

Razvrstitev kolokacij v slovarskem vmesniku: uporabniške prioritete

Špela ARHAR HOLDT

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

The Collocations Dictionary of Modern Slovene contains a large amount of collocations, which were automatically extracted from the reference corpus of written Slovene Gigafida. In the dictionary interface, the collocations can be listed according to logDice (default setting) and other parameters. In this paper we focus on the so-called collocational cream of the crop, the information that is shown as the main, most easily accessible and noticeable part of an entry. This information summary is a sort of a preview of the entry and the point of departure for further exploration in the dictionary. The criteria of the users for this cream of the crop have been identified by using a survey, which was completed by 457 participants, comprising of proofreaders, translators, teachers and other users. In a questionnaire, the participants first made a selection of collocations of the headwords *belina* and *izolirati*, and then described the criteria and the preferences used for ordering. Nearly all the participants named frequency as the main criterion when making the cream of the crop selection. Also often used criteria were understandability of the collocate, the presence in the general language (not rare or terminological), and the typicality in a certain syntactic structure. The analysis of actual orderings of the participants also confirms that their decisions are more similar to the ordering by frequency than by the logDice association score. Based on these and similar other findings we have formed research hypotheses that can be tested on a larger sample of dictionary material.

Keywords: Collocations Dictionary of Modern Slovene, sorting collocations, user preferences

1 Uvod

Z razvojem digitalnega medija in strojno podprte metodologije je (tudi) na področju uporabnega jezikoslovja na voljo vedno obširnejša količina jezikovnih virov in podatkov. Podatkovno obilje, s katerim se srečujemo prvič v zgodovini, s sabo prinaša nova raziskovalna in razvojna vprašanja. Na eni strani se pojavljajo izzivi in rešitve na področju pridobivanja podatkov iz jezikovnih virov ter njihovega strukturiranja, razvrščanja, označevanja in opisovanja v strojno berljivih slovarskih bazah, na drugi strani vprašanja, kako količinsko bogate in raznorodne jezikovne informacije predstaviti jezikovni oz. uporabniški skupnosti. V tej točki se digitalno slovaropisje, podprto s spoznanji vmesniškega oblikovanja ter uporabniških raziskav, trudi zagotoviti prijetno uporabniško izkušnjo z digitalnimi slovarskimi viri, ki naj bi omogočali hiter dostop do jezikovnih informacij, njihovo preglednost, razumljivost, povezanost, napredne možnosti iskanja, filtriranja, razvrščanja in izvažanja.

V tem kontekstu se prispevek posveča kolokacijskim podatkom za slovenščino, ki so na voljo v Kolokacijskem slovarju sodobne slovenščine (Kosem idr. 2018a). Slovar je uporabnikom prosto na voljo prek spletnega slovarskega vmesnika,¹ kot odprto dostopna slovarska baza pa na repozitoriju CLARIN.SI (Kosem idr. 2019). Osrednje vprašanje prispevka je razvrstitev kolokacij v slovarskem vmesniku, pri čemer izhajamo iz predpostavke, da slovarski uporabniki pri svojem delu potrebujejo hiter vpogled v najrelevantnejše kolokacijske informacije o določeni besedi, ne pa zgolj poljubno urejenega seznama vseh možnih, čeprav legitimnih, kolokatorjev. Kolokacijski "jagodni izbor" nam pomeni podatke, ki jih kažemo na prvem oz. najhitreje dostopnem, najbolj vidnem mestu kolokacijskega slovarja in so izbrani tako, da uporabniku podajo učinkovito izhodišče za nadaljnje delo. Kaj so kriteriji za jagodni izbor po mnenju potencialnih uporabnikov slovarja, ugotavljamo s pomočjo ankete, ki jo je izpolnilo 457 sodelujočih.

Raziskava je nastala pod okriljem projekta KOLOS – Kolokacije kot temelj jezikovnega opisa: semantični in časovni vidiki (ARRS

1 <https://viri.cjvt.si/kolokacije/slv/>

J6-8255), ki se je posvečal tudi merjenju kolokativnosti, mdr. evalvaciji oz. določitvi optimalnih statističnih metod za namene različnih ciljnih uporabnikov.² Vsebinsko se prispevek povezuje z uporabniško evalvacijo vmesnika Kolokacijskega slovarja sodobne slovenščine, ki je bila pod okriljem istega projekta opravljena med predstavniki različnih uporabniških skupin (Pori idr. 2020; 2021): sodelujoči so ocenjevali preglednost, intuitivnost, funkcionalnost vmesnika, torej možnosti, ki jih ima glede razvrščanja kolokacij v vmesniku na voljo uporabniška skupnost. Raziskavo, ki jo predstavljamo v tem poglavju monografije, zanima korak prej: kako naj kolokacijske podatke uredijo razvijalci slovarja, da bo izhodiščna postavitev v slovarskem gestu skladna z uporabniškimi željami in pričakovanji.

V prispevku najprej opredelimo raziskovalno področje, v katerem se študija umešča, nadaljujemo s predstavitevjo anketnega vprašalnika, metapodatkov sodelujočih jezikovnih uporabnic in uporabnikov ter izbranih rezultatov ankete. Ker je podatkov preveč za celovito navedbo v prispevku, so v pregledni tabelarični obliki odprto dostopni prek spletne strani projekta Kolos, za preverljivost in ponovljivost raziskave pa je na voljo tudi celotni anketni vprašalnik.³

2 Raziskovalna izhodišča

2.1 Uporabniške raziskave na področju digitalnega slovaropisja

Slovaropisne uporabniške raziskave imajo v evropskem prostoru relativno dolgo tradicijo, čeprav so se v primerjavi z drugimi vrstami slovarskih raziskav pojavile pozno. Po nekaj desetletjih utemeljevanja je področje zacvetelo z vstopom slovarjev v digitalni svet, ki je znatno razširil nabor metodoloških možnosti za identifikacijo uporabniških navad, potreb in želja, kar opisujejo npr. Tarp (2009), Welker (2003a; 2003b), Lew in De Schryver (2014). Predvsem pristop z

² Spletna stran projekta: <https://www.cjvt.si/kolos/>.

³ Vprašalnik: <https://www.cjvt.si/kolos/wp-content/uploads/sites/9/2021/03/Anketni-vprasanik-Razvrstitev-podatkov-v-kolokacijskem-slovarju.pdf>, podatki: <https://www.cjvt.si/kolos/wp-content/uploads/sites/9/2021/03/Anketni-rezultati-Razvrstitev-podatkov-v-kolokacijskem-slovarju.xlsx>.

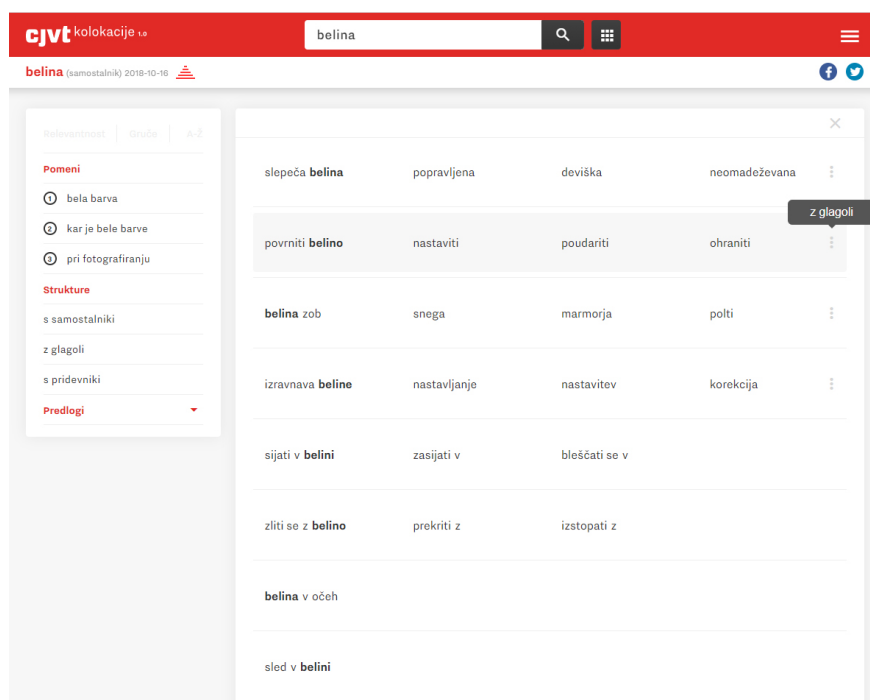
anketiranjem, ki v sodobnem času postaja skoraj izključno spletno, je bil v literaturi tudi že kritično ovrednoten (Bogaards 2003; Tarp 2009; Bergenholtz 2011), čemur so sledile jasneje zasnovane in metodološko preiščene anketne študije (npr. Lorentzen in Theilgaard 2012; Müller-Spitzer 2014; Kosem idr. 2018b), ki jim sledimo tudi v pričujočem prispevku. Idejno izhajamo iz funkcijske teorije slovaropisja, po kateri »so slovaropisni izdelki po temeljnih značilnostih enaki kateremu koli drugemu človeškemu orodju, zasnovani so namreč – oz. bi morali biti – za zadovoljevanje določene vrste človeških potreb« (Fuentes-Olivera in Tarp 2014: 45, prevod Š. A. H.).

Po opozorilih, da akutno primanjkuje empiričnih podatkov o slovenskih slovarskih uporabnikih (npr. Stabej 2009; Logar 2009), in načrtih, kako to vrzel odpraviti (Gorjanc idr. 2017), so v zadnjih letih uporabniške raziskave v porastu tudi pri nas, še zlasti v povezavi z odzivnimi jezikovnimi viri, ki uporabnike vključujejo ne le kot naslovnike, ampak tudi kot sodelujoče pri razvoju digitalne jezikovne infrastrukture za slovenščino. K že omenjenim uporabniškim raziskavam, ki se dotikajo Kolokacijskega slovarja sodobne slovenščine (Pori idr. 2020; 2021), je mogoče dodati anketno raziskavo Arhar Holdt (2020), ki identificira odnos potencialnih slovarskih uporabnikov (N = 671) do novosti odzivnega koncepta, kot so stalno posodabljanje, digitalni format, povezave na korpus, uporabniško vključevanje, vsebnost napak v strojno pridobljenih podatkih in podobno. Anketo, ki jo predstavljamo v pričujočem prispevku, je mogoče razumeti kot naslednji korak v pridobivanju uporabniške povratne informacije s pomočjo spletnega anketiranja, pri čemer je interes to pot usmerjen v izboljšavo specifičnega jezikovnega vira. Za ta namen so bile nujne določene prilagoditve metodologije, čemur se posvečamo v razdelku 3.

