

Ocenjevanje uporabniško dodanih sopomenk v Slovarju sopomenk sodobne slovenščine – pilotna študija

Magdalena Gapsa*

* Filozofska fakulteta, Univerza v Ljubljani
Aškerčeva 2, 1000 Ljubljana
magdalena.gapsa@ff.uni-lj.si

Povzetek

V prispevku opisujem prvi korak uporabniške raziskave, znotraj katere bodo različne strokovne skupine ocenjevalcev presojale o relevantnosti določenih uporabniško dodanih sopomenk v *Slovarju sopomenk sodobne slovenščine*. V okviru raziskave želim preveriti, ali se ocene strokovnjakov, kot so lektorji, prevajalci in učitelji, razlikujejo od ocen slovaropiscev ter kako enovite so ocene znotraj posamezne skupine. Osredotočam se na potek in rezultate prvega sklopa ocenjevanja, ki ga je kot testna množica izvedla skupina študentov. Ta korak je služil tudi kot preizkus navodil in izbranih orodij za lažje načrtovanje dela preostalih predvidenih skupin ocenjevalcev. Navajam ugotovitve glede relevantnosti sopomenskega gradiva po presoji študentske skupine, kjer so zlasti zanimive mejne kategorije »pogojno« sprejemljivega gradiva, sledijo identificirane šibke točke zasnovane raziskave ter rešitve, ki bodo vključene v nadaljnji potek ocenjevanja.

Evaluation of User-Added Synonyms in the Thesaurus of Modern Slovene – a Pilot Study

The paper describes the first step of a user research in which various expert groups of evaluators will assess the relevance of certain user-added synonyms in the Thesaurus of Modern Slovene. Part of the research is to check whether the evaluations of experts such as proofreaders, translators and teachers differ from those of lexicographers, and how consistent the assessments are within each group. The main focus is on the process and results of the first set of assessments carried out by a group of students as a test set. This step also served as a test of the instructions and tools chosen to facilitate the planning of the work of the remaining intended groups of evaluators. The results are then presented in terms of the relevance of the synonymous material assessed by the group of students, with the borderline categories of "conditionally" acceptable material being of particular interest, followed by the weaknesses identified in the research designed and the solutions and improvements that will be incorporated into the further assessment process.

1. Uvod

S pojavom digitalnega medija se na področju jezikoslovja in strojne obdelave naravnega jezika spreminjajo tako potrebe kot tudi priložnosti, ki se kažejo zlasti kot možnost avtomatiziranega (hitrejšega, enostavnejšega in cenejšega) posodabljanja jezikovnih podatkov in opisov, večja povezljivost med podatki različnih vrst, neomejen prostor za njihov prikaz ter vključevanje širše skupnosti v proces priprave slovarjev¹ itd. V prispevku se osredotočam na slednje, torej možnost doprinosna širše jezikovne skupnosti, natančneje možnost, da slovarski uporabniki dodajajo sopomensko gradivo v *Slovar sopomenk sodobne slovenščine*² (Arhar Holdt et al., 2018, v nadaljevanju tudi *Sopomenke*), s čimer je povezano vprašanje morebitne spremembe v pogledih na sopomenskost. Na podlagi uporabniškega gradiva je možno opazovanje, kako sopomenskost dojemajo in občutijo uporabniki, zlasti v razmerju do slovaropiscev, ki podajajo končne odločitve o vključitvi sopomenskega gradiva v referenčne jezikovne vire.

Prispevek temelji na raziskovalnem vprašanju iz doktorske disertacije z naslovom *Sopomenskost v Slovarju sopomenk sodobne slovenščine in izbranih različicah wordneta*,³ o prispevku širše jezikovne skupnosti k pogledom na sopomenskost. V disertaciji predpostavljam,

da se pogled strokovne oz. širše jezikovne skupnosti razlikuje od pogleda slovaropiscev, vendar ta potencialni drugačni pogled jezikovne skupnosti lahko bistveno pripomore pri gradnji novih oz. nadgradnji obstoječih jezikovnih virov. To hipotezo bom preverila z analizo ocen sopomenskosti izbranega nabora uporabniško dodanega gradiva, ki ga ocenjujejo različne strokovne skupine (naštete v nadaljevanju). Ocene bom najprej primerjala znotraj posameznih skupin, nato pa tudi med skupinami.

Namen prispevka je predstaviti izsledke prvega, testnega oz. pilotskega ocenjevanja uporabniško dodanih sopomenk, ki ga je izvedla skupina šestih študentov jeziko(slo)vnihi smeri. Evalvacijska naloga, posredovana tej skupini, je imela dva glavna namena: (I) priprava gradiva za ocenjevanje uporabniških sopomenk, preizkus modela, orodij in navodil ter morebitne dopolnitve oz. prilagoditve le-teh ter (II) zbiranje povratnih informacij za načrtovanje nadaljnjega obsega in izvedbe ocenjevanja. Raziskava se deloma povezuje s projektom *Sopomenke in Kolokacije 2.0 – SoKol, Nadgradnja temeljnih slovarskih virov in podatkovnih baz CJVT UL*,⁴ ki ga med leti 2021 in 2022 financira Ministrstvo za kulturo Republike Slovenije. Glavni cilj projekta je prenovitev *Slovarja sopomenk sodobne slovenščine* ter *Kolokacijskega slovarja sodobne slovenščine*. Projekt je omogočil dostop do študentov jezikoslovja z dobrim poznavanjem *Slovarja sopomenk*

¹ V slovenskem prostoru se tematike dotika monografija *Slovar sodobne slovenščine: problemi in rešitve* (Gorjanc et al., 2017). Podrobneje o vlogi uporabnikov v procesu priprave slovarjev in načinih sodelovanja z njimi sta pisala npr. A. Abel in C. Meyer (2013).

² *Slovar sopomenk sodobne slovenščine*: <https://viri.cjvt.si/sopomenke/slv/> (dostop: 6. 5. 2022).

³ Doktorska disertacija nastaja v okviru raziskovalnega programa *Jezikovni viri in tehnologije za slovenski jezik* (številka programa P6-0411) in jo v letih 2019–2023 financira Javna agencija za raziskovalno dejavnost Republike Slovenije. Mentorira jo zn. sod. dr. Špela Arhar Holdt.

⁴ Spletna stran projekta SoKol: <https://www.cjvt.si/sokol/> (dostop: 3. 5. 2022).

sodobne slovenščine ter izkušnjami z označevanjem pomensko povezanih podatkov.

2. Opis vira in pregled področja

Slovar sopomenk sodobne slovenščine, ki ga je leta 2018 objavil Center za jezikovne vire in tehnologije Univerze v Ljubljani, je prvi primer novega slovaropisnega koncepta, t. i. odzivnega slovarja (Arhar Holdt et al., 2018). Njegova glavna značilnost je, da se slovar stalno odziva na spremembe v jeziku ter na potrebe uporabnikov. Za namene tega prispevka je najbolj pomembna značilnost ta, da je uporabnikom omogočeno sodelovanje v procesu nastajanja slovarja, saj se podatki spreminjajo glede na aktivnost ter pripombe skupnosti, hkrati pa ta lahko prispeva k čiščenju nerelevantnih ali napačnih podatkov.⁵

Množičenje za namene slovaropisja je sicer znana praksa. Množica, ki jo želimo vključiti v ocenjevanje, ne potrebuje posebnih predznakov ali izobrazbe, saj so tudi uporabniki jezika, ki niso strokovnjaki s področja, dovolj nadarjena, ustvarjalna in učinkovita skupina, ki zmore reševanje manj zahtevnih oz. bolj rutinskih nalog, strokovnjaki pa se lahko po vključitvi množičenja osredotočijo na kompleksnejše oz. bolj analitične naloge (Kosem et al., 2013, str. 46; Čibej et al., 2015, str. 70–71). Množičenje je lahko izredno učinkovito in zanesljivo – odgovori oz. ocene nestrokovne skupnosti se skorajda ne razlikujejo od zlatega standarda oz. odgovorov, ki so jih podali slovaropisci, kar je bilo že leta 2008 dokazano z uporabo orodja *Amazon Mechanical Turk* (AMT) (Snow et al., 2008, str. 257–258), še zlasti, če zagotovimo zadostno količino ocenjevalcev (prim. Nicolas et al., 2021). Na tej predpostavki temelji vključevanje skupnosti v razvoj *Slovarja sopomenk sodobne slovenščine*.

