

Slovenščina 2.0: “Slovnične raziskave za jezikovni opis”

Uredniki: Špela Arhar Holdt, Polona Gantar, Vojko Gorjanc, Robert Grošelj

Tehnični urednik: Damjan Popič

Založnik: Trojina, zavod za uporabno slovenistiko

Za založnika: Iztok Kosem

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

COBISS.SI-ID=298688512

ISBN 978-961-93594-6-4 (pdf)

Izdajatelj: [Trojina, zavod za uporabno slovenistiko](#)

Soizdajatelj: [Filozofska fakulteta, Univerza v Ljubljani](#)

Ljubljana, 2018

E-pošta: slovenscina2.0@trojina.org

Format: elektronski

Jezik: slovenščina in angleščina

Vključenost v podatkovne baze:

- [Directory of open access journals](#) (DOAJ)
- [Digitalna knjižnica Slovenije](#) (IdLib)
- [COBISS.SI/COBIB](#)
- [BrillOnline Linguistic Bibliography](#)

Izid publikacije je finančno podprla [Javna agencija za raziskovalno dejavnost Republike Slovenije](#) iz naslova razpisa za sofinanciranje domačih znanstvenih periodičnih publikacij.

SLOVENŠČINA 2.0: »SLOVNIČNE RAZISKAVE ZA JEZIKOVNI OPIS«

Špela ARHAR HOLDT

Center za jezikovne vire in tehnologije

Polona GANTAR

Filozofska fakulteta Univerze v Ljubljani

Vojko GORJANC

Filozofska fakulteta Univerze v Ljubljani

Robert GROŠELJ

Filozofska fakulteta Univerze v Ljubljani

Arhar Holdt, Š., Gantar, P., Gorjanc, V., Grošelj, R. (2018): Slovenščina 2.0: »Slovnične raziskave za jezikovni opis«. Slovenščina 2.0, 6 (2): i–iv.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.i-iv>.

Letošnja tematska številka revije *Slovenščina 2.0* vsebuje sedem razprav o izbranih slovničnih pojavih, ki so jih pripravili raziskovalke in raziskovalci z različnih slovenskih institucij. Poleg razprav je na voljo zapis strokovnega panela, na katerem so sodelujoči (Kozma Ahačič, Irena Krapš Vodopivec, Simon Krek, Marko Stabej, Rok Žaucer in Špela Arhar Holdt) razmišljali o bodočem slovničnem opisu za slovenski jezik. Panel slikovito predstavlja duha časa, v katerem tematska številka nastaja: digitalna sodobnost zahteva ne samo vsebinske posodobitve slovničnega opisa za slovenščino, ampak tudi odprto dostopnost, strojno berljivost in povezljivost slovničnih podatkov, kot tudi njihovo premišljeno zasnovanost za različne vrste (strokovnih in uporabniških) potreb. Zato tudi ni naključje, da zbrani prispevki v luči novih izzivov na slovnično problematiko gledajo aplikativno, empirično in metodološko inovativno. Prijazno vabljeni k branju.

Špela Arhar Holdt in **Jaka Čibej** predstavljata metodologijo za avtomatsko pridobivanje oblikoslovnih vzorcev iz leksikona besednih oblik Sloleks. Metoda razvrsti oblikoslovne paradigme glede na spremenljivi in nespremenljivi del oblike, z ročnim pregledom pa so vzorci urejeni in v strojno berljivi obliki pripravljeni za nadaljnjo rabo. Dodana vrednost opisanega dela so natančno opredeljene nedoslednosti, pomanjkljivosti in napake leksikonskih podatkov, ki bodo ob prihajajoči posodobitvi tega izredno uporabnega jezikovnega vira lahko odpravljene.

Kaja Dobrovoljc piše o formulaičnih besednih nizih, tj. neprekinjenih nizih dveh ali več besed z izstopajočo pogostostjo pojavljanja, ne glede na njihovo strukturno, pomensko ali zaznavno celovitost. Avtorica predstavi metodologijo strojnega pridobivanja nizov, nato pa podatke kategorizira in opiše glede na njihovo zgradbo ter funkcijo. Prispevek je pomemben, ker utemeljuje jezikoslovni interes za tovrstno gradivo, ki je v preteklosti kljub visoki pogostosti v jezikovni rabi ostajalo na obrobju jezikovnega opisa.

Nataša Logar podprto z izčrpnim pregledom področja in na veliki količini korpusnih zgledov razišče pojem besedotvorne produktivnosti. Avtorica izluščene samostalniške kandidate za izbranih 10 samostalniških pripon ročno selekcionira, opredeli njihovo produktivnost po uveljavljeni Baayenovi formuli in jih analizira. Ključnega pomena so smernice, kako produktivost upoštevati pri slovničnem opisu, kot tudi konkretna opozorila na stilistične, zvrstne, fonološke in številne druge vidike, ki jih je pri pripravi nove slovnice smiselno upoštevati.

Agnes Pisanski Peterlin in **Tamara Mikolič Južnič** s primerjavo gradiva pisnega in govornega referenčnega korpusa raziskujeta primere jezikovne rabe, pri katerih se v vlogi osebka pojavlja osebni zaimsek. Čeprav je takšna raba za slovenščino netipična, se pojavlja tako v pisnem, še pogosteje pa govorjenem jeziku, in kot ugotavljata avtorici, pogosto služi posebnemu pragmatičnemu namenu, po večini za razdvoumljanje in poudarjanje. Analiza razkrije, da je

tematika rabe osebnega zaimka kot osebk v slovenščini mnogo kompleksnejša, kot bi lahko sklepali iz dosedanjih opisov.

Eva Pori in **Iztok Kosem** se slovnične tematike dotikata prek analize kolokacij, in sicer tistega dela kolokacijskega gradiva, ki nastopa v besednozveznih strukturah s prislovi. Prispevek predstavlja rezultate jezikoslovne evalvacije strojno izluščenega kolokacijskega gradiva in odpira empirično podprto razpravo o tem, kje so meje slovarsko relevantne kolokacije kot koncepta. To vprašanje je ključno za nadgradnjo avtomatskega luščenja kolokacijskih kandidatov za slovenščino, ki bo neposredno podprla razvoj Kolokacijskega slovarja sodobne slovenščine in podobnih jezikovnih virov.

Mojca Smolej se posveča dvema skupinama odvisniških stavkov, ki so bili v dosedanjih obravnavah pogosto prezrti: izhodiščno nestavnočlenskim odvisnikom in členkovnim odvisnikom. Razprava, ki s pomočjo korpusnega gradiva opiše navedeni skupini odvisnikov, opozarja na dejstvo, da tudi vprašanja, ki so v preteklosti uživala veliko jezikoslovne pozornosti, niso nujno v celoti odgovorjena. Celovit izkorpusni pristop k pregledu odvisnikov lahko naslovi identificirane vrzeli, ki so pereče tudi za jezikovno didaktiko.

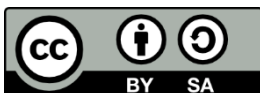
Jerica Snoj podaja slovaropisni pogled na besednovrstno kategorizacijo leksikalnega gradiva, ki je v idealnem stanju skladna s slovnično teorijo, saj sta slovar in slovnica v prizmi leksikogramatike različni področji enega samega kontinuuma. Avtorica ob primerih za povedkovnik, členek in izdeležniške tvorjenke prikaže razvoj besednovrstnega kategoriziranja v izbranih slovenskih slovarjih in na osnovi tega podaja dragocene smernice za besednovrstno obravnavo v prihodnjih slovenskih leksikalnih virih.

SLOVENŠČINA 2.0: “GRAMMAR IN LINGUISTIC DESCRIPTION”

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/>



NOVA SLOVNICA: KJE SMO IN KAM GREMO

Špela ARHAR HOLDT, Kozma AHAČIČ, Irena KRAPŠ VODOPIVEC, Simon KREK, Marko STABEJ, Rok ŽAUCER, Helena DOBROVOLJC, Vojko GORJANC, Polona GANTAR

Arhar Holdt, Š., Ahačič, K., Krapš Vodopivec, I., Krek, S., Stabej, M., Žaucer, R., Dobrovoljc, H., Gorjanc, V. in Gantar, P. (2018): *Nova slovnica: kje smo in kam gremo. Slovenščina 2.0, 6 (2): 1–32.*

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.1-32>.

6. junija 2018 je na Inštitutu Jožef Stefan¹ potekal dogodek, na katerem so bili javnosti predstavljeni cilji in prvi rezultati projekta *Nova slovnica sodobne standardne slovenščine: viri in metode* (ARRS J6-8256). Namen projekta, ki ga natančneje opredeljuje predstavljena spletna stran,² je razviti jezikoslovno metodologijo za računalniško podprto analizo sodobne slovenščine, kakršna je zajeta v referenčnih besedilnih korpusih slovenskega jezika. Z novo metodologijo bodo pripravljene baze jezikovnih podatkov, ki bodo po koncu projekta skupnosti odprto na voljo za raziskave, gradnjo jezikovnih priročnikov ter učnih gradiv, razvoj jezikovnotehnoloških orodij ipd.

Omenjeno projektno financiranje izdelave kakršnegakoli zaključnega izdelka sicer ne pokriva, vendar že priprava podatkovnih baz zahteva premisleke o trenutnih prioritetah slovenskega prostora. Med cilji za prihodnost je brez dvoma sodobni slovnici opis, ni pa še konsenza, kako naj bo slednji oblikovan, da bo odgovoril na (različne) potrebe sodobne družbe. Da odpremo razpravo, smo na projektnem dogodku organizirali strokovni posvet, opredeljen z naslednjimi vprašanji: kdo so deležniki, ki bi lahko projektne rezultate uporabljali; na kaj moramo pri pripravi paziti, da bodo podatki optimalno uporabni; kakšno oz. katero slovnico potrebujemo najprej; katere so metodološke in logistične premise njene priprave; kje je trenutno slovensko

¹ <https://ijs.si/ijsw>

² <http://slovnica.ijs.si/>

slovničarstvo in kakšen razvoj si lahko obetamo; kakšne so potrebe po slovnicih pri različnih uporabniških skupinah ter kaj bi trenutne vrzeli najboljše naslovilo.

Na posvetu so sodelovali: **dr. Kozma Ahačič** (Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU), **dr. Irena Krapš Vodopivec** (Škofijska gimnazija Vipava), **dr. Simon Krek** (Laboratorij za umetno inteligenco, Inštitut Jožef Stefan), **dr. Marko Stabej** (Oddelek za slovenistiko, Filozofska fakulteta Univerze v Ljubljani) in **dr. Rok Žaucer** (Center za kognitivne znanosti jezika, Univerza v Novi Gorici). Vsak izmed sodelujočih je dobil čas za predstavitev uvodnih tez oz. vprašanj, sledila je razprava, ki sem jo moderirala **dr. Špela Arhar Holdt** (Center za jezikovne vire in tehnologije, Univerza v Ljubljani).

Zapis je po posnetku dogodka³ pripravil **Haris Agović**. Pri zapisu so bile izjave za namene lažjega branja skladenjsko prilagojene značilnostim pisnega jezika, vendar po principu minimalne intervencije. Govorke in govorci so opis pregledali in predlagali (nevsebinske) popravke za boljše razumevanje besedila. Zapis začnemo z nastopom prvega panelista.

Kozma Ahačič: Hvala za besedo. Najprej bi v svoji uvodni besedi opozoril na tri sestavne dele slovnice vseh vrst, ki se jih mora po mojem ločevati, ko se govori o preteklih, sodobnih in prihodnjih slovnicih. Prvi del je teoretično izhodišče slovnice. Manj ko je slovnica, ki izhaja iz korpusa in jezikovnotehnološke obdelave, teoretično obremenjena, bolj ko so kategorije preproste in jasno razmejene, lažja je njena izdelava, podatki pa bolj zanesljivi. To kažejo izkušnje, tudi iz tujine. Slabost teoretično manj obremenjene slovnice pa je, da ne rešuje mejnih primerov v jeziku in služi bolj za izhodišče kot za popoln opis jezika. Kljub temu se mi zdi primernejši prvi korak za korpusno slovnico.

Drugi del, ki bi ga izpostavil, je realno jezikovno gradivo. Realno jezikovno gradivo skoraj praviloma zastavljeni teoretični okvir postavlja pod vprašaj.

³ http://videlectures.net/novaslovnica2018_ljubljana/

Moja temeljna izkušnja⁴ (in to na precej bolj osnovni ravni, kot se je loteva pričujoči projekt) je, da pri pripravi slovnice niso problem posamezna poglavja in izolirana vprašanja, ampak je problem celota. Noben jezikovni pojav namreč ne more ostati izven slovnice, ker slovnica ne opisuje dela jezika, ampak jezik v celoti. Nek slovnični opis se mora torej znotraj sebe iziti za vse primere, ki živijo v določenem jeziku. Odločitev o tem, ali bomo v slovnici na primer na ravni besednih vrst operirali s pojmi, kot so členek, povedkovnik in tako naprej, ali pa bomo vse to označevali kot prislove, je tako stvar teoretičnega izhodišča. Na ravni gradiva pa nas to ne reši zagate, kako opisati nenavadno vedênje posamičnih skupin besed ne glede na njihovo kategorialno uvrstitev. Prav tako nas večja ali manjša vključitev pojmovanja konverznosti v teoretični okvir ne odreši razreševanja tega problema na ravni gradiva, saj se bo v gradivu isti leksem v različnih pozicijah lahko vedel popolnoma različno. Manjša obremenjenost s teoretičnimi razmisleki na ravni teoretičnega izhodišča slovnice tako ne pomeni manj dela s slovničnim opisom, ampak zgolj preložitev oziroma odložitev problema z ene ravni na drugo. Če sem navedel samo nekaj preprostih oblikoslovnih problemov, se na ravni skladnje in besednega reda ti problemi prav neverjetno namnožijo.