2.2 Razvrstitev kolokacij v Kolokacijskem slovarju sodobne slovenščine

Pred nadaljevanjem razprave je treba nekaj besed nameniti trenutnemu prikazu kolokacijskih podatkov v Kolokacijskem slovarju

sodobne slovenščine. Vstopni zaslon posameznega gesla trenutno (različica 1.0) prikaže besednozvezne strukture, v katerih se iztočnica pojavlja, za vsako od struktur pa so prikazani tudi (do) štiri kolokatorji (Slika 1). Na levem robu ekrana je na voljo filter, s katerim je strukture mogoče filtrirati glede na besedno vrsto. Gesla na višji stopnji slovaropisne pregledanosti v filtru ponujajo tudi indikatorje pomenov, ki jih iztočnica (lahko) izkazuje. Na vstopnem zaslonu si uporabnik izbere strukturo, ki ga zanima, in tako nadaljuje z raziskovanjem kolokacijskih podatkov v slovarju (na Sliki 1 je denimo izbrana druga struktura v seznamu; klik na ikono treh pik vodi do več kolokacijskih podatkov).⁴



Slika 1: Vstopni zaslon pri iztočnici *belina*.

4 Vstopne zaslone za iztočnici *belina* in *izolirati*, ki ju obravnavamo v tem prispevku, si je mogoče ogledati na povezavah: <https://viri.cjvt.si/kolokacije/slv/headword/48924#> in <https://viri.cjvt.si/kolokacije/slv/headword/1684#>. Pri prvi iztočnici so na voljo tudi pomenski indikatorji, ki so pridobljeni iz Leksikalne baze za slovenščino (Gantar 2015).

Kot opisujejo Kosem idr. (2018a), so kolokacije avtomatsko pridobljene iz referenčnega korpusa pisne slovenščine Gigafida (Logar idr. 2012). Pri izvozu kolokacij iz korpusa se v podatkovni bazi zabeleži statistika logDice (Rychlý 2008), kot tudi pogostnost kolokacije v korpusu. Pri razvrstitvi podatkov na vstopnih zaslonih se trenutno upošteva kolokacijska moč: pri posamezni strukturi so izpisani tisti štirje kolokatorji, ki imajo najvišjo vrednost logDice. Načeloma je izpis padajoč glede na vrednost, vendar se vrstni red lahko spremeni tako, da na prvo mesto pride tista od štirih besed, ki je najkrajša. Prva kolokacija je namreč v tem delu vmesnika izpisana v celoti in pri uporabi kratkih kolokatorjev lahko ostane izpisana v eni vrstici, kar je z vidika vmesniškega oblikovanja preferenčno. Na drugi strani je vrstni red samih struktur v trenutni različici slovarja univerzalen in sledi vnaprej pripravljenemu seznamu.

Ker je interpretacija anketnih rezultatov v nadaljevanju tega prispevka v veliki meri povezana s statističnimi podatki o kolokatorjih besed *belina* in *izolirati*, v Tabelah 1 in 2 predstavljamo podatke, ki so vključeni v Kolokacijski slovar sodobne slovenščine. V prvem stolpcu je navedena struktura, kot je opredeljena v slovarju. Kadar opis strukture v slovarju manjka, smo dodali zvezdico * in opisali strukturo po analogiji z obstoječimi opisi. Sledi število kolokatorjev (ki predstavlja kolokacijsko bogatost strukture); mesto oz. zaporedna številka strukture v slovarskem geslu; nato pa so kolokatorji naštetih padajoče glede na logDice, pri čemer sta v oklepaju navedeni tako statistična vrednost logDice kot korpusna pogostnost. Tisti kolokatorji, ki so del trenutnega vstopnega zaslona pri posameznem geslu, so v tabeli prikazani s krepkim tiskom.

Tabela 1: Kolokatorji za samostalnik *belina* v Kolokacijskem slovarju sodobne slovenščine.

Struktura	Število kolokat.	Mesto v slovarju	Kolokator (logDice / pogostnost v korpusu Gigafida)
s samostalniki v roditeljski / s samostalniki v vseh sklonih ⁵	32	3. in 4.	izravnava (8,38 / 67); nastavljanje (7,17 / 49); nastavitev (7,03 / 201); korekcija (5,94 / 22); uravnavanje (5,49 / 17); ravnovesje (5,43 / 36); določanje (5,39 / 46); prilagajanje (5,25 / 41); zajemanje (4,99 / 8); popravljjanje (4,09 / 5); kanček (3,28 / 7); odtonek (2,27 / 8); določitev (2,14 / 7); košček (1,51 / 5); temperatura (1,46 / 15); stopnja (1,06 / 26); videz (1,05 / 8); simbol (0,96 / 5); nivo (0,96 / 5); zob (5,1 / 70); snega (4,65 / 85); marmorja (4,6 / 8); polti (4,2 / 9); perila (3,62 / 17); platna (3,61 / 19); papirja (2,98 / 49); stene (2,56 / 25); zidu (2,03 / 12); kamna (1,97 / 12); obleke (1,38 / 11); kože (0,9 / 12); obraza (0,67 / 10)
s pridevniki v vseh sklonih	21	1.	popravljen (8,22 / 79); slepeča (8,03 / 19); deviška (7,6 / 38); neomadeževana (7,59 / 17); snežna (7,55 / 197); bleščeča (6,75 / 63); brezmadežna (6,53 / 14); nenatančna (6,41 / 12); nedolžna (6,18 / 64); neskončna (6,02 / 36); nedotaknjena (5,72 / 12); opojna (5,54 / 8); prosojna (5,23 / 9); brezhibna (5,1 / 11); sijoča (5,06 / 10); nežna (4,29 / 26); popolna (3,84 / 62); čista (3,83 / 61); mlečna (3,79 / 13); sveža (2,67 / 20); naravna (2,44 / 34)
z glagoli + predlog 'v' *+ tožilnik	7	9.	odeti (6,62 / 22); oviti (4,18 / 7); zazreti se (4,01 / 8); potopiti se (2,6 / 5); obleči (1,79 / 6); zaviti (1,13 / 4); spremeniti (-1,71 / 6)
s pridevniki v vseh sklonih + predlog 'v'	4	14.	odet (7,58 / 35); ovit (5,27 / 5); potopljen (5,16 / 6); oblečen (1,77 / 4)

⁵ V slovarju so skupaj prikazane kolokacije tipa *izravnava beline* (19 kolokacij) in *belina zob* (13 kolokacij), kar je bilo pri dosedanjih evalvacijah že identificirano kot problematično in bo v prihodnji različici odpravljeno. Za večjo preglednost kolokatorje v tabeli ločujemo v dve skupini. Napačno je tudi poimenovanje strukture »s samostalniki v vseh sklonih«, ker gre pri obeh skupinah za samostalnike z roditeljsko obliko.

Struktura	Število kolokat.	Mesto v slovarju	Kolokator (logDice / pogostnost v korpusu Gigafida)
z glagoli	4	2.	nastaviti (4,43 / 27); povrniti (2,68 / 13); poudariti (1,41 / 13); ohraniti (0,75 / 11)
z glagoli + predlog 'v' *+ mestnik	3	5.	zasijati (5,23 / 9); bleščati se (4,1 / 5); sijati (3,78 / 8)
*z glagoli + predlog 's/z'	3	6.	zliti se (2,9 / 5); prekriti (2,05 / 4); izstopati (1,21 / 4)
*s pridevniki v vseh sklonih + predlog 's/z'	2	12.	prekrit (4,12 / 5); obseden (3,62 / 4)
*s predlogom 'v' + samostalnik	1	7.	očeh (-0,79 / 5)
*s samostalniki + predlog 'v'	1	8.	sled (1,94 / 9)
*s samostalniki + predlog 'na'	1	11.	pogled (-0,72 / 8)
*z glagoli + predlog 'po'	1	10.	drseti (None / None) ⁶
*z glagoli + predlog 'od'	1	13.	bleščati (4,11 / 5)

6 Pri urejanju slovarskega gesla je bila ena od kolokacij oz. struktur dodana ročno na osnovi pregleda posodobljenega referenčnega korpusa Gigafida 2.0 (Krek idr. 2020). Celovita posodobitev podatkov z novim korpusom (tudi statistik logDice in korpusne pogostnosti) se načrtuje za slovarsko različico 2.0.

Tabela 2: Kolokatorji za glagol *izolirati* v Kolokacijskem slovarju sodobne slovenščine.

Struktura	Število kolokat.	Mesto v slovarju	Kolokator (logDice / pogostnost v korpusu Gigafida)
s samostalniki v tožilniku	54	1.	<p>podstrešje (8,66 / 18); protein (8,03 / 12); bakterijo (8 / 19); spojino (7,92 / 11); steno (7,6 / 37); radij (7,52 / 6); gen (7,42 / 18); alkaloid (7,29 / 5); strop (7,22 / 10); virus (7,02 / 13); glivo (6,97 / 4); vlaknino (6,96 / 4); učinkovino (6,88 / 7); Hrvaško (6,76 / 7); celico (6,68 / 24); inzulin (6,65 / 4); molekulo (6,51 / 5); povzročitelja (6,48 / 4); substanco (6,37 / 5); snov (6,37 / 28); beljakovino (6,33 / 4); kabino (6,3 / 5); antibiotik (6,22 / 5); zid (6,05 / 11); fasado (5,97 / 5); ostrešje (5,96 / 4); skladišče (5,83 / 5); streho (5,81 / 14); notranjost (5,64 / 8); hrup (5,56 / 5); tla (5,48 / 10); zvok (5,37 / 11); Nemčijo (5,3 / 4); sestavino (5,27 / 7); cev (4,98 / 4); bolnika (4,8 / 5); stavbo (4,79 / 6); posameznika (4,77 / 4); hišo (4,12 / 21); zapis (4,06 / 4); element (3,96 / 6); telo (3,55 / 9); objekt (3,34 / 5); stran (3,25 / 10); ploščo (3,08 / 4); prostor (2,56 / 13); dom (2,42 / 4); državo (2,35 / 6); skupino (2,35 / 5); Slovenijo (2,24 / 4); del (1,51 / 7); sistem (1,36 / 4); vrsto (1,35 / 5); človeka (1,01 / 4)</p>
s prislovi	28	2.	<p>toplotno (11,36 / 131); zvočno (10,3 / 64); socialno (8,49 / 22); mednarodno (7,03 / 7); politično (5,78 / 12); popolnoma (5,15 / 70); dodatno (5,08 / 51); povsem (4,96 / 76); rekoč (4,08 / 6); primerno (3,92 / 9); ustrezno (3,49 / 9); pravilno (3,39 / 11); temeljito (3,26 / 9); prvič (3,15 / 35); dobro (2,77 / 101); bolj (2,44 / 43); treba (2,08 / 58); odlično (2,03 / 6); preveč (1,99 / 8); težko (1,55 / 20); uspešno (1,18 / 7); najprej (1,16 / 16); takoj (1,1 / 10); pogosto (1,07 / 11); hkrati (1,02 / 6); torej (0,5 / 6); nato (0,32 / 8); lahko (0,28 / 79)</p>