Presojanje o (ne)sopomenskosti besed je bilo v široko razumljenem digitalnem slovaropisju mnogokrat uporabljeno še zlasti ob (nad)gradnji in čiščenju različnih wordnetov, npr. ruskega, kjer ocenjevalci presojujejo o (ne)pravilnostih ter sami sestavljajo in popravljajo sinsete (Braslavski et al., 2014) ali češkega, za katerega je bil razvit poenoten sistem prijavljanja napak, ki jih uporabniki odkrijejo, ob tem pa lahko tudi predlagajo popravek (Horák in Rambousek, 2018). V slovenskem prostoru pa so ocenjevalci s pomočjo orodja za množičenje sloWCrowd (Tavčar et al., 2012) presojali, ali so avtomatsko pridobljeni predlogi sopomenski in ali spadajo v predvideni sinset ter tako pomagali odpraviti napake v slovenskem wordnetu (Fišer et al., 2014). Ocene skupnosti, tudi presojanje o (ne)sopomenskosti besed, so uporabne tudi širše, npr. pri evalvaciji natančnosti vektorskih vložitev za povezane oz. sorodne besede (prim. Schnabel et al., 2015, str. 301–303).

Ob tem se odpira vprašanje, ali se s širjenjem skupine sodelujočih širi oz. spreminja tudi sam pogled na gradivo, ki (naj) ga slovar sopomenk prinaša. Slovaropisci za identifikacijo sopomenskosti sledijo vnaprej izbranim (včasih tudi dokaj strogim) jezikoslovnim izhodiščem, slovarski uporabniki pa lahko o sopomenskosti presojajo precej bolj subjektivno, in sicer z vidika "uporabnosti" ali

"relevantnosti" predloga za njihovo delo (npr. predloge tipa *brat – sorojenec*, *avto – osebno vozilo* itn., so ocenjevalci večinoma opredelili kot relevantne, čeprav gre za drugo relacijo).⁶ Potrebno je opozoriti, da presojanje o podobnosti dveh besed oz. sopomenskosti tudi za slovaropisce nikakor ni lahka in nedvoumna naloga, saj univerzalne definicije sopomenskosti ni, sam koncept pa je zelo širok in tesno povezan s kontekstom in okoliščinami rabe, hkrati pa ga različni raziskovalci drugače interpretirajo in opisujejo (gl. npr. Snoj, 2019, str. 13–41; Vidovič Muha, 2013, str. 172–183; Zorman, 2000, str. 20–48).

Večina definicij sopomenke opredeljuje kot besede, ki imajo identičen pomen, vendar različno formo (Zgusta, 1971, str. 89), poudarja se tudi razlikovanje besed z istim pomenom po njihovi stilni ali zvrstni vrednosti (Toporišič, 1992, str. 294). V literaturi prevladujeta dva glavna pogleda: sopomenke so le besede, ki imajo popolnoma isti pomen (popolna sopomenskost) ali pa besede, katerih pomen je zelo podoben (delna sopomenskost). Popolna sopomenskost je zelo redka, saj krši načelo jezikovne ekonomičnosti oz. gospodarnosti, pogosta pa je delna sopomenskost (prim. Hock, 1991, str. 283; Snoj et al., 2016, str. 5; Vidovič-Muha, 2013, str. 175; Zgusta, 1971, str. 89), ki se najpogosteje kaže v primeru prenesenih pomenov, izposojenk iz tujih jezikov, arhaizmov in ekspresivnega besedišča, največ sopomenk pa imajo besede, ki so rabljene prav v prenesenem ali s kolokacijami povezanem pomenu (Apresjan, 2000, str. 37). V slovenskem prostoru je delna sopomenskost razumljena kot del stilistike in ne semantike, zato je bila dolgo, še zlasti v slovaropisju, obravnavana predvsem popolna sopomenskost, sopomenke v SSKJ pa imajo tudi normativno vlogo (usmerjajo od zaznamovanega proti nezaznamovanemu), ob izidu *Sinonimnega slovarja slovenskega jezika* (2016) je bilo več pozornosti namenjene tudi delni sopomenskosti (Vidovič Muha, 2013, str. 180; Snoj et al., 2016, str. 6). *Slovar sopomenk sodobne slovenščine* nudi nov okvir, saj osvetljuje vlogo in vrednost konteksta s prikazom kolokacij ali povezavo na korpusne zglede, obenem pa nudi tudi možnost, da uporabniki dodajo gradivo, ki po njihovem mnenju v slovarju manjka. Na podlagi skoraj 1.000 uporabniško dodanih sopomenskih predlogov želimo ponovno odpreti vprašanje razumevanja sopomenskosti in preveriti, ali se je le-to spremenilo z nastankom in razvojem digitalnih jezikovnih virov, zlasti odzivnega slovarja..

3. Predviden potek in izsledki raziskave

Cilj znotraj doktorske raziskave je, da evalvacijo poleg študentov opravijo tudi predstavniki drugih skupin sodelujočih: slovaropisci (kot najbolj specializirani strokovnjaki s področja), prevajalci, lektorji, učitelji slovenščine in ljubiteljski raziskovalci jezika brez jezikoslovne izobrazbe (kot predstavniki širše jezikovne skupnosti). Interesne skupine so bile določene na podlagi tipologije ciljnih skupin uporabniških raziskav (gl. Arhar Holdt, 2015, str. 142–146), kjer so slovarski uporabniki

⁵ Preostale glavne značilnosti slovarja so npr. (a) je dostopen le v digitalni obliki ob upoštevanju potreb, pogojev in prednosti le-te, hkrati pa nikoli ni zaključen, saj se podatki stalno spreminjajo in prilagajajo trenutnemu jezikovnemu stanju, (b) slovarska baza nastaja z uporabo naprednih računalniških metod, kar uporabnikom hitro ponuja veliko količino odprto in prosto dostopnih jezikovnih podatkov, ki so relevantni, a še neprečiščeni,

(c) sopomenski podatki so povezani z besedilnim kontekstom s pomočjo kolokacij, korpusnih zgledov in povezav na korpus ter (č) slovar in slovarska baza sta prosto in odprto dostopni pod ustrezno licenco (prim. Arhar Holdt et al., 2018, str. 404; Čibej in Arhar Holdt, 2019, str. 339–340).

