

V ISKANJU SLOVARSKO RELEVANTNE KOLOKACIJE NA PRIMERU STRUKTUR S PRISLOVI

Eva PORI

Filozofska fakulteta Univerze v Ljubljani

Iztok KOSEM

Filozofska fakulteta Univerze v Ljubljani

Pori, E., Kosem, I. (2018): V iskanju slovarsko relevantne kolokacije na primeru struktur s prislovi. Slovenščina 2.0, 6 (2): 154–185.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.154-185>.

Prispevek na osnovi analiz kolokacijsko produktivnih struktur s prislovi prikaže, kje in katera so problematična mesta posamezne prislovne strukture, ki so potrebna natančnejše jezikoslovne obravnave in diskusije. Analize jezikovne evalvacije označevalcev kolokacijskih struktur s prislovi predstavljajo model jezikoslovnega evalviranja vseh kolokacijsko produktivnih struktur znotraj procesa izdelave Kolokacijskega slovarja sodobne slovenščine in pomenijo izhodišče za nadaljnjo razpravo o kolokativnosti oz. opredelitvi slovarsko relevantnih kolokacij ter za izboljšave in nadgradnjo avtomatskega luščenja kolokacijskih kandidatov. Raziskava, ki se osredotoča na doslej premalo raziskano področje besednih zvez s prislovi, je izrazito leksikogramatična in kot taka relevantna pri razmisleku o sodobnem slovničnem jezikovnem opisu.

Ključne besede: leksikografija, semantika, kolokativnost, kolokacijski slovar sodobne slovenščine, prislov

1 UVOD

Pojma kolokacija in kolokativnost sta v jezikoslovju znana že več kot pol stoletja, če ne drugače, vsaj po slavni in večkrat citirani izjavi J. R. Firtha (1957): »Besedo spoznaš po njeni okolici«. ¹ Ključni poudarek izjave je v tem, da sta pomen in pojavnost besede odvisna od konteksta. Tovrstni pristop lahko

¹ »You shall know a word by the company it keeps.«

zasledimo v Hallidayevi funkcijski slovnici (1966), Melčukovih leksikalnih funkcijah (1996, 1998), Sinclairjevih na korpusu temeljčih kolokacijskih študijah (1987, 1991), slovnici vzorcev (Hunston in Francis 2000) in teoriji leksikalnega proženja (Hoey 2005).

Jezikoslovci pri opredeljevanju kolokacij uporabljajo različne kriterije, od statističnega (Firth 1957; Halliday 1961; Benson in dr. 1986; Sinclair in dr. 2004 etc.), skladijskega (Cowie 1994), leksikalnega (Firth 1957; Halliday 1961; Hausmann 1984; Sinclair in dr. 2004 etc.) in semantičnega (Howarth 1996, Melčuk 1998). Omenjeni kriteriji niso izključujoči – podrobna analiza dejansko pokaže, da so pogosto prekrivni oziroma se medsebojno dopolnjujejo.

Za naše izhodišče lahko vzamemo opredelitev Bartsch (2004), ki pravi, da so kolokacije leksikalno in/ali pragmatično povezane sopojavitve vsaj dveh leksikalnih enot, ki sta med seboj v neposrednem skladijskem razmerju, pri čemer je treba dodati, da se sopojavnost identificira na podlagi določene statistične vrednosti² (in ne zgolj pogostosti).

Kolokacije imajo zelo pomembno vlogo v jezikovnem opisu, saj so kolokatorji pogosto uporabljeni kot izhodišče pri identifikaciji pomenov, poleg tega pa so ključni pri oblikovanju pomenskih opisov. Z vidika pomena je semantična preferenca najrelevantnejša od vseh ravni abstrakcije razmerij med leksikalnimi enotami,³ saj pomeni izhajajo iz leksikalnih nizov pogostih kolokatorjev, ki jim je skupna določena semantična lastnost (Stubbs 2002: 449). Nedavni trendi v leksikografiji največ pozornosti posvečajo prav avtomatizaciji tistih segmentov jezikovnega opisa, ki so povezani s kolokacijami

² Poznamo različne statistične metode za merjenje kolokacijske jakosti, npr. t-score, MI (razmerje vzajemnosti), MI3, log-likelihood (logaritemska verjetnost), Diceov koeficient itd. (Manning in Schütze 1999).

³ Prva raven abstrakcije razmerij med leksikalnimi enotami so kolokacije; ostale tri ravni so še koligacija (razmerje med besedo in slovnimi kategorijami), semantična preferenca (semantični nizi kolokatorjev) in semantična prozodija (konotativni pomeni besede pri rabi s tipičnimi kolokatorji; Sinclair 1996, 1998).

in zgledi (npr. Kilgarriff in Rychlý 2010; Rundell in Kilgarriff 2011). Kot poudarjajo Cook in dr. (2013: 50), je »bistvena ugotovitev dosedanjih raziskav na tem področju ta, da avtomatizacija postopkov ne samo skrajša postopek leksikalne analize, ampak tudi izboljša njeno kakovost«.

Ravno na področju leksikografije je bilo v preteklih letih veliko narejenega z vidika identifikacije kolokacij v slovenskem jeziku. Tu velja izpostaviti postopke za avtomatsko luščenje kolokacij in njihovih zgledov (gl. Gantar in dr. 2015, 2016, Kosem in dr. 2013), ki predstavljajo temeljni del izdelave Slovarja sodobnega slovenskega jezika (Krek in dr. 2013, Gorjanc in dr. 2015) in Kolokacijskega slovarja sodobne slovenščine (Kosem in dr. 2018).⁴ S Kolokacijskim slovarjem sodobne slovenščine je tesno povezan raziskovalni projekt *Kolokacije kot temelj jezikovnega opisa: semantični in časovni vidiki* (KOLOS; J6-8255), ki se sicer osredotoča na temeljne raziskave kolokacij v slovenskem jeziku, je pa eden od njegovih ciljev tudi priprava metodoloških opisov za prenos izsledkov v (leksikografsko) prakso.

Kolokacijski slovar je dejansko najustreznejši vir za opredeljevanje slovarsko relevantnih kolokacij, vendar pa nam pri tem obstoječi uveljavljeni kolokacijski slovarji niso v veliko pomoč. Ti slovarji so bili namreč zasnovani za tiskano obliko, kar se pozna tako v omejenem naboru iztočnic kot obsegu gesel in struktur. Kot primer kolokacijskega slovarja, ki je temeljil na ozkih oz. strogih kriterijih za vključitev gradiva v slovar, navedimo angleški kolokacijski slovar Macmillian, pri katerem so po načelu, da naj ne bi imele statistično zelo relevantnih kolokatorjev, avtorji izločali iztočnice, kot je npr. *hiša, kupiti, dober*.⁵ Nove metode ter možnosti luščenja in predstavitve leksikalnih

⁴ Nov jezikovni vir za slovenščino Kolokacije 1.0: Kolokacijski slovar sodobne slovenščine je uradno objavljen in dostopen na <http://viri.cjvt.si/kolokacije/slv/>. Nastaja kot produkt skupnega projekta Centra za jezikovne vire in tehnologije Univerze v Ljubljani, Centra za uporabno jezikoslovje pri zavodu Trojina ter sodelavcev programske skupine »Slovenski jezik – bazične in kontrastivne raziskave« Filozofske fakultete Univerze v Ljubljani.

⁵ <http://www.macmillandictionaries.com/features/how-dictionaries-are-written/macmillan-collocations-dictionary/>

podatkov, ki jih je v leksikografijo prinesel tehnološki napredek, ter s tem povezane potrebe in navade uporabnikov, terjajo novo opredelitev kolokacij za leksikografske namene oz. njihovem podajanju uporabnikom.