Struktura	Število kolokat.	Mesto v slovarju	Kolokator (logDice / pogostnost v korpusu Gigafida)
predlog 'od' + s samostalniki v roditeljski strukturi	12	10.	okolice (10,55 / 24); sveta (9,75 / 64); okolja (9,63 / 13); dogajanja (8,38 / 6); skupnosti (8,07 / 6); vrstnika (7,77 / 4); izjave (7,72 / 5); Evrope (7,67 / 6); družbe (7,49 / 12); družine (7,4 / 6); medija (7,4 / 4); človeka (6,49 / 10)
predlog 'od' + s samostalniki v roditeljski strukturi	11	8.	tkiva (9,49 / 9); droge (9,17 / 5); rude (9,1 / 5); bakterije (8,96 / 4); krvi (8,93 / 10); rastline (8,82 / 12); zarodka (8,63 / 4); možganov (8,34 / 4); celice (7,45 / 6); življenja (5,11 / 4); vira (4,36 / 4)
*z glagoli v nedoločniku ⁷	9	9.	uspjeti (3,72 / 62); poskušati (2,23 / 19); nameravati (2 / 8); želeti (1,59 / 1); dati (1,56 / 6); skušati (1,52 / 11); pomagati (1,46 / 5); hoteti (1,18 / 7); znati (1,11 / 9)
predlog 'v' + s samostalniki v mestniku	5	3.	laboratoriju (5,51 / 8); politiki (3,82 / 4); sobi (3,33 / 6); obliki (2,95 /); letu (-1,21 / 5)
z zanikanimi glagoli v nedoločniku ⁸	4	12.	želeti (2,34 / 10); uspjeti (2,14 / 6); smeti (1,3 / 12); moči (1,29 / 41)
predlog 'z' + s samostalniki v orodniku	4	4.	volno (9,94 / 7); fasado (9,32 / 5); materialom (5,89 / 4); oljem (4,75 / 5)
*predlog 'pred' + s samostalniki v orodniku	2	11.	svetom (8,25 // 13); novinarjem (7,71 // 8)
*predlog 'na' + s samostalniki v mestniku	1	5.	otoku (3,56 // 5)
*predlog 'za' + s samostalniki v tožilniku	1	6.	čas (2,75 // 5)
*predlog 'na' + s samostalniki v tožilniku	1	7.	način (1,1 // 12)

3 Zasnova ankete

Prednost spletnih anket je, da so enostavne in poceni za pripravo ter diseminacijo, zato je z njimi mogoče doseči veliko število potencialnih

⁷ V slovarju je struktura napačno poimenovana »z glagoli + predlog Inf-GBZ«.

⁸ V slovarju je struktura napačno poimenovana »z zanikanimi glagoli + predlog Inf-GBZ«.

sodelujočih.⁹ Pristopi, ki sodelujoče sprašujejo po mnenju, so po drugi strani pomanjkljivi, ker se vprašani (lahko) opredeljujejo le do poznanega stanja (Müller-Spitzer 2014: 169) in ker sodelujoči pri anketah lahko vprašanja narobe interpretirajo, nanje nimajo odgovora, v ponujenih odgovorih podzavestno iščejo tiste, ki so po njihovi presoji bolj všečni in podobno (Groves idr. 2004: 209–226). Anketo smo zato zasnovali tako, da so za izbrani besedi – samostalnik *belina* in glagol *izolirati* – sodelujoči morali razvrščanje kolokacijskih struktur in kolokatorjev najprej opraviti sami, šele nato smo jih vprašali po prioritetah in mnenju. Na tak način so anketiranci pri odgovorih lahko izhajali iz lastne izkušnje s problemom in ne iz predvidevanj, kako bi različne rešitve v slovarju (lahko) funkcionirale. Strukturo anketnega vprašanja, ki je v celoti na voljo na spletni strani projekta, prikazuje Tabela 3.

Tabela 3: Struktura in vsebina anketnega vprašalnika.

Razdelek vprašalnika	Vsebina
Uvod	<ul style="list-style-type: none"> • Uvodni nagovor in poljudna definicija pojma »kolokacija« s primeri. • V1: Samoopredelitev razumevanja: »Ali se vam zdi, da dovolj dobro razumete pojem kolokacija, da lahko nadaljujete z reševanjem ankete?«
Priklic kolokacij (po spominu)	<ul style="list-style-type: none"> • V2: »Vpišite tri kolokacije z besedo <i>belina</i>, ki vam najprej padejo na pamet.« • V3: »Vpišite tri kolokacije z besedo <i>izolirati</i>, ki vam najprej padejo na pamet.«
Razvrščanje struktur	<ul style="list-style-type: none"> • V4: Na ekranu je 6 kolokacijskih struktur s samostalnikom <i>belina</i>. Sodelujoči/-a izbere tri skupine, ki bi jih v slovarju rad/-a videl/-a na prvem mestu in jih nato razvrsti glede na relevantnost (po lastni presoji). • V5: Na ekranu so 4 kolokacijske strukture z glagolom <i>izolirati</i>. Sodelujoči/-a izbere tri skupine, ki bi jih v slovarju rad/-a videl/-a na prvem mestu in jih nato razvrsti glede na relevantnost (po lastni presoji).

⁹ Pri delu smo uporabili zmožljiv spletni servis za anketiranje 1KA (<https://www.1ka.si/>), ki zbrane podatke tudi uredi in vizualizira, kar preprečuje napake obdelave.

Razdelek vprašalnika	Vsebina
Razvrščanje kolokacij	<ul style="list-style-type: none"> • Napoved nalog za razvrščanje kolokacij in navodila. • V6-V12: Na ekranu je nabor (okrog 20) kolokacij z besedo <i>belina</i> ali <i>izolirati</i>. Sodelujoči/-a izbere 5 kolokacij, ki bi jih v slovarju rad/-a videl/-a na prvem mestu in jih nato razvrsti glede na relevantnost (po lastni presoji).
Opredelitev kriterijev	<ul style="list-style-type: none"> • V13: Sodelujoči/-a opredeli, katere kriterije za razvrščanje kolokacij je upošteval/-a in kateri kriteriji pri odločanju niso igrali vloge. • V14 in V15: Sodelujoči/-a opredeli, ali je opazil/-a, da nastopata besedi <i>belina</i> in <i>izolirati</i> v različnih pomenih, in ali je to igralo vlogo pri razvrščanju.
Dodatni komentarji	<ul style="list-style-type: none"> • V16: Odprto vprašanje, kjer je mogoče vnesti morebitne dodatne komentarje.¹⁰
Metapodatki	<ul style="list-style-type: none"> • V17: Starost. • V18: Status oz. zaposlitev. • V19: Opredelitev uporabniške skupine. • V20: Dosežena izobrazba.

Pri diseminaciji ankete smo ciljali na izbrane skupine potencialnih slovarskih uporabnikov (Arhar Holdt idr. 2016): lektorje, prevajalce ter učitelje slovenščine kot prvega (na različnih stopnjah šolanja) in kot drugega ali tujega jezika. Pripadnost uporabniškim skupinam smo v vprašalniku identificirali (V19 v Tabeli 3), nismo pa sodelovanja v anketi nikakor omejevali. Zastopanost ciljnih skupin v vzorcu je zato višja, ne pa izključna in uravnotežena, zato rezultati niso uporabni za posploševanje na celotno populacijo slovarskih uporabnikov. Dodaten zadržek pri posploševanju je tudi, da spletne ankete, ki temeljijo na prostovoljnem sodelovanju, običajno pritegnejo sodelujoče, ki jih tematika bolj zanima in so zato lahko do vsebine atipično nekritični ali kritični. V nadaljevanju prispevka rezultate interpretiramo naštetim omejitvam primerno.

Anketa je bila aktivna med 22. 5. in 19. 9. 2018. V tem času jo je delno izpolnilo 820 sodelujočih, končalo pa 457 sodelujočih. Kompleksnost ankete je povzročila precejšen osip sodelovanja: najvišji upad je bil na začetku ankete, med samo izvedbo nalog razvrščanja

¹⁰ Zaradi omejenega prostora v prispevku teh komentarjev ne analiziramo, po večini gre za pojasnila sodelujočih glede odločitev v anketi ali spodbudne besede glede Kolokacijskega slovarja sodobne slovenščine.

pa je obupalo cca. 100 sodelujočih. V preverbi razumevanja pojma kolokacija (V1 v Tabeli 3) je 28 sodelujočih (3 %) odgovorilo, da teme ne razumejo dovolj, da bi z anketo nadaljevali.¹¹ Vseeno je treba omeniti, da po avtomatski oceni programa za spletno anketiranje 1ka anketa še vedno spada med enostavne in srednje dolge; tako ocenjeni kot dejanski čas reševanja je bil cca. 10 minut.

Ker gre za prvo predstavitev raziskave, rezultate v nadaljevanju prikazujemo opisno, izpostavimo pa tudi možnosti za nadaljnje analize in raziskovalne hipoteze, ki bi jih bilo smiselno preveriti na večji količini gradiva. Pri tem je treba upoštevati, da so bili metapodatki v vprašalnik vključeni nazadnje (V17–20 v Tabeli 3), zato bo za primerjalne statistične analize mogoče uporabiti le do konca izpolnjene ankete, pri katerih sodelujoči teh vprašanj obenem niso preskočili (okrog 415 anket, gl. razdelek 4.4).