Poleg tega ima vsak jezik vsaj nekaj oblikoslovnih in skladenjskih primerov, ki jih težko objektivno kategoriziramo. Značilen primer za slovenščino sta besedi *lahko* in *rad* ter njuna skladenjska raba. Z vidika leksikalnega opisa se nam takšne zagate hitro pokažejo, ko je treba celotno slovensko leksiko uvrstiti v posamezne besedne vrste, ji določiti konverznost, skladenjske vloge, stalnost besednih zvez, dvojničnost, potem pomen, ki je vezan na posamezne slovnične kategorije, pomen v predikativni rabi in seveda slovnični učinek, ki ga povzročajo tako imenovane slovnične vrste (kot so na primer vezniki). Realno jezikovno gradivo nas torej ne glede na teoretični okvir postavlja pred primere, ko se moramo odločiti. Uporabnik, ki bo slovnico uporabljal kot priročnik ali

⁴ Priprava dveh šolskih slovníc, Kratkoslovnica: slovenska slovnica za osnovno šolo (2018) in Slovnica na kvadrat: slovenska slovnica za srednjo šolo (2018).

kot osnovo za nadaljnje raziskovanje, bo namreč osredotočen predvsem na mejne in nejasne primere, ne na primere, ki jih lahko razreši že na podlagi osnovnega jezikovnega znanja. Prav zato je golo oblikoslovje pregibnih besednih vrst del slovnice, ki ga je relativno najlažje dobro opisati, saj na tej ravni praviloma vemo, kaj iščemo.

Zadnji, tretji del, ki bi ga izpostavil, je vprašanje zvrstnosti. Če to vprašanje ni razrešeno na začetku (in tu je za operativno delo povsem vseeno, ali operiramo s pojmom standardnega ali s pojmom knjižnega jezika), nas bo v vprašanje zvrstnosti prisililo gradivo samo. Že samo na ravni standardnega ali knjižnega jezika bomo namreč lahko opazovali zelo različne glasoslovne, oblikoslovne, besedotvorne in skladenjske uresničitve, ki jih bomo lahko pojasnjevali samo ob ustrezni vezanosti na tip besedil, v katerih jih najdemo. Če gremo proti prihodnji slovnici, bi sklenil samo še s tem, da moramo paziti, da ne bi bila takšna slovnica zgolj stranski produkt jezikovnotehnoloških potreb in možnosti – da bi se slovnico dejansko delalo kot osnovni cilj že od začetka. Hvala.

Špela Arhar Holdt: Najlepša hvala, lahko kar Irena.

Irena Krapš Vodopivec: Hvala, lepo pozdravljeni, hvala za besedo. Jaz se bom zadeve lotila veliko manj znanstveno oziroma čisto iz prakse. Sedemnajst let že učim slovenščino na gimnaziji in zato bom v zvezi z današnjo temo izhajala s tega področja. Začela bi s tem: če govorimo o gimnazijah, še vedno velja predpostavka, da na gimnazije prihajajo najboljši učenci. Ob sicer množičnejšem vpisu je odličnosti nekoliko manj, pa vendar je podatek pomemben, ker gimnazijci na koncu opravljajo maturo iz slovenščine. Čeprav gimnazija seveda predstavlja le segment v izobraževalni vertikali, so rezultati na maturi tako ali drugače pokazatelj razvoja jezikovne zmožnosti skozi celotno vertikalo. Rezultati na maturi iz materinščine pa niso spodbudni. Če pogledamo lanske rezultate (po RIC-ovih podatkih), so bili gimnazijci pri slovenščini le malo nad povprečno oceno. To se pravi približno dobro, nekaj čez. Ampak govorimo o vseh treh ravneh ocenjevanja. Če izpustimo ustno ocenjevanje, če

gremo samo k eksternemu, pisnemu ocenjevanju, so rezultati še nižji.

Kako je to povezano z današnjo temo? Od začetkov svojega dela do danes sem pri jezikovnem pouku zamenjala že nekaj gradiv, prenovljen je bil učni načrt, zares spremenilo pa se po moji presoji ni prav veliko. Slovnicične in pravopisne vsebine se še vedno podajajo v obliki tradicionalnega opisa, in sicer je glavni razlog pomanjkanje časa ob pripravi na maturo; o tem bom nekoliko več povedala kasneje. Vedno pogosteje se mi ob pregledovanju, popravljanju besedil, ki jih tvorijo dijaki, potrjuje opažanje, da pišejo v knjižnem jeziku najpogosteje tako, kot so navajeni govoriti v svojem narečju oziroma neknjižnih sporazumevalnih okoliščinah – v smislu pretvarjanja tega v knjižni jezik. Tisti z nekoliko več jezikovnega čuta so pri pretvarjanju uspešnejši. Pravopisno-slovnicični opis (tak, kot je zdaj v šoli) pa k temu po mojih izkušnjah ne prispeva dovolj.

Pri pouku sama poskušam upoštevati dejstvo, da so dijaki rojeni govorniki slovenščine, da v sporazumevalni resničnosti živijo, jo sooblikujejo, pa čeprav v neknjižnih zvrsteh. Zato mislim, da bi morali pri poučevanju maternega jezika, predvsem pri slovnicično-pravopisnem opisu, izhajati iz dejstva, da se mladi jezika ne učijo v celoti na novo in da opis jezika za izboljšanje njihove sporazumevalne zmožnosti ni tako bistven. Seveda ga je treba smiselno vključiti v proces učenja – po moji presoji s precej bolj individualiziranim pristopom. Torej z veliko tvorjenja besedil v obeh prenosnikih in z ozaveščanjem napak ob svojih besedilih in pomnjenjem pravil ob tem. Zdi se mi, da gre pri pouku materinščine pravzaprav za korekcijo že naučenega. To je po moji presoji težji proces kot prvo pomnjenje, s čimer se srečujejo dijaki pri učenju tujega jezika. Nezanemarljivo je dejstvo, da je narečno obarvani sleng prvi jezik dijakov oziroma edini jezik, s katerim se mladi zares poistovetijo. V knjižni slovenščini jim je komuniciranje tuje, v obeh prenosnikih. Mogoče malo bolj naravno v pisni slovenščini, absolutno pa ne v govorjeni. Avtentičnih priložnosti za to znotraj šolskega procesa ni veliko in seveda ne tako želene notranje motivacije. Pri pouku tujega jezika je notranje motivacije pri dijakih absolutno več. Poleg

tega, da se od začetka učijo knjižne zvrsti, imajo tudi veliko več avtentičnih možnosti za uporabo jezika.

Pred leti sem zato poskusila izvajati nekoliko prilagojeno izpolnjevanje učnega načrta: manj frontalnih slovničnih opisov in veliko več samostojnega dela, več tvorjenja različnih besedil in kvalitetno individualizirano povratno informacijo. Težava je seveda čas, ampak moji dijaki po tej spremembi na maturi dosegajo enake – tukaj pa operiram s številkami, ne govorim na pamet – enake ali boljše rezultate, kot so jih prej. To se pravi, sposobni so se sami opremiti z orodji, da poiščejo opise jezika, ko je potrebno; kot mi sami povedo (spet v anketah), to tudi hitro pozabijo. Po tem procesu pa, vsaj tako pravijo, odhajajo z gimnazije s precej večjo samozavestjo pri rabi knjižne slovenščine, kar je seveda cilj učnega načrta, z védenjem, da zmorejo komunicirati v knjižnem jeziku, in z zavedanjem, da je učenje maternega jezika vsaj toliko pomembno in smiselno kot učenje tujega jezika. Kar pa po mojem v današnji dobi, v 21. stoletju, pri razvijajočih se mladostnikih pomeni pravzaprav uspeh.

Kaj vidim kot osrednje težave v zvezi s slovničnim opisom v šoli, torej v gimnaziji? Če sem konkretna: po mojem je glavna težava izhajanje iz predpostavke, da dijaki nimajo jezikovnega predznanja. To še posebej velja pri gimnazijcih, ker celotni slovnični opis poslušajo že v osnovni šoli. Tu bi lahko čas pridobili. [Druga težava je] prepričanje, da bodo dijaki samozavestni govorci slovenščine, če bodo obvladali pravila. Korelacija med pravili in rabo je precej nerealna; tudi o tem bom, upam, kakšno besedo lahko rekla še kasneje. Tretja težava: prenatrpanost učnega načrta za slovenščino v gimnazijah, absolutno. In zato premalo časa za tisto, kar je bistveno, torej za samostojno tvorjenje, za rabo jezika in seveda ob tem tudi pomnjenje pravil. Sprašujem se, ali se v resnici z dijaki skozi štiri leta pripravljamo na maturo ali se pač učimo jezika oziroma poskušamo vzpostaviti možnost, da bo dijak postal samozavesten tvorec v knjižnem jeziku. To dvojje se namreč zelo pogosto žal izključuje, zato se vedno znova vračam k temu, da bi bilo treba premisliti o smiselnosti tovrstnega preverjanja znanja. Ker je jasno, kam se bo pouk

usmeril, če bo treba izbirati. Usmerili se bomo v pripravo na maturo, ne pa v razvoj zmožnosti in rabe knjižnega jezika, zato ker je tak čas. Dijak, ki bo imel slabše rezultate na maturi, se ne bo mogel vpisati na željeno fakulteto in tako naprej.

Zaradi tega je v šoli še vedno veliko preveč teorije. Kljub temu da je po prenovi učnega načrta načeloma pristop izbesedilen, je še vedno veliko, veliko premalo prakse. Torej ni avtentičnih možnosti za to, da bi opis uporabili. Veliko več je tega pri tujem jeziku. Težavo vidim še v tem, da je podajanje snovi pri mnogih drugih predmetih v gimnazijah pogosto v nestandardnem jeziku. To spet pomeni eno avtentično možnost za srečevanje s knjižnim jezikom manj. Težava je, da je današnji pouk manj uniformiran, da se učitelji, profesorji borijo za dijake, zato se vzpostavlja precej manj formalne okoliščine in to pomeni tudi hiter zdrs v nestandardni jezik. In še zadnje: težava je ponovno in še enkrat čas. Ker imajo dijaki zelo različno razvito sporazumevalno zmožnost, bi bilo treba vsakemu posebej podati kvaliteto povratno informacijo o njegovem tvorjenju. Na ta način bi se seveda učil tudi pravil. Hvala lepa.

Špela Arhar Holdt: Hvala, super. Izvoli.

Simon Krek: Vseeno moram poudariti par stvari z izhodišča projekta, ki danes organizira ta dogodek. In sicer, da ne bo kakšne dileme – to ni projekt, ki bi proizvedel novo slovnico slovenščine. To je projekt, katerega cilj je proizvesti podatke, na katerih je mogoče narediti slovnico v kateremkoli teoretskem okviru želimo. Na nek način tukaj lahko izpostavim projekt, kot je Survey of English Usage⁵ iz petdesetih let, ki je imel empirično izhodišče: »Ne bomo delali angleške slovnice, če ne vemo, kakšni so v resnici empirični podatki.« To so naredili s takratnimi možnostmi in iz tega je nastala Quirkova slovnica.⁶ V zdajšnjem času, se pravi leta 2018, imamo neskončno več možnosti za empirično preverjanje vseh jezikovnih pojavov in to je v resnici tudi namen –

⁵ <http://www.ucl.ac.uk/english-usage/about/index.htm>

⁶ https://en.wikipedia.org/wiki/A_Comprehensive_Grammar_of_the_English_Language

da jih organiziramo na način, ki je potem dostopen vsem.

Drugi del, ki ga je pomembno poudariti, je ta: ne glede na to, kar si kdorkoli misli o čemerkoli, je treba omogočiti računalniško analizo vseh jezikovnih pojavov, ki jih kdorkoli producira v slovenščini z intenco, da ga nekdo razume. To je drugačno izhodišče od tega, če rečemo, da bomo na nek način preverili sistemsko pravilnost tvorjenih stavkov. Znati moramo interpretirati čisto vsak stavek, ki ga nekdo skomunicira komurkoli, tudi računalniku. Ta del bo v resnici tudi pokrit, tukaj ni težave, čeprav je seveda veliko dela. Ta del je pomemben za slovenščino tukaj in zdaj. Če samo nakažem, za kaj gre. Imeti moramo formalizirano slovnico, ki je računalniško berljiva. Kar se tiče skladnje, imamo ta hip denimo model odvisnostne drevesnice. Prihaja iz praške šole, prihaja iz češčine, to je okej, ker je vseeno tudi slovansko. To je del, ki ga je treba pokriti ne glede na to, kaj bomo potem naredili s slovnico za človeške konzumente.

Zadnja stvar, ki jo želim poudariti, je to, kar sem že omenil v prejšnjem delu diskusije, ko smo predstavljali,⁷ in sicer, da imamo na nek način dva loka, ki ju nočemo ločevati med sabo. En je leksikogramatični. Če vzamemo in analiziramo neke pojave, v resnici ni dobro, da [jih] ločujemo na nekaj, kar naj bi bila strogo skladnja ali strogo morfologija, s semantiko se pa pač ne bomo ukvarjali, ker so drugi ljudje, ki se ukvarjajo s tem, recimo leksikografi. V tem smislu je treba reči, da v bistvu ni slovarja brez slovnice. Kdorkoli se bo lotil jezikovnega opisa slovenščine za 21. stoletje, bo moral imeti tudi slovnične kategorije, ki po možnosti temeljijo na empiričnih podatkih tega, kar proizvajajo ljudje, da bi komunicirali z drugimi. To je izhodišče: da imamo en lok. Recimo, da si jezikoslovci želijo imeti ločene vrtilčke in se ukvarjati s frazeologijo, ki je čisto nekaj drugega od kolokacij, ki je čisto nekaj drugega od vezljivosti in tako naprej. To je en in isti lok, ki ga je Sinclair opisal z 'idiom principle' – v smislu organizacije jezika zato, da bi lahko komunicirali pomen. Ene stvari, ki nosijo pomen, so večje, druge so manjše. Mogoče gre do morfema

⁷ http://videlectures.net/novaslovnica2018_ljubljana/

ali do frazeologije na drugi strani, ampak tega ni mogoče ločevati na neke trde dele, s katerimi se radi ukvarjamo. Toliko o tem.