⁶ Predpostavljam, da se bodo v določenih kategorijah gradiva oz. odločitve pokazale skupne točke, v drugih pa razlike.

načeloma pripadniki (vsaj) ene skupine: (I) uporabniki, ki slovarje uporabljajo v procesu izobraževanja (npr. študenti in učitelji slovenščine),⁷ (II) uporabniki, ki slovarje uporabljajo v poklicne namene (npr. slovaropisci, prevajalci, lektorji in učitelji) ter (III) uporabniki, ki slovarje uporabljajo za prostočasne aktivnosti (npr. ljubiteljski raziskovalci jezika).

Tipologija je bila uporabljena tudi v raziskavi odnosa uporabnikov do novosti v *Sopomenkah*, kjer so bile v rezultatih najbolj zastopane skupine lektorji, prevajalci, učitelji slovenščine na različnih ravneh izobraževanja, pisci različnih vrst besedil (npr. beletristika, strokovno in znanstveno pisanje, kreativno pisanje, novinarstvo, blogerstvo ipd.) ter ljubiteljski raziskovalci jezika (prim. Arhar Holdt, 2020, str. 477). Iz tega lahko sklepamo, da so te skupine najbolj zainteresirane za *Sopomenke* (in sopomenske podatke na splošno), hkrati so tudi relevantne in reprezentativne. Ker je za preverbo hipoteze potrebno tudi mnenje slovaropiscev, je skupina piscev⁸ nadomeščena s slovaropisci.⁹ Njihovi odgovori bodo analizirani znotraj skupine, hkrati pa bodo služili kot referenčna evalvacija sopomenskih parov, s katero bodo primerjani odgovori vseh ostalih skupin.

Na podlagi izsledkov ocenjevanja po skupinah in primerjav odgovorov med njimi želim pridobiti empirično podlago o željah in pričakovanjih uporabnikov, ki bo z aplikativnega vidika podlaga za pripravo smernic za uredniške protokole, ki bodo uporabljeni pri nadgradnjah slovarja. Z znanstvenega vidika pa bodo odgovori podlaga za definiranje sopomenskosti v luči odzivnih digitalnih jezikovnih virov. Prvi sklop uporabniške raziskave, ki so ga opravili študenti, pa je poleg prej omenjenih ciljev služil še preizkusu zasnove raziskave in odkrivanju šibkih točk ter zbiranju povratnih informacij za načrtovanje nadaljnjega obsega in izvedbe.

4. Gradivo in metoda

4.1. Gradivo

Uporabniška raziskava temelji na delu podatkov, ki so podatkovni vzorec za doktorat, in sicer seznamu 546 samostalnikov, ki se pojavijo tako v podatkovni bazi *Slovarja sopomenk sodobne slovenščine* (Krek et al., 2018) in *sloWNeta* (Fišer, 2015) kot tudi v *Leksikalni bazi za slovenščino* (Gantar et al., 2013) in v *Velikem slovensko-madžarskem slovarju*, kjer so samostalniki opremljeni z oznakami semantičnih tipov (Kosem in Pori, 2021).¹⁰

⁷ Študenti, zlasti jezikovnih smeri, so na prehodu med izobraževanjem in poklicno rabo, podobno učitelji, ki slovarje uporabljajo v poklicne namene, vendar je njihov poklic vezan na izobraževalni proces.

⁸ V primeru piscev bi bilo najtežje pridobiti koherentno skupino, ki bi pokrivala različne prej naštetje žanre, po drugi strani preostale skupine zadoščajo potrebi po predstavniki skupine, ki slovarje uporablja v poklicne namene.

⁹ Treba se je zavedati, da so slovaropisci zaradi svoje izobrazbe in specializiranosti zelo atipična uporabniška skupina za slovarske raziskave (prim. Arhar Holdt, 2015, str. 140), vendar prav to na tem mestu služi namenu raziskave.

¹⁰ Slednja dva vira sta upoštevana, saj želim v (preostalih) analizah v okviru doktorske disertacije upoštevati tudi korpusno osnovan pomenski opis (potencialno) sopomenskega gradiva.

¹¹ Slovarska baza *Sopomenk*, ki je dostopna v okviru repozitorija CLARIN.SI, ne vsebuje uporabniško dodanih sopomenk. Slednje

Za to raziskavo so bili podatki za doktorsko disertacijo dodatno opremljeni s podatki o uporabniško dodanih sopomenkah v *Sopomenkah* na podlagi internega izvoza podatkov iz 18. 11. 2021.¹¹ Tako sem pridobila seznam 307¹² iztočnic, ki imajo vsaj eno uporabniško dodano sopomenko oz. 976 sopomenskih parov. Nekateri uporabniški vnosi (68 parov) so vsebovali dodatna pojasnila in opombe v oklepajih, največkrat opombe o zaznamovanosti predloga, npr. *arheolog – žličkar (šalj., pog.)*, *bonbon – cukrer (neknj.)*, *klient – kunt (nar.)*, *preteklost – prtljaga (ekspresivno)*, *stopnica – štenga (nižje pog.)*. Ker ocenjevalcem nisem želela sugerirati odgovorov, so bile tovrstne opombe odstranjene. Zabeleženih je bilo tudi 5 primerov, kjer so uporabniki znotraj oklepajev dodajali bolj kontekstualne razlage in ne kvalifikatorje. Ti primeri so bili brez sprememb vključeni v končni nabor za ocenjevanje, saj sem s tem želela preveriti reakcijo ocenjevalcev na tovrstne oznake. Gre za predloge: *interier – ambient (v zaprtem prostoru)*, *kmet – kmet (šahovska figura)*, *koncentracija – (velika/majhna) vsebnost, priloga – priponka (k e-pismu)* in *torbica – (torbica) pismo*. Z odstranitvijo opomb je prišlo do podvojevanja sopomenskih parov,¹³ zaznala sem 4 take primere, ki so bili prav tako izločeni iz seznama za ocenjevanje. Ta je v končni fazi obsegal 972 parov.

4.2. Navodila ocenjevalcem

Ocenjevalci so prejeli nagovor s kratkim pojasnilom, da se ocenjevanje izvaja v okviru doktorske raziskave in kakšne podatke želim zbrati. Bili so naprošeni, naj med ocenjevanjem ne uporabljajo drugih jezikovnih virov in priročnikov. Navedeno je bilo, da je naloga sestavljena iz dveh obveznih delov: preglednice s sopomenskimi pari, kjer bodo podajali svoje odgovore in morebitne pripombe, ter vprašalnika, kjer bodo podajali demografske podatke o sebi ter povratne informacije o evalvacijski nalogi sami. V primeru dvomov so udeleženci lahko zastavljali dodatna vprašanja po e-pošti.

Glavno navodilo ocenjevalcem je bilo odgovoriti na vprašanje: »Ali sta besedi v paru sopomenki?«. Vsak sopomenski par so lahko uvrstili v eno izmed štirih kategorij, oz. izbrali enega izmed štirih možnih odgovorov, in sicer DA, NE, POGOJNO DA ter NISEM PREPRIČAN/NE VEM. Odgovor DA je bil predviden za primere, ko so bili prepričani, da sta besedi sopomenki, odgovor NE za primere, ko so bili prepričani, da besedi nista sopomenki ter kadar je šlo za očitne napake oz.

izvozimo iz slovarskega vmesnika s pomočjo prilagojene skripte, kar omogoča, da so uporabniško dodani podatki ažurni.

¹² Ne razlikujem med iztočnicami z veliko in malo začetnico. V naboru gradiva za nalogo je to samo en primer, in sicer *zemlja* in *Zemlja*, ki je tukaj obravnavan kot ena iztočnica.