4 Rezultati z diskusijo

4.1 Priklic kolokacij po spominu

Naštevaje kolokacij po spominu (V2 in V3 v Tabeli 3) smo v anketo vključili za dodatno preverbo, ali sodelujoči koncept kolokacije razumejo, in oceno, v kolikšni meri se tak kolokacijski priklic prekriva s kolokacijskim naborom, ki ga ponuja Kolokacijski slovar sodobne slovenščine. Predvidevali smo, da se bodo nekatere kolokacije v odgovorih sodelujočih pojavljale pogosteje, in zanimalo nas je, kolikšne so v teh primerih statistične mere in korpusna pogostnost.

Tabeli 4 in 5 predstavljata rezultate. V prispevek vključujemo tiste kolokacije, ki se v podatkih pojavijo s pogostnostjo 10 ali več.¹² Takšnih je pri obeh iztočnicah primerljivo število: 29 za *belina* in 30

11 Ta filter smo vključili, da lahko pri interpretaciji odgovorov računamo s tem, da so sodelujoči osnovne pojme (vsaj po lastni oceni) ustrezno razumeli. Razumevanje posredno preverjata tudi vprašanji V2 in V3.

12 Ker so sodelujoči v odgovore pisali tako kolokacije kot posamezne kolokatorje, smo rezultate uredili tako, da smo posamezne kolokatorje razširili v kolokacije (npr. *neskončna* – *neskončna belina*). Pri tem smo ohranili morebitne razlike v slovničnem številu (npr. *izolirati človeka* – *izolirati ljudi*), kot kaže Tabela 5. Če bi seštevali vse primere v edninski obliki, bi se v tabelo prebila še kakšna dodatna kolokacija, vendar so za analizo in razmisleke o prikazu kolokacij v slovarju zanimive tudi tovrstne razlike v kategorialnih lastnostih.

za *izolirati*. V prvem stolpcu tabel navajamo kolokacijo, sledi podatek, kolikokrat se je pojavila v anketnih odgovorih, ustrezajoča struktura in statistike (oboje povzemamo iz Tabele 1 in 2). Dodali smo tudi mesto, ki ga iz korpusa pridobljena kolokacija zaseda na seznamu, če je slednji urejen padajoče glede na logDice ali glede na korpusno pogostnost. Kolokacije, ki se že zdaj pojavljajo na vstopnem zaslonu slovarskega gesla (gl. razdelek 2.2), so v tabeli prikazane s krepkim tiskom.

Tabela 4: Najpogostejše uporabniško naštete kolokacije s samostalnikom *belina*.¹³

Uporabniško naštete kolokacije z <i>belina</i>	Št. v anketi	Ustrezajoča struktura	logDice	pogostnost	Mesto glede na logDice	Mesto glede na pogostnost
snežna belina	344	s prid. v vseh sklonih	7,55	197	5 od 21	1 od 21
čista belina	236	s prid. v vseh sklonih	3,83	61	18 od 21	6 od 21
popolna belina	71	s prid. v vseh sklonih	3,84	62	17 od 21	5 od 21
belina snega	61	s samost. v rodilniku	4,65	85	2 od 13	1 od 13
bleščeča belina	60	s prid. v vseh sklonih	6,75	63	6 od 21	4 od 21
neskončna belina	52	s prid. v vseh sklonih	6,02	36	10 od 21	8 od 21
belina zob	46	s samost. v rodilniku	5,1	70	1 od 13	2 od 13
belina perila	45	s samost. v rodilniku	3,62	17	5 od 13	6 od 13
sveža belina	37	s prid. v vseh sklonih	2,67	20	20 od 21	11 od 21
slepeča belina	37	s prid. v vseh sklonih	8,03	19	2 od 21	12 od 21
očesna belina	27	s prid. v vseh sklonih				ni v slovarju
belina papirja	24	s samost. v rodilniku	2,98	49	7 od 13	3 od 13
zobna belina	23	s prid. v vseh sklonih				ni v slovarju
zaslepljujoča belina	23	s prid. v vseh sklonih				ni v slovarju
Janez Belina ¹⁴	23	s samost. v vseh sklonih				ni v slovarju
sijoča belina	19	s prid. v vseh sklonih	5,06	10	15 od 21	19 od 21
belina stene	17	s samost. v rodilniku	2,56	25	8 od 13	4 od 13
zimska belina	17	s prid. v vseh sklonih				ni v slovarju
belina neba	16	s samost. v rodilniku				ni v slovarju
belina obraza	15	s samost. v rodilniku	0,67	10	13 od 13	11 od 13

¹³ Prvo kolokacijo je vneslo 642 sodelujočih, drugo 623 in tretjo 595 sodelujočih.

¹⁴ Gre za osebno lastno ime znanega televizijskega lika.

Uporabniško naštete koloka- cije z <i>belina</i>	Št. v anketi	Ustrežajoča struktura	logDice	pogost- nost	Mesto glede na logDice	Mesto glede na pogostnost
nebeška belina	15	s prid. v vseh sklonih				ni v slovarju
bela belina	14	s prid. v vseh sklonih				ni v slovarju
brezhibna belina	13	s prid. v vseh sklonih	5,1	11	14 od 21	18 od 21
prostrana belina	12	s prid. v vseh sklonih				ni v slovarju
neomadeževa- na belina	11	s prid. v vseh sklonih	7,59	17	4 od 21	13 od 21
belina dneva	11	s samost. v rodilniku				ni v slovarju
nedolžna belina	11	s prid. v vseh sklonih	6,18	64	9 od 21	3 od 21
velika belina	10	s prid. v vseh sklonih				ni v slovarju
brezmadežna belina	10	s prid. v vseh sklonih	6,53	14	7 od 21	14 od 21

V Tabeli 5 se pojavljajo primeri, kjer so sodelujoči pri vpisovanju kolokatorjev oz. kolokacij upoštevali slovnično število, in sicer pri primerih *izolirati* [*bolnika / bolnike; steno / stene; človeka / ljudi*]. Pri teh primerih smo v drugi stolpec dodali klicaj, ki opozarja, da je celostno teh primerov več.

Tabela 5: Najpogostejše uporabniško naštete kolokacije z glagolom *izolirati*.¹⁵

Uporabniško naštete koloka- cije z <i>izolirati</i>	Št. v anketi	Ustrežajoča struktura	logDice	pogost- nost	Mesto glede na logDice	Mesto glede na pogostnost
izolirati hišo	216	s samost. v tožilniku	4,12	21	39 od 54	4 od 54
izolirati bolnika	83 (!)	s samost. v tožilniku	4,8	5	36 od 54	31 od 54
dobro izolirati	72	s prislovi	2,77	101	15 od 28	2 od 28
toplotno izolirati	55	s prislovi	11,36	131	1 od 28	1 od 28
zvočno izolirati	52	s prislovi	10,3	64	2 od 28	6 od 28
izolirati streho	44	s samost. v tožilniku	5,81	14	28 od 54	8 od 54
izolirati osebo	39	s samost. v tožilniku				ni v slovarju
izolirati stene	35 (!)	*s samost. v tožilniku – množina		v slovarju samo ednina	(5 od 54)	(1 od 54)

15 Prvo kolokacijo je vneslo 576 sodelujočih, drugo 570 in tretjo 556 sodelujočih.

Uporabniško naštete koloka- cije z <i>izolirati</i>	Št. v anketi	Ustrezajoča struktura	logDice	pogost- nost	Mesto glede na logDice	Mesto glede na pogostnost
popolnoma izolirati	32	s prislovi	5,15	70	6 od 28	5 od 28
izolirati človeka	28 (!)	s samost. v tožilniku	1,01	4	54 od 54	40 od 54
izolirati žico	28	s samost. v tožilniku				ni v slovarju
izolirati sobo	27	s samost. v tožilniku				ni v slovarju
izolirati prostor	27	s samost. v tožilniku	2,56	13	46 od 54	9 od 54
izolirati se ¹⁶	26	/				ni v slovarju
izolirati steno	26 (!)	s samost. v tožilniku	7,6	37	5 od 54	1 od 54
izolirati stano- vanje	24	s samost. v tožilniku				ni v slovarju
izolirati kabel	21	s samost. v tožilniku				ni v slovarju
izolirati fasado	20	s samost. v tožilniku	5,97	5	25 od 54	34 od 54
izolirati stavbo	19	s samost. v tožilniku	4,79	6	37 od 54	26 od 54
izolirati virus	17	s samost. v tožilniku	7,02	13	10 od 54	10 od 54
izolirati ljudi	15 (!)	*s samost. v tožilniku – množina	v slovarju samo ednina		(54 od 54)	(40 od 54)
izolirati okna	15	*s samost. v tožilniku – množina				ni v slovarju
izolirati sebe	15	/				ni v slovarju
izolirni trak ¹⁷	14	/				/
izolirati električno	13	s samost. v tožilniku				ni v slovarju
izolirati bolnike	12 (!)	*s samost. v tožilniku – množina	v slovarju samo ednina		(36 od 54)	(31 od 54)
izolirati pacienta	12	s samost. v tožilniku				ni v slovarju
povsem izolirati	12	s prislovi	4,96	76	8 od 28	4 od 28
izolirati cev	10	s samost. v tožilniku	4,98	4	35 od 54	47 od 54
izolirati posame- znika	10	s samost. v tožilniku	4,77	4	38 od 54	46 od 54

Med najpogostejšimi anketnimi vnosi je kot problematične v smislu razumevanja koncepta kolokacije najti samo *izolirni trak*, *izolirati [sebe / se]* ter *Janez Belina*, vendar tudi zadnja dva primera

16 *Izolirati se* in *izolirati sebe* v slovar nista vključeni kot kolokaciji; *izolirati se* bi bila potencialno iztočnica.

17 V ta odgovor so šteti tako primeri, ko je sodelujoči napisal zvezo *izolirni trak*, kot primeri s samo besedo *trak*. Pri teh vnosi gre za asociacijo, ne pa dejansko kolokacijo z glagolom *izolirati*.

ne nasprotujeta poenostavljeni definiciji, ki je bila podana v uvodu ankete.¹⁸ V obeh tabelah je opaziti, da najpogostejši oz. najpogostejša odgovora glede zastopanosti močno izstopata, nato pogostnost postopoma pada. Hipoteza, da obstajajo primeri, ki jih po spominu navaja večina sodelujočih, se za dani dve iztočnici torej potrdi in bi jo bilo zanimivo preveriti tudi na večji količini gradiva. Podobno velja za strukture: v tabelah je najti samo po dve različni strukturi (izstopa tudi sicer atipičen primer *Janez Belina*): pri *belina* 20 kolokacij sodi med strukture z ujemalnimi pridevniki (npr. *nežna belina*) in 8 z določujočim desnim samostalnikom v roditniku (npr. *belina snega*); pri *izolirati* se 22 primerov pojavlja s samostalnikom v tožilniku (npr. *izolirati hišo*) in 5 z določujočim prislovom (npr. *toplotno izolirati*).