Marko Stabej: Hvala lepa. Hvala za povabilo. Kot se mi vidi, nisem slovničar, sem samo simpatizer. Vidi se pa tudi, da rad zajemam z veliko žlico, prebava pa malo slabše deluje. To je prispodoba za to, kar čutim o sodobnem slovničarstvu. Zdi se mi, da podatkov ne more biti nikoli premalo, ker je bila to največja pomanjkljivost preteklih obdobij, deloma objektivna, deloma ideološka. Odsev te ideologije se kaže v tem, da je še vedno v obtoku termin *gradivo*. 'Gradivo' ni isto kot 'podatki'. Gradivo je nekaj, kar v jeziku najdeš, s čimer in na čemer lahko gradiš svoje teoretske predpostavke. Zdi se nam, da je normalna povratna zanka, da s tem potem realnost pojasnjuješ. Ampak zgodovina slovenske slovnice je v tem, da se to ne zgodi. V tem smislu se mi zdi, da moramo imeti pri slovnici ... Bi bilo fajn, no, če bi imela naša interdisciplinarno zasnovana znanstvena skupnost maksimalni program, ki ga je postavil že Bezljaj. Jaz sem bral Bezlaja leta '64 v eseju *Blišč in beda slovenskega jezika*, kjer pravi, da bolj kot česarkoli Slovenci potrebujemo slovnice. Ne slovnice standardnega jezika, ne slovnice posameznih narečij, ampak slovnice, ki bi teoretsko pojasnila sleherno izjavo v slovenskem jeziku. To je seveda maksimalni program, da te zaboli glava, če samo pomisliš, kako bi bilo to možno. Ampak brez maksimalnega programa ostajamo v nekem parcialnem opisu, kjer ne vemo, ali predpisujemo, opisujemo, kaj opisujemo, ali se s tem, da opisujemo, komu zdi že, da predpisujemo. Vse to, kar se danes pravzaprav dogaja.

Zanimivo, da je to pisal '64, ko se je Toporišičeva slovnica že pripravljala in je bila dolgo časa – mi smo bili navajeni nekako privzeti – odrešitev vsega, kar je bilo problematično. Na to ste nekateri iz publike že opozarjali: takega hibridnega jezikoslovnega dela, kot je Toporišičeva slovnica, zlepa ne najdeš, ker je normativno, deskriptivno, ne vem, kaj vse. Pravzaprav logičen nasledek je, da je bil slovnični opis izrazito parcializiran. Postavlja se mi naslednje vprašanje, analogno z recimo zgodovinopisjem, ki ugotavlja, da velike zgodbe (velike zgodovine, dvanajst tomov, zgodovina francoske države) pravzaprav

niso možne, ne da bi temeljito lagal. Ker veš toliko podatkov, toliko nasprotujočih si silnic, da če govoriš samo zgodovino zmagovalca, pač ne govoriš resnice. Ne rekonstruiraš resnice, ampak rekonstruiraš zgodbo, ki ti paše kot zmagovalcu. Tudi zgodba slovničarja kot neke vrste avtoritarnega obvladovalca te zgodbe – to si je Toporišič vzel, vsi vemo. Ampak zgleda, da mu je tako uspelo (ali pa tako neuspelo), da si tega vsaj do zadnjega časa ni nihče niti v sanjah upal prevzeti.

Če lahko povem malo notranje informacije: tudi na Katedri za slovenski knjižni jezik na Filozofski fakulteti Univerze v Ljubljani⁸ smo v preteklih dvajsetih, petindvajsetih letih na več raznih sejah oziroma sestankih (zdaj že nekaj časa ne) imeli na sporedu, kaj če bi se lotili pisanja slovnice. Ampak ker smo se seveda zavedali metodološke, teoretske različnosti, dlje kot do tega vprašanja pravzaprav ni nikoli prišlo. Ker je bilo vprašanje: na podlagi česa, komu in zakaj pišeš to veliko zgodbo.

Zdi se mi, da se je tega dobro zavedati in s tega stališča se mi zdi projekt [Nova slovnica sodobne standardne slovenščine] res odličen začetek. Podatkovje mora biti tukaj, potem pa se izluščijo take ali drugačne zgodbe. Samo ena velika zgodba po mojem ni možna. To se vidi v šolski slovnici (imam srečo, da prakse ne poznam več kot toliko). Tudi za šolsko slovnico rečemo: »Aha, to pa je žanr.« Ampak ali je to žanr nečesa, kar naj bi pomagalo pri mejnih primerih, kot je rekel kolega Ahačič? Da ne veš, kaj, pa pogledaš v slovnico in rečeš: »Aha, zdaj pa vem.« Zakaj bi to pravzaprav počel? Marsikdaj se lahko zgodi, da obstaja v realnosti A in B. In pogledaš v slovnico, ki pravi: »Raje B.« Pa ne veš, zakaj »raje B«. Ker se slovničarju kot pripovedovalcu tako zdi bolje ali ker je bolj utemeljeno v skladu s tradicijo? To je ena stvar. Po drugi strani pa je moj občutek in moja starševska izkušnja, da otrok najprej rabi metajezikovni nabor oznak, s katerimi jezik sploh lahko napade. Ne da bi hotel imeti ta jezik opisan, ampak hoče nečemu, za kar mi vemo, da je pridevnik, reči pridevnik. In to sploh

⁸ <https://www.slov.si/>

ni slovnica, ampak nabor metajezikovnih oznak. Ki se jih je zdaj, kot vemo, treba izogibati, da ne bi prekmalu rekli česa premalo dokončnega o katerikoli stvari. Pri čemer dokončne resnice – to pa je, hvala bogu, zadnjih deset let prineslo – ni. To bi jaz rekel. Hvala lepa.

Rok Žaucer: Hvala. Jaz si bom pomagal, nisem tak govorec, da bi znal govoriti na pamet. Dobili smo dva sklopa stvari, o katerih naj razmislimo, in jaz sem splošnejši sklop vzel za podlago za kratko [predstavitev], drugo pa sem si napisal na plonk ceglč. Eno od splošnih vprašanj, na katero naj bi odgovarjali na posvetu, je: *Kakšno oziroma katero slovnico potrebujemo najprej?* Ideja je, če obstajajo, če se da določiti nacionalne prioritete in potrebe. Kar se tega tiče, mi je s stališča generativnega jezikoslovca jasno, da ne glede na moje potrebe rabi slovenščina orodja, s katerimi se jo da omogočiti za boljšo uporabo v digitalnem okolju. Je za to potrebna slovnica? Če je, rabimo tako slovnico najprej. Ne vem pa, da bi za take stvari slovnica bila potrebna – da bi Google, preden je naredil prevajalnik, naredil slovnico in tako naprej. Pri vsem ostalem se mi zdi, da je [odgovor], kaj se najbolj rabi, odvisen od tega, koga bomo vprašali. Kakorkoli bomo obračali, bomo prišli do tega, da bo zgodovinski jezikoslovec rekel: »Jaz rabim pač to.« Konec koncev lahko reče: »Splošno slovnico imamo, zgodovinske slovnice nimamo, dajmo narediti to najprej.« Teoretični jezikoslovec (kjer 'teoretični' pomeni samo cilj pristopa, ne da teoretični jezikoslovci niso empirični jezikoslovci) bo rekel, da rabimo razlagalno opisno slovnico. Učitelju slovenščine kot tujega jezika, recimo specifično za govorce kitajščine, bo prišlo prav nekaj drugega in tako naprej.

Širši konsenz bomo, po mojem mnenju, težko dobili in v nekem smislu nisem prepričan, da je blazno nujen. Seveda zmeraj pomaga, ampak nujen je, če bo eno ali drugo ali tretje ministrstvo reklo: »V naslednjih dvajsetih letih financiramo eno samo stvar.« Potem rabimo širši konsenz. Do neke mere nedavna zgodovina kaže, da se da dobiti tudi neodvisno financiranje, se pravi neodvisno od modela, kot sem ga zdaj omenil. Kozma ima šolsko slovnico, ki ni bila financirana na tak način. V tem smislu so učbeniki za slovenščino kot tuji

jezik, kot jih dela Center za slovenščino kot drugi ali tuji jezik,⁹ Rada Lečič, tudi na te stvari se da gledati kot na nek tip slovnice. Razmišljamo lahko tudi o odprtem wikiju, ki ga lahko dopolnjujemo vsi. Nekaj takega se je usedlo na UNG-jev¹⁰ strežnik. Izključno samo zametek in ne izgleda, da bo lahko hitro karkoli ratalo.

Če pogledamo tuje prakse, obstajajo tudi generativno zasnovane knjige, ki jim lahko rečemo slovnica, recimo Cambridge Syntax Guides¹¹ obstajajo za velike in majhne jezike, za recimo valižanščino, za islandščino. To je financiral Oxford University Press, do neke mere, vsaj izdajo. Ostalo pa je v rednem financiranju, jezikoslovca univerza plačuje, da je jezikoslovec na tisti univerzi. Peter Jurgec ima narejen velik del za fonologijo slovenščine, ki mu jo je financirala Nizozemska znanstvena fundacija¹² in tako naprej. Mislim, da ni nujno, neobhodno, da imamo nek širši konsenz.

Druga stvar je še: *Kje je trenutno slovensko slovničarstvo in kakšen razvoj si lahko obetamo.* Mi je malo težko odgovoriti. Lahko bi rekel, da mogoče niti ne vemo. Vemo, kako zgleda tradicionalna slovenska slovnica, ampak če bomo dobili še eno, pri tem ne bomo govorili o razvoju. Kako bo zgledala slovnica, ki bo res močno jezikovnotehnološko osnovana? Po mojem še ne vemo. Težko rečemo, kje smo, kar se tega tiče. Kako bi zgledala razlagalno-opisna generativno zasnovana slovnica, tudi ne vemo, ker tega nimamo. Lahko sklepamo po primerih tistih parcialnih The Syntax of Welsh¹³, pa od Petra Jurgca fonologija in podobno, ampak težko bi bilo oceniti. Izgleda, da bo sočasno nastajalo več slovnice, ki bodo odgovarjale na različne potrebe različnih uporabnikov. Konec koncev je Kozma to nekoč že rekel, čeprav mogoče z drugim poudarkom, ampak do istega pridemo.

⁹ <http://centerslo.si/>

¹⁰ <http://www.ung.si/sl/>

¹¹ <http://admin.cambridge.org/nr/academic/subjects/languages-linguistics/grammar-and-syntax/series/cambridge-syntax-guides>

¹² Nizozemska znanstvena fundacija (NWO): <https://www.nwo.nl/en>

¹³ <https://doi.org/10.1017/CBO9780511486227>

Kakšne so potrebe po slovničnih podatkih v raziskovalni skupnosti in izobraževanju ter kaj bi trenutne težave najbolj rešilo? Spet pridemo do tega, da odvisno, koga vprašamo. Če vprašate mene, kakšen tip slovnice nam lahko pomaga z vidika ciljev teoretičnega jezikoslovja: ja, splošna opisna slovnica posameznega jezika, v tem primeru slovenščine – seveda nam lahko pomaga. Teoretično jezikoslovje jo bo jemalo za nabor podatkov, s katerimi odgovarja na teoretična vprašanja, in v istem smislu bo dobrodošel tudi projekt, ki bo proizvajal take podatke. Teoretično zasnovana razlagalna slovnica je absolutno lahko koristna mojemu tipu jezikoslovca, ker jo lahko uporabljam ne samo kot podatke, ampak že tudi za analizo. Še ena stvar, ki je vsekakor potreba v raziskovalni skupnosti, je vključevanje slovenskih podatkov v mednarodne baze slovničnih zgradb, kot je recimo World Atlas of Language Structures¹⁴ in podobne stvari. Terraling¹⁵ je še konkretno za skladnjo. In podobno.

Prva stvar (ki sem jo dal na konec) je: *Kdo so deležniki, ki bi lahko projektne rezultate uporabljali?* Če bojo korpusi nadgrajeni, bomo nove podatke seveda lahko uporabljali praktično vsi sinhroni jezikoslovci in uporabljala [jih] bo lahko širša javnost. Ena stvar, ki je omenjena v opisu projekta, je »na tem osnovana jezikovna analiza«; kar se tega tiče, pa vem premalo, da bi lahko sodil, kdo jo bo lahko uporabljal in ali jo bom jaz lahko uporabljal. To je to, kar imam povedati, hvala.

Špela Arhar Holdt: Jaz sem vas opazovala in sem hitro videla, da si zapisujete ob nastopih, tako da sem prepričana, da imate pripravljenih že kar nekaj replik. Bi pa vseeno rada najprej dala besedo Simonu kot vodji projekta in izpostavila dve stvari, ki se mi zdi, da sta se tukaj nakazovali, pa bi jih bilo mogoče dobro najprej razčistiti. Prva stvar je vprašanje celote proti parcialnosti in vprašanje, ki je po moje tukaj ležeče spodaj: ali lahko korpusni pristop, ali lahko empirija, kot si prej rekel, v resnici ponudi celovitost, ki jo lahko kdorkoli za kakršenkoli namen od slovničnega opisa pričakuje. Se ti zdi, da je to zdaj

¹⁴ <https://wals.info/>

¹⁵ <http://test.terraling.com/>

možno in na kakšen način je možno? Druga stvar je pa mogoče na ravni same priprave kakršnegakoli produkta, kjer smo ponovno govorili o celovitosti napram parcialnosti. Zdaj imamo dve izkušnji. Imamo enega zelo prominentnega slovničarja – mislim kot arhetip v našem prostoru –, ki samodejno prevzame nalogo, ko se ta pojavi, in jo potem tudi nekako celovito reši; in na drugi strani izkušnjo, o kateri je prej govorila Amália¹⁶ in ki se je Oddelek za slovenistiko očitno ne lotuje – in je tudi jasno in razumljivo, zakaj ne: nabor jezikoslovcev, ki so specializirani za različna področja in potem skupaj menda sposobni kreirati celoto, če je to v resnici sploh mogoče. Se pravi, kakšna je rešitev v smislu ekipnega dela?