Ob tem velja poudariti, da se pri opredeljevanju kolokacije za jezikovni opis neizogibno dotikamo tudi slovničnega opisa, saj gre za jezikovni fenomen leksikogramatične narave. Skladenjska komponenta je, poleg statistične in pomenske, namreč inherentni del same kolokacije. Ko govorimo o kolokacijah kot besednih sopojavitvah oz. zvezah besed, govorimo o besednozvezni skladnji ter s tem prehajamo na raven slovničnega fenomena.⁶ Lahko pa pričakujemo, da zaradi večje osredotočenosti na pomenski vidik prekrivnost ne bo vedno popolna – kar ne bo relevantno za slovarske namene, bo mogoče relevantno za slovnični opis.

Pričujoči prispevek najprej predstavlja metodo evalvacije kolokacijskih kandidatov, ki izkorišča pristope množičenja, in rezultate s tem povezane analize rezultatov na izbranem vzorcu 333 lem. V nadaljevanju se prispevek osredotoči na strukture s prislovi, ki smo jih uporabili kot testno množico za opredeljevanje slovarsko relevantnih kolokacij. Na konkretnih primerih so prikazana tako neproblematična kot problematična mesta posamezne prislovne strukture, s katerimi se tudi kaže pomembnost razlikovanja med statistično prepoznano in slovarsko relevantno kolokacijo. V sklepu povzamemo glavne ugotovitve raziskave in relevantnost opisane metode ter pripravljene podatkovne množice za ostale jezikoslovne in jezikovnotehnološke raziskave.

2 ANALIZE KOLOKACIJSKIH STRUKTUR

Raziskava, ki se osredotoča na pogosto premalo in slabo raziskano področje slovničnega opisa – skladenjske vzorce na ravni prislova oz. besednih zvez s prislovi, je bila izpeljana za leksikografske namene, odpira pa vprašanja, ki niso

⁶ Več o povezanosti pomenskih in slovničnih lastnosti leksikalnih jezikovnih elementov oz. o kolokacijah kot pomensko-skladenjskem pojavu gl. v Gantar 2015.

zanimiva zgolj leksikografsko, ampak so leksikogramatična in na ta način tudi slovnična, torej relevantna pri razmisleku o sodobnem slovničnem jezikovnem opisu.

2.1 Metodologija

Namen analiziranja kolokacijskih struktur je bil opredeliti kolokacijo kot semantično relevantno sopojavitev dveh (ali več) besed ter s tem tudi razliko med slovarsko relevantnimi kolokacijami in statistično prepoznanimi oz. šibkejšimi kolokacijami, ki ne opravljajo semantične funkcije in so posledično nerelevantne za kolokacijski slovar. Izhajali smo iz empiričnega induktivnega raziskovalnega pristopa od spodaj navzgor (od konkretnih primerov k posplošitvam), ki se ne podreja obstoječim jezikovnim vzorcem in konceptom, pač pa upošteva jezik kot naravni pojav in ga v skladu s tem opazuje v vseh segmentih njegove jezikovno-govorne pojavnosti.

Za analizo smo se odločili uporabiti metodo množičenja, ki je bila v slovenskem prostoru že uporabljena v leksikografskih projektih (Gorjanc in dr. 2015, Kosem in dr. 2018, Arhar Holdt in dr. 2018). Bistvo pristopa je, da je vsaka mikronaloga ločena enota, ki posamezniku ne sme vzeti veliko časa, dokončen pregled vseh rešenih mikronalog pa potem pokaže obseg medsebojnega ujemanja označevalcev pa tudi njihove interne doslednosti pri označevanju podatkov istega tipa, v našem primeru kolokacij določene skladske strukture. S pomočjo pilotne množičenjske naloge, v kateri so označevalci jezikoslovci – člani projektne skupine označevali avtomatsko izluščene kolokacijske kandidate iz korpusa Gigafida⁷ (Logar Berginc in dr. 2012), smo na ta način preverjali njihova stališča, kaj je in kaj ni slovarsko relevantna kolokacija. S pilotno nalogo smo želeli predvsem preveriti, ali lahko na podlagi ozkega nabora ponujenih odgovorov Da, Ne in Ne vem in nekih osnovnih navodil, s katerimi so označevalci ocenjevali kolokacijske kandidate, pridemo

⁷ <http://www.clarin.si/info/orodja-in-servisi/> Opomba se ne nadaljuje na naslednji strani? (opomba se nadaljuje na naslednji strani)

do jasnih opredelitev, kaj je slovarsko relevantna kolokacija.⁸

Naloga ocenjevanja kolokacijskih kandidatov se je odvijala v odprtokodni platformi za množičenjske naloge Pybossa.⁹ Pri vsaki nalogi so imeli označevalci na voljo kolokacijskega kandidata in njegov zgled, izluščen z orodjem GDEX za slovenščino (Kosem in dr. 2011; Kosem in dr. 2013; Kosem in dr. 2015), ki med drugim skuša identificirati zglede, ki kolokacijo prikazujejo v čim bolj tipičnem kontekstu. Označevalci so pri ocenjevanju kolokacijskih kandidatov lahko izbirali med ponujenimi možnostmi na seznamu oz. imeli na voljo 3 odgovore: DA, NE, NE VEM. Označevalcem je bila ponujena tudi podopcija odgovora DA, in sicer »DA (slab zgled)«, za katero naj bi se odločali v primerih, ko je bila kolokacija sicer legitimna, zgled pa neustrezen, predvsem zato, ker je bil nejasen oz. jezikovno ali pomensko slab. Označevalci so skupaj označili približno 8.800 kolokacijskih kandidatov v 226 različnih skladenjskih strukturah, pri čemer smo za vsakega od kandidatov zahtevali po 3 odgovore, kar je pomenilo, da vsi označevalci niso označili vseh kandidatov. Ujemanje označevalcev je bilo v razponu 42–76 %, v povprečju 62 % kolokacij sta se v odgovoru strinjala dva označevalca, Cohenova kapa je bila 0,35. Opazne so bile razlike med različnimi strukturami, tj. pri nekaterih strukturah so se označevalci precej bolj strinjali o tem, kaj je oziroma ni slovarsko relevantna kolokacija, kot pa pri drugih.

Po nalogi smo poleg analize podatkov opravili tudi razgovore z označevalci, ki so opozorili na različne pomanjkljivosti pristopa oz. naloge, izpostavljene pa so bile predvsem sledeče:

- premajhen nabor potencialnih odgovorov glede na obliko podatkov. Na odločitve označevalcev o legitimnosti kolokacije je namreč vplivala sama oblika, ki včasih ni ustrezala prevladujoči obliki, podani tudi v

⁸ Na tej točki smo bili tudi še odprti za možnost množičenja kolokacij med širšo javnostjo, če bi pilotna raziskava pokazala potencial za to.

⁹ <https://pybossa.com>

zgledu, npr. kolokator ni bil v množini;

- premajhna heterogenost iztočnic na račun širokega nabora skladskejskih struktur. Posledično ni bilo znano, kakšen vpliv imajo na opredeljevanje kolokacije različne lastnosti iztočnic, kot so večpomenskost, povratnost ipd.;
- vprašljivost vloge navodil. Označevalci so namreč komentirali, da bi bilo dejansko bolje označevati brez njih, na podlagi lastnih znanj in predstav o kolokacijah, ter se usklajevati kasneje;
- vsi podatki v eni nalogi. Označevalci so opozorili, da so morali biti zelo pozorni na preskoke na novo strukturo, ker informacija o strukturi ni bila nikjer eksplicirana.

V nadaljevanju smo tako posvetili več pozornosti pripravi nabora iztočnic, in sicer smo za zagotovitev večje reprezentativnosti in heterogenosti pri izbiri iztočnic uporabili različne kriterije (npr. besedna vrsta, večpomenskost, izvor, (ne)števnost, pogostost v korpusu Gigafida ipd.). Končni vzorec je vseboval 333 iztočnic, od tega 154 samostalnikov, 73 glagolov, 81 pridevnikov in 25 prislovov). Ocenjevanje kolokacijskih kandidatov se je ponovno odvijalo v platformi Pybossa, vendar tokrat niso bile vse strukture zajete v eni nalogi, pač pa je bila za vsako strukturo pripravljena ločena naloga. Poudarek novega eksperimenta je bil predvsem na tem, da je ocenjevanje kolokacijskih kandidatov temeljilo na lastnem pojmovanju kolokacije in da se kolokativnost (tako temeljno kot slovarsko) opredeli na podlagi analize rezultatov.