¹³ Slovarski vmesnik ima sicer preprosto varovalko, ki uporabnikom preprečuje ponovni vnos že dodanega predloga, vendar temelji na prepoznavi znakov in dovoljuje vnos tako alfanumeričnih kot nealfanumeričnih znakov, npr. oklepajev. Ko uporabnik obstoječemu sopomenskemu predlogu doda opombo, sistem to prepozna kot nov vnos. V mojem vzorcu se je to zgodilo štirikrat, in sicer dvakrat znotraj gesla *babica*, kjer sta bila predloga *nona* in *nona (lokalno)* ter *oma* in *oma* ;), znotraj gesla *živina*, kjer sta bili predlagani *živad* in *živad (star.)* ter znotraj gesla *nakup*, kjer sta bili predlagani *kupilo* in *kupilo (star.)*.

zatipkane besede. Odgovor POGOJNO DA je bil predviden za pare, ko so ocenjevalci sicer menili, da gre za sopomenki, vendar so hkrati videli tudi omejitve oz. imeli pomisleke, dvome, npr. da sta besedi sopomenki samo v določenem pomenu, kontekstu, ena ali obe besedi sta zaznamovani itn. Odgovor NISEM PREPRIČAN/NE VEM je bil predviden za pare, ko niso poznali ene ali obeh besed v paru, pomena ene ali obeh besed v paru ali niso bili prepričani, ali so težko podajali svoje mnenje. Pri vsakem paru je bila možnost dodajanja opomb, ki so bile zahtevane pri odgovoru POGOJNO DA, zaželeno pri NISEM PREPRIČAN/NE VEM in možne pri drugih odgovorih.

Ker je eden glavnih ciljev raziskave preverjanje, kaj ocenjevalci razumejo kot relevantno sopomensko gradivo, so bila navodila, v izogib sugeriranju odgovorov, zelo splošna. Zato »sopomenka« ni natančneje definirana, možni odgovori so vsebovali le kratek opis, ne pa tudi primerov. Prav tako ni bilo navodil, kam umestiti mejne primere.

4.3. Ocenjevanje

Sopomenski pari so bili ocenjevalcem posredovani v obliki tabele, ki je bila dostopna kot Googleva preglednica.¹⁴ Datoteka je bila sestavljena iz dveh listov. Prvi list je obsegal skrajšano verzijo navodil, da so jih ocenjevalci imeli vedno pri roki, drug list pa seznam 972 sopomenskih parov za ocenjevanje. V prvem stolpcu tabele je zaporedna številka para, v drugem iztočnica, v tretjem pa predlagana sopomenka, npr. *vonj – vzduh, stigma – brazda, reforma – sprememba, pošta – sporočila, dopust – vakance*. Celice v teh stolpcih so bile zaklenjene v izogib namernim in nenamernim spremembam podatkov. V četrtem stolpcu so ocenjevalci iz spustnega seznama (v izogib zatipkom) izbirali enega izmed štirih odgovorov. Zadnji, peti stolpec, je bil predviden za komentarje in opombe ocenjevalcev. To je tudi edini stolpec, kjer so lahko prosto vnašali podatke. Ocenjevalci so do podatkov dostopali po principu en ocenjevalec – ena preglednica, da odgovori drugih ocenjevalcev ne bi vplivali na posameznikove odločitve.

4.4. Vprašalnik

Ocenjevalci so dobili tudi povezavo do spletnega vprašalnika, ki je bil sestavni del ocenjevanja. Vprašalnik je bil sestavljen in dostopen v spletnem orodju za anketiranje Ika.¹⁵ V prvem delu vprašalnika so sodelujoči odgovarjali na vprašanja o sebi: starost, zaposlitveni status, izobrazba (jezikoslovna ali ne), načini, na katere se z jezikom ukvarjajo in glavna področja, ki so zanje najbolj pomembna v zvezi z jezikom. V drugem delu so odgovarjali na vprašanja, povezana z evalvacijo samo: koliko časa so potrebovali, ali so imeli pri reševanju kakršne koli težave, ali so bila navodila jasna in ali so pri njih kaj pogrešali. Vprašalnik je bil dostopen brez omejitev, ocenjevalci so si lahko vprašanja vnaprej ogledali, njihovi odgovori so se sproti shranjevali, da so lahko npr. najprej podali podatke o sebi, informacije o nalogi pa naknadno.

5. Rezultati

¹⁴ Googleva preglednica od ocenjevalcev ne zahteva posebne strojne opreme, hkrati pa se vneseni odgovori sproti shranjujejo, zato naloge ni bilo potrebno reševati brez prekinitve.

¹⁵ Spletno orodje za anketiranje Ika: <https://www.ika.si/> (dostop: 5. 5. 2022).

Skupina pilotskih ocenjevalcev je obsegala 6 študentov. Dostop do preglednic je bil študentom dodeljen 15. 2. 2022, ocenjevanje so lahko začeli takoj in si ga prilagodili glede na druge obveznosti. Prvi študent je obvestilo o zaključenem ocenjevanju podal 16. 2. 2022, zadnji pa 8. 3. 2022. Vse zelene odgovore sem pridobila v roku treh tednov. Pridobljene podatke je mogoče razdeliti v dva glavna sklopa, in sicer ocene vzorca sopomenskih parov in odgovore, pridobljene z vprašalnikom.

5.1. Izsledki ocenjevanja

Vsi odgovori, ki so jih dali ocenjevalci, so bili združeni v tabele z uporabo programa MS Excel. Prva tabela je obsegala podatke o izbranem odgovoru (brez opomb), kar je omogočilo preverjanje ujemanja oz. enotnosti ocenjevalcev. V drugi tabeli so bile zabeležene podane opombe. Te so bile ročno pregledane in dodeljene v eno izmed kategorij, ki so se oblikovale med pregledovanjem: samo v določenem pomenu ali kontekstu, zaznamovano, neznana beseda ali pomen besede, nad- oz. podpomenka, razlaga ter drugo (npr. nepravilno črkovanje, pomenske nianse, druga medbesedna razmerja, nenavadne oblike besed neujemanje besednih vrst, redkost rabe itn.). V primerih, kadar so ocenjevalci opredelili tudi vrsto zaznamovanosti (npr. ljubkovalno, pogovorno, zastarelo itn.), so bili tudi ti podatki ohranjeni.¹⁶ Številčni podatki o opombah so predstavljeni v tabeli 1. 914 parov je imelo pripisano vsaj eno izmed šestih kategorij opomb, 435 je imelo pripisani vsaj dve kategoriji, 75 parov vsaj tri kategorije in 3 pari so imeli pripisane štiri kategorije opomb. Tretja tabela je obsegala združene podatke o ujemanju oz. enotnosti ocenjevalcev ter o že kategoriziranih opombah.

Kategorija	Število
samo v določenem pomenu/kontekstu	406
zaznamovano	375
neznana beseda ali pomen besede	266
nad- oz. podpomenka	182
razlaga	65
drugo	122
skupaj opomb	1.416

Tabela 1: Številčna razporeditev kategorij opomb.

Popolnega ujemanja, kjer je vseh šest ocenjevalcev podalo isti odgovor, je bilo zelo malo: le 34 znotraj seznama 972 parov oz. približno 3,5 % celotnega nabora. 17 parov je vseh šest ocenjevalcev prepoznalo kot nedvomno sopomenske (6 odgovorov DA), 5 parov kot pogojno sopomenskih (6 odgovorov POGOJNO DA), 5 parov kot nedvomno nesopomenskih (6 odgovorov NE) ter 7 parov kot neznanne oz. neopredeljive (6 odgovorov NISEM PREPRIČAN/NE VEM). Bistveno več je bilo večinskega ujemanja med ocenjevalci, kjer izstopa samo en odgovor. Takih parov je bilo skupno 132 oz. približno

¹⁶ Podrobnejša analiza dejanskih opomb in komentarjev, ki so jih podali ocenjevalci, presega okvirje in namen tega prispevka, je pa zagotovo relevantna in zanimiva, tudi z vidika razumevanja sopomenskosti, zato bo naslovljena v prihodnosti.