Simon Krek: Bom odgovoril tudi na to vprašanje. Najbrž je treba za začetek nekaj razumeti. Takoj, ko s podatki naredimo karkoli izven tega, da samo zapišemo, kaj je nekdo napisal ali rekel – dodamo neke meta[podatke] –, takoj pademo v slovnični opis. Temu se enostavno ni mogoče izogniti. Treba je razumeti, da ko pademo v to past, se vsaj pri slovenščini znajdemo na natezalnici med dvema poloma. En [pol] je opis, ki je neprilagojen za to, da bi lahko naredil oziroma uporabil primerljive kategorije za drugi del te natezalnice, ki je mednarodni. Ko moraš svoje korpuse ali jezikovne oznake ali karkoli vključiti v mednarodni prostor, padeš v to, da te nekdo vpraša: »Imate samostalnik?« – »Okej, imamo.« – »Imate tole?« – »Ja, imamo.« – »Kaj pa tole?« – »Hm, mogoče imamo, mogoče pa nimamo,« in tako naprej. Ampak to v tem prostoru v bistvu ni dopuščeno. Lahko rečeš: »Tole bom uporabil na ta način, to bom uporabil na ta način.« Moram reči, da je mednarodni vidik na nek način olajšanje, ker izstopiš iz popolnoma enojezične situacije, kjer nad tabo visi veliki lik slovničarja in nikamor ne moreš. Če imaš na drugi strani prisilo, kjer nekdo reče: »Aha, frazni glagoli.« Frazni glagoli v angleščini, ne vem, kje. Mi rečemo: »Jah, nimamo fraznih glagolov, kaj hočemo.« Se pa začneš

¹⁶ Vabljen predavatelj na projektnem dogodku: dr. Amália Mendes (Centro de Linguística da Universidade de Lisboa): *Planning and Writing the Comprehensive Grammar of Portuguese*.

spraševati o vseh teh fenomenih in padeš v situacijo, kjer je mednarodna primerljivost oziroma uporabnost kategorij zapoved. Nujnost je. Ne samo ker imamo to radi, ampak ker je treba slovenščino procesirati na isti način kot vse ostale jezike. To je del, kjer se ni mogoče izogniti računalniški slovnici, ki ima pedigre v znanstveni slovnici – ki pa ni čisto isti, ampak je prej ovira kot ne.

Na drugi strani bom reagiral na ta, bom rekel, 'naj cveti sto cvetov' – imamo podatke, vsak vzame posamezni teoretski okvir in dela, kar hoče. Imamo lok med nujno mednarodno računalniško slovnico, ki je povezana (hkrati pa ne) z bivšo ali neobstoječo znanstveno slovnico, ki je po nujnosti povezana tudi s pedagoško slovnico; to je stvar, ki jo prostor sam zahteva, ker imamo šole, imamo učitelje slovenščine, nekaj morajo povedati o slovenščini. Se pravi, pedagoško slovnico rabimo. Po mojem ne bi bilo preveč pametno, da se to raztrešči. Ko računalnik procesira, je to čisto nekaj drugega. Osebkū rečemo ena, dva, tri, štiri.¹⁷ V šoli se še vedno učijo o osebkih, ni pa čisto isto in nobeden točno ne ve, kaj je eno in kaj je drugo. Kar po mojem pomeni, da je potreben napor za to, da se situacijo razume in se vsaj v osnovi vzpostavi konsenz okrog osnovnih stvari. Potem lahko seveda gre vsak gojit svoje cvetove na svojo stran, ampak del, ki je trdno jedro 21. stoletja, dokler še imamo izobraževanje, je po mojem nujen. Toliko bi jaz imel o tem povedati, da ne bom preveč nakladal.

Špela Arhar Holdt: Ok, lahko, ja, kar.

Kozma Ahačič: Samo na eno stvar bi se odzval, čisto tako mimogrede. Jaz mislim, da so časi velikega slovničarstva že mimo. Vsaj v mlajši generaciji jih ni več čutiti. Mislim, da smo kljub vsemu vedno bolj v času, ko ljudje nimajo časa, da bi o jeziku ne vem kako razmišljali in zato želijo čim lažje rešitev vprašanj. Ne gre pa več za to, da bi nekdo s pozicije moči, kot je bilo pred dvajsetimi, tridesetimi leti, karkoli skušal zapovedovati. Mislim, da se je tu ozračje bistveno sprostilo. Mogoče bi opozoril še na eno stvar, ki se mi zdi zelo pomembna, in je nanjo opozoril Marko, pa mimogrede tudi Irena. Slovnični opis v osnovni šoli

¹⁷ <http://www.slovenscina.eu/tehnologije/razclenjevalnik>

(pa če se ga v osnovni šoli ne naučijo do konca, tudi v srednji šoli) dejansko nima primarne funkcije učenja slovenščine, ampak ima primarno funkcijo jezikovnega opisa na splošno. Glede na to, da govorimo slovensko, se ga pač aplicira na slovenščino. Kljub vsemu se mi zdi zelo pomembno, da ne pozabimo, da namen pouka slovenščine ne more biti samo to, da bomo bolj kompetentno uporabljali slovenščino, ampak tudi to, da se spoznamo z jezikoslovjem kot znanostjo, da dobimo nek osnovni uvod v jezikoslovje.

Špela Arhar Holdt: Lahko, razen, če bi ti na tole repliko?

Irena Krapš Vodopivec: Čisto kratka replika na to zadnje. Se strinjam, da opis jezika v šolah mora biti. Strinjam se z Markom, ki je rekel, da morajo biti stvari poimenovane, in mislim, da se ponovno vrača, da so tudi v osnovni šoli slovnični pojmi poimenovani. To je bila ena izmed stvari, ki jih jaz tudi ves čas poskušam in zagovarjam. Se pravi, osnoven slovnični opis, definirani slovnični pojmi, neko temeljno znanje, ki bi ga v osnovni šoli morali usvojiti in ga v gimnaziji in srednjih šolah mogoče malenkost nadgraditi, sicer pa utrjevati. Ampak moja misel je bila: iz rabe. Ne da bi popolnoma opustili slovnični opis, ampak da bi ob svojih izdelkih ta opis spoznavali. To je bila moja misel. Druga stvar pa je, da se na nek način strinjam, da mora biti opis jezika zaradi opisa samega, kot je rekel Kozma, vendar je temu namenjenega veliko, veliko preveč časa. In potem zmanjka za primarni cilj, ki ga vidim. Samo to, to sem mislila.

Špela Arhar Holdt: Ok, a lahko kar Roku?

Rok Žaucer: Več stvari, ampak vsako čisto na kratko. Pedagoška slovnica: ja, nikakor nisem rekel, da ne rabimo pedagoške slovnice. Hočem samo reči, da neka osnova itak obstaja. [Od nedavnega] obstajata tudi dve pedagoški slovnici, ampak obstajajo učbeniki, ki se uporabljajo, in ti operirajo ravno z osnovnimi stvarmi. Starejši mogoče malo bolj eksplicitno s 'pridevniki' in 'samostalniki', ki se jih pač ne bomo znebili in ni razloga, da se jih poskušamo znebiti. Nikakor ne bi hotel reči, da tega ne rabimo. To rabimo in to tudi obstaja. Potencialno se lahko reče: v taki meri, da je povezljivo s poučevanjem tujih jezikov, in imamo

osnovo že fajn omejeno.

Kar se utilitarizma tiče, je vsekakor nekaj, kar mi gre na živce. [Da bi bil] pouk slovenščine namenjen samo učenju tega, da se boš naučil govoriti. Ideje (kar je Kozma omenil) sploh ni. Vsaj v osnovni šoli, pa tudi v srednji šoli ne. Če pomislite na biologijo: nobeden ne bo rekel, da ne rabimo biologije, ker še zmeraj ne tečemo hitreje ali kaj podobnega, čeprav nas učijo, kako deluje človeško telo. Tega ne razumem.

In še empiričnost. Malo se tudi pretirava s tem, da se delamo, kot da je zdaj vse novo in se dela empirično, včasih se pa ni. Toporišičeva slovnica je bila empirična in SSKJ je bil korpusno narejen. V drugih časih, na drugačen način – da ne eno ne drugo ne izhaja iz nobenih podatkov, pa ne moremo reči. Danes nam je pač omogočeno marsikaj drugega in lahko delamo na super boljše načine. Ampak delali so pa že Pleteršnik pa Tesnière s svojo dvojino. So pač nabirali podatke, delali so korpuse.

Špela Arhar Holdt: Ok, jaz bi Marku dala zdaj besedo.

Marko Stabej: Bom počrl besedo, da se ne bom oglasil. Je res, seveda niso delali čisto na pamet. Ampak delali so z gradivom, ki je bilo večinoma že obdelano. Dekontekstualizirano gradivo od marsikje. Pri Pleteršniku je znano, da pravzaprav na koncu niti ni vedel, od kod je marsikaj. Še posebej pa ne, iz katerih tekstov je. Ta dekontekstualizacija slovničnega opisa je še zdaj problem. Na nek način jo gledam pravzaprav pri Ahačičevih slovnica. Seveda je dobro, da podobno kot pri biologiji izveš, kaj so kodoni in kaj so nevemkaj, tudi pri jezikoslovju izveš. Ampak če pogledamo objektivno: smeri v biologiji, ki bi o isti stvari trdile zelo diametralno nasprotni stvari, kolikor jaz vem, ni. V jeziku pa so in tudi v javnem življenju se zadnje čase kaže, kako je zelo hitro lahko podcenjevati – upravičeno podcenjevati – publiko in ji prodati marsikaj, kar bi ob poprejšnji kritičnosti zavrnila. Da rečem konkretno, če učencem rečeš: »Prilastkov odvisnik. Prilastkov odvisnik dobimo tako, da iz stavčnega člena tvorimo stavek.« Čemu služi tako védenje samo po sebi? Jaz nisem napisal

slovnice in je tudi nikoli ne bom. Nisem zmožen tega, ker preveč kompliciram. Ampak [vsebina] v tem smislu: Kako lahko izrazimo čas v izjavljanju? Lahko tako, da imamo neke specializirane izraze, včasih pa uporabimo v stavku in temu rečemo odvisniki.

To je zdaj parcialno – vprašanje metode, vprašanje tradicije in tako naprej. Ampak se mi zdi, da če je bil od leta '98 naprej pouk slovenščine nominalno funkcionalističen (v resnici je bil afunkcionalističen, ampak to je stvar didaktike in ocene), zdaj ni čas, da gremo nazaj v formalizacijo in rečemo: »Aha, to je seznanjenje, vstop v jezikoslovno znanost.« Jaz tega ne vidim kot v vstop v jezikoslovno znanost. Niti slučajno ne. Vstop v vsako znanost je pač čudenje, ne pa ponujanje vzorcev, da si pravzaprav omejiš svojo radovednost na način, da drugače niti ne znaš gledati. In to je korpusno jezikoslovje definitivno veliko bolj približalo: da vidiš, ne kako je že nekdo opisal. [Ker da je] Toporišič na gradivu nastal – Rok, večina je prepisana iz gradiv prejšnjih slovnici in to vsi vemo. On je sicer govoril, da je normiral ta jezik na podlagi izobraženca iz središča, srednjih let, nek demografski [kriterij], ampak ni bilo jasno, kje se to vidi. Nikjer ni niti serije izjav. Ni izmišljeno, ampak ni pa to tip podatkov, kot ga ponuja korpusno jezikoslovje.

Ki je pa grozno, saj se dejansko ne moreš znati, to je pa res. Brez vmesnikov je večina nas, ki smo simpatizerji tega jezikoslovja, bosih. Če ne najdeš vodnika po peklu podatkov. To je res in mislim, da je tudi tukaj treba konvergenco graditi bolj, čeprav si Društvo za jezikovne tehnologije¹⁸ od leta '98 (ali še prej) prizadeva za to konvergenco. Živimo v svetu, ko smo na Oddelku za slovenistiko opazili, da če hočemo literati in jezikoslovci prijaviti skupen projekt, moramo prijaviti interdisciplinarni projekt. Ja, ker literarna veda in jezikoslovje nista ista veda. Pa tako naprej. Skratka, zdi se mi, da je en vidik slovničarstva tudi to, da dva svetova prideta skupaj na operativen način.

Špela Arhar Holdt: Jaz imam [ta svoj mikrofoni], hvala. Prej, ko si govoril, si

¹⁸ <http://www.sdjt.si/wp/>

omenil, da podatkov ni nikoli premalo. Si zdaj hotel povedati, da je podatkov lahko preveč? Je to tvoje mnenje?

Marko Stabej: Občan Marko Stabej, ki zjutraj gleda v računalnik, [meni, da] je podatkov preveč. Ker če mora čez dve uri govoriti, kaj je tako in se zdi na korpusnih podatkih, pa podatki podrejo vse hipoteze, ki jih imaš za neko stvar – jih je preveč. Na načelni ravni jih pa ne more biti preveč. Če lahko iz svoje pedagoške izkušnje, ki je glede uporabe korpusa razmeroma šibka, ampak: študentje začnejo uživati, če lahko [brskajo] po sinhronem korpusu Gigafida,¹⁹ hitro grejo tudi v IMP²⁰ pogledat, kako je bilo včasih. Zdaj se že pritožujejo, da od leta '18 do leta 1990 ni nobenega korpusnega podatka, kje pa naj tam dobimo podatke. Ja, ne vem, kaj, zgradite jih. Nastane potreba, da ne govoriš na pamet in da ne govoriš: »To je že v Pletersniku in vemo, da je bilo prej že v Pohlinu, pa Pohlin je prepisal iz Belostenca,« pa to, ono, tretje. To je bila včasih nit podatkov, zdaj pa je podatkov o marsičem premalo in ostanem pri tem, da jih je zmeraj premalo.

Špela Arhar Holdt: Okej. Ja, kar.