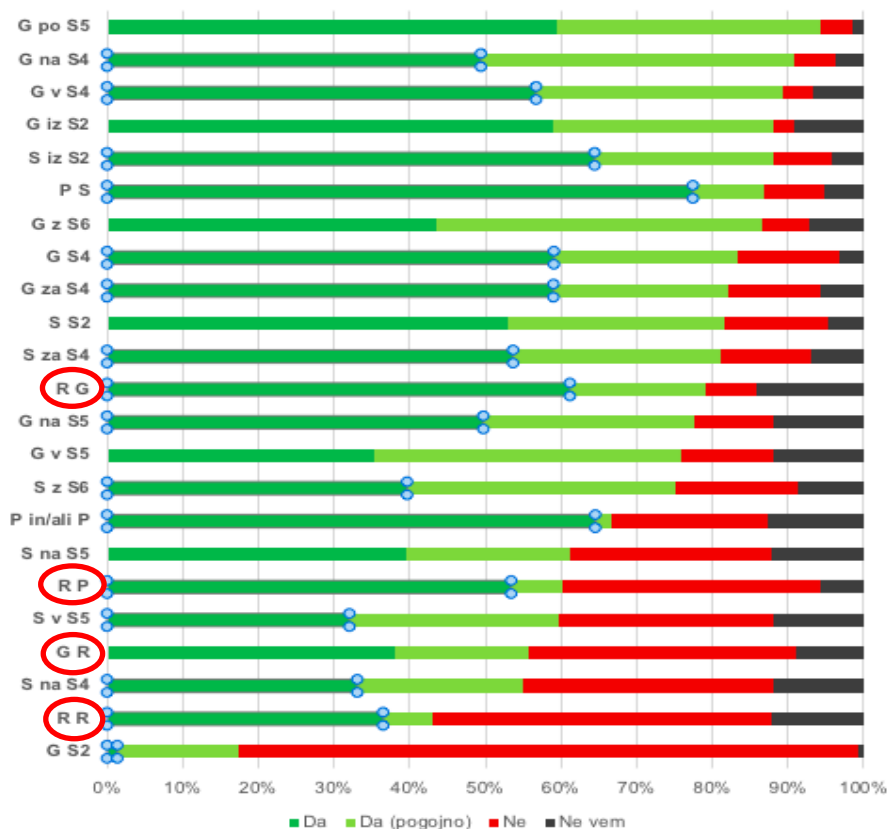
Označevalci jezikoslovci so še vedno izbirali med 3 krovnimi odgovori (DA, NE, NE VEM), a so jim bile ponujene podopcije:

- Množina (podopcija DA) za primere, ko je bila kolokacija sicer ustrezna, manjkal pa je kolokator v množini; npr. *tihotapljena cigareta* → *tihotapljene cigarete*.
- *Si/Se* (podopcija DA) pri glagolskih strukturah, ko je povratni osebni

ali svojilni zaimek obvezen, npr. *ogledati prestolnico* → *ogledati si prestolnico*.

- Največji (podopcija DA) pri pridevnikih in prislovih, ki so v kolokaciji vedno v primerniški ali presežniški obliki, npr. *znatno lahek* → *znatno lažji*.
- Razširjena kolokacija (podopcija DA), ki ob sebi predvideva dodaten element; npr. *dnevno brezplačno* → *4-krat dnevno brezplačno*.
- Zgled Ne, Kolokacija Morda za primere, ko zgled ne potrjuje kolokacije, čeprav je sama kolokacija videti povsem legitimna, npr. *doktorski študent* → *na doktorski (stopnji) pa 15 študentov*.
- Fraze, ko ne gre za kolokacijo, ampak za del fraze, npr. *ne mešati jabolk* → *ne mešati jabolk in hrušk*.
- Struktura (podopcija NE), za primere, kjer je šlo za napako pri oblikoskladenjskem označevanju korpusa (npr. prekrivnost prislova s pridevniško obliko: *medtem ko je **grobo mleti** sladkor najboljši*).

Skupno je bilo ocenjenih 17.576 kolokacijskih kandidatov v 143 različnih skladenjskih strukturah. Razporeditev odgovorov označevalcev glede na skladenjsko strukturo prikazuje Slika 2 (prikazanih je 23 struktur z največ kolokacijskimi kandidati).



Slika 2: Prikaz deležev odgovorov označevalcev glede na skladenjsko strukturo kolokacije.

Strukture z največjim deležem DA (vključno s podopcijami) so bile:

- glagol + [po] + samostalnik v mestniku (G po S5): *poseči po cigareti*;
- glagol + [na] + samostalnik v tožilniku (G na S4): *plezati na jambor*;
- glagol + [v] + samostalnik v tožilniku (G v S4): *prevesti v francoščino*;
- pridevnik + samostalnik (pbz0 + sbz0 = P S): *televizijska*

cenzura.

Strukture z največjim deležem NE in NE VEM pa so bile:

- glagol + samostalnik v rodilniku (gbz sbz2 = G S2): *primanjkovati goveda, angažirati izvedenca* (tožilnik, ne rodilnik!);
- prislov + glagol (rbz gbz = R G): *dolgo vreti, kako odrezati*;
- glagol + prislov (gbz rbz = G R): *redko obiskovati, boleti enako, prebiti tam*;
- prislov + prislov (rbz rbz = R R): *kje natančno, lahko brezplačno*;
- prislov + pridevnik (rbz pbz0 = R P): *vsestransko angažiran, dnevno sklenjen [promet]*.

Analiza je, kot pri pilotni nalogi, izpostavila različne ravni ujemanja med označevalci glede na skladijsko strukturo, tj. pri nekaterih strukturah so bila razhajanja precej večja, kar je nakazovalo na njihovo problematičnost z vidika opredeljevanja kolokativnosti. Slika 3 kaže podatke za deset skladijskih struktur z največ kolokacijskimi kandidati na ravni strinjanja, deleža kolokacij, pri katerih so se vsi trije označevalci strinjali v odgovoru, ter deleža razhajanj, kjer so upoštevani kolokacijski kandidati, pri katerih sta bila vsaj dva od treh odgovorov označevalcev nasprotujoča (DA in NE) ali pa sta bila dva od treh odgovorov NE VEM. Vidimo lahko, da so deleži razhajanj višji pri strukturah s predlogi, v nekoliko manjši meri pa tudi pri strukturah s prislovi (v krepkem tisku).

<i>struktura</i>	<i>ujemanje (Cohenova kapa)</i>	<i>delež kolokacij s popolnim ujemanjem</i>	<i>delež razhajanj (vsaj en DA in en NE ali dva NE VEM)</i>
P + S	0,42	78 %	11,8 %
S + S ₂	0,45	63 %	13,3 %
G + S ₄	0,46	73 %	16,1 %
R + G	0,37	63 %	19,5 %
R + P	0,46	64 %	15,9 %
S + [v] + S ₅	0,35	50 %	39,0 %
G + R	0,47	61 %	19,4 %
G + [v] + S ₅	0,33	46 %	26,4 %
S + [s, z] + S ₆	0,42	56 %	23,3 %
G + [s, z] + S ₆	0,46	61 %	10,2 %

Slika 3: Prikaz ujemanja oz. razhajanj odgovorov označevalcev glede na skladijsko strukturo (prvih deset struktur po številu kolokacijskih kandidatov).

Strukture s prislovi so tako zaradi relativno visokih deležev odgovorov NE VEM na eni strani in dokaj visokih deležev razhajanj (a vseeno ne previsokih) v odgovorih označevalcev na drugi predstavljale zelo dobro testno množico za opredeljevanje slovarsko relevantne kolokacije.