13,5 % nabora gradiva. V 50 primerih je 5 ocenjevalcev izbralo odgovor DA, v 46 POGOJNO DA, v 19 NE in v 17 NISEM PREPRIČAN/NE VEM.

Skupaj je bilo parov z visokim ujemanjem ocenjevalcev 166 oz. 17 % nabora gradiva. 67 parov (40 %) je bilo umeščenih v kategorijo DA, 51 parov (31 %) v kategorijo POGOJNO DA, 24 parov (14,5 %) v kategorijo NE, preostalih 24 parov (14,5 %) pa je bilo uvrščenih v

kategorijo NISEM PREPRIČAN/NE VEM. Razporeditev ocen v štiri predvidene kategorije prikazuje Tabela 2.

Odgovor	Popolno ujemanje	Večinsko ujemanje	Skupaj
DA	17 parov	50 parov	67 parov
POGOJNO DA	5 parov	46 parov	51 parov
NE	5 parov	19 parov	24 parov
NISEM PREPRIČAN/NE VEM	7 parov	17 parov	24 parov
Skupaj	34 parov	132 parov	166 parov

Tabela 2: Številčna razporeditev odgovorov.

V primerih, kjer so se ocenjevalci strinjali (vsi so izbrali isti odgovor), je bilo skupno podanih 22 opomb za 15 sopomenskih parov. V 132 primerih, kjer so se ocenjevalci večinoma strinjali (izstopal je en odgovor), je bilo skupno za 109 parov podanih 158 opomb. V primeru opombe iz kategorije drugo so ocenjevalci največkrat navajali pomenske nianse, črkovanje oz. zapis, redkost rabe, prevzete besede itn. Razporeditev opomb po kategorijah in številčni podatki so prikazani v Tabeli 3, zaradi večje preglednosti in lažje primerjave je ohranjeno zaporedje kategorij iz Tabele 1.

Kategorija	Popolno ujemanje	Večinsko ujemanje
samo v določenem pomenu/kontekstu	3	37
zaznamovano	5	55
neznana beseda ali pomen besede	11	24
nad- oz. podpomenka	1	20
razlaga	0	2
drugo	2	20
skupaj opomb	22	158
skupaj parov	15	109

Tabela 3: Številčna razporeditev kategorij opomb ob popolnem in večinskem ujemanju odgovorov.

Med pari, ki so jih sodelujoči označevali kot sprejemljive (DA), so najpogosteje opozarjali, da sta besedi sopomenski samo v enem pomenu ali kontekstu, npr. *dilema – težava*, *identiteta – osebnost*, *koncentracija – osredotočenost*, *privilegij – ugodnost*, *stigma – zaznamovanost*. Pogosto se pojavljajo tudi opombe glede zaznamovanosti besedišča, npr. *beluš – asparagus* (citatno), *cedilo – cedilka* (pogovorno), *morilec – krvnik* (zastarelo), *pes – kuža* (pogovorno), *strpnost – potrpežljivost* (pogovorno), da sta besedi nad- in podpomenka (npr. *avto – osebno vozilo*, *avtomobil – osebno vozilo*, *brat – sorojenec*, *poroka – ženitev*, *kašelj –*

pokašljevanje), da ne poznajo besed ali pomenov besed (*modrček – nedrc*, *oklevanje – obiranje*, *rit – zadnja plat*), da gre za razlago (*jok – pretakanje solz*) ter drugo (npr. *dež – dežne kaplje*: mero- oz. holonimija, *elita – veljaki* in *elita – pomembneži*: neujemanje slovničnega števila, *prerok – profet*: nenavadna oblika, *sestra – sorojenka*: redka raba). Zanimivo, da so par *brat – sorojenec* ocenjevalci občutili kot razmerje nad- oz. podpomenskosti, pri paru *sestra – sorojenka* pa je en ocenjevalec opozoril na redkost rabe, drugih komentarjev ni bilo.

Med pari, ki so jih sodelujoči označevali kot pogojno sprejemljive (POGOJNO DA), se najpogosteje pojavljajo primeri zaznamovanega besedišča, npr. *avto – kripa* (slabšalno), *deček – mulec* (slabšalno, negativni odnos), *juha – župca* (pogovorno, manjšalnica), *krema – maža* (pogovorno), *zadrga – fršlus* (pogovorno, narečno). Pogosti so tudi primeri, kjer sta besedi sopomenski samo v določenem pomenu oz. kontekstu, npr. *izkušnja – dogodivščina*, *kaos – štala*, *jesen – starost*, *posluh – čut*, *preteklost – prtljaga*, ter kjer gre za nad- in podpomenke, npr. *alkohol – etanol*, *aorta – arterija*, *avto – prevozno sredstvo*, *fotoaparat – digič*, *priseljensec – tujec*. Zaznani so tudi pari, kjer so ocenjevalci navedli, da ne poznajo besed ali pomenov besed, npr. *koder – krauželj*, *pivo – pirček*, *rit – prdulja*, *telovnik – lajbič*, ter kjer so jim pripisali druge opombe, npr. *pogum – jajca*: manjka del zveze, *policija – murja* in *rit – guza*: tujka.

Med pari, ki so jih sodelujoči označevali kot nesprejemljive (NE), so najpogosteje navajali, da gre za besedi, ki sta sopomenski samo v določenem pomenu oz. kontekstu, npr. *ljubezen – življenjski tok*, *stopnica – terasa*, *živina – blago*. Pojavljale so se tudi opombe glede zaznamovanosti besedišča, npr. *čarovnica – čudežnica* (pozitiven odnos), *nedelja – teden* (zastarelo), da ocenjevalci besede ali pomena besed ne poznajo (*čik – žvečilni gumi* in *čik – žvečilka*),¹⁷ *laboratorij – pospeševalnik* je predlagana sopomenka bolj razlagalne narave (*rekreacija – raztezne vaje in vaje za moč*), da gre za nad- in podpomenki (*projekcija – podatek*) ter druge opombe (*davek – dan*: nepravilno črkovanje, *nedelja – teden*: tujka).

Pri vseh parih, kjer so ocenjevalci izbrali odgovor NISEM PREPRIČAN/NE VEM, je bila zabeležena

¹⁷ Med opombami na ta dva para so ocenjevalci opozarjali tudi, da sicer besedi sta sopomenski v določenem pomenu oz. kontekstu,

vendar so ju označili kot nesopomenska, kar so utemeljili, da za študentsko generacijo *čik* pomeni izključno cigareto.

opomba, da gre za besede ali pomene, ki jih ocenjevalci ne poznajo, npr. *čarovnica – bela žena, civilist – legist, obrok – rata, stranka – kunt, zaliv – olmun*. Dodatno so se pojavljale tudi opombe, da gre za zaznamovane besede, npr. *avto – gare (slabšalno), kašelj – brehanje (pogovorno), koder – loken, krona – dika (arhaično), zdravilo – arcnije* (zastarelo, arhaično), da sta besedi sopomenski samo v enem pomenu ali kontekstu (npr. *postava – geštel*), da gre za nad- in podpomenke (*torbica – nabočnica*) ter drugo, npr. *srajca – košilja*: nepravilno črkovanje, *zdravilo – biofarmacevtik*: dvomi o dejanski rabe besede, *zdravilo – arcnije*: neujemanje slovničnega števila.