Irena Krapš Vodopivec: Lahko še jaz repliciram, saj bom na kratko. Zdi se mi, da je pomembno, da se vprašamo, kaj bi radi, glede na vse, kar je bilo do zdaj povedano. Kaj bi radi v šoli oziroma pri pouku slovenščine dosegli? Prej je bilo rečeno, da potrebujemo tudi opis zaradi opisa samega. Spet bom čisto konkretna: eno generacijo sem analizirala oziroma statistično preverila, koliko so uspešni na posameznih področjih. Če gledamo samo sposobnost reprodukcije opisa jezika, so uspešni blizu prav dobre ocene, če govorim o ocenah. Takoj ko gremo k nalogam višje taksonomske stopnje, kakršne so na maturi, smo za celo oceno ali še več nižje. Govorim o povprečni oceni ene generacije skozi vsa štiri leta, [to] je nek vzorec. Ko gremo k eseju, še k nadgradnji vsega skupaj, smo nekje na dobri oceni. To je generacija na naši

¹⁹ <http://www.gigafida.net/>

²⁰ <http://nl.ijs.si/imp/>

srednji šoli oziroma gimnaziji. Podatki na slovenski ravni, sem pa že rekla: lansko leto so bili rezultati porazni. Če gledamo primerjalno s tujimi jeziki, za eno oceno nižje. S tem, da moramo upoštevati, da govorim o tujih jezikih na višji ravni, ker je tudi slovenščina na višji ravni. In ob védenju, da je bilo za pozitivno pri slovenščini potrebno doseči 47, v določenih primerih celo samo 38 točk od 100 (se pravi, v procentih). Pri tujih jezikih pa 50 %. Vprašati se moramo, kaj pravzaprav želimo v tistih parih urah, ki jih profesorji oziroma učitelji slovenščine pač imamo.

Samo še eno stvar, npr. konkretno v oblikoslovju. V drugem letniku gremo ponovno skozi celo oblikoslovje. (Imam tudi sina v osnovni šoli in poznam potek osnovnošolskega šolanja, ker sem tam učila. Tudi tam spoznajo praktično vse temeljne pojme oblikoslovja.) Ker vodim mape svojih dijakov, lahko rečem, da vem, katere napake skozi štiri leta delajo. Recimo: *ki*, *kateri*. Dijak ve, kdaj se uporablja ta zaimsek, kdaj enega, kdaj drugega. V esejih pa seveda to dosledno uporablja narobe. Ko sem jih vprašala (spet je v ozadju anketa – anketirala sem 150 dijakov), ali razmišljate o tem, katero pravilo uporabljate, ko tvorite besedilo, je bil odgovor skoraj 100 % negativen. Ja, absolutno slovnični opis, ampak jaz pravim: iz lastne rabe oziroma iz lastnega tvorjenja in res temeljni slovnični pojmi. Hvala lepa.

Špela Arhar Holdt: Se opravičujem, ampak imam še eno podvprašanje, ker si odprla točno to, kar sem hotela poizvesti, in sicer: se ti zdi, da smo kot raziskovalci, kot akademiki, ki v resnici nimamo dostopa do šolskega sveta, nimamo dostopa do tvojih izkušenj, sposobni pripraviti pedagoško slovnico, ki bo odgovarjala na te potrebe? Ali je to nekaj, kar bi morali v bistvu učitelji sami pripravljati – če gre za vprašanje uporabnika, se pravi nekaj, kar mora izhajati od uporabnika, ki ga mi ne poznamo. Kaj je po tvojem mnenju rešitev?

Irena Krapš Vodopivec: Ja, absolutno povezava s praktiki. Vedno povezava s praktiki. Tudi pri maturi, ki se jo oblikuje, bi morala biti povezava s praktiki. Raziskovalci in učitelji praktiki, to bi moralo biti razmerje 50:50. To je to. Še to

bi rekla, ko govoriva o tem, kdo bi moral. Vprašati se moramo o našem šolskem sistemu: ni morda vendarle treba premisliti o načinu, kako preverjamo znanje? To mene ves čas muči. Kako preverjamo znanje slovenščine znotraj šolskega sistema in seveda na tej grozni končni postaji mature. Slovenist bi moral razčlembo neumetnostnega besedila, se pravi celotni slovnični, pravopisni in pravorečni opis, rešiti v tridesetih minutah. To je ekvivalentno devetdesetim minutam, ki jih ima na voljo dijak. Nimam sicer reprezentativnega vzorca, imam pa nekaj profesorjev slovenščine, ki so to rešili. Redkokateri 100 %, ker zmanjka časa tudi profesorju slovenščine. Dala sem pa [v reševanje] določenemu številu drugih profesorjev (spet ni reprezentativen vzorec, delam na tem), se pravi strokovnjakom drugih področjih, in to univerzitetno izobraženih strokovnjakov. Verjemite, da smo blizu med 20 in 30 % rešenega testa. Moramo se torej vprašati, kaj je sploh namen takega preverjanja in šele posledično, kako spet spremeniti – če že – karkoli v šolskem sistemu pri podaji. Hvala lepa.

Špela Arhar Holdt: Zastavim vprašanja, ali je kakšna replika? V zvezi s temo, o kateri sta prej govorila Marko in Irena, se mi zdi, da imaš ti največ izkušenj s pripravo slovnice, ki že živijo v razredih. Tukaj je bil narejen korak, ki je na nek način mogoče neintuitiven, zanima me tvoje mnenje. V odsotnosti – če se strinja – sodobne znanstvene slovnice je bila pripravljena pedagoška slovnica in ta pedagoška slovnica je morala na nek način nagovoriti tudi neke vrzeli. Ti si recimo uporabljal tudi korpusno gradivo iz korpusa Šolar, kar je korak proti uporabnikom. In zdaj me zanima, če gre za spremembo koncepta, ki je vseeno nadgradnja? Kaj se zgodi, če tvoj koncept preslikamo nazaj na znanstvena izhodišča? Kaj se ti izkaže, kakšna je po tvojem mnenju logična nadaljnja [slovnična] pot?

Kozma Ahačič: Jaz bi tudi še enkrat poudaril ločevanje med podatki (da ne bom vznemirjal) in teoretičnim opisom. Bistvo vsake slovnice je namreč interpretacija podatkov in potem abstraktizacija te interpretacije, zato da podatke lažje razumemo. Če pogledamo v zgodovino in če se ozremo v svet, je

praviloma razlika med šolskimi in znanstvenimi slovnici velika. Ni nujno, da je to neposredno povezano, se pa strinjam, da je uporabno, če govorijo šolske in znanstvene slovnice vsaj v osnovnih pojmovnih konceptih isti jezik. Slovnica je namreč tudi osnova za izdelavo vseh ostalih priročnikov, zato ker z nekim kvalifikatorjem lahko povemo tudi v slovarju več, kot če bi to morali razlagati na dolgo. Skratka, da bi imela šolska slovnica vpliv na znanstveno slovnico, bi bilo neumno. Dejstvo pa je, da če bo nastala znanstvena slovnica, bodo vse šolske slovnice zajemale iz gradiva, iz podatkov, ki jih bo znanstvena slovnica zbrala, in seveda tudi iz vseh ostalih dostopnih podatkov. Hkrati bi opozoril še na potrebo po neki srednji poti. To je slovnica, ki ob samem opisu opisuje podatke še z vidka norme. To se mi zdi potrebno zaradi tistih uporabnikov, ki normativne podatke pričakujejo in ki nimajo dovolj znanja, da bi se čutili sami sposobni presoje ali pa presoje niti ne želijo. Skratka, splošna slovnica za vsakogar (ne le za mlade, za učenje) bi bila zagotovo tudi zelo dobrodošlo dopolnilo širše zasnovane znanstvene slovnice, za katero je pa nemogoče, da jo napiše en sam – v današnjem času in z današnjimi podatki.

Špela Arhar Holdt: Samo to bi še rada vprašala, ker zdaj gremo počasi, da imajo ljudje malo več časa za debato. Rekli smo, da je ena stvar, o kateri bi se danes res radi pogovarjali, kdo so deležniki, kdo bo uporabljal te podatke. Če ta projekt pripravi podatke, te vprašam kot vodjo programske skupine:²¹ se ti zdi to nekaj, kar bi na inštitutu lahko uporabljali za pripravo nove slovnice? Ali se ti zdi, da je vseeno?

Kozma Ahačič: Verjetno ne samo za pripravo nove slovnice. Tudi pri vseh ostalih raziskavah, pri vseh ostalih aktivnostih. Vsaki novi interpretirani podatki so v slovenskem prostoru, kjer tega res ni veliko, velik plus. Še toliko večji, ker je tega dejansko premalo.

Špela Arhar Holdt: Simon, imaš repliko?

²¹ <https://isjfr.zrc-sazu.si/sl/programi-in-projekti/slovenski-jezik-v-sinhronem-in-diahronem-razvoju-1#v>

Simon Krek: Ja, imam par stvari za povedati. Začel bom pa od zadnjega konca. Do zdaj v resnici nisem slišal, da bi kdorkoli v tem raziskovalnem prostoru deklariral, da bo naredil novo slovnico. Tako da prvič slišim, da bo ta raziskovalni program dejansko naredil slovnico. Mislim, to me veseli.

Špela Arhar Holdt: Če je bilo to rečeno, ja.

Simon Krek: Ne samo, a ne.

Kozma Ahačič: Ni bilo rečeno.

Simon Krek: Aha, ni bilo rečeno. Ker je bilo rečeno: »Ne samo za slovnico, ampak tudi za druge stvari.« Jaz sem nekako razumel, da je to potem prva stvar ob vseh ostalih zadevah. Okej, potem to ne drži. Se pravi, se nihče v resnici še ni javil, da bi naredil slovnico.

Druga stvar je ta. Meni se zdi, da govorimo malo mitološko, v smislu: »Imamo šolsko slovnico, potem je znanstvena slovnica, ki je čisto nekaj drugega in zapletena in grozno, in delaš jo trideset let in tako naprej.« Treba se je vprašati, kaj naj bi ločilo šolsko in znanstveno slovnico. Po mojem mnenju enostavno to, kar znanost je – da je osnovana na empiriji in na tem, da so ponovljivi preizkusi, podatki in tako naprej. Se pravi, da lahko z gotovostjo rečeš, da če bo kdo ponovil tvoj poizkus, recimo da bo rekel: »Zdaj bom pa tam našel ta in ta slovnčni pojav,« ga bo v podatkih verjetno našel. Če ne, je tisti znanstvenik lagal. To je znanstvena slovnica, vsaj po mojem mnenju naj bi bila. Od pedagoške pa jo loči to, da se v šolskem prostoru ni treba ukvarjati s preverjanji preizkusov. In v resnici morajo vse šolske slovnice lagati, zato ker idealizirajo neko situacijo in vse robne pojave zanemarjajo, da bi podali osnovno informacijo o vseh jezikovnih pojavih. To je na nek način pozitivno laganje z dobrim namenom, zato da sporočaš in v edukacijskem sistemu pokriješ neko področje. Tako jaz razumem to razliko.

In zdaj smo dobili še eno slovnico, ki se ji reče 'normativna slovnica'. Na kar bi reagiral: to zame absolutno ni srednja pot, ampak je ekstremna pot. Ker je (vsaj

v našem prostoru) bilo vprašanje, kje se pravzaprav norma dogaja. Do zdaj smo bili prepričani, da se dogaja znotraj pravopisa, pa znotraj pravil, ki pokrivajo še malo slovnice, potem se malo dogaja po slovarjih. Že SSKJ je bil deklariran kot informativno-normativni slovar, zdaj pa je novi SSKJ sploh deklariran na ta način. Kar pomeni, da bomo imeli vse troje: normativno slovnico, informativno-normativni slovar in še pravopis kot ultimativno normo. S tem po mojem začnemo na hujši način ločevati jezikovno realnost in to, kar je predpisano.

Kozma Ahačič: Jaz mislim, da ta pogled izhaja iz časov, ko je bila norma nekaj, kar je oteževalo življenje. Mislim, da so ti časi mimo. Nihče več, razen nekaj zapoznelih jezikoslovcev, ne bo več nikogar javno napadal, če bo normo zavestno kršil. Namen norme v sodobnem času je samo – ne normirati, ampak povedati, kakšna raba se zdi ob izbiri različnih možnosti najmanj zaznamovana, najbolj primerna za splošnega uporabnika jezika. Nikakor pa ni mišljena norma kot nek zakon, že dolgo. Saj imamo tu tudi nekaj pravopiscev, ki to že več let razlagajo in se je že zelo prijel. Obsedenost z jezikoslovcem, ki rohni, ni več sodobno stanje.

Simon Krek: Samo eno vprašanje. Če imamo institucijo, ki dobi za svoje delovanje približno milijon evrov na leto in vodja te inštitucije razlaga o brezvestnih uporabnikih slovenščine, [tega] ne morem videti drugače kot ločevanje zrnja od plev – šestih uporabnikov jezika, ki ga znajo uporabljati, in vseh ostalih, ki po njem tacajo na vse mogoče načine. To ni pozicija, ki bi bila karkoli, ampak je pozicija znotraj določene družbe v določenem prostoru v določenem času.

Špela Arhar Holdt: Če je replika, bi z veseljem. Ni? Rok, tebe sem tudi videla nekaj odkimavati. Splošno odkimavanje [*sme*h].

Rok Žaucer: Kar se tiče norme, mislim, da je mogoče pogled, ki ga je Kozma zagovarjal, v jezikoslovju. V družbi na splošno nikakor ne in žal tudi nikoli ne bo. Tudi v jezikovno bolj liberalnih deželah obstaja norma kot nekaj, kar je pri

delu govorcev nekaj, na podlagi česar se druge blati. To je obstajalo in bo zmeraj obstajalo – pri jeziku, pri oblačenju in pri vseh drugih stvareh.