2.2 Analize struktur s prislovi

Analiza je zajela vse kolokacijsko produktivne strukture s prislovi, ki so bile v skupini izluščenih podatkov za 333 gesel:

- prislov + glagol (rbz gbz = R G): *hudo zboleti*;
- glagol + prislov (gbz rbz = G R): *boleti enako*;
- prislov + prislov (rbz rzb = R R): *blazno glasno*;

- pridevnik + prislov (pbz0 + prislov = P R): *dostopen brezplačno*;
- prislov + pridevnik (rbz pbz0 = R P): *pretežno aluminijast*;
- prislov in/ali prislov (rbz in/ali rbz = R in/ali R): *burno in glasno*;
- prislov + samostalnik s predložno zvezo (rs = R [predlog] S): *brezplačno na razpolago*.

Izhajali smo iz jezikovne evalvacije jezikoslovcev označevalcev oz. njihovih končnih odločitev glede ocene o ustreznosti (slovarski relevantnosti) posameznega kolokacijskega kandidata v množičenjski nalogi. Vse ocenjene kolokacijske kandidate znotraj posamezne strukture smo najprej razvrstili v skupine glede na odločitev o ustreznosti ali neustreznosti kolokacije:

- odločitev DA (brez razhajanja, možne različne podopcije DA);
- prevladujoča odločitev DA (z razhajanjem);
- odločitev NE.

Zanimalo nas je, na katerih mestih so se jezikoslovci označevalci glede kolokativnosti strinjali in kje so bila v njihovih odločitvah razhajanja. Pri analizi smo se predvsem osredotočili na krovne opredelitve DA, NE, NE VEM, smo pa pri podrobnejših analizah podskupin odgovorov upoštevali tudi podopcije znotraj krovnih opredelitev. To velja zlasti za odgovore Struktura, ki smo jih želeli beležiti ločeno od ostalih odgovorov NE, saj gre za napake označevanja, ki za naše namene niso bile relevantne, so pa relevantne za nadaljnja prizadevanja izboljševanja oblikoskladenjskih označevalnikov besedil. Podobno velja za odgovore Fraze, saj so ti kolokacijski kandidati mogoče relevantni za pripravo postopkov za detekcijo (krajših) frazeoloških enot.¹⁰

¹⁰ Kot precej problematični kolokacijski kandidati so se izkazali tisti s (prevladujočim) odgovorom Razširjena kolokacija. Gre namreč za precej heterogeno skupino, zato smo se odločili, da teh kandidatov ne obravnavamo v okviru analize, opisane v tem prispevku, temveč bomo opravili analizo vseh razširjenih kolokacij v vseh strukturah hkrati.

Na podlagi vseh teh analitičnih prijemov se je pokazalo, katere vrste prislovov najdemo kot kolokatorje v teh treh skupinah, nadalje pa predvsem, v kolikšni meri in kdaj so posamezne strukture z vidika vključevanja v slovar problematične oz. neproblematične.

2.2.1 STRUKTURA RBZ GBZ

V pregledanem vzorcu so bile tipične kolokacije strukture RBZ GBZ, pri katerih ni bilo razhajanja v kolokacijskih opredelitvah označevalcev (DA), kolokacije z načinovnimi prislovi (*brezplačno prejeti*), prislovi mere ali stopnje (*blazno boleti*), prislovi kratnosti (*mnogokrat slišati*) ter časovnimi prislovi (*dnevno videvati*). Redki v tej DA so bili kolokacijski kandidati s primerniško in presežniško obliko prislova *bolj* in *najbolj*, ravno tako se je v vzorcu redko pojavljal elativ (*najbolj* pogoste so bile zveze s *preveč* – npr. *preveč boleti*, *deževati* – *predolgo* in *prezgodaj*). Pri zvezah lastnostnega prislova mere in načinovnega prislova je šlo za prislov mere (in ne načina) večinoma tudi v primerih kot *pošteno jezen* 'precej jezen'.

Z NE so označevalci ocenili večinoma kolokacijske kandidate:

- z vprašalnimi prislovi: kam, kako, kdaj, kaj (kako plezati: *Druge je mogoče videti, kako plezajo nezavarovani*);
- s kazalnimi prislovi: tako, toliko (toliko ohladiti: *pustimo, da se čaj toliko ohladi, da je le še mlačen*);
- s *kar* (v poudar(jal)ni oz. členkovni vlogi): *kar gnati: me je kar gnalo naprej*; *kar obupati: bi kar obupal, če se ne bi prav na koncu zgodil*;
- s prislovi *zato*, *nato*, *potem* (pri čemer je treba ločevati med prislovi in vezniki): *zato angažirati: Za obrambo igralca so zato angažirali prave znanstvene kapacitete*; *potem prevajati: razmišlja v angleškem jeziku in potem prevaja v slovenščino*; *nato barvati: ter šele nato barvamo*.

Z oceno NE so se označevalci opredelili tudi do kolokacijskih kandidatov s prislovi v povedni rabi oz. povedkovodoločilni vlogi s tipično nepregibnostjo,

obvezno vezljivostjo in pogosto vezavo z nedoločniki (predvsem naklonski prislovi tipa *treba*, *potrebno*, *lahko*): *Za prometno varnost bi bilo potrebno angažirati nove strokovnjake; Dijaki so lani (letos ne) lahko obiskovali.*

Razhajanja v odločitvah (DA, NE, NE VEM) smo identificirali pri kolokacijskih kandidatih:

- s časovnimi prislovi: danes, zjutraj, zvečer, letos, lani, takoj (takoj prevesti: *sem ga dal takoj prevesti v nemščino*; lani zboleti: *ki je lani zbolela za rakom na dojki*);
- s prislovi kratnosti: malokrat; enkrat, dvakrat; ponovno, znova, zopet in zaporedja: prvič, drugič, četrtič (dvakrat poplaviti: *je mesto dvakrat poplavila Savinja*; četrtič zbrati: *Srednja vas je že četrtič zbrala za poln kombi oblačil*);
- z načinovnimi prislovi: tako, težko (težko prevajati: *jo je težko prevajati v besede*);
- prislovi mere ali stopnje: tako, večinoma (tako boleti: *so jo roke tako bolele, da jih ni mogla dvigniti nad glavo*);
- s primerniško in presežniško obliko prislova bolj in najbolj (tip s si/se): najbolj (se) angažirati (*se je bolj angažiral in sodeloval v akcijah*);
- s kazalnimi prislovi: tu, tukaj, tam (tu gnezdi: *saj tu gnezdi okrog 140 vrst ptičev*).

V obravnavanem vzorcu smo pri napakah strukture identificirali primere, v katerih gre za prislove, ki modificirajo pridevnike/deležnike (npr. *medtem ko je grobo mleti sladkor najboljši za masleno testo; naj bi /.../ zahtevali tudi deklaracije za doma pridelano zelenjavo*). Med ostalimi napakami strukture so se pojavljali primeri napak v oblikoskladenjskem označevanju:

- prepoznavanje samostalnikov (pogosto srednjega spola ednine imenovalnika) kot prislovov ali glagolov: *enostavno zavežite konec niti in odvečno **blago odrežite**; Špagete denite v **obilo vrele** slane vode*;

- prepoznavanje samostalnikov kot glagolov zaradi nepravilne oblike, neustrezne besedne vrste: *do pomilovanja vas, ki pišete tako **neprimerne komentirate***;
- prepoznavanje zaimkov kot prislovov: *(je) na sodišču zatrjevala, da je očimu **vse odpustila**; Zdaj, končno, **vse vre** na plano.*

2.2.2 STRUKTURA GBZ RBZ

Tipične kolokacije strukture GBZ RBZ so bile kolokacije z načinovnimi prislovi (*povedati natančno*), prislovi kratnosti (*narediti naenkrat*) ter časovnimi prislovi (*popiti dnevno*). Kolokacijski kandidati s primerniško in presežniško obliko prislova *bolj* in *najbolj*, *manj* in *najmanj*, *več* in *največ* so bili redki, tudi elativ je v vzorcu nastopal zanemarljivo malokrat (*boleti preveč, premalo*).