Prekrivnost po spominu naštetih kolokacij s korpusno pridobljenimi je nižja od pričakovanega. Od 29 najpogosteje naštetih kolokacij s samostalnikom *belina* jih je v slovarju 18 (62 %), od 30 z glagolom *izolirati* pa 19 (63 %), pri čemer je treba upoštevati, da se nekateri od teh primerov v korpusu pojavljajo, niso pa dosegli frekvenčnih pragov za umestitev v slovar. Tipičen tak primer je *zimska belina*, ki se v korpusu Gigafida pojavi 20-krat, v Gigafida 2.0 pa 21-krat, ali *izolirati hišo*, ki se v Gigafida 2.0 pojavi 15-krat. Nadaljnje analize na večji količini gradiva bodo pomagale opredeliti, ali in v katerih primerih bi sprememba parametrov lahko pozitivno vplivala na reprezentiranost podatkov v slovarju. Pri tem gre kot izziv izpostaviti zlasti kolokacije, ki so izvorno terminološke in v referenčnem korpusu redke, vendar očitno prisotne v pojmovnem svetu sodelujočih, kot npr. *izolirati* [*žico, kabel, elektriko*].

Kot protiprimer je omeniti kolokacije, ki se pojavljajo v slovarju, ni pa jih med najpogostejšimi po spominu naštetimi primeri. Pregled širšega nabora uporabniških vnosov pokaže, da se večina primerov najde med redkeje naštetimi: pri samostalniku *belina* najdemo *mlečna* (9 odgovorov), *deviška* (8), *nežna* (8), *prosojna* (7), *naravna*

18 Ubeseditev: »Kolokacije so besede, ki v jeziku tipično nastopajo skupaj, npr. *hitra vožnja, varna vožnja, spretna vožnja* itd. Kolokacije so po zgradbi različne, npr. glagol + samostalnik (*nadaljevati vožnjo, učiti se vožnje*), pridevnik + samostalnik (*hitra vožnja, varna vožnja*), samostalnik + samostalnik (*inštruktor vožnje, vožnja avtomobila*) itd.«

(2), *nedotaknjena* (1) in *opojna* (1 odgovor). Ne pojavita pa se med uporabniškimi vnosi dva terminološka primera s področja fotografije, *popravljen belina* in *nenatančna belina*, čeprav glede na logDice in korpusno pogostnost kotirata precej visoko (Tabela 1).

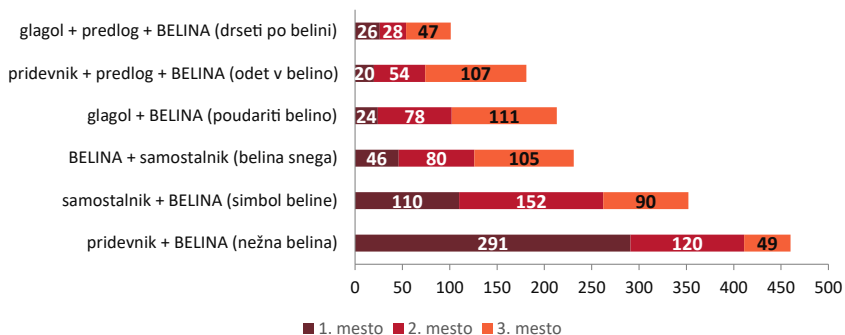
Pregled statističnih mer pokaže, da je v podatkih nekoliko večji delež takih kolokacij, ki so v seznamu korpusne pogostnosti više kot v logDice. Pri *belina* je takih primerov 11, pri *izolirati* 10 – v primerjavi s 7 in 4 primeri, kjer je kolokacija višje glede na logDice. Zlasti kolokacije, ki jih je naštevalo največ sodelujočih (*snežna belina*, *čista belina*, *izolirati hišo*), so na seznamu pogostnosti bistveno više, kot so na seznamu, urejenem glede na logDice.

4.2 Izbira in razvrščanje besednozveznih struktur

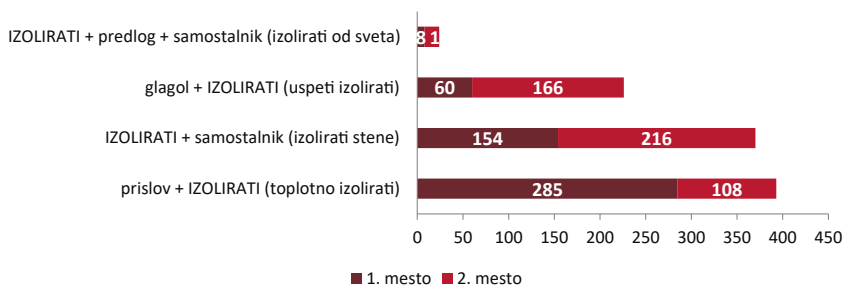
Naslednji korak ankete je vključeval selekcijo in razvrščanje kolokacijskih struktur (V4 in V5 v Tabeli 3). Za vsako od obeh besed je bilo naštetih nekaj struktur¹⁹ (6 za *belina* in 4 za *izolirati*), ki so se v vprašalniku prikazale v naključnem vrstnem redu. Sodelujoči so izbrali polovico struktur, ki bi jih glede na lastne preference želeli videti na prvem mestu slovarja, v drugem koraku pa so izbrane strukture tudi razvrstili. Rezultate prikazujeta Slika 2 in Slika 3, kjer barve nakazujejo, kolikokrat je bila posamezna struktura izbrana za prvo, drugo ali tretje mesto slovarja.

Rezultati kažejo, da bi na prvo mesto slovarja sodelujoči po večini umestili zveze samostalnika z levim pridevniškim prilastkom (npr. *nežna belina*), sledita obe strukturi z samostalnikom v rodilniku (*simbol beline*, *belina snega*). Pri zvezah z glagolom *izolirati* so prepričljivo na prvem mestu zveze s prislovom (*toplotno izolirati*) in samostalnikom v sklonu (*izolirati stene*). Vsaka od struktur je bila (tudi za prvo mesto) izbrana vsaj nekajkrat, vendar se kažejo trendi, ki potrjujejo hipotezo, da je mogoče identificirati večinske preference uporabniške skupnosti glede razvrstitve kolokacijskih struktur.

19 Natančno členjene strukture, kot so navedene v Kolokacijskem slovarju sodobne slovenščine, smo za vprašalnik združili v robustnejše skupine. V vprašalniku so bili za boljšo predstavo na voljo tudi konkretni primeri različnih kolokacij, ki ustrezajo posamezni strukturi.



Slika 2: Razvrstitev kolokacijskih struktur za samostalnik *belina*.



Slika 3: Razvrstitev kolokacijskih struktur za glagol *izolirati*.

Za vtis, kako se razvrstitev struktur primerja z ostalimi razpoložljivimi podatki, navajamo Tabelo 6. Strukturam iz ankete je pripisano, na katero mesto jih razvrščajo sodelujoči (Slika 2 in 3) in na katero mesto se uvrščajo glede na: (a) pojavljanje med kolokacijami, ki so najpogostejše priklicane po spominu (Tabeli 4 in 5); (b) kolokacijsko bogatost, tj. število kolokacij v posamezni strukturi (Tabela 1 in 2); (c) skupno frekvenco teh kolokacij v referenčnem korpusu (Tabela 1 in 2).

Primerjava pokaže, da je uporabniška razvrstitev dokaj skladna s priklicem po spominu, z izjemo dejstva, da se med najpogostejše naštetimi po spominu strukture tipa *simbol beline* ne pojavljajo, pri razvrščanju pa so sodelujoči ta tip zveze umestili na drugo mesto v slovarskem prikazu. Druga zanimiva ugotovitev, ki bi jo bilo smiselno

preveriti na večji količini gradiva, je, da v primerjavi s korpusnimi statistikami uporabniške preference nekoliko slabše vrednotijo predložne zveze. Ta podatek nakazuje težnjo sodelujočih k prioretiziranju krajših, dvodelnih zvez s polnopomenskimi elementi. V splošnem podatki pokažejo, da je za dana dva primera skupna frekvenca kolokacij v korpusu skladnejša s preferencami sodelujočih v raziskavi kot število kolokacij v strukturi, kar je mogoče nadalje testirati in uporabiti za dinamično razvrščanje struktur znotraj slovarskih iztočnic.

Tabela 6: Primerjava razvrstitve kolokacijskih struktur.

	Uporabniška razvrstitev	Uporabniški priklic po spominu	Število kolokacij v obravnavani strukturi	Skupna frekvenca kolokacij v korpusu
BELINA				
nežna belina	1. mesto	1. mesto	1. mesto (21)	1. mesto (805)
simbol beline	2. mesto	-	2. mesto (19)	2. mesto (578)
belina snega	3. mesto	2. mesto	4. mesto (13)	3. mesto (339)
poudariti belino	4. mesto	-	6. mesto (4)	5. mesto (64)
odet v belino	5. mesto	-	5. mesto (6)	6. mesto (59)
drseti po belini	6. mesto	-	3. mesto (15)	4. mesto (98)
IZOLIRATI				
toplotno izolirati	1. mesto	2. mesto	3. mesto (28)	1. mesto (891)
izolirati stene	2. mesto	1. mesto	1. mesto (54)	2. mesto (469)
uspeti izolirati	3. mesto	-	4. mesto (13)	4. mesto (237)
izolirati od sveta	4. mesto	-	2. mesto (37)	3. mesto (322)

4.3 Izbira in razvrščanje posameznih kolokacij

V nadaljevanju ankete je sledilo sedem nalog, pri katerih so sodelujoči s seznama kolokacij izbrali tiste, ki po njihovem mnenju najbolj sodijo na prvo mesto v slovarju (V6–V12 v Tabeli 3). Sodelujoči so za vsako od danih struktur izbrali po pet kolokacij in jih razvrstili. Že pred reševanjem nalog so bili tudi obveščeni, da bodo po razvrščanju kolokacij opredelili kriterije za svoje odločitve.