Pri idealizaciji in laganju mi je prišlo na misel: pedagoška slovnica je v nekem smislu poenostavljanje in v tem smislu laganje. Ampak veliko stvari je v nekem smislu laganje in v nekem smislu idealizacija. Čomskijanstvo je nekaj, čemur se je zelo očitala idealizacija – da jemlje govorca in ne upošteva njegove *real-time* jezikovne produkcije, in to je bil velik problem. Korpusi delajo veliko idealizacijo. Če delamo slovnico na podlagi korpusa, bomo izpovprečili in dobimo nekaj, za kar lahko rečemo: »To je standard, zato ker je povprečje.« Ampak idealizacija je v tem smislu, da če bomo vprašali posameznega govorca: »Ali sprejmeš to kot nekaj, kar je v slovenščini okej?« bomo dobili stvari, kjer bodo rojeni govorce slovenščine rekli: »Ne.« Povsod pač pridemo do zadržkov. Ne pravim in nikakor ne mislim, da je zato treba korpusse vreči skozi okno. Niti pod razno, korpusi so zelo uporabni. In korpusnih podatkov je premalo; tudi v obstoječih slovenskih korpusih sem iskal stvari, ki jih nisem našel, pa sem jih brez težav našel z Googlom. Zdaj je milijonkrat oziroma neskončno boljše, kot je bilo v časih Pleteršnika, kot je bilo rečeno. Kar pa ne pomeni, da Pleteršnik ni tudi delal empirično.

Špela Arhar Holdt: To se mi zdi super, zdaj moram preveriti pri publiki – malo smo čez, ampak če ful uživate, mogoče gremo lahko še eno rundo replik. Uživanje lahko nakažete z razbrzdanim ploskanjem. [smeh] Vseeno moram še vprašati, ker absolutno nismo odgovorili, kam gremo, sploh ne. Povedali smo malo o tem, kje smo, malo o tem, kam bi bilo dobro zaviti. Zdaj bom vsakega posebej izzvala z istim vprašanjem: kdo bi po vašem mnenju moral uporabiti podatke, ki jih bomo naredili, zato da pripravi konkreten slovnični opis? Kakšna bi morala zgledati ekipa in kako bi moralo to potekati? Ampak na hitro. Zanima me. Simon, lahko ti začneš, ni problem.

Simon Krek: Kar mi izgleda zdaj po tej debati (eno uro pa petnajst minut) – če zelo na kratko povzamem: pedagoški del lahko pustimo komercializmu in

voluntarizmu, ker bodo wikiiji nastajali, založbe bo pedagoški del zanimal, ker lahko dobijo povrnjene stroške za pedagoške slovnice. Računalniški del se bo zgodil po nujnosti sam po sebi, v resnici. Za znanstveni del pa se nobeden ne javlja in nikogar ta hip ne briga. To bi bil moj povzetek.

Špela Arhar Holdt: Okej, Rok?

Rok Žaucer: Kdo bi to moral uporabiti?

Špela Arhar Holdt: Ja, koga bi ti, recimo, videl v tem prostoru? Kdo bi bil, kaj bi bilo optimalno, kaj se ti zdi?

Rok Žaucer: Prva stvar je spet vprašanje, kakšno slovnico. Če se to uporablja za izdelavo nove slovnice, o kakšni slovnici govorimo. Če rečemo, da je zadnja v slovenščini napisana slovnica, ki se jo tipično označuje za znanstveno slovnico (pa nočem reči, da ni, ampak ni mi čisto jasno, kaj koncept znanstvene slovnice je), stara že skoraj petdeset let, potem je mogoče čas, da se naredi nekaj novega. Kdo bi to lahko naredil? Če drug ne bo, bi se po neki službeni dolžnosti, kar se mene tiče, moral tega lotiti ZRC. Službeni dolžnosti v smislu, [da ima] v družbi pač tako pozicijo, da bi to bilo prav. S tem, ko sem jim to naložil, nočem reči: »Saj imate finance, nimate financ.« V to se ne vtikam. Čisto možno, da bi za to rabili druge finance, ampak če kdo to je – to je tam.

Špela Arhar Holdt: V eni sekundi sem bila prepričana, da se boste javili - generativa [*smeh*]. Okej, Kozma.

Kozma Ahačič: Zanimivo je, da se niti ruska korpusna slovnica ne dela z namenskimi financami in mislim, da bo tudi odločitev za slovensko znanstveno slovnico verjetno v večji meri sad odločitve neke skupine ljudi, da gre v to na prostovoljni bazi, kot pa sad nekega projekta; ker spraviti slovnico v projektni razpis, tudi če se ga razvleče na pet let, je operativno nemogoče. Sploh, če bi pri njej sodelovalo več ljudi. Jaz ne bom rekel, naj naredijo drugi. Rekel bom: če bi imel čas, ko bom imel čas, bi se tega z veseljem lotil. Ampak zaenkrat, v naslednjih letih, ne vidim te možnosti. Bi pa rekel še to. Zdi se mi zelo prav za

naš prostor, da skušamo vstopati vanj z rezultati, ne s sanjanjem, kaj bi bilo dobro, sicer mine deset, dvajset, trideset let in se dejansko nič ne premakne. S tega vidika mislim, da bo ta projekt v vsakem primeru plus.

Irena Krapš Vodopivec: Kakšno slovnico naj bi rabili? Sem prav razumela?

Špela Arhar Holdt: Ne, kdo jo bo naredil.

Irena Krapš Vodopivec: Bom šla s tega, kakšno. Mislim, da rabimo eno slovnico, ki naj bi nastala na podlagi korpusnih podatkov. Tukaj bi se po moji presoji združili vsi, tudi če so bili različnih mnenj, zato ker je v resnici idealizacija (kot je rekel Rok) manjša, če izhajamo iz ene ogromne baze podatkov, kot če izhajamo iz petih jezikoslovcev, ki se odločajo za najbolj optimalen opis. Se pravi: eno slovnico in potem pustiti ... Profesor slovenščine naj bi bil po končani fakulteti dovolj izobražen, da bi bil sposoben interpretirati to slovnico in tudi naučiti dijake jo v pravi meri uporabljati. Če bi bilo po moje, ne bi preveč ločevala na znanstveno in pedagoško in tako naprej slovnico.

Špela Arhar Holdt: Hvala, še Marko.

Marko Stabej: A sprašuješ o standardni ali kakšni?

Špela Arhar Holdt: O naslednji. Sprašujem o prioriteti in o tem, kdo.

Marko Stabej: Jaz bi tako rekel. Seveda se prepoznam v tem mečkaštvu, deset let razmišljam, pa nič ne narediš. Ampak [vprašljiv] je tudi statement, da nekaj narediš, potem pa pogledaš: »Kaj sem pa pravzaprav naredil?« Tako se marsikaj *naredi*, ampak je vprašanje, kaj se s tem *stori*. Meni se zdi (razbralo se je, da ni samo moje mnenje tako): ideja o znanstveni slovnici z velikim Z trenutno ne more biti drugega kot neplemenita laž, če že lahko rečemo, da je pedagoška slovnica plemenita laž. Kar kaže vtis, da jezikoslovci ne znamo narediti znanstvene slovnice brez povezovanja z drugimi strokami, če to hočemo. Zato [sprašujem], kakšno. Standardnega jezika mogoče še in bi se strinjal z Ireno, da jo lahko vidiš kot kontinuum. Da pa pojasni jezikovno dejavnost v raznolikosti, ko enkrat imaš razgled nad to raznolikostjo ... Ne vem,

zdej imamo tudi spontani govor, [za tem] ko je jezikoslovje dolgo časa z gnusom zavračalo kakršenkoli podatek in poskrbelo, da ga ja ne bi bilo, da ne bi mi [raziskovali] jezika ulice in tako naprej. Če rečeš: »Nekaj bom predstavil kot znanstveno slovnico,« padeš v naslednjo jezikovnonačrtovalno past. Staro past SSKJ-ja, da nekaj mora imeti vsaka slovenska družina, ampak nikoli nihče razen jezikoslovcev, lektorjev pa mogoče še koga notri ne pogleda. Imeti znanstveno slovnico, ki bo znanstvena slovnica z zelo velikim Z, pa ne bo pravzaprav ničesar znanstveno pojasnila in ne bo naredila ničesar dobrega za uporabni namen, ker tega ne more, kot je rekel Rok ... So zelo različne potrebe različnih publik, zato se mi zdi, da se nima smisla delati, da to bo.

Špela Arhar Holdt: Okej, hvala, zdaj pa publika, prepričana sem, da je ogromno vprašanj ... Vem, da ste utrujeni, izgledate ful utrujeni ... Jaz sem kar uživala, moram priznati; ni bilo tako zelo grozno. Ampak zdaj je priložnost, da pač poveste oziroma vprašate, komentirate ... Ja, Vojko.

Vojko Gorjanc: Jaz bom pravzaprav povedal svoje mnenje in se ne bom strinjal s kolegico Ireno Krapš Vodopivec, da rabimo eno slovnico. Meni se zdi, da (kar je kolega Žaucer omenil na začetku), rabimo veliko različnih. Predvsem slovenski prostor rabi veliko različnih, zato ker je navajen na en opis, ki v trenutku deluje avtoritativno – in se ga kot avtoritativnega tudi zlorablja. Zato se mi zdi, da je prvi korak k različnosti in šele potem k neki skupnosti. Zdi se mi, da je to boljši princip, kot pa načrtovanje nekega velikega projekta – ki bi si ga pravzaprav želel, ampak se mi zdi, da bi bilo tako za vse nas boljše, oddahnili bi si od tega vsemogočnega, absolutnega.

In drugo, mogoče samo na kratko o vprašanju slovnice, jezikovne kompetence, šole, spoznavanja jezikoslovja. To res lahko gre z roko v roki, samo če se pametno tega lotimo. Izgradnja jezikovne kompetence nikoli ne izključuje jezikoslovja, je pa vprašanje točno to, kot je rekel kolega Stabej: katerega in kakšnega jezikoslovja. Kar je kot jezikoslovje trenutno razumljeno v slovenskem prostoru (tudi v slovenskem šolskem prostoru), je nekaj, kar je veliko bolj

konzervativno, kot je bilo v času usmerjenega izobraževanja. V času usmerjenega izobraževanja je bil šolski prostor v resnici veliko bolj odprt za poststrukturalne kritične pristope in nas je naučil tudi kritičnega mišljenja. Kakršnegakoli že takrat, vendarle kritičnega mišljenja. In če si kaj želim o jezikoslovju, je ravno kritično mišljenje in kritični pristop – zato, da opremimo dijake in kasneje študente z jezikoslovnimi znanji, da bodo k vsemu v svetu znali humanistično pristopati, kritično. Zato me veseli izhodišče današnjega celotnega dogodka, to je predstavitev portugalske slovnice, kjer je kolegica Amália Mendes čisto na začetku pokazala strukturo slovnice in obsežno poglavje, ki je sociolingvistično. Ki je zelo kvaliteten, kritičen sociolingvistični opis stanja portugalsčin, in je del slovnice. Kar pomeni, da ko govorimo o slovnici, pravzaprav lahko govorimo o marsičem; in hkrati, ko govorimo o jezikoslovju, se moramo zavedati, o čem vse govorimo. Premalokrat govorimo o vsem mogočem. Govorimo o samo enem tipu jezikoslovja in samo to imamo v glavi, ko govorimo, da bomo naučili dijake tudi nekaj jezikoslovja. Jaz si takega jezikoslovja, tudi potem, ko pridejo na študij, pravzaprav niti ne želim. Veliko raje bi jih imel opremljene s popolnoma drugimi znanji in védenji.

Špela Arhar Holdt: Hvala. Če ima kdo od panelistov slučajno repliko ali kakšen dodaten komentar na to izjavo ... [*prekinitev v posnetku*]

Helena Dobrovoljc: [*prekinitev v posnetku, zapis po spominu govorke*] Mogoče je moja replika manjša zastranitev od sodobnih načrtov, a se mi zdi pomembna. Ko smo ob 90-letnici rojstva Jožeta Toporišiča pripravljali razstavo o njegovem delu, smo predstavili tudi korespondenco, ki do sedaj še ni bila objavljena. Med pismi najdemo tudi prošnjo Antona Bajca, akademijskega slovnicarja in kasneje enega vodilnih slovaropiscev pri SSKJ, z vabilom, da bi se Toporišič [*nadaljevanje zapisa po posnetku*] vendarle pridružil ekipi na SAZU, ki bi morala pripraviti novo slovnico, saj je njihova²² zastarela in ob vsej leksikalni ekspanziji, ki jih je z medijsko revolucijo dosegla, enostavno niso bili

²² <https://fran.si/slovnice-in-pravopisi/42/1956-bajec-kolaric-rupel-solar>

več sposobni v modelu predvojne breznikovske slovnice opisati novega sistema. Toporišič je to, kot veste, zavrnil, tudi zato, ker se je bal, kako bo lahko v okviru kolektivnega dela na SAZU s svojimi idejami tako prodril kot individualist lahko na Filozofski fakulteti. Bil je sam avtor in ne del kolektiva. Je pa zanimivo, da na SAZU-ju potem te želje niso uresničili mimo Toporišiča in da se je celotna strokovna javnost zadovoljila z individualno slovnico, ki pač skozi prizmo enega človeka kaže na jezikovni sistem. Vendar je kljub drugačnim mnenjem, ki prevladujejo, Toporišičeva slovnica izredno šibko normativna, kar se izrecnih prepovedi tiče. On na primer pravi: »Slovničarji že dolgo prepovedujejo [ta pojav].« Ne napiše: »Jaz tudi prepovedujem,« ampak opisuje prevladujoče. Vsi vemo, da so bile veliko bolj izrecno normativne vse nadaljnje aplikacije te slovnice. V veliki želji po – kako bi se reklo – kodifikacijski vnemi, je to potem šlo in v slovarje in v pravopise in tudi v različne lektorske priročnike in še dlje in še dlje. Meni se zdi razmejitev med opisnim in normativnim danes drugotna in to odsevajo tudi novi koncepti, ki smo jih pravopisci že predstavili. [Ne gre] delati »znanosti« iz tega, da imamo nekje dvojnico npr. *Zidanega Mostu* ali *Zidanega Mosta*, temveč ločiti med slovničnimi dejstvi in stilističnimi skladenjskimi napotki.