Označevalci so se z oceno NE opredeljevali večinoma do kolokacijskih kandidatov z vprašalnimi prislovi v vezniški vlogi in do primerov, v katerih prislov ni bil pomensko vezan na glagol, temveč na sledeči pridevnik, samostalnik ali prislov, v določenih primerih pa je bil prislov pomensko vezan na glagolsko dejanje, izraženo s sledečim nedoločnikom. Ker je bil seznam tovrstnih opredelitev obsežen, navajamo le nekaj primerov, ki ponazarjajo specifičen tip problema, to je t. i. nanašalnost naprej:

- s primerniškimi in presežniškimi oblikami prislova *bolj/najbolj*, *manj/najmanj*, *več/največ*: obetati bolj: *se tudi vam obeta bolj sproščeno, pa tudi romantično obdobje*;
- s prislovom *kar*, čim: prevesti kar: *da svetopisemska besedila prevedejo kar najbolj v skladu z izvirnikom*; odrezati čim: *da si bodo odrezali čim večji kos tržne pogače*;
- z načinovnim prislovom *natančno*: dati natančno: *se je dalo natančno določiti*; hoteti natančno: *pa hočejo natančno vedeti*;

- tudi zveze z glagolom *obetati*: *obetati izjemno: tako da se obeta izjemno zanimiv finale.*; *obetati nadvse: se obeta nadvse zanimiv in prijeten večer.*

Razhajanj v odločitvah (DA, NE, NE VEM) je bilo malo, identificirali pa smo jih pri ocenjevanju kolokacijskih kandidatov:

- s časovnimi prislovi: dnevno, dvakrat, spomladi, sinoči, včeraj, danes (komentirati danes: *kot so nekateri komentirali danes*);
- s prislovi kratnosti: dvakrat, malokrat (zboleti dvakrat: *zato ne morete zboleti dvakrat*);
- s kazalnimi prislovi: tu, tukaj, tam (komentirati tukaj: *zato komentiramo tukaj, kjer je zastoj*);
- s primerniški in presežniškimi oblikami prislova bolj/manj: boleti manj: *je ženske bolelo še manj kot moške.*

Pri napakah strukture je šlo za napake oblikoskladenjskega označevanja zaradi prekrivnosti enakopisnih oblik (najpogosteje pridevnikov in prislovov):

- zaznavanje pridevnikov kot prislovov (*zato se **obeta pestro** in zanimivo dogajanje*; v zvezah s *kratko*: **napišite kratko** in *nagajivo seksi sporočilo*; pri čemer v določenih primerih vseeno gre za prislov: *in opazil, da **nosi kratko** pristrizeno brado*);
- zaznavanje samostalnikov (pogosto srednjega spola ednine imenovalnika) kot prislovov: *ne pomeni le biti sposoben **ljubiti žensko***;
- zaznavanje samostalnikov kot glagolov: **vlada načelno podpira**;
- zaznavanje zaimkov kot prislovov: *ne bom **komentirala, kar ste povedali***.

2.2.3 STRUKTURA RBZ RBZ

Tipični kolokacijski kandidati strukture RBZ RBZ so bile kolokacije s prislovi *natančno, burno, vestno*, pri katerih tudi ni bilo razhajanja v odločitvah oz.

kolokacijski opredelitvi (DA). Šlo je večinoma za zveze prislova mere ali stopnje (*dokaj natančno*) in načinovnega prislova (*strmo pokonci*), pri čemer za prislov mere (in ne načina) večinoma tudi v primerih kot *pošteno jezno* 'precej jezno'.

V primerih, ko je bila prevladujoča odločitev označevalcev NE, so se najpogosteje pojavljali kolokacijski kandidati s prislovi v povedni rabi oz. povedkovnodoločilni vlogi s tipično nepregibnostjo in pogosto vezavo z nedoločniki: LAHKO: *igralci lahko izjemoma za določene udarce uporabljajo*; MOGOČE, MOŽNO: *je mogoče brezplačno sneti iz interneta; bo ... možno brezplačno zamenjati*; TREBA: *se je bilo treba pošteno potruditi*; pri čemer je treba ločevati med prislovom v povedni rabi in 'navadnim' časovnim prislovom kot npr. NAJPREJ: *je najprej kratko odgovorila*.

Med ostalimi kolokacijskimi kandidati, ocenjenimi z NE, so bili še kandidati:

- s časovnimi prislovi (pri čemer je treba ločevati med časovnimi prislovi in vezniki): nato, potem in hkrati (potem doma: *Najprej je bil v bolnišnici, potem doma*; hkrati vestno: *z infrardečo svetlobo osvetluje voznikov obraz in hkrati vestno opazuje*) in
- s prislovi kratnosti: enkrat, dvakrat, nekajkrat (dvakrat pošteno: *se je proti večeru dvakrat pošteno zazibalo*).

Razhajanja v odločitvah (DA, NE, NE VEM), znotraj prevladujoče odločitve DA, smo identificirali pri kolokacijskih kandidatih:

- s prislovi kratnosti: kolikokrat, večkrat; dvakrat, trikrat (dvakrat zaman: *zatem pa kar dvakrat zaman poskušal sestaviti novo koalicijo*);
- s kazalnimi prislovi: tod, tukaj (tod doma: *je tod doma tudi prava gostoljubnost in prijaznost*);
- s prislovi mere ali stopnje: kar, res, zares, toliko (kar pošteno: *(so) že kar pošteno načeti*).

Znotraj prevladujoče odločitve DA so bila razhajanja tudi pri opredeljevanju do

primerov kolokacijskih kandidatov, v katerih so se pojavljali:

- vprašalni prislovi: kje, kam, kako, kdaj, kaj (kako vestno: *v katerega bomo lahko vpisovali, kako vestno so izpolnjevali zapovedane naloge*);
- primerniške in presežniške oblike prislovov: bolj burno: *bo vse precej bolj burno*; najbolj vestno: *bi gospodarila z našo zemljo kar najbolj vestno*;
- časovni prislov: vedno, danes, tudi tip odslej, doslej (vedno natančno: *da bi kolesa lahko vedno natančno sledila zamišljeni liniji*).

Kot najbolj problematično se je pokazalo ločevanje enakopisnih oblik pridevnikov in prislovov, pojavljalo pa se je tudi prekrivanje prislovov s samostalniki in zaimki. Napake oblikoskladenjskega označevanja so bile sledeče:

- označevanje pridevnikov kot prislovov (npr. *omogoča **izjemno natančno** in varno plovbo*, tudi v bolj težavnih primerih (npr. *da je v življenju **vse brezplačno***); pridevniške oblike se lahko prekrivajo z osnovno prislovno obliko (npr. *(življenje) je **res kratko***; ali primerniško prislovno obliko (npr. *(krilo) sme biti kvečjemu **malo krajše***);
- označevanje samostalnikov (pogosto srednjega spola ednine imenovalnika) kot prislovov: *ki vam bo **blago brezplačno** dostavljal na dom*;
- označevanje zaimkov kot prislovov: ***kar načeloma** lahko izrabimo za zapis informacije*.

2.2.4 STRUKTURA RBZ IN/ALI RBZ

Tipične kolokacijski kandidati strukture RBZ IN/ALI RBZ so bile kolokacije z načinovnimi prislovi, pri katerih tudi ni bilo razhajanja v odločitvah oz.

kolokacijski opredelitvi (DA). V vseh primerih je šlo za zveze dveh lastnostnih načinovnih prislovov (*vestno in odgovorno*).

Primerov, kjer je bila odločitev NE, skorajda ni bilo, navedemo lahko primer kolokacijskega kandidata z lastnostnim načinovnim prislovom *pošteno*: transparentno in pošteno (*ker pri nas bo tako vse transparentno in pošteno*), kjer gre za pridevniško obliko v povedkovodoločilni vlogi.

Razhajanja označevalcev (DA, NE, NE VEM) so se pokazala pri opredelitvah do kolokacijskih kandidatov:

- s časovnimi prislovi: spomladi (spomladi in [poleti, jeseni, pozimi]: *pozno spomladi in poleti obilno cvetijo*);
- s prislovi zaporedja: četrtič (četrtič in/ali [petič, tretjič, zadnjič]: *so gostovali že četrtič ali petič zapored*).