Rezultati razvrščanja so v celoti na voljo na projektni spletni strani. V prispevku navajamo samo najbolj kolokacijsko bogate strukture in tiste kolokacije, ki so jih sodelujoči izbrali za prva mesta v slovarju, torej kolokacijski jagodni izbor po mnenju sodelujočih.²⁰ V Tabeli 7 in Tabeli 8 je pri vsaki posamezni kolokaciji navedeno, kolikokrat so jo sodelujoči izbrali za prvo, drugo, tretje, četrto ali peto mesto v slovarskem prikazu. Izbor primerjamo s kolokacijami, ki so jih sodelujoči najpogosteje navajali po spominu ter z jagodnima izboroma, ki ju ponujata razvrstitev glede na logDice in korpusno pogostnost.

Tabela 7: Jagodni izbor kolokacij s samostalnikom *belina* po mnenju sodelujočih v anketi.

Kolokacije	1	2	3	4	5	Skupaj	Najpogosteje po spominu	Top 5 glede na logDice	Top 5 glede na pogostnost
pridevnik + BELINA									
snežna belina	261	97	46	34	24	462	da	da	da
bleščeča belina	107	115	66	48	39	375	da	ne	da
brezhibna belina	45	50	46	44	45	230	da	ne	ne
naravna belina	21	43	66	55	38	223	ne	ne	ne
sijoča belina	7	30	58	42	63	200	da	ne	ne
BELINA + samostalnik									
belina zob	75	119	89	63	43	389	da	da	da
belina snega	193	73	49	31	33	379	da	da	da
belina perila	40	43	53	55	50	241	da	da	ne
belina papirja	25	34	37	47	46	189	da	ne	da
belina kože	39	32	40	35	35	181	ne	ne	ne
samostalnik + BELINA									
odtenek beline	105	69	50	37	28	289	ne	ne	ne
kanček beline	98	42	53	41	29	263	ne	ne	ne
videz beline	44	52	48	43	39	226	ne	ne	ne
določanje beline	49	47	38	35	33	202	ne	ne	da
nastavitev beline	48	46	34	26	25	179	ne	da	da
glagol + predlog + BELINA									
odeti se v belino	87	77	51	33	17	265	ne	da	da

²⁰ Navajamo po 5 kolokacij, ki so jih sodelujoči najpogosteje izbrali za prioriteten prikaz v slovarju. V primeru, ko za peto mesto tekmujeta dve zvezi, smo v tabelo vključili obe.

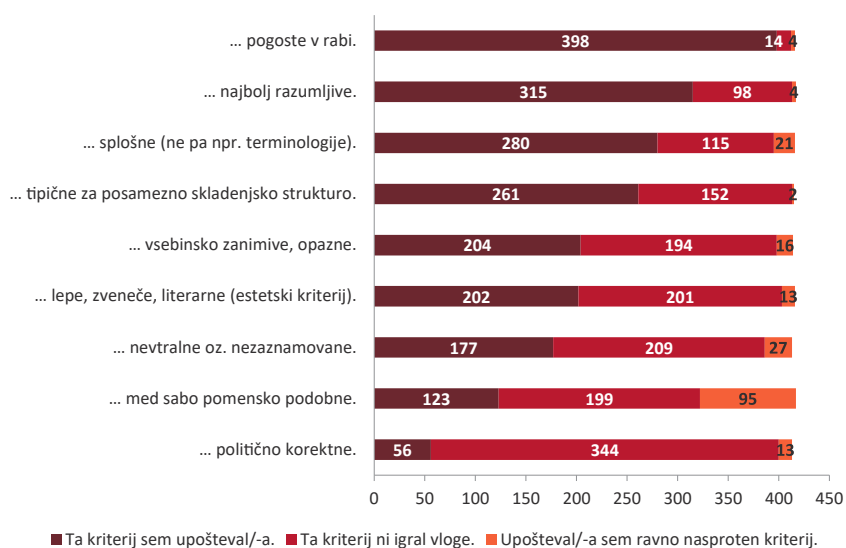
Kolokacije	1	2	3	4	5	Skupaj	Najpogostejše po spominu	Top 5 glede na logDice	Top 5 glede na pogostnost
ziti se z belino	18	33	39	58	70	218	ne	ne	ne
bleščati se od beline	110	25	21	20	23	199	ne	da	ne
zazreti se v belino	23	30	52	58	34	197	ne	ne	da
izginjati v belini	16	32	28	30	40	146	ne	ne	ne
potopiti se v belino	19	31	31	29	36	146	ne	ne	ne
Skupaj 'Da'							8 od 21	7 od 21	9 od 21

Tabela 8: Jagodni izbor kolokacij z glagolom *izolirati* po mnenju sodelujočih v anketi.

Kolokacije	1	2	3	4	5	Skupaj	Najpogostejše po spominu	Top 5 glede na logDice	Top 5 glede na pogostnost
prislov + IZOLIRATI									
zvočno izolirati	65	114	73	72	50	374	da	da	da
toplotno izolirati	122	91	72	59	28	372	da	da	da
socialno izolirati	17	36	67	57	70	247	ne	da	ne
dobro izolirati	114	29	31	33	24	231	da	ne	da
od dodatno izolirati	35	60	36	33	41	205	ne	ne	ne
IZOLIRATI + samostalnik									
izolirati bolnika	110	59	49	35	36	289	da	ne	ne
izolirati hišo	88	62	47	46	23	266	da	ne	da
izolirati fasado	39	45	35	19	20	158	da	ne	ne
izolirati človeka	43	31	27	18	25	144	da	ne	ne
izolirati stene	14	29	21	33	27	124	da	da	da
IZOLIRATI + predlog + samostalnik									
izolirati od družbe	67	66	39	25	24	221	ne	ne	da
izolirati pred vlago	60	40	30	40	46	216	ne	ne	ne
izolirati od sveta	36	47	41	52	39	215	ne	da	da
izolirati od ljudi	29	48	49	47	26	199	ne	ne	ne
izolirati pred hrupom	46	47	39	22	20	174	ne	ne	ne
Skupaj 'Da'							8 od 15	5 od 15	7 od 15

Kot velja za predhodne rezultate, se tudi pri jagodnem izboru kolokacij korpusna pogostnost izkaže kot mera, ki je (nekoliko)

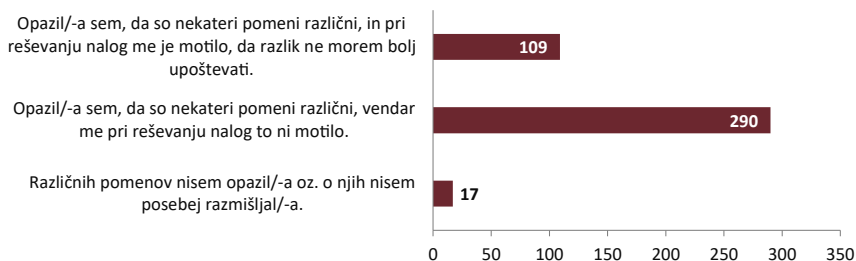
skladnejša s preferencami sodelujočih v anketi, kar je mogoče v nadaljevanju testirati na večji količini gradiva in uporabiti pri razvrščanju podatkov v slovarju. Da je pogostnost v jezikovni rabi – oziroma vtis sodelujočih glede pogostnosti – dejansko igrala pomembno vlogo pri razvrščanju, pa povsem neposredno pokažejo odgovori na vprašanje v anketi, kjer so sodelujoči opredelili kriterije, po katerih so kolokacije izbirali. Kot kaže Slika 4, so skoraj vsi sodelujoči opredelili, da so izbirali kolokacije, ki so pogoste v rabi. Sledijo kriteriji razumljivosti, splošnosti in tipičnosti.



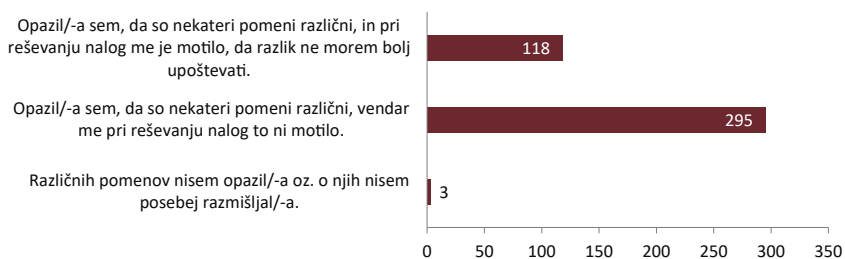
Slika 4: »Pri razvrščanju sem izbiral/-a kolokacije, ki so ...« (N = 417).

Ker v anketi kolokacijski podatki niso bili urejeni glede na različne pomen besed *belina* in *izolirati* – takšno je namreč tudi trenutno stanje pri večini gesel v Kolokacijskem slovarju sodobne slovenščine – smo na koncu ankete sodelujoče vprašali, koliko jih je pri razvrščanju motilo, da pomenov ne morejo upoštevati. Rezultati, ki jih kažeta Sliki 5 in 6, so primerljivi za obe obravnavani besedi. Večina sodelujočih je razlike v pomenu opazila, moteče so bile za približno tretjino in nemoteče za približno dve tretjini sodelujočih. Rezultate

je mogoče uporabiti za nadaljnje analize, zlasti v luči temeljne značilnosti odzivnih slovarjev, da v prvi, avtomatsko pripravljene fazi informacij, ki zahtevajo ročni leksikografski pregled, po večini ne prinašajo.



Slika 5: Ne/upoštevanje pomenov pri samostalniku *belina* (N = 416).



Slika 6: Ne/upoštevanje pomenov pri glagolu *izolirati* (N = 416).

4.4 Anketni vzorec

Na koncu ankete smo zbrali še metapodatke o sodelujočih v raziskavi. Kot omenjeno (razdelek 3), je kompleksnost ankete povzročila precejšen osip sodelovanja. Za primerjalne statistične analize bo zato mogoče uporabiti približno 415 anket, ki jih predstavljamo v nadaljevanju.

Kot kažejo Tabele 9, 10 in 11, je večina sodelujočih starih med 26 in 55 let, z dokončano univerzitetno izobrazbo, zaposlenih v javnem sektorju, samozaposlenih ali študentov.