Moje osebno stališče je, da je problem poststrukturalne slovnice vse bolj problem [opisa] nenadzorovane rabe jezika, torej usklajevanje pisnega jezika z govorjenim. To nam očitno povzroča težave, in še dobro je, da je tako. Potrebujemo pa slovnico, ki bo ob verodostojnih podatkih ponudila utrditev slovničnih konceptov, ne pa slovnične terminologije. In tukaj se mi zdi, da velja: kolikor jezikoslovcev, toliko različnih definicij. [Polstavčni prilastek, polstavek in npr. povedkovnik ...] Kdo od jezikoslovcev danes to [heterogenost] obvladuje? Se pa vendarle nenehno ponavlja, utrjuje. Vprašanje slovničnega koncepta, mislim, je temeljno vprašanje tudi za pedagoško slovnico. In različne slovnice, ki služijo različnim namenom. Tukaj se strinjam z Vojkom. Hvala.

Špela Arhar Holdt: Hvala za mnenje in mogoče tudi implicitno vprašanje, če se je kdo pripravljen odzvati. Ne? Ekipa je utrujena, vsi želijo domov ... Če ima

še kdo kakšno vprašanje, z veseljem, bom kar vztrajala. Edina možnost je pobeg [smeh]. Ja, Polona.

Polona Gantar: Jaz bi imela dva komentarja na izjavi. Prva je bila mogoče (vsaj jaz sem jo tako razumela) izražena kot bojazen, da govorimo o slovnici, ki je stranski produkt jezikovnotehnoloških potreb. Meni se zdi, da je to zelo preprosto dojemanje te smeri. Izkušnje, kar jih imam, odkar delam s korpusi, pa tudi zdaj pri pripravi podatkov znotraj tega projekta, so ravno obratne. Takrat, ko je jezikoslovec zadovoljen s formaliziranim opisom, ki ga ponudi računalničarju, takrat je ta opis dober. Torej je smer ravno obratna. Drugi komentar, ki je s tem na nek način povezan, je to, kar že tudi srečujem v tem prostoru: konsenz, da je korpusni pristop pravzaprav samo metodologija, ne pa teorija. No, jaz se s tem ne strinjam. Zlasti na področju pomenskih raziskav je ravno korpusni pristop ponudil nove teoretične rešitve. Postavljanje leksikalne gramatičnosti v prvi plan je prav gotovo tak teoretični pristop. Samo toliko.

[zaključek posnetka]

ZAHVALA

Izvedba panela in njegov zapis sta nastala v okviru projekta ARRS J6-8256 *Nova slovnica sodobne standardne slovenščine: viri in metode*. Projekt sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

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/>



OBLIKOSLOVNI VZORCI V LEKSIKONU SLOLEKS: IZHODIŠČNI NABOR ZA SAMOSTALNIKE

Špela ARHAR HOLDT

Center za jezikovne vire in tehnologije

Jaka ČIBEJ

Inštitut "Jožef Stefan"

Arhar Holdt, Š., Čibej, J. (2018): Oblikoslovni vzorci v leksikonu Sloleks: izhodiščni nabor za samostalnike. Slovenščina 2.0, 6 (2): 33–66.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.33-66>.

Prispevek predstavlja prvi korak k dopolnjevanju leksikona Sloleks z oblikoslovnimi vzorci, in sicer na primeru samostalnikov. Vzorci so v prvem koraku strojno pridobljeni iz leksikona na osnovi izbranih razločevalnih lastnosti (oblikoskladenjskih oznak in spremenljivih delov besednih oblik). Sledi ročno razvrščanje, v katerem (a) ločimo sistemsko in v rabi utemeljene vzorce od primerov, ki se pojavljajo spričo šuma pri strojnem luščenju in nedoslednosti v leksikonu Sloleks; (b) uredimo skupine glede na vsebovanost in sorodnost podatkov; (c) poiščemo in natančneje opredelimo variantnost, tako pri standardnih kot nestandardnih oblikah; (č) načrtamo korake za nadaljnji razvoj programa in nadgradnjo leksikona. Rezultat je izhodiščni nabor formaliziranih oblikoslovnih vzorcev za (občno- in lastnoimenske) samostalnike, ki prinaša 10 skupin za moški spol, 9 skupin za ženski spol in 8 skupin za srednji spol. Priprava nabora vzorcev je razkrila številne možnosti za izboljšavo leksikona, strojno namenski pogled na pregibanje pa priložnosti za dopolnitev slovničnega opisa slovenščine. V nadaljevanju dela bodo vzorci pripravljeni tudi za preostale besedne vrste in dopolnjeni s korpusnim gradivom. Končna nomenklatura bo vpisana v bazo leksikona Sloleks, v obliki strojno berljivih vzorcev pa bo objavljena tudi na repozitoriju Clarin.si.

Ključne besede: Sloleks, leksikon besednih oblik, oblikoslovni vzorci, samostalnik, slovenščina

1 LEKSIKON SLOLEKS

Sloleks je odprtodostopni leksikon besednih oblik za slovenščino,¹ ki poleg osnovne oblike besede vsebuje nabor pregibnih oblik, podatke o pogostosti leme in pregibnih oblik iz referenčnega pisnega korpusa Gigafida (Logar in dr. 2012), zbir standardnih in nestandardnih oblikoslovnih variant ter povezave na besedotvorno sorodne besede. Leksikon je bil pripravljen v projektu Sporazumevanje v slovenskem jeziku² po specifikacijah v (Erjavec in dr. 2008), kot je opisano v (Arhar 2009). Trenutno je na voljo popravljena in dopolnjena različica 1.2 (Dobrovoljc in dr. 2015a), ki jo z vidika namena, formata, vsebine in nadaljnjega razvoja natančno opredeljujejo (Dobrovoljc in dr. 2015b).

Sloleks je bil kot vir že večkrat uporabljen pri razvoju jezikovnotehnoloških orodij za obdelavo slovenščine, v prvi vrsti za oblikoskladenjsko označevanje slovenskih besedil (Grčar in dr. 2012, Ljubešić, Erjavec 2016, Čibej in dr. 2016), pa tudi denimo za modernizacijo historičnih slovenskih besed (Scherrer, Erjavec 2013) in normalizacijo slovenskih tvitov (Ljubešić in dr. 2014), za avtomatsko napovedovanje stopnje (ne)standardnosti spletnih besedil (Ljubešić in dr. 2015), za avtomatsko generiranje besednih oblik s pomočjo strojnega učenja (Rejc 2017) ter za luščenje terminologije iz forumskih zapisov (Vintar 2015) in besedil s področja borznega posredništva (Pollak, Božinovski 2014). Leksikon je pogosto rabljen tudi med jezikovnimi uporabniki, ki prek vmesnika iščejo odgovore na jezikovne zadrege, povezane z oblikoslovjem (Dobrovoljc 2015); s tega vidika je bil prepoznan kot dragocen pripomoček za uporabo pri pouku slovenščine (Stritar, Dobrovoljc 2013). V literaturi pa je bilo tudi že opozorjeno, da Sloleks potrebuje nadgradnjo. Ob razvoju koncepta za slovar sodobnega slovenskega jezika (Gorjanc in dr. (ur.) 2015) so bili med informacijskimi dopolnitvami, ki bi tako razvojni kot uporabniški skupnosti najbolj koristile, mdr. izpostavljeni formalizirani oblikoslovni vzorci

¹ Leksikon je dostopen prek vmesnika na spletni strani: <http://www.slovenscina.eu/sloleks>, kot baza pa v repozitoriju Clarin.si: <http://hdl.handle.net/11356/1039> (Dobrovoljc in dr. 2015a).

² Projekt je potekal med leti 2008 in 2013, spletna stran: www.slovenscina.eu.

(Dobrovoljc in dr. 2015b: 95):

Eno najpomembnejših vprašanj, povezanih tako s širitvijo kot reevalvacijo obstoječih oblikoslovnih leksikonov za slovenščino, je izdelava nabora strojno berljivih vzorcev pregibanja besed v slovenskem jeziku, ki bi omogočil validacijo pregibnih paradigem iztočnic v obstoječih priročnikih, pripisovanje paradigem novim leмам ter razvoj metod za njihovo samodejno prepoznavanje v besedilnih korpusih /.../

Odprtodostopni nabor vzorcev, pripravljen namensko za strojno obdelavo slovenščine, bi torej omogočil nadgradnje leksikona in drugih jezikovnih virov ter večjo natančnost označevanja oz. pridobivanja jezikovnih podatkov iz besedilnih korpusov. Na drugi strani bi vključitev vzorcev in lem, ki se po določenem vzorcu pregibajo, v leksikonski vmesnik uporabnikom ponudila pregled nad besediščem, ki se oblikoslovno obnaša primerljivo, kar je izrednega pomena za jezikovno didaktiko – na ravni usvajanja slovenščine kot prvega in tudi drugega oz. tujega jezika. Priložnost za želeno dopolnitev leksikona je prinesel projekt 'Nova slovnica sodobne standardne slovenščine: viri in metode',³ ki ima med cilji tudi razvoj metodologije za slovnični opis slovenščine na ravni oblikoslovja in besedotvorja. Namen prispevka je predstaviti prve projektne rezultate: metodologijo luščenja in ročnega urejanja oblikoslovnih vzorcev za samostalnike, pri čemer je pozornost usmerjena tudi v identifikacijo nalog za vsebinske izboljšave Sloleksa.

2 STROJNA PREDPRIPRAVA PODATKOV

2.1 Metodološko izhodišče

Nabor oblikoslovnih vzorcev za slovenščino, ki ga predstavlja prispevek, temelji na podatkih Sloleksa samega. Iz baze leksikona so s pomočjo v te namene pripravljenega programa pridobljeni kandidati za oblikoslovne vzorce, skupaj s

³ Spletna stran projekta: <http://slovnica.ijs.si/>. Projekt (J6-8256) finančno podpira ARRS (2017–2020), vodja je Simon Krek.

pripadajočim besediščem in opredeljenimi razločevalnimi lastnostmi (oblikoskladenjske oznake in spremenljivi deli besednih oblik). Metoda temelji na smernicah (Dobrovoljc in dr. 2015b: 95–99), ki opredeljujejo tri glavne kriterije za kategorizacijo: (a) vzorci morajo biti strojno berljivi, (b) za optimalno procesiranje pisnega jezika je smiselno ločevati oblikoslovno in naglasno raven in (c) pristop mora temeljiti na jezikovni rabi. Kot je utemeljeno v navedenih smernicah, je za učinkovito strojno obravnavo k vzorcem nujno pristopiti formalistično, z identifikacijo razločevalnih značilnosti iz gradiva samega. Šele v drugem koraku nastopi jezikoslovna obravnava, pri kateri upoštevamo jezikovnosistemske značilnosti. Kot bo razvidno v nadaljevanju (pogl. 5.3), postavi vidik strojnega procesiranja v ospredje druga vprašanja kot jezikovnosistemski pristop, vendar prav sprememba zornega kota prinaša številne novosti, ki jih je mogoče uporabiti tudi za izboljšavo jezikovnega opisa.

Metoda kljub strojnim izhodiščem ni jezikoslovno nepodprta, saj Sloleks temelji na ročnih vnosih, ki na eni strani upoštevajo referenčne priročnike za slovenščino, na drugi pa odstopa od jezikovnega standarda, ki so bili prepoznani pri razvoju slovničnega pregledovalnika Besana.⁴ Strojno razvrščeni rezultati, še bolj pa primeri, ki ostanejo po procesu nerazvrščeni, nakazujejo mesta, ki se jim je pri prihodnjem razvoju leksikona treba posebej posvetiti. Za predstavljeno delo je zato ključen ročni pregled strojno pridobljenega gradiva, ki na primeru samostalnikov⁵ postavlja načela za obravnavo ostalih pregibnih besednih vrst, opredeljuje nadaljnji razvoj programa za luščenje in ob razumevanju sestave leksikona opiše njegova šibka mesta ter korake za njegovo nadgradnjo (pogl. 5.1 in 5.2).

Vnaprej je treba opozoriti, da v prispevku navedeni seznam ni dokončen. Šele

⁴ Ta (kot tudi Sloleks) temelji na leksikalni zbirki Ases, ki je predstavljena v (Arhar, Holozan 2009). Sama metodologija ročnih vnosov v bazo, ki je ena od osnov za Sloleks, v literaturi še ni bila podrobneje predstavljena, jo pa na kratko povzemajo (Dobrovoljc in dr. 2015b).

⁵ Samostalniki so bili izbrani za izhodišče, ker so v leksikonu Sloleks najpogostejše zastopana besedna vrsta – po podatkih iz (Dobrovoljc in dr. 2015b: 84) predstavljajo slabih 54 % iztočnic.

luščenje novega gradiva iz referenčnega korpusa bo omogočilo odločitve, ki jih zgolj na osnovi leksikonskega gradiva ni mogoče dokončno sprejeti. Rezultate je torej treba videti kot korak v razvojnem delotoku – oblikovanje izhodiščnega nabora za luščenja podatkov, s katerimi je nato mogoče nabor nadgrajevati.

2.2 Pridobivanje vzorcev iz leksikona

Pri strojnem pridobivanju vzorcev smo izhajali iz seznama lem, ki so zabeležene v leksikonu, natančneje skupkov leme in oznake za besedno vrsto, npr. *korak_S*, s čimer smo ločili besednovrstno raznolike enakopisne leme (npr. *lev_S* in *lev_P*). V prvem koraku smo za vsakega od tovrstnih skupkov iz leksikona izluščili vse zabeležene besedne oblike in njihove oblikoskladenjske oznake.⁶ Ker v tabelaričnem formatu leksikona Sloleks oblike niso vedno razvrščene v predvidenem vrstnem redu (npr. ednina, dvojina, množina) oz. se red med različnimi lemmi zaradi različnega števila oblik lahko razlikuje, smo v drugem koraku seznam oblik razvrstili glede na kanonični vrstni red oblikoskladenjskih oznak (pri glagolih npr. po osebi od prve do tretje in po številu od ednine do množine, na koncu še velelnik in neosebne glagolske oblike). Primer za samostalnik prikazuje Tabela 1.