Glede na popolno in večinsko ujemanje je bilo v kategoriji DA ali POGOJNO DA umeščenih skupaj 118 izmed 166 parov ali 71 %, v kategoriji NE in NISEM PREPRIČAN/NE VEM je bilo umeščenih po 24 izmed 166 parov oz. po 14,5 %. V kategoriji DA ali POGOJNO DA torej v sprejemljivo oz. relevantno gradivo, so ocenjevalci načeloma umeščali pare, kjer so tudi opozarjali, da gre za zaznamovane besede, npr. *babica – starejša gospa* (pogovorno, zastarelo, pozitiven odnos), *debelost – zašpehanost* (slabšalno, pogovorno, negativen odnos), *kmet – seljak* (slabšalno, negativen odnos), *novinar – pisun* (slabšalno, negativen odnos), *steklenica – flaša* (pogovorno), da sta besedi sopomenski samo v enem pomenu ali kontekstu (npr. *blago – capa, izrazoslovje – izrazje, legenda – štorija, rit – zahrbtnož, žarnica – sijalka*) ter nad- oz. podpomenke (npr. *kovanec – novič, nakup – fasunga*). V ti kategoriji so bili umeščeni tudi pari, kjer so ocenjevalci opozarjali na npr. pomenske nianse, redkost rabe ali slovnična neujemanja (npr. predlagana sopomenka je v množini), npr. *cedilo – sito, stereotip – predsodek, pes – štirinožni prijatelj*. V kategorijo NE, torej nesprejemljivo oz. nerelevantno gradivo, so najpogosteje spadale tujke, nepravilno zapisane besede ter besedi, ki bi lahko bili sopomenki samo v enem pomenu ali kontekstu, ampak na ta pogoj v primeru večinskega odgovora NE je opozarjal le en ocenjevalec, npr. *davek – dan, nedelja – teden, živina – blago, stopnica – terasa, projekcija – podatek*. V kategorijo NISEM PREPRIČAN/NE VEM, torej gradivo, ki zahteva dodaten in podroben pregled, so ocenjevalci načeloma umeščali pare, kjer besed ali pomenov niso poznali ali kjer so ocenjevalci menili, da se besede sploh oz. redko uporablja, npr. *avto – sinhronka, cigareta – španjoleta, fotografija – heliotipija, moka – mlevina, zdravilo – biofarmacevtik, torbica – nabočnica*. Pri večinskem ujemanju sta le dva para dobila opombo, da je predlagana sopomenka razlagalne narave, od tega je bil en par (*jok – pretakanje solz*) opredeljen kot sprejemljiv (večina odgovorov DA), drug (*rekreacija – raztezne vaje in vaje za moč*) pa kot nesprejemljiv (večina odgovorov NE).

5.2. Podatki o ocenjevalcih in nalogi

V prvem delu vprašalnika sem zbirala podatke o ocenjevalcih. Vsi ocenjevalci v pilotni skupini spadajo v starostno skupino 20–30 let, najmlajši je rojen leta 2001, najstarejši pa leta 1995. Ker je šlo za študentsko populacijo, so vsi ocenjevalci navedli, da študirajo, večina je navedla tudi, da je jezik osrednji predmet njihovega študija. Le en študent je opredelil, da jezik ni v ospredju njegovega študija, ker študira filozofijo. V naslednjem vprašanju sem spraševala, na katerih področjih ima jezik zanje osrednjo vlogo, možnih je bilo več odgovorov. Na voljo so imeli tri odgovore, in sicer da jih jezik zanima, ker se z njim

pretežno ukvarjajo v procesu izobraževanja, da jezik uporabljajo v poklicne namene ali da se z jezikom ukvarjajo zgolj ljubiteljsko. Vsi so označili, da se z jezikom ukvarjajo v izobraževalnem procesu, polovica je dodatno označila, da jezik uporabljajo tudi v poklicne namene. Nato so opredeljevali največ tri področja oz. dejavnosti, kjer je jezik v ospredju njihovega zanimanja, ki so jih izbirali iz ponujenega seznama devetih možnosti. Najpogostejši odgovor je bil raziskovanje oz. študij jezika (5 odgovorov), lektoriranje (4), prevajanje (3), poučevanje slovenščine (2) ter predavanje jezikoslovnih predmetov na višji oz. univerzitetni ravni in tvorjenje besedil (po 1 odgovor). Nihče ni izbral področja leksikografije in ljubiteljskega raziskovanja jezika, drugih odgovorov prav tako ni bilo. V naslednjem vprašanju so morali ocenjevalci izbrati le eno izmed prej navedenih področij oz. dejavnosti, ki je za njih glavno oz. najbolj relevantno. Kot glavno so raziskovanje oz. študij jezika izbrali trije, po en je navedel lektoriranje, prevajanje in tvorjenje besedil.

Sledila so vprašanja o nalogi. Najprej so ocenjevalci opredeljevali, koliko ur jim je reševanje naloge vzelo. V povprečju so študenti za izpolnitev preglednice potrebovali približno 6 ur, najhitrejši jo je rešil v treh urah, najpočasnejši pa v enajstih. Vsi so zatrili, da so bila navodila jasno opredeljena. Le en študent je navedel, da je imel med reševanjem naloge težave, in sicer da veliko besed ni poznal, zato se je težko opredeljeval do potencialne sopomenskosti. Ocenjevalci so imeli tudi možnost podati svoje pomisleke, opazke in komentarje, ki niso bili zajeti v navodilih. Te so podali trije ocenjevalci, ki so navedli, da bi si želeli, da bi bila kategorija POGOJNO DA bolj opredeljena, da niso bili prepričani, v katero skupino naj uvrstijo nad- in podpomenke, razlage besed in neuveljavljene tujke ter da so pogrešali možnost preverbe sopomenk v drugih virih, saj bi s tem lahko podali boljše odgovore, vendar hkrati razumejo, zakaj jih niso smeli uporabljati. Dodatnih komentarjev niso imeli.

6. Diskusija

Glede ustreznosti uporabniško dodanih sopomenk se je izkazalo, da je bilo nedvomno nesopomenskega gradiva, ki so ga prispevali uporabniki, zelo malo. Primeri, kjer so se odgovori ocenjevalcev popolnoma ujemali, so majhen del vzorca (34 parov oz. približno 3,5 %), kar je bilo sicer predvidljivo glede na obseg podatkov in število ocenjevalcev. Nekoliko več je bilo primerov, kjer je izstopal odgovor enega ocenjevalca (132 parov oz. približno 13,5 % nabora). Skupaj je torej parov z večinskim ujemanjem 166 oz. 17 % nabora. Znotraj tega je parov, kjer je izstopajoč odgovor iz nasprotnega pola (npr. vsi odgovori NE in en POGOJNO DA, vsi odgovori DA in en NE), približno ena tretjina (42 parov). Veliko večino teh parov so ocenjevalci ocenili kot sprejemljive. Parov, ki so bili večinsko umeščeni v kategorijo NE ali NISEM PREPRIČAN/NE VEM, je bilo le 24 izmed 166 oz. 14,5 %. To kaže, da so uporabniško dodani predlogi sopomenk načeloma relevantni in konstruktivni, saj je bilo skupaj v kategoriji DA in POGOJNO DA umeščenih 118 izmed 166 parov (71 %), kjer so se odgovori ocenjevalcev večinsko ujemali, četudi gre za primere, ki bistveno presegajo tradicionalno jezikoslovno dožemanje sopomenskosti. Ta ugotovitev je v skladu z raziskavo iz leta 2020, kjer se je po analizi uravnoteženega dela vzorca 1.662 sopomenk (največ 10 predlogov na uporabnika) izkazalo, da je okoli