Med napake strukture bi lahko uvrstili dva primera, ki izkazujeta tipično povedkovnodoločilno rabo pridevnikov oz. povedkovniško vlogo pridevnikov in ne prislovov, pri čemer gre za pridevniške oblike, ki se prekrivajo z osnovno prislovno obliko (npr. (*spremno besedilo*) je karseda **kratko in preprosto**; *ni tako **silovito in strumno***). Pojavljajo pa se tudi primeri, ko prislovi določajo, modificirajo pridevnike/deležnike, pri katerih je treba ločevati navadne prislovne zveze od zloženek dveh pridevnikov; kratko in dolgo: *Starejšim bolnikom včasih predpišemo mešanico **dolgo in kratko** delujočega inzulina dvakrat na dan*.

2.2.5 STRUKTURA PRISLOV + SAMOSTALNIK S PREDLOŽNO ZVEZO

Tipične kolokacije strukture prislov + [predlog] + samostalnik so bile kolokacije s krajevnim prislovom *doma*, časovnimi prislovi (*sinoči, spomladi*), tudi z načinovnimi prislovi (*brezplačno*), pri katerih ni bilo razhajanja v odločitvah oz. kolokacijski opredelitvi (DA).

Razhajanja v odločitvah (DA, NE, NE VEM) smo identificirali pri različnih

kolokatorjih, ki ob sebi predvidevajo odprti niz lastnoimenskih samostalnikov¹¹ (v vzorcu so se najpogosteje vezali s prislovom *doma, sinoči*):

- kolokacijski kandidati s časovnimi prislovi, različne strukture (najpogostejše RBZ v sbz5, RBZ na sbz5): SINOČI v [Atenah, Bruslju, Celju]: *sta se sinoči v Atenah pomerila*; SINOČI v [hotelu, galeriji, gledališču]; SPOMLADI v [Ženevi, Ljubljani];
- kolokacijski kandidati s primernišskimi ali presežniškimi oblikami prislova: manj, več, največ; precej, veliko: manj alkohola: *ter pijte manj alkohola in kave*.

V primerih, kjer je bila odločitev NE oz. STRUKTURA, je šlo za napake lematizacije, do katerih pride zaradi prekrivnosti enakopisnih oblik prislovov z drugimi besednimi vrstami (najpogosteje s pridevniki, tudi s samostalniki). Kot primer napačne lematizacije navedimo primer, kjer sta pri kolokatorju možni dve lemi in je bila izbrana napačna imenovalniška oblika prve leme namesto imenovalniške oblike druge leme: četrtič očke (*ker je postal že četrtič očka*).

2.2.6 STRUKTURA PBZ0 RBZ

Med tipičnimi kolokacijami strukture PBZ0 RBZ, ki so jih označevalci ocenili z Da, so bili kolokacijski kandidati z lastnostnimi (*postavljen pokonci*) in časovnimi prislovi (*objavljen spomladi*) ter krajevnim prislovom *doma* (*narejen doma*).

Odločitve NE oz. STRUKTURA smo identificirali v primerih neustreznega strukturnega razmerja, kolokacijskih kandidatov z napačno lematizacijo in lastnoimenskimi kolokatorji ali v primerih t. i. nanašalnosti naprej, ko se prislov ni nanašal na dotični pridevnik:

¹¹ Lastnoimenski samostalniki so se kot potencialni kolokatorji izkazali za precej problematično kategorijo na ravni vseh struktur in tako niso nadalje pokriti v sklopu tega prispevka, saj so del ločene obsežne analize.

- [glaven, lep] DOMA: *so v glavnem doma v sredozemskem območju; najlepših doma narejenih božično-novoletnih voščilnic;*
- [Cankarjev, Šeškov] DOMA: *je v organizaciji Cankarjevega doma potekala, kjer bodo vrata Šeškovega doma odprta;*
- DEBEL [najmanj, komaj, manj, nekaj, pol]: *naj bi bile debele najmanj pet centimetrov; sposoben pošteno: če niso sposobni pošteno delat.*

Razhajanja v označevalnih odločitvah (DA, NE, NE VEM) smo identificirali pri pomensko praznih kolokacijskih kandidatih: s časovnim prislovom *spomladi* (znan *spomladi: bo verjetno znano spomladi*) in krajevnim prislovom *doma* (navajen doma: *kot smo navajeni doma*).

2.2.7 STRUKTURA RBZ PBZO

Tipične kolokacije strukture RBZ PBZO, ki so jih označevalci ocenili z DA, so bile kolokacije s prislovi mere ali stopnje (*strašno jezen*).

Razhajanja v označevalnih odločitvah (DA, NE, NE VEM) smo identificirali pri kolokacijskih kandidatih:

- s pomensko praznejšimi prislovi mere ali stopnje: tako, res, večinoma (večinoma enosmeren: *ceste so večinoma enosmerne*);
- s prislovi primerniške ali presežniške oblike bolj/najbolj (*tip še bolj, vse bolj*): (bolj debel: *vse bolj in bolj debelo plast zraka*; bolj ekološki: (*je postal še bolj ekološki in še varčnejši*).

Odločitve NE oz. STRUKTURA smo identificirali predvsem v primerih neustreznega strukturnega razmerja med posameznimi kolokacijskimi sestavinami: uspešno doktorski (tudi dobro kolesarski, premalo učiteljski): *V primeru uspešno zaključenega doktorskega študija*.

2.3 Diskusija

Pregled analiz jezikovne evalvacije označevalcev kolokacijskih struktur s prislovi je pokazal, katera so z vidika opredeljevanja slovarske kolokativnosti problematična mesta posamezne prislovne strukture, ki so potrebna kasnejše obravnave in natančnejše jezikoslovne diskusije.

Pomensko polnejše prislovne kolokatorje so označevalci prepoznali kot slovarsko relevantne pri vseh obravnavanih prislovnih strukturah, kar potrjuje njihova enotnost pri opredeljevanju do kolokacijskih kandidatov (DA), kot je prikazana v Tabeli 1.

Vrsta prislova	R G	G R	R R	P R	R P	R in/ali R	R [predlog] S
lastnostni	brezplačno, natančno, ...	brezplačno, natančno ...	strmo, trdno	pokonci	pokončno	vestno, pošteno, pokončno	brezplačno, izjemoma
merni	hudo, blazno, pošteno, močno, znatno	preveč	premalo, dokaj, karseda	/	pošteno, strašno, znatno, hudo, močno	/	nič, precej, četrť, pol
kratnostni	mnogokrat, velikokrat, večkrat	naenkrat, pogosto	/	/	/	/	/
časovni	dnevno, dolgo, kratko, nenehno	dnevno, kratko, občasno	zgodaj, predolgo	/	/	/	/
stopnjevalni	bolj, najbolj	/	/	/	/	/	/
krajevni	doma	doma	/	doma	/	/	doma

Tabela 1: Prikaz označevalnih odločitev (DA – brez razhajanj) glede na posamezne skupine prislovov po posameznih prislovnih strukturah.¹²

Analiza prislovnih kolokacijskih struktur je pokazala, da so bili označevalci enotni tudi pri opredeljevanju do kolokatorjev in kolokacijskih kandidatov, ki

¹² Zaradi omejenosti s prostorom v tabeli prikazujemo primere osamljenih prislovov, brez tipičnega kolokatorja ali iztočnice. Prislove, ki so bili v strukturah lahko kolokatorji ali iztočnice, skupaj s tipičnimi kolokatorji navajamo v preostalem delu diskusije.

so jih ocenili kot slovarsko nerelevantne (NE). To je bilo v primerih:

- znotrajbesedilne referenčnosti: komentirati kako (*ne bom komentiral, kako so pripravljani*); nato barvati (*jih šele nato barvamo*);
- nanašalnosti naprej: dati natančno (*se je dalo natančno določiti* → *natančno določiti*); zboleti kar (*je zbolelo kar šest poštarjev* → *kar šest*); obetati izjemno (*se obeta izjemno zanimiv finale* → *izjemno zanimiv*);
- napak strukture oz. napačnega avtomatskega označevanja strukture: odvečno blago (*zavežite konec niti in odvečno blago odrežite* → *samostalnik, ne prislov*); kar gnati (*me je kar gnalo naprej* → *členek, ne prislov*); uspešno doktorski (*v primeru uspešno zaključenega doktorskega študija*); pokončno volanski (*zelo pokončno postavljen volanski obroč*).

