na vseh ravneh: sopomenke, kolokacije, korpusni zgledi, kar lahko vodi v napačne interpretacije slovarske vsebine. V naslednji različici SSSS želimo zato uvesti indikator stopnje gesla³ in dodati v predstavitev slovarja opozorila o dometu in posledicah metodologije ter razlago korakov, po katerih se slovar razvija.

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 pohitrilo 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.

Kot smo poudarili v prispevku, je izražanje negativnega odnosa večkrat 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 uporabili 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 bomo izvedli že pod okriljem projekta Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL, kjer je med cilji tudi nadgradnja SSSS z 2.000 pomensko členjenimi gesli, ki bodo imela slovaropisno pregledane in razvrščene sopomenke, kolokacije ter korpusne zglede.

V nadaljnje premisleke glede sovražnega in grobega besedišča znotraj koncepta odzivnega slovarja bi bilo smiselno celoviteje vključiti vidike okoliščin rabe. Zanimivo bi bilo denimo 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.

8. Zahvala

Projekt Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL v letih 2021–2022 financira Ministrstvo za kulturo Republike Slovenije. Raziskovalna programa št. P6-0411 (Jezikovni viri in tehnologije za slovenski jezik) in št. P6-0215 (Slovenski jezik – bazične, kontrastivne in aplikativne raziskave), sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

³ Po zgledu Kolokacijskega slovarja sodobne slovenščine (KSSS), ki z ikono petstopenjske piramide uporabniku na jasen in ekspliciten način posreduje informacijo o razvoju ter različnih stopnjah izdelanosti slovarskih gesel (Kosem et al., 2018b).

9. Literatura

- 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): 465–482. <https://doi.org/10.31724/rihij.46.2.1>
- Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Apolonija Gantar, Vojko Gorjanc, Bojan Klemenc, Iztok Kosem, Simon Krek, Cyprian Laskowski in Marko Robnik Šikonja. 2018. Thesaurus of Modern Slovene: By the Community for the Community. V: J. Čibej, V. Gorjanc, I. Kosem in S. Krek, ur., *Proceedings of the 18th Euralex International Congress: Lexicography in Global Contexts*, str. 401–410. Znanstvena založba Filozofske fakultete, Ljubljana. <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1>
- Polona Gantar. 2015. *Leksikografski opis slovenščine v digitalnem okolju*. Znanstvena založba Filozofske fakultete, Ljubljana. <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/download/62/138/2602-1?inline=1>
- Vojko Gorjanc. 2005. Neposredno in posredno žaljiv govor v jezikovnih priročnikih: diskurz slovarjev slovenskega jezika. *Družboslovne razprave*, 21(48): 197–209.
- Vojko Gorjanc. 2017. *Nije rečnik za seljaka*. Biblioteka XX vek, Beograd.
- Geoffrey Hughes. 2011. *Political Correctness: A History of Semantics and Culture*. Wiley-Blackwell, MA.
- Boris Kern. 2015. Politična korektnost v slovaropisju. V: D. Zuljan Kumar in H. Dobrovoljc, ur., *Zbornik prispevkov s simpozija 2013*, str. 144–154, Založba Univerze, Nova Gorica.
- Iztok Kosem, Júlia Čeh Bálint, Vojko Gorjanc, Anna Kolláth, Attila Kovács, Simon Krek, Sonja Novak-Lukanovič in Jutka Rudaš. 2018a. *Osnutek koncepta novega velikega slovensko-madžarskega slovarja*. Univerza v Ljubljani, Filozofska fakulteta, Ljubljana. <https://www.cjvt.si/komass/wp-content/uploads/sites/17/2020/08/Osnutek-koncepta-VSMS-v1-1.pdf>
- Iztok Kosem, Simon Krek, Polona Gantar, Špela Arhar Holdt, Jaka Čibej in Cyprian Laskowski. 2018b. Kolokacijski slovar sodobne slovenščine. V: D. Fišer in A. Pančur, ur., *Jezikovne tehnologije in digitalna humanistika*. Znanstvena založba Filozofske fakultete, Ljubljana. http://www.sdjt.si/wp/wp-content/uploads/2018/09/JTD-H-2018_Kosem-et-al_Kolokacijski-slovar-sodobne-slovenscine.pdf
- 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: I. Kosem et al., ur., *Proceedings of eLex 2017: Lexicography from Scratch*, str. 93–109, Leiden, Netherlands. <https://elex.link/elex2017/wp-content/uploads/2017/09/paper05.pdf>
- Simon Krek, Cyprian Laskowski, Marko Robnik Šikonja, Iztok Kosem, Špela Arhar Holdt, Polona Gantar, Jaka Čibej, Vojko Gorjanc, Bojan Klemenc in Kaja Dobrovoljc. 2018. Thesaurus of Modern Slovene 1.0, Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1166>
- Nataša Logar, Nina Perger, Vojko Gorjanc, Monika Kalin Golob, Neža Kogovšek Šalamon in Iztok Kosem. 2020.

- Raba slovarjev v slovenski sodni praksi. *Teorija in praksa*, 57:89–108.
https://www.fdv.uni-lj.si/docs/default-source/tip/tip_pos_2020_logar_idr.pdf?sfvrsn=0
- Rosamund Moon. 2014. Meanings, Ideologies, and Learners' Dictionaries. V: A. Abel et al., ur., *Proceedings of the XVI EURALEX International Congress: The User in Focus*, str. 85–105, Bolzano/Bozen. Institute for Specialised Communication and Multilingualism.
https://euralex.org/elx_proceedings/Euralex2014/euralex_2014_004_p_85.pdf
- Andrej Motl in Veronika Bajt. 2016. *Sovražni govor v Republiki Sloveniji: Pregled stanja*. Mirovni inštitut, Ljubljana.
<https://dlib.si/stream/URN:NBN:SI:DOC-F2YZP2RB/c117f4c6-8fe9-437d-8c64-5b7987a856b6/PDF>
- Brankica Petković in Neža Kogovšek Šalomon. 2007. *O diskriminaciji: Priročnik za novinarje in novinarke*. Mirovni inštitut, Ljubljana.
<https://www.mirovni-institut.si/wp-content/uploads/2014/08/Prirocnik-o-diskriminaciji-final-all.pdf>
- Dušan Rebolj. 2008. Uporabnejša opredelitev politične korektnosti. V: S. Autor in R. Kuhar, ur., *Politična (ne)korektnost*. Mirovni inštitut, Ljubljana, str. 4–15.
<https://www.mirovni-institut.si/wp-content/uploads/2014/08/nestrpnost-6.pdf>
- SSKJ. 2014. *Slovar slovenskega knjižnega jezika: Uvod*. Druga, dopolnjena in deloma prenovljena izdaja. Inštitut za slovenski jezik Frana Ramovša ZRC SAZU, Ljubljana.
<https://fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika>
- Ada Vidovič Muha. 2013. *Moč in nemoč knjižnega jezika*. Znanstvena založba Filozofske fakultete, Ljubljana.