70 % uporabniško dodanih predlogov konstruktivnih in hkrati nezaznamovanih, okoli 20 % konstruktivnih in zaznamovanih ter le dobrih 6 % odstotkov nekonstruktivnih oz. zlonamernih (prim. Arhar Holdt in Čibej, 2020, str. 6). Ocene le ene sodelujoče skupine razumljivo niso in ne smejo biti zadostna podlaga za generalizacijo, a vendar so podatki glede relevantnosti uporabniško dodanega gradiva spodbudni. Posebno pozornost bo potrebno nameniti kategoriji NISEM PREPRIČAN/NE VEM, saj so vsi v to kategorijo umeščeni potencialni sopomenski pari dobili opombo, da ocenjevalci ne poznajo besede ali pomena besede. To ne pomeni, da so tovrstni predlogi nerelevantni, bodo pa v procesu posodobitev *Sopomenk* zahtevali več pozornosti s strani urednikov, npr. natančnejše iskanje korpusnih zgledov, uporabo dodatnih virov za preverjanje dejanske rabe itn.

Na podlagi povratnih informacij iz vprašalnika in korespondence s študenti, ki so se name obračali med ocenjevanjem, je razvidno, da je pred nadaljnjim ocenjevanjem treba dopolniti navodila ter podrobneje pojasniti ozadje in cilje raziskave. Študenti so prejeli le zelo kratko pojasnilo, da se ocenjevanje izvaja v okviru doktorske naloge ter kakšen je glavni cilj, brez podrobnejših opisov in pojasnitev. V navodilih so sicer dobili tudi informacijo, naj pri ocenjevanju ne uporabljajo drugih jezikovnih virov, a brez obrazložitve, zakaj to ni zaželeno. Iz opomb, ki so jih dajali, je razvidno, da so nekateri to navodilo kršili, saj so bile pogoste opombe tipa: *v Gigafidi sem zasledil_a, Iz Gigafide je razvidno, Ne v Franu ne v Gigafidi nisem zasledil_a*, ali so v tabelo celo pripenjali povezave na druge priročnike. To je zelo verjetno posledica pomanjkanja utemeljitve, zakaj to ni zaželeno. Veliko vprašanj, zastavljenih po e-pošti, je spremljala izjava, da želijo nalogo "pravilno rešiti". Možna razlaga za to je, da so študenti vajeni na ocenjevanje odgovorov po principu prav–narobe, v opisu naloge pa ni bilo izrecno navedeno, da ni pravih ali napačnih odgovorov oz. da so vsi odgovori pravilni, saj sprašujemo po njihovem mnenju. Ta navedba bo mogoče manj relevantna za ostale predvidene skupine ocenjevalcev, a vendarle se kaže kot smiselna dopolnitev opisa naloge in njenega namena.

V navodilih so študenti dobili informacijo, da gre za ocenjevanje uporabniško dodanih sopomenk, vendar brez razlage, kakšni vse predlogi se lahko na seznamu pojavijo. Pogosta so bila vprašanja, kaj narediti z nad- oz. podpomenkami, neuveljavljenimi tujkami ali popačenkami ter predlogi, ki so bolj razlagalne narave. Na podlagi povratnih informacij se kot smiselna dopolnitev kaže tudi dodatni opis, katere vrste podatkov lahko ocenjevalci pričakujejo na seznamu. Na drugi strani so bili zatipki oz. očitne napake izpostavljeni kot primeri, ki se smatrajo za nerelevantne oz. nesopomenske, vendar jih niso vedno vsi ocenjevalci umestili v kategorijo NE (npr. par *Zemlja – e*, kjer gre za očitno napako, je eden izmed ocenjevalcev umestil v kategorijo NISEM PREPRIČAN/NE VEM). Kot so ocenjevalci izpostavili sami, so pogrešali natančnejšo opredelitev kategorije POGOJNO DA. V navodilih za ostale skupine uporabnikov je zato treba natančneje opredeliti, da v ta sklop spadajo pari, kjer ocenjevalci sicer lahko povedo nekaj glede sopomenskosti, a ta ni nedvomna in bi želeli ob sopomenskem paru dodatne informacije.

7. Zaključek in naslednji koraki

V prispevku sem opisala načrt raziskave, s katero želim nasloviti raziskovalno vprašanje doktorske disertacije, in sicer da se pogled strokovne oz. širše jezikovne skupnosti razlikuje od pogleda slovaropiscev, vendar je ta potencialni drugačni pogled jezikovne skupnosti lahko uporaben in pomemben za razvoj virov. Predstavila sem tudi potek evalvacijske naloge, ki so jo kot testna množica opravili študenti. S tem sem želela preizkusiti sestavljena navodila in izbrana orodja ter določiti časovno-finančni obseg in zahtevnost naloge, kar mi bo pomagalo pri načrtovanju dela in rekrutaciji preostalih predvidenih skupin ocenjevalcev.

Na podlagi odgovorov in povratnih informacij v pilotni skupini se je ocenjevanje izkazalo kot izvedljivo. Izbrani orodji, in sicer Googlova preglednica za evalvacijo sopomenskih parov ter spletno orodje za anketiranje Ika, sta se izkazala kot ustrezna, enostavna za uporabo ter finančno in časovno vzdržna.¹⁸ Potrebno bo izboljšati navodila za ocenjevalce, hkrati se zdi smiselno ocenjevalcem podrobneje pojasniti kontekst raziskave in namen ocenjevanja (pridobitev njihovega subjektivnega mnenja, ne "pravih" odgovorov). Naslovljena so bila problematična mesta v navodilih, navodila za ocenjevalce pa ustrezno preoblikovana in dopolnjena za naslednje predvidene skupine.

Predvideno je, da isti seznam v ocenjevanje dobi še 5 skupin ocenjevalcev, in sicer slovaropisci, poklicni prevajalci, lektorji, učitelji slovenščine in ljubitelji jezika brez jezikoslovne izobrazbe. V času pisanja prispevka poteka rekrutacija sodelujočih, podatki pa naj bi bili pridobljeni do poletja 2022. Sledijo analize rezultatov znotraj skupin, nato pa še primerjalno med skupinami.

Čeprav prvi rezultati še niso primerni za generalizacijo, ponujajo dober uvid v dileme pri presojanju sopomenskosti uporabniško dodanega gradiva. Spodbudno je, da je bilo (vsaj po prvem koraku raziskave) veliko uporabniško dodanega gradiva ocenjenega kot relevantnega. V primeru, da ocene drugih predvidenih sodelujočih skupin prinesejo podobne rezultate, bo to ugotovitev mogoče upoštevati pri nadaljnjem razvoju sopomenskih virov za slovenščino, v smer širitve in bogatenja z novimi podatki. Hkrati se potrjujejo predhodna spoznanja, da so uporabniški predlogi v kar največji meri konstruktivni in dobronamerni, kar je ključno za delovanje in nadaljnji razvoj odzivnih slovarjev.