Izhodišče za nadaljnjo razpravo o pomenski relevantnosti skupin prislovov pa predstavljajo semantično oslajljene skupine prislovnih kolokatorjev, pri katerih je bilo identificiranih največ razhajanj v odločitvah označevalcev:

- prislovi kratnosti in pogostnosti: *enkrat, pogosto, znova*;
- prislovi zaporedja: *prvič, drugič*;
- časovni prislovi (deikti): *takoj, takrat, dnevno, letno*;
- kazalni prislovi (deikti): *tukaj, tam*;
- prislovi mere: *tako, toliko*;
- pomensko praznejši prislovi mere: *večinoma, kar, okoli, res*;
- vprašalni: *kje, kam, kaj, kako*;
- stopnjevalni prislovi: *bolj, najbolj*.

Tabela 2 kaže, da so se razhajanja v zgoraj omenjenih skupinah prislovov pojavljala pri različnih strukturah.

Vrsta prislova	R G	G R	R R	P R	R P	R in/ali R	R [predlog] S
lastnostni	težko	takole	/	dosegljiv, dostopen	/	/	/
merni	tako, večinoma	tako	kar, res, zares, tako, toliko	/	tako, res, večinoma	/	precej, veliko, premalo, preveč
kratnostni/ zaporednostni	enkrat, pogosto, znova, drugič	dvakrat, malokrat	dvakrat, trikrat, kolikokrat	/	/	četrtič	četrtič
časovni	takoj, letos, lani ...	takoj, naprej	vedno, danes, doselj, odslej	spomladi	dnevno	spomladi	sinoči, spomladi, dnevno
stopnjevalni	bolj, najbolj (tip s se/si)	manj	bolj, najbolj (tip precej bolj, kar najbolj)	/	bolj, najbolj (tip še bolj)	/	/
kazalni	tu, tukaj, tam	tu, tukaj, tam	od, tukaj	/	/	/	/
vprašalni	/	/	kje, kam, kako, kdaj, kaj	/	/	/	/

Tabela 2: Prikaz primerov prislovov (kot kolokatorjev ali iztočnic) po skupinah v različnih strukturah z označevalno odločitvijo DA – z razhajanjem.

Zgoraj omenjene skupine prislovov tako ravno zaradi pomenske nerelevantnosti kolokacij, v katerih se pojavljajo, ne sodijo med slovarsko relevantne kolokacije. Nasprotno pa je kompatibilnost tovrstnih kolokatorjev, tj. njihova medsebojna usklajenost in gramatična relacijskost, zanimiva za obravnavo na ravni slovničnega opisa, ki se izraziteje usmerja v analizo in opis sintakse konkretnih slovničnih relacij.

Na osnovi analiz kolokacijsko produktivnih struktur s prislovi bi lahko sklenili, da so bili kot slovarsko nerelevantni prepoznani kolokacijski kandidati s prislovi v deiktični vlogi (*tukaj boleti, določiti tam*), v vezniški vlogi (*prepričati, kaj* [je res]; *komentirati, kako* [so pripravljani]), prislovi modalnosti (*treba angažirati, lahko ohladiti*) in kolokacijski kandidati s pomensko šibkimi ali oslabljenimi prislovi (*večinoma doma*).

Posamezni primeri kolokacijskih kandidatov ali tipi kolokatorjev kažejo, da se je o njihovi slovarski vključenosti treba odločati na ravni posamezne strukture. Tovrstni primer so prislovi, ki lahko nastopajo v vlogi intenzifikatorja (tip *kar*

pošteno [načeti]) ali pa semantično manj relevantni vlogi poudarnosti oz. členkovnosti (tip *kar prekiniti*). Podobne jezikoslovne obravnave so potrebne širše skupine števnosti, kot je kratnost ali zaporedje (zaporednostni prislovi), ki jih zaradi raznolike semantične relevantnosti ne moremo strukturno omejiti (*četrtič doktorirati* proti *stokrat povedati*).

3 SKLEP

Analize jezikoslovne evalvacije struktur s prislovi, ki so sledile stopnjam v procesu izdelave celovitega kolokacijskega opisa slovenskih besed, tj. avtomatsko izluščenim kolokacijskim podatkom in pilotni množičenjski nalogi, so se izkazale za zelo učinkovit način opredeljevanja ne samo slovarsko relevantne kolokacije, temveč prek identifikacije nerelevantnih kolokacijskih kandidatov (npr. napake strukture) tudi statistično relevantne kolokacije. Kot se izkaže, je za opredeljevanje kolokacije manj bistveno ugotoviti, kaj kolokacija je; precej pomembneje je opredeliti, kaj kolokacija ni.

Takšna analiza predstavlja model nadaljnjega jezikoslovnega evalviranja in opredelitev kolokacije znotraj projekta KOLOS ter s tem povezane opredelitve slovarsko relevantne kolokacije, ki bo lahko neposredno aplicirana v posodobitev Kolokacijskega slovarja sodobne slovenščine. Učna množica s 17.576 označenimi kolokacijskimi kandidati je pomembna tudi za ostale dejavnosti projekta KOLOS: za uvrščanje kolokatorjev v gruče, opredelitve podobnosti oz. razlik sopomenk prek kolokacij in nenazadnje za proučevanje kolokacijskih trendov skozi čas. Posamezne skupine prislovov oz. relacije kolokatorjev, ki so bile prepoznane kot slovarsko nerelevantne (npr. prislovi modalnosti), pa predstavljajo izhodišče za naslavljanje in temeljitejši premislek leksikalnih jezikovnih elementov na ravni slovničnega opisa.

Podatki, pridobljeni na podlagi opravljenih analiz, bodo koristni za nadaljnje podrobnejše analize za različne jezikoslovne in jezikovnotehnološke potrebe. Na podlagi identificiranih napak strukture bo tako mogoče izboljšati oblikoskladenjsko označevanje, kar bi pomenilo izboljšanje avtomatskega

luščenja za problematične strukture (kot je npr. prepoznavanje roditelja v strukturi glagol + samostalnik v roditelju), mogoče pa bo tudi nadgraditi obstoječe definicije v slovnici besednih skic za luščenje kolokacijskih kandidatov. Podatki bodo zelo uporabni tudi pri nadaljnjem vključevanju in obravnavi novih struktur, prvotno izločenih zaradi precejšnjega šuma (npr. osebek + glagol, npr. *dež pada, jagode zorijo*). Vse ugotovitve bo mogoče implementirati v leksikografski delotok in na ta način izboljšati podatke ne samo v Kolokacijskem slovarju sodobne slovenščine, temveč tudi v drugih (slovarskih) virih, ki vsebujejo kolokacijske informacije.

ZAHVALA

Znanstveno-raziskovalno delo, ki ga predstavlja prispevek, je nastalo pri projektu 'Kolokacije kot temelj jezikovnega opisa: semantični in časovni vidiki' (št. J6-8255), ki ga med 2017 in 2020 sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Znanstvenoraziskovalno delo, ki ga predstavlja prispevek, je nastalo pri projektu Nova slovnica sodobne standardne slovenščine: viri in metode (št. J6-8256), ki ga med 2017 in 2020 sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Avtorja se zahvaljujeva tudi podpora infrastrukturnih programov Centra za jezikovne vire in tehnologije Univerze v Ljubljani in Centra za uporabno jezikoslovje pri zavodu Trojina.

LITERATURA

Arhar Holdt, Š., Čibej, J., Dobrovoljc, K., Gantar, P., Gorjanc, V., Klemenc, B., Kosem, I., Krek, S., Laskowski, C., Robnik Šikonja, M. (2018): Thesaurus of Modern Slovene: By the Community for the Community. V J. Čibej, V. Gorjanc, I. Kosem in S. Krek (ur.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*: 401–410.