Tabela 9: »V katero starostno kategorijo spadate?« (N = 417)

Odgovori	Frekvenca
1 (do 15 let)	0
2 (od 16 do 25 let)	62
3 (od 26 do 35 let)	102
4 (od 36 do 45 let)	119
5 (od 46 do 55 let)	82
6 (od 56 do 65 let)	42
7 (66 let ali več)	10

Tabela 10: »Kakšen je vaš status oziroma glavno področje angažiranja (dela, zaposlitve)?« (N = 416)

Odgovori	Frekvenca
1 (dijak/-inja)	1
2 (študent/-ka)	69
3 (samozaposlen/-a)	63
4 (zaposlen/-a v javnem sektorju (uprava, šolstvo, zdravstvo, sociala, kultura ...))	194
5 (zaposlen/-a v neprofitnem sektorju (društva, združenja ...))	8
6 (zaposlen/-a v podjetju)	43
7 (vzdrževanje gospodinjstva)	2
8 (upokojen/-a)	18
9 (v iskanju zaposlitve)	13
10 (na porodniškem ali starševskem dopustu)	2
11 (Drugo:)	3

Tabela 11: »Kakšna je vaša najvišja dosežena izobrazba?« (N = 415)

Odgovori	Frekvenca
1 (Nedokončana osnovna šola)	0
2 (Osnovna šola)	2
3 (Poklicna šola)	1
4 (Štiriletna srednja šola)	50
5 (Višja šola)	9
6 (Visokošolski strokovni študij)	1

Odgovori	Frekvenca
7 (Visoka šola)	9
8 (Univerzitetni študij)	240
9 (Magisterij)	49
10 (Doktorat)	50
11 (Specializacija)	4

V zadnjem vprašanju so sodelujoči opredelili, v kakšnem kontekstu se srečujejo z jeziko(slo)vnimi vprašanji, izbrali so lahko več odgovorov. Prednjačijo posameznice in posamezniki, ki se ukvarjajo z lektoriranjem ali/in prevajanjem, velik delež je tudi aktivnih v izobraževanju in piscev različnih vrst besedil. Podatki kažejo, da je zajem zelenih ciljnih skupin (gl. razdelek 3) ustrezen in da so podatki primerni za nadaljnje primerjalne statistične analize.

Tabela 12: »Če igra, je igral ali bo igral pri vašem delu jezik posebno vlogo, označite vse ustrezajoče kategorije.« (N = 405)

Odgovor (možnih je bilo več izbir)	Frekvence	% – Veljavni (N = 405)
Lektoriranje	223	55 %
Prevajanje	193	48 %
Poučevanje slovenščine kot 1. jezika – osnovna šola	69	17 %
Poučevanje slovenščine kot 1. jezika – srednja šola	46	11 %
Poučevanje slovenščine kot 2. ali tujega jezika	55	14 %
Predavanje jezikoslovnih predmetov na višji / univerzitetni ravni	38	9 %
Novinarstvo	38	9 %
Marketing	37	9 %
Pravo	10	2 %
Bibliotekarstvo	29	7 %
Javna uprava	40	10 %
Administracija	37	9 %
Menedžment	20	5 %
Beletristika	29	7 %
Kreativno pisanje, blogerstvo	84	21 %

Odgovor (možnih je bilo več izbir)	Frekvence	% – Veljavni (N = 405)
Strokovno in znanstveno pisanje	105	26 %
Jezikoslovne raziskave	68	17 %
Leksikografija	29	7 %
Ljubiteljsko raziskovanje jezika	99	24 %
Drugo:	20	5 %

5 Zaključek in prihodnje delo

Raziskava potrjuje, da se je pri razvoju Kolokacijskega slovarja sodobne slovenščine smiselno posvetiti razvrščanju kolokacij glede na relevantnost za uporabniško skupnost in opredelitvi parametrov za prikaz kolokacijskega »jagodnega izbora«: kolokacijskih struktur in kolokatorjev, ki omogočijo hiter vpogled v besedišče, ki je po izbranih kriterijih najbolj relevantno, in ponujajo prvi vtis o (sicer količinsko zelo obsežnem) kolokacijskem gradivu. Rezultati kažejo, da preference oz. preferenčne tendence med sodelujočimi v anketi obstajajo in so metodološko opredeljive. Delo s specifičnimi kolokacijami je sodelujočim v raziskavi omogočilo vpogled v konkretne podatke in urejevalno izkušnjo z njimi, kar je bil predpogoj za zanesljivejše opredeljevanje kriterijev in želja za razvrščanje kolokacij v slovarju. Identificirane preference (Slika 4) so torej posledica konkretne izkušnje z gradivom, kar je močna točka raziskave, na drugi strani pa tako natančna in kvalitativna poglobitev v gradivo (lahko) zajame izredno omejen nabor iztočnic. Izsledke raziskave je zato mogoče uporabiti zlasti za identifikacijo perečih raziskovalnih vprašanj in oblikovanje raziskovalnih hipotez, kot sledi v nadaljevanju.

Izkazuje se, da so – vsaj za obravnavane podatke – preference sodelujočih prekrivnejše s korpusno pogostnostjo kot kolokacijsko jakostjo. V nadaljevanju je mogoče to hipotezo preveriti na večji količini gradiva in posledično podatke urediti po pogostnosti, kot je bilo denimo preizkušeno pri pripravi baze za Estonski kolokacijski slovar (Kallas idr. 2015). Pogostnost je mogoče vključiti kot del kombiniranega pristopa, ki upošteva mesto kolokacije v seznamu, urejenem

tako glede na statistično jakost kot tudi pogostnost, kot je bilo pri luščenju kolokacij za slovenščino preizkušeno v Gantar idr. (2016: 215). Kombinirani pristopi na ravni luščenja so posledica spoznanja, da zanašanje na eno samo statistično mero povzroči izpad določenih relevantnih kolokacij. Vprašanje, ki se ga dotikamo v pričujočem prispevku pa je, kako mere kombinirati tudi pri urejanju (karseda celovito izluščenih) podatkov, da na prva mesta v slovarju pridejo kolokacije, ki jih uporabniki tam želijo in pričakujejo.

Rezultati kažejo obstoj kolokacij, ki jih po spominu navaja večji delež sodelujočih: za dani iztočnici v uporabniških odgovorih močno izstopajo kolokacije *snežna belina*, *čista belina* in *izolirati hišo*, nato pogostnost odgovorov strmo pada. Primerljive podatke, vendar za veliko večji nabor iztočnic, bo mogoče pridobiti s pomočjo aplikacije Igra besed. Aplikacija podpira model odzivnega slovarja z elementom igrifikacije, ki je namenjena čiščenju in urejanju avtomatsko pridobljenega kolokacijskega ter sopomenskega gradiva (Arhar Holdt idr. 2020). V enem od igralnih modulov igralke in igralci po spominu vpisujejo kolokacije za izbrane iztočnice, kar bo omogočilo nadaljnje analize v prispevku zastavljenega vprašanja. Hipoteze, ki jih postavlja pričujoča raziskava, so: v jeziku obstajajo kolokacije, ki jih po spominu navaja velik delež jezikovne skupnosti; korpusna frekvenca je boljši napovedovalec teh primerov kot statistična jakost; tovrstno besedišče je skoraj vedno splošno, ne terminološko; v vrhu po pogostnosti je manjši nabor primerov, tj. eden ali dva.

Primerjava po spominu naštetih kolokacij s korpusno pridobljenimi podatki posredno evalvira izbrane frekvenčne prage in mere, ki so kriterij za umestitev kolokacije v slovar. Primeri, kot je po spominu pogosto navajana *zimsko belina*, se pojavljajo v korpusu, ne pa v slovarju, zato so lahko indikator za spust frekvenčnega praga za vključitev gradiva. Na drugi strani podatki jasno osvetlijo problematičnost terminologije v referenčnem korpusu: na eni strani so v korpusu redke, vendar po spominu pogosto navajane zveze, kot npr. *izolirati* [*žico, kabel, električno*], na drugi strani se v slovarju na najbolj vidnih mestih pojavljajo terminološke kolokacije, kot npr. [*popravljen, nenatančen*] *belina*, ki jih sodelujoči po spominu ne

naštejejo (pa tudi v primeru, ko jih vidijo pred sabo, ne izbirajo za prva mesta slovarja).²¹

Preferenčne tendence sodelujočih se kažejo tako na ravni posameznih kolokatorjev kot kolokacijskih struktur. Pri slednjih je zanimivo, da v nasprotju s korpusnimi statistikami sodelujoči za prva mesta slovarja redko izbirajo predložne zveze: preferenca so dodelne zveze s polnopomenskimi elementi. Podatki tudi pokažejo, da je za obravnavani iztočnici skupna frekvenca kolokacij v korpusu skladnejša s preferencami sodelujočih kot število kolokacij v posamezni strukturi, kar je pri nadgradnji slovarja mogoče uporabiti za (dinamično) prioretiziranje struktur znotraj slovarskih iztočnic.

Tudi pri razvrščanju posameznih kolokacij se korpusna pogostnost izkaže kot mera, ki je skladnejša s preferencami sodelujočih, še bolj pa v prid upoštevanju pogostnosti priča dejstvo, da so jo skoraj vsi sodelujoči opredelili kot kriterij, po katerem so kolokacije izbirali. Pri tem je seveda treba razumeti, da gre za *predvideno* pogostnost, ki je glede na rezultate raziskave tesno povezana s splošnostjo, kot je bilo izpostavljeno že zgoraj. Poudariti je treba, da takšna homogenost pri izbiri kateregakoli od ponujenih kriterijev ni bila pričakovana, je pa zato toliko bolj pomembna za nadaljnje premisleke. Rezultat namreč posredno nakazuje, da sodelujoči jagodnega izbora ne vidijo kot didaktično priložnost za usvajanje manj znanega besedišča, ampak se preferenca pravzaprav bliža korpusno osnovanemu jezikovnemu opisu, ki na prvo mesto postavlja pogoste (tudi razumljive, splošne – lahko bi torej rekli *znane*) kolokacije, s tem pa (najbrž) kompenzira tudi za pomenske informacije, ki jih slovar trenutno prinaša le mestoma. Redkejša, zaznamovane kolokacije so posledično stvar usmerjenega iskanja po slovarju, bodisi na klik prek jagodnega izbora ali pa neposredno prek iskalnega okenca. Dokaj intuitivno se zdi, da uporabniki za izhodišče želijo empirično določeno, stabilno,

21 Vključitev terminološkega gradiva je seveda povezana s slovaropisnimi odločitvami in pristopi. Za Kolokacijski slovar sodobne slovenščine je značilno, da nastane v prvem koraku strojno, nato se ročno nadgrajuje. Gesla, ki imajo ročno urejeno pomensko členitev (kot velja za geslo *belina*), zahtevajo ponazoritev vsakega identificiranega pomena – tudi terminološkega – s kolokacijami, kar (lahko) dodatno spodbudi vključitev v korpusu redkejšega gradiva v slovar.

prototipsko podatkovno jedro, možnosti za nadgradnjo jezikovnega znanja (svojega ali v znanja učencev) pa raje naslavlja sami, saj so te potrebe individualne, jezikovna vprašanja pa kontekstno odvisna. Pri nadaljnjem razvoju bi bilo to mogoče upoštevati in uporabnikom ponuditi več kontrole nad urejanjem podatkov za didaktične namene, bodisi v slovarju bodisi v specializiranih orodjih v podporo jezikovnemu razvoju. Vendar bi bilo v prvem koraku koristno zagotoviti dodatne uporabniške raziskave, mdr. empirično definirati različne scenarije uporabe Kolokacijskega slovarja sodobne slovenščine.