8. Zahvala

Prispevek je nastal v okviru raziskovalnega programa Jezikovni viri in tehnologije za slovenski jezik (številka programa P6-0411), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije.

9. Literatura

Andrea Abel in Christian M. Meyer. 2013. The dynamics outside the paper: user contributions to online dictionaries. V: *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*, str. 179–94. Trojina, Institute for Applied Slovene Studies in Eesti Keele Instituut.

Jurij Apresjan. 2000. *Systematic Lexicography* (prev. Kevin Windle). Oxford University Press, Oxford.

¹⁸ Obe orodji sta namreč brezplačni za uporabo tako za ocenjevalce kot za raziskovalce ter ne zahtevata predznanja oz.

dodatnega usposabljanja, posebne strojne opreme ali dodatne registracije.

- Špela Arhar Holdt. 2015. Uporabniške raziskave za potrebe slovenskega slovaropisja: prvi korak. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek, ur., *Slovar sodobne slovenščine: problemi in rešitve*, str. 136–49. Znanstvena založba Filozofske fakultete.
- Špela Arhar Holdt. 2020. How Users Responded to a Responsive Dictionary: the Case of the Thesaurus of Modern Slovene. *Rasprave Instituta za hrvaški jezik i jezikoslovje*, 46(2): 465–82. doi:10.31724/rihjj.46.2.1
- Špela Arhar Holdt in Jaka Čibej. 2020. Rezultati projekta “Slovar sopomenk sodobne slovenščine: Od skupnosti za skupnost”. V: *Zbornik konference Jezikovne tehnologije in digitalna humanistika, 24. – 25. september 2020, Ljubljana, Slovenija*, str. 3–9. Inštitut za novejšo zgodovino.
- Špela Arhar Holdt, Jaka Čibej, Kaja Dobrovoljc, Polona 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: *Proceedings of the XVIII EURALEX International Congress, Lexicography in Global Contexts, 17-21 July 2018, Ljubljana*, str. 401–10. Znanstvena založba Filozofske fakultete.
- Pavel Braslavski, Dmitry Ustalov in Mikhail Mukhin. 2014. A Spinning Wheel for YARN: User Interface for a Crowdsourced Thesaurus. V: *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics*, str. 101–104. Association for Computational Linguistics. doi: 10.3115/v1/E14-2026
- Jaka Čibej in Špela Arhar Holdt. 2019. Repel the syntruders! A crowdsourcing cleanup of the thesaurus of modern Slovene. V: *Electronic lexicography in the 21st century: Smart lexicography. Proceedings of the eLex 2019 conference, 1–3 October 2019, Sintra, Portugal*, str. 338–56. Lexical Computing CZ s.r.o.
- Jaka Čibej, Darja Fišer in Iztok Kosem. 2015. The role of crowdsourcing in lexicography. V: *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of eLex 2015 Conference, 11-13 August 2015, Herstmonceux Castle, United Kingdom*, str. 70–83. Trojina, Institute for Applied Slovene Studies in Lexical Computing Ltd.
- Darja Fišer, Aleš Tavčar in Tomaž Erjavec. 2014. sloWCrowd: A crowdsourcing tool for lexicographic tasks. V: *Proceedings of the Ninth International Conference on Language Resources and Evaluation. LREC'14*, str. 3471–75. European Language Resources Association (ELRA).
- Darja Fišer. 2015. *Semantic lexicon of Slovene sloWNet 3.1*. Repozitorij raziskovalne strukture CLARIN.SI, <http://hdl.handle.net/11356/1026>
- Polona Gantar, Simon Krek, Iztok Kosem, Mojca Šorli, Polonca Kocjančič, Katja Grabnar, Olga Yerošina, Petra Zaranšek in Nina Drstvenšek. 2013. *Leksikalna baza za slovenščino 1.0*. Repozitorij raziskovalne strukture CLARIN.SI. Pridobljeno iz <http://hdl.handle.net/11356/1030>
- Vojko Gorjanc, Polona Gantar, Iztok Kosem in Simon Krek, ur. 2017. *Slovar sodobne slovenščine: problemi in rešitve*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani, Ljubljana. doi:10.4312/9789612379759
- Hans Henrich Hock. 1991. *Principles of Historical Linguistics* (druga, razširjena in dopolnjena izdaja izd.). Mouton de Gruyter, Berlin, New York.
- Aleš Horák in Adam Rambousek. 2018. Wordnet Consistency Checking via Crowdsourcing. V: *Proceedings of the XVIII EURALEX International Congress, Lexicography in Global Contexts, 17–21 July 2018, Ljubljana*, str. 1023–29). Znanstvena založba Filozofske fakultete.
- Iztok Kosem in Eva Pori. 2021. Slovenske ontologije semantičnih tipov: samostalniki. V: I. Kosem, ur., *Kolokacije v slovenščini*, str. 159–202. Znanstvena založba Filozofske fakultete Univerze v Ljubljani, Ljubljana. doi:10.4312/9789610605379
- Iztok Kosem, Polona Gantar in Simon Krek. 2013. Automation of lexicographic work: an opportunity for both lexicographers and crowd-sourcing for both lexicographers and crowd-sourcing. V: *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, 17-19 October 2013*, str. 32–48. Trojina, Institute for Applied Slovene Studies in Eesti Keele Instituut.
- 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*. Repozitorij raziskovalne strukture CLARIN.SI., <http://hdl.handle.net/11356/1166>
- Lionel Nicolas, Lavinia Aparaschivei, Verena Lyding, Christos Rodosthenou, Federico Sangati, Alexander König in Corina Forascu. 2021. An Experiment on Implicitly Crowdsourcing Expert Knowledge about Romanian Synonyms from Language Learners. V: *Proceedings of the 10th Workshop on NLP for Computer Assisted Language Learning*, str. 1–14. LiU Electronic Press.
- Tobias Schnabel, Igor Labutov, David Mimno in Thorsten Joachims. 2015. Evaluation methods for unsupervised word embeddings. V: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, str. 298–307. Association for Computational Linguistics. doi: 10.18653/v1/D15-1
- Jerica Snoj. 2019. *Leksikalna sinonimija v Sinonimnem slovarju slovenskega jezika*. Založba ZRC, ZRC SAZU, Ljubljana.
- Jerica Snoj, Martin Ahlin, Branka Lazar in Zvonka Praznik. 2016. *Sinonimni slovar slovenskega jezika*. Založba ZRC, ZRC SAZU, Ljubljana.
- Rion Snow, Brendan O'Connor, Daniel Jurafsky in Andrew Y. Ng. 2008. Cheap and Fast—But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks. V: *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, 25-27 October 2008, Honolulu, Hawaii, USA*, str. 254–63. Omnipress Inc.
- Tavčar, Aleš, Darja Fišer in Tomaž Erjavec. 2012. sloWCrowd: orodje za popravlanje wordneta z izkoriščanjem moči množic. V: *Zbornik Osme konference Jezikovne tehnologije*, str. 197–202. Inštitut Jožef Stefan.
- Jože Toporišič. 1992. *Enciklopedija slovenskega jezika*. Cankarjeva založba, Ljubljana.
- Ada Vidovič-Muha. 2013. *Slovensko leksikalno pomenoslovje*. Znanstvena založba Filozofske fakultete, Ljubljana.

Ladislav Zgusta. 1971. *Manual of Lexicography*.
Academia, Publishing House of the Czechoslovak
Academy of Sciences, Praga.

Marina Zorman. 2000. *O sinonimiji*. Znanstveni inštitut
Filozofske fakultete, Ljubljana.