- Ljubljana University Press, Faculty of Arts. Dostopno prek: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/2991-1.pdf> (15. 12. 2018).
- Bartsch, S. (2004): *Structural and Functional Properties of Collocations in English: A Corpus Study of Lexical and Pragmatic Constraints on Lexical Co-occurrence*. Tübingen: Gunter Narr.
- Benson, M., Benson, E., Ilson, R. (1986): *The BBI combinatory dictionary of English: A guide to word combinations*. Amsterdam: John Benjamins.
- Cook, P., Lau, J. H., Rundell, M., McCarthy, D., Baldwin, T. (2013): A lexicographic appraisal of an automatic approach for detecting new word senses. V *Electronic lexicography in the 21st century: thinking outside the paper*: 49–65. Estonia: Proceedings of the eLex conference.
- Cowie, A. (1994): Phraseology. V R. Asher in J. Simpson (ur.): *The Encyclopedia of Language and Linguistics*. Vol. 6. Oxford: Pergamon.
- Firth, J. R. (1957): *Modes of Meaning. Papers in Linguistics*. London: Oxford University Press: 1934–51.
- Gantar, P., Krek, S., Kosem, I., Šorli, M., Grabnar, K., Pobirk, O., Zaranšek, P. in Drstvenšek, N. (2012): *Leksikalna baza za slovenščino* Ljubljana: Ministrstvo za izobraževanje, znanost, kulturo in šport. Dostopno prek: <http://www.slovenscina.eu/spletni-slovar/leksikalna-baza>; <https://www.clarin.si/repository/xmlui/handle/11356/1030> (22. 11. 2018).
- Gantar, P., Kosem, I., Krek, S. in Gorjanc, V. (2015): Collocations dictionary of Slovene: challenge for automatization and crowdsourcing. G. Corpas Pastor in dr. (ur.): *Computerised and Corpus- based Approaches to Phraseology: Monolingual and Multilingual Perspectives*. Europhras, Malaga.
- Gantar, P. (2015): *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete.

- Gorjanc, V., Gantar, P. Gantar, Kosem, I. in Krek, S. ur. (2015): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana, Univerza v Ljubljani, Filozofska fakulteta.
- Gantar, P., Kosem, I. in Krek, S. (2016): Discovering Automated Lexicography: The Case of the Slovene Lexical Database. *International Journal of Lexicography*, 29(2): 200–225.
- Halliday, M. A. K. (1961): Categories of the theory of grammar. *Word*, 17: 241–292.
- Halliday, M.A.K. (1966): Lexis as a linguistic level. V C. Bazell idr. (ur.): *In Memory of J.R. Firth*:148–162. London: Longman.
- Hausmann, F. (1984): Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortwendungen. *Praxis des neusprachlichen Unterrichts*, 31.1: 395–406.
- Hoey, M. (2005): *Lexical Priming: A new theory of Words and Language*. London: Routledge.
- Hunston, S., Francis, G. (2000): *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. John Benjamins.
- Kilgarriff, A., Rychlý, P. (2010): Semi-automatic Dictionary Drafting. V G.-M. de Schryver (ur.): *A Way with Words: A Festschrift for Patrick Hanks*: 299–312. Kampala: Menha Publishers.
- Kolokacije 1.0: Kolokacijski slovar sodobnega slovenskega jezika*. Ljubljana: CJVT UL. Dostopno prek: <https://viri.cjvt.si/kolokacije/slv/#> (22. 11. 2018).
- Kosem, I., Gantar, P. in Krek, S. (2013): Automation of lexicographic work: an opportunity for both lexicographers and crowd-sourcing. V I. Kosem idr. (ur.): *Electronic lexicography in the 21st century: thinking outside the paper*: 32–48. Ljubljana: Trojina, Institute for Applied Slovene Studies; Tallinn: Eesti Keele Instituut.

- Kosem, I., Krek, S., Gantar, P., Arhar Holdt, Š., Čibej, J. in Laskowski, C. (2018): *Kolokacijski slovar sodobne slovenščine*. Dostopno prek: <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/download/120/214/3152-1?inline=1> (29. 10. 2018).
- Krek, S., Kosem, I., Gantar, P. (2013): *Predlog za izdelavo Slovarja sodobnega slovenskega jezika*. Dostopno prek: http://www.sssj.si/datoteke/Predlog_SSSJ_v1.1.pdf (18. 11. 2018).
- Logar Berginc, 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*. Ljubljana: Trojina, zavod za uporabno slovenistiko, Fakulteta za družbene vede.
- Manning, C. D. in Schütze, H. (1999): *Foundations of statistical natural language processing*: Chap. 5. Collocations. Cambridge, Massachusetts: The MIT Press.
- Mel'čuk, I. (1996): Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. V L. Wanner (ur.): *Lexical Functions in Lexicography and Natural Language Processing*: 37–102. Amsterdam: Benjamins Academic Publishers.
- Mel'čuk, I. (1998): Collocations and Lexical Functions. V A. P. Cowie (ur.): *Phraseology. Theory, Analysis, and Applications*: 23–53. Oxford: Clarendon Press.
- Rundell, M., Kilgarriff, A., (2011): Automating the creation of dictionaries: where will it all end? V F. Meunier (ur.): *A Taste for Corpora. A tribute to Professor Sylviane Granger*: 257–281. Benjamins.
- Sinclair, John, Mc., (1987): *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London and Glasgow: Collins ELT.
- Sinclair, J. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (1996): The search for units of meaning. *Textus IX*: 75–106.
- Sinclair, J. (1998): The lexical item. V E. Weigand (ur.): *Contrastive Lexical Semantics*: 1–24. Amsterdam: John Benjamins, pp.
- Sinclair, J. (2004): *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7/2: 215–44.

IN THE SEARCH OF LEXICOGRAPHICALLY RELEVANT COLLOCATION: THE EXAMPLE OF GRAMMATICAL RELATIONS CONTAINING ADVERBS

This paper presents the results of the analysis of grammatical relations that focussed on identifying not only collocations relevant for lexicographic purposes, but also problematic areas that need further investigation on both lexicographic and grammatical level. In the initial study, collocation candidates for a wide selection of grammatical relations for a heterogeneous sample of 333 lemmas have been automatically extracted from the Gigafida reference corpus of Slovene. A group of linguists then annotated the relevance of collocation candidates, examining both collocations and their examples of use, and their answers were analysed for agreement. The findings were that relations such as adjective + noun, noun + noun in gerund, and some relations verb + preposition + noun exhibited high agreement and large shares of approved collocation candidates. On the other hand, grammatical relations containing adverbs proved to be among the ones where disagreement or uncertainty of linguists-annotators was the highest. Consequently, it was decided that these adverbial relations should be analysed first as a sample set in testing our bottom-up approach to determining which collocation candidates are lexicographically relevant.

Further analysis has shown that the decision on the relevance of collocation candidates for dictionary purposes needs to be made separately for each relation, and groups of adverbs within it. An example of semantically less relevant group proved to be adverbs functioning as intensifiers or having a semantically less relevant role of a participle. Even more problematic is a group of numeral adverbs (*once, twice...*) which have different levels of semantic relevance (e.g. *četrtrič doktorirati* 'to receive a PhD for the fourth time' versus *stokrat povedati* 'to say something a hundred times') and thus cannot be delimited on a group level within a particular grammatical relation.

The data from the analyses described in this paper will enable further detailed

analyses, in particular a description of each grammatical relation from the perspective of its collocationality. In addition, bad collocation candidates that are the result of errors in morphosyntactic annotation will enable the improvement of sketch grammar and relatedly the quality of automatic extraction output. Furthermore, we intend to use existing findings in order to improve the results of grammatical relations that have been initially excluded from the automatic extraction procedure due to a high percentage of noise.

Keywords: lexicography, semantics, collocationality, collocations dictionary of Modern Slovene, adverb

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-
Deljenje pod enakimi pogoji 4.0 Mednarodna.

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0
International.

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