Kljub kompleksnosti vprašalnika ocenjujemo, da je za dani namen dobro zasnovan in pridobljeni podatki primerni za nadaljnje statistične analize. Za veljavnost raziskave je bilo nujno zagotoviti, da sodelujoči ustrezno interpretirajo pojem kolokacije. Anketa vključuje preverbo razumevanja dveh vrst, kar se je izkazalo za ustrezno. Pregled metapodatkov kaže, da so bile ciljne uporabniške skupine pri anketiranju ustrezno zajete in da podatki omogočajo primerjalne analize za cca. 415 anket. V nadaljevanju bi bilo dragoceno statistično analizirati in opredeliti morebitne korelacije med kriteriji za razvrščanje ter starostjo, statusom ali poklicnim interesom sodelujočih. Takšen vpogled bi omogočil boljše razumevanje preferenc in potreb posameznih uporabniških skupin in zagotovil morebitne adaptacije podatkovnega prikaza tudi na ravni slovarskega vmesnika.

Zahvala

V prispevku so opisani rezultati, ki so nastali v okviru projekta *Kolokacije kot temelj jezikovnega opisa: semantični in časovni vidiki* (J6-8255) ter programskih skupin P6-0411 – *Jezikovni viri in tehnologije za slovenski jezik* in P6-0215 – *Slovenski jezik – bazične, kontrastivne in aplikativne raziskave*, ki jih financira Javna agencija za raziskovalno dejavnost Republike Slovenije.

Reference

Arhar Holdt, Š., Kosem, I. in Gantar, P. (2016): Dictionary user typology: the Slovenian case. V T. Margalitadze in G. Meladze (ur.): *Lexicography and linguistic diversity: proceedings of the XVII EURALEX International*

- Congress: 179–187*. Tbilisi: Ivane Javakhishvili Tbilisi State University. Dostopno prek: <https://euralex.org/publications/dictionary-user-typology-the-slovenian-case/> (9. 3. 2021).
- Arhar Holdt, Š., Logar, N., Pori, E. in Kosem, I. (2020): "Game of Words": Play the Game, Clean the Database. V Z. Gavriilidou, M. Mitsiaki in A. Fliatouras (ur.): *Lexicography for inclusion: EURALEX XIX: Congress of the European Association for Lexicography: Vol. 1*: 41–49. Dostopno prek: https://euralex2020.gr/wp-content/uploads/2020/11/EURALEX2020_ProceedingsBook-p041-049.pdf (9. 3. 2021).
- 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.
- Bergenholtz, H. (2011): Access to and presentation of needs-adapted data in monofunctional internet dictionaries. V P. A. Fuertes-Olivera in H. Bergenholtz (ur.): *E-lexicography: the Internet, digital initiatives and lexicography*: 30–45. London in New York: Continuum.
- Bogaards, P. (2003): Uses and users of dictionaries. V P. Van Sterkenburg (ur.): *A Practical Guide to Lexicography*: 26–33. Amsterdam in Philadelphia: John Benjamins.
- Fuertes-Olivera, P. A. in Tarp, S. (2014): *Theory and practice of specialised online dictionaries: Lexicography versus terminography*. Berlin in New York: de Gruyter.
- Gantar, P. (2015): *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. <https://doi.org/10.4312/9789612377922>.
- Gantar, P., Kosem, I. in Krek, S. (2016): Discovering automated lexicography: the case of Slovene lexical database. *International journal of lexicography*, 29 (2): 200–225. <https://doi.org/10.1093/ijl/ecw014>.
- Gorjanc, V., Gantar, P., Kosem, I. in Krek, S. (ur.) (2017): *Slovar sodobne slovenščine: problemi in rešitve* (1. izd.). Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Dostopno prek: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/book/15> (9. 3. 2021).
- Groves, M. R., Fowler, F. J. Jr., Couper, M. P., Lepkowski, J. M., Singer, E. in Tourangeau, R. (2004): *Survey methodology*. Hoboken (NJ): J. Wiley.
- Kallas, J., Kilgarrieff, A., Koppel, K., Kudritski, E., Langemets, M., Michelfeit, J., Tuulik, M. in Viks, Ü. (2015): Automatic generation of the Estonian

- Collocations Dictionary database. V I. Kosem, M. Jakubiček, J. Kallas in S. Krek (ur.): *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 Conference*: 11–20. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd. Dostopno prek: https://elex.link/elex2015/proceedings/eLex_2015_01_Kallas+etal.pdf (9. 3. 2021).
- Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J. in Laskowski, C. A. (2018): Collocations dictionary of modern Slovene. V J. Čibej, V. Gorjanc, I. Kosem in S. Krek (ur.): *Proceedings of the 18th EURALEX International Congress: lexicography in global contexts*: 989–997. Ljubljana: Ljubljana University Press, Faculty of Arts. Dostopno prek: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1> (9. 3. 2021).
- Kosem, I., Gantar, P., Krek, S., Arhar Holdt, Š., Čibej, J., Laskowski, C., Pori, E., Klemenc, B., Dobrovoljc, K., Gorjanc, V. in Ljubešič, N. (2019): *Collocations Dictionary of Modern Slovene KSSS 1.0*. Slovenian language resource repository CLARIN.SI, <http://hdl.handle.net/11356/1250>.
- Kosem, I., Lew, R., Müller-Spitzer, C., Ribeiro Silveira, M., Wolfer, S., Dorn, A., Gurrutxaga, A., ... Nesi, H. (2018). The image of the monolingual dictionary across Europe: Results of the European survey of dictionary use and culture. *International Journal of Lexicography*, 32 (1): 92–114. doi: 10.1093/ijl/ecy022.
- Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešič, N., Kosem, I. in Dobrovoljc, K. (2020): Gigafida 2.0: the reference corpus of written standard Slovene. V N. Calzolari (ur.): *Twelfth International Conference on Language Resources and Evaluation (LREC 2020)*: 3340–3345. Paris: ELRA – European Language Resources Association. Dostopno prek: <http://www.lrec-conf.org/proceedings/lrec2020/LREC-2020.pdf> (9. 3. 2021).
- Lew, R. in de Schryver, G.-M. (2014): Dictionary users in the digital revolution. *International Journal of Lexicography*, 27 (4): 341–359.
- Logar Berginc, N. (2009): Slovenski splošni in terminološki slovarji: za koga? V M. Stabej (ur.): *Infrastruktura slovenščine in slovenistike. Obdobja 28*: 225–231. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani. Dostopno prek: <https://centerslo.si/simpozij-obdobja/zborniki/obdobja-28/> (9. 3. 2021).
- Logar, N., Grčar, M., Brakus, M., Erjavec, T., Arhar Holdt, Š. in Krek, S. (2012): *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES*:

- gradnja, vsebina, uporaba* (1. izd.). Ljubljana: Trojina, zavod za uporabno slovenistiko: Fakulteta za družbene vede.
- Lorentzen, H. in Theilgaard, L. (2012): Online dictionaries – how do users find them and what do they do once they have? V R. Vatvedt Fjeld in J. M. Torjusen (ur.): *Proceedings of the 15th EURALEX International Congress*: 654–660. Oslo: Universitetet i Oslo, Institutt for lingvistiske og nordiske studier.
- Müller-Spitzer, C. (ur.) (2014): *Using Online Dictionaries*. Berlin in Boston: De Gruyter Mouton.
- Pori, E., Čibej, J., Kosem, I. in Arhar Holdt, Š. (2020): The Attitude of Dictionary Users towards Automatically Extracted Collocation Data: A User Study. V I. Kosem in P. Gantar (ur.): *Kolokacije v leksikografiji: trenutne rešitve in izzivi za prihodnost [tematska številka]*. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*: 8 (2): 168–201. doi: 10.4312/slo2.0.2020.2.168-201.
- Pori, E., Čibej, J., Kosem, I. in Arhar Holdt, Š. (2021): Evalvacija uporabniškega vmesnika Kolokacijskega slovarja sodobne slovenščine. V I. Kosem (ur.): *Kolokacije v slovenščini*: 235–268. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Rychlý, P. (2008): A Lexicographer-Friendly Association Score. V P. Sojka in A. Horák (ur.): *RASLAN 2008: Recent Advances in Slavonic Natural Language Processing: Second Workshop in Recent Advances in Slavonic Natural language Processing*: 6–9. Brno: Masarykova Univerzita.
- Stabej, M. (2009): Slovarji in govorcji: kot pes in mačka? *Jezik in slovstvo*, 54 (3–4): 115–138.
- Tarp, S. (2009): Reflections on Lexicographical User Research. *Lexikos*, 19 (1): 275–296.
- Welker, H. A. (2013a): Methods in Research of Dictionary Use. V R. H. Gouws, U. Heid, W. Schweickard in H. E. Wiegand (ur.): *Recent Developments with Focus on Electronic and Computational Lexicography [dodatna številka]*. *Dictionaries. An International Encyclopedia of Lexicography*: 540–547. Berlin, New York: De Gruyter Mouton.
- Welker, H. A. (2013b): Empirical Research into Dictionary Use since 1990. V R. H. Gouws, U. Heid, W. Schweickard in H. E. Wiegand (ur.): *Recent Developments with Focus on Electronic and Computational Lexicography [dodatna številka]*. *Dictionaries. An International Encyclopedia of Lexicography*: 531–540. Berlin, New York: De Gruyter Mouton.