Lema čolnar	<i>Ednina</i>	<i>Dvojina</i>	<i>Množina</i>
Oznake in oblike	Somei: čolnar	Somdi: čolnarja	Sommi: čolnarji
	Somer: čolnarja	Somdr: čolnarjev	Sommr: čolnarjev
	Somed: čolnarju	Somdd: čolnarjema	Sommd: čolnarjem
	Sometd: čolnarja	Somdt: čolnarja	Sommt: čolnarje
	Somem: čolnarju	Somdm: čolnarjih	Sommm: čolnarjih
	Someo: čolnarjem	Somdo: čolnarjema	Sommo: čolnarji

Tabela 1: Razvrščene besedne oblike kot podlaga za identifikacijo vzorca.

V tretjem koraku smo za referenčno točko vzeli najkrajšo (oz. prvo najkrajšo)

⁶ Oblikoskladenjske oznake sistema JOS navajava s predpostavko, da njihovo pojasnjevanje ni potrebno. O sistemu označevanja je mogoče več prebrati na <http://nl.ijs.si/jos/msd/html-sl/index.html> ter v (Erjavec, Krek 2008).

besedno obliko in jo strojno primerjali z vsemi ostalimi oblikami v seznamu, s čimer smo identificirali nespremenljivi del besede, ki je vsem oblikam skupen. V Tabeli 2 so spremenljivi deli obarvani rdeče.

Lema čolnar	<i>Ednina</i>	<i>Dvojina</i>	<i>Množina</i>
Oznake in oblike	Somei: čolnar	Somdi: čolnarja	Sommi: čolnarji
	Somer: čolnarja	Somdr: čolnarjev	Sommr: čolnarjev
	Somed: čolnarju	Somdd: čolnarjema	Sommd: čolnarjem
	Sometd: čolnarja	Somdt: čolnarja	Sommt: čolnarje
	Somem: čolnarju	Somdm: čolnarjih	Sommm: čolnarjih
	Someo: čolnarjem	Somdo: čolnarjema	Sommo: čolnarji

Tabela 2: Identifikacija nespremenljivega in spremenljivega dela besednih oblik.

V zadnjem koraku smo vsem oblikam odstranili nespremenljivi del besede in tako pridobili zaporedje spremenljivih delov. Tabela 3 prikazuje vzorec za lemo čolnar, pod isti vzorec pa spada še 1.152 občnoimenskih lem, npr. *direktor*, *davkar*, *enoceličar* in *guverner*.

Primer vzorca	<i>Ednina</i>	<i>Dvojina</i>	<i>Množina</i>
Oznake in spremenljivi deli oblik	Somei: -Ø	Somdi: -ja	Sommi: -ji
	Somer: -ja	Somdr: -jev	Sommr: -jev
	Somed: -ju	Somdd: -jema	Sommd: -jem
	Sometd: -ja	Somdt: -ja	Sommt: -je
	Somem: -ju	Somdm: -jih	Sommm: -jih
	Someo: -jem	Somdo: -jema	Sommo: -ji

Tabela 3: Strojno pridobljen pregibni vzorec za lemo čolnar.

Pri luščenju smo upoštevali tudi oblike, ki so bile v leksikonu označene kot nestandardne (npr. *hči* v tožilniku ednine). Kot take smo jih označili tudi v končnem izpisu pregibnega vzorca, in sicer tako, da smo jim pripisali znak #. Tako nestandardne kot standardne variantne oblike smo ločili z znakom |. Primer podatkov za lemo *hči* prikazuje Tabela 4.

Vzorec za lemo hči	<i>Ednina</i>	<i>Dvojina</i>	<i>Množina</i>
Oznake in spremenljivi deli oblik	Sozei: -era_# -er_# -i	Sozdi: -eri	Sozmi: -ere
	Sozer: -ere	Sozdr: -era -er	Sozmr: -era -er
	Sozed: -eri	Sozdd: -erama	Sozmd: -eram
	Sozet: -er -ero_# -i_#	Sozdt: -eri	Sozmt: -ere
	Sozem: -eri	Sozdm: -erah	Sozmm: -erah
	Sozeo: -erjo	Sozdo: -erama	Sozmo: -erami

Tabela 4: Strojno izluščen vzorec za lemo *hči* z variantnimi (standardnimi in nestandardnimi) oblikami.

Vsakemu tako pridobljenemu vzorcu smo pripisali identifikacijsko številko ter seznam vseh lem, ki mu pripadajo. Končni izpis strojno izluščenih vzorcev je bilo tako mogoče razvrščati po produktivnosti (tj. številu lem, ki se pregibajo po določenem vzorcu), po besedni vrsti in po vsebnosti nestandardnih (#) ali variantnih (|) prvin.

2.3 Pridobivanje tipskega primera

Kot tipski primer oz. zgled za posamezni vzorec smo strojno izvozili podatek o pogostosti posameznih lem v korpusu Gigafida in znotraj vzorca poiskali tisto z najvišjo absolutno pogostostjo. Ta sicer preprosti postopek za veliko večino vzorcev prinese dobre rezultate, tj. splošno in predvidoma razumljivo besedišče, ki ustrezno reprezentira leme določenega vzorca, npr. *čas*, *predsednik*, *delo*, *življenje*, *država*, *možnost* itd. Samo v sedmih primerih smo za tipski primer izbrali lemo, ki ni bila prva po pogostosti. Razlogi so bili, da: (a) predlagani primer ni bil skladen s pisnim standardom (*studijo* zamenjamo s *pončo*); (b) predlagani primer ni bil intuitivno skladen z opisom vzorca (*živelj*, ki naj bi predstavljal enega od vzorcev za nežive samostalnike, zamenjamo z *žajbelj*); (c) predlagani primer (glede na subjektivne ocene) ni bil dovolj besednovrstno asociativen (*hvala* zamenjamo z *nafta*); (č) pri predlaganem primeru so obstajali dvomi glede obstoja alternativne možnosti pregibanja (*Maze* – *Mazeja/Mazeta* zamenjamo z *Brezigar*); ali (d) je bil predlagani

primer kako drugače sporen (*mami*, ki glede na Sloleks obstaja samo v ednini, zamenjamo z *madame*).⁷

3 JEZIKOSLOVNA ANALIZA IN UREJANJE V VZORCE

3.1 Namen in domet urejanja

Namen ročnega urejanja strojno pridobljenih podatkov je: (a) identificirati sistemsko in v rabi utemeljene vzorce in jih ločiti od rezultatov, ki se pojavljajo spričo šuma pri strojnem luščenju in nedoslednosti v leksikonu Sloleks; (b) urediti vzorce v skupine glede na vsebovanost in sorodnost podatkov; (c) poiskati in natančneje opredeliti variantnost, tako pri standardnih kot nestandardnih oblikah; (č) načrtovati korake za nadaljnji razvoj programa in nadgradnje leksikona. Naloga, ki bo opravljena po prerazvrščanju podatkov, pripravi vzorcev za ostale besedne vrste in dodatnem luščenju iz korpusa Gigafida, bo vključevala doslednejše hierarhično urejanje pridobljenih skupin in dokončno poimenovanje posameznih vzorcev.

Analiza je potekala v tabelarni obliki. Za vsak vzorec so bili stolpično uvoženi spremenljivi deli vzorca skupaj z oblikoskladenjskimi oznakami, dodan je bil podatek o številu lem, ki ustrezajo vzorcu, njihov izpis ter tipski primer za vzorec. Ob ročnem pregledu rezultatov je bil v tabelo pripisan kratek opis vzorca z informacijami iz referenčnih jezikovnih priročnikov, ki dodatno pojasnjujejo posamezne dileme ali odločitve.⁸ Primer urejenih podatkov za eno izmed skupin samostalnikov ženskega spola je v Tabeli 5. Na razlike med vzorci opozarjajo siva polja v tabeli.

Vzorec	SZ-1	SZ-1-ednina	SZ1-množina	SZ-1-j
--------	------	-------------	-------------	--------

⁷ Družbeno občutljivost jezikovnih priročnikov (o izzivih trenutnega stanja piše npr. Gorjanc 2017) je nujno zagotavljati na vseh stopnjah njihove priprave. Pri navedenem vzorcu sicer med petimi zajetimi lemmi ni bilo najti dobrega tipskega primera in verjetno je, da bo v nadaljevanju vse gradivo premeščeno med primere, ki se pregibajo v vseh slovničnih številih.

⁸ Glavni vir je bila Slovenska slovnica (Toporišič 2004), pri posameznih dilemah pa so bili podatki preverjeni tudi v slovarju SSKJ2 (Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja, www.fran.si, dostop oktober 2018).

Tip	<i>država</i>	<i>hvala</i>	<i>finance</i>	<i>alineja</i>
Število lem	12.092	865	58	15
Leme	abdikacija, abdukcija, abeceda, abecednica ...	aerobika, aerodinamika, aerostatika, afrikanistika ...	atmosferilije, bakanalije, bermudke, bikinke ...	alineja, aloa, boa, gloria, goa ...
Oblike	Sozei: -a	Sozei: -a		Sozei: -a
	Sozer: -e	Sozer: -e		Sozer: -e
	Sozed: -i	Sozed: -i		Sozed: -i
	Sozet: -o	Sozet: -o		Sozet: -o
	Sozem: -i	Sozem: -i		Sozem: -i
	Sozeo: -o	Sozeo: -o		Sozeo: -o
	Sozdi: -i			Sozdi: -i
	Sozdr: -Ø			Sozdr: -j
	Sozdd: -ama			Sozdd: -ama
	Sozdt: -i			Sozdt: -i
	Sozdm: -ah			Sozdm: -ah
	Sozdo: -ama			Sozdo: -ama
	Sozmi: -e		Sozmi: -e	Sozmi: -e
	Sozmr: -Ø		Sozmr: -Ø	Sozmr: -j
	Sozmd: -am		Sozmd: -am	Sozmd: -am
	Sozmt: -e		Sozmt: -e	Sozmt: -e
	Sozmm: -ah		Sozmm: -ah	Sozmm: -ah
	Sozmo: -ami		Sozmo: -ami	Sozmo: -ami
Opis	Osnovni vzorec za ženske samost. na -a, v rod. -e., prim. 1. ženska sklanj. (SS 289).	Podvzorec SZ-1 za edn. – verjetno legitimna raba dv. in mn. (preveriti korpus). Tipski primer ni najbolj intuitiven.	Podvzorec SZ-1 za mn.	Vzorec SZ-1 za primere, ko se beseda konča na zev (SS str. 290 umešča pod premene osnove).

Tabela 5: Ročno razvrščanje strojno pridobljenih vzorcev.

3.2 Načela urejanja vzorcev

Pri razvrščanju vzorcev so bila oblikovana in uporabljena naslednja načela:⁹

- a) **Načelo nepopravljanja:** Kadar je strojno pridobljeni rezultat posledica leksikonskih pomanjkljivosti ali težav luščenja, se v tabeli rezultatov posebej označi, da gre za problem, sami podatki v leksikonu pa se v tem koraku še ne popravljajo. Glede na vrsto težave določimo, kakšna vrsta rešitve je potrebna v naslednjem koraku. Primere prinaša poglavje 5.2.
- b) **Načelo vsebovanosti:** Vzorce, ki so omejeni na posamezno slovnično število, umestimo kot podvzorec ustrezajočih celotnih vzorcev. Primer kaže Tabela 5, kjer tip *država* predstavlja celovito deblo, tip *hvala* navaja samo edninske in tip *finance* samo množinske oblike.¹⁰
- c) **Načelo sorodnosti:** Če se določen vzorec od drugega v jeziku pogostega vzorca loči v posamezni značilnosti, ki je jasno določljiva in omejena na posamezno obliko (pri čemer pa ne gre za variantnost), ga umestimo v isto skupino. Primer je umestitev tipa *alinea* ob tip *država* (Tabela 5). Načelo sorodnosti tipično uporabljamo za primere, kjer sta vzorca v pregibnih oblikah prekrivna, ločuje pa ju končnica leme (npr. umestitev tipa *dodo* ali *kamikaze* v skupino SM-1, gl. Tabela 7). Nastanek skupin je sekundaren v procesu urejanja: najprej se določi relevantnost posameznih strojno pridobljenih vzorcev, skupina pa nastane, ko so identificirani sorodni vzorci.
- d) **Načelo pogostosti:** V jeziku zelo pogoste in tipične značilnosti vedno

⁹ Pri tem uporabljamo naslednja poimenovanja: *vzorec* je nabor oblikoskladenjskih oznak in spremenljivih delov oblik (kot prikazuje Tabela 3); *podvzorec* je delček drugega, obstoječega vzorca, ki pa je omejen (npr. samo na množino); *skupina* je nabor vzorcev, ki so si po značilnostih zelo podobni oz. sorodni, razlike med njimi pa so jasno opredeljene in zamejene; *izjema* je vzorec, za katerega je mogoče z gotovostjo predvidevati, da je zelo omejen in v sodobnem jeziku ni produktiven (npr. *otrok*).

¹⁰ V podatkih se pojavljajo edninski in množinski podvzorci (ne pa tudi druge možne različice oz. kombinacije), v redkih primerih še podvzorci, ki pokrivajo posamezne oblike (npr. tip *EUR*, ki se sklanja z ničtimi končnicami, ima beleženo obliko samo za imenovalnik ednine).

povzročijo delitev na ločene vzorce. Po tem načelu stopijo v ospredje kot razločevalne denimo kategorija živosti pri moških samostalniki, preglašenost končnic pri moškem in srednjem spolu ter izpuščanje polglasnika osnovne oblike oz. vrivanje samoglasnika pri vseh spolih. Načelo je pomembno, ker se tradicionalno te značilnosti obravnavajo kot premene osnovnih paradig, kar vpliva na strukturiranost in vsebino jezikovnega opisa. Razlike, ki jih (tudi za opis) prinese načelo pogostosti, natančneje prikazuje poglavje 5.3.

- e) **Načelo produktivnosti:** Kot izjeme opredelimo vzorce, ki so izrazito omejeni na posamezne besede, obenem pa je mogoče predvideti, da v sodobnem jeziku niso produktivni (npr. vzorec za samostalnika *otrok* ali *kri*).¹¹ Če je vzorec potencialno produktiven, ga ohranimo v naboru, četudi ima nizko zastopanost (npr. (pod)vzorci, ki pokrivajo posamezne samostalnike *mozeg*, *bezeg* in *mezeg*).¹²
- f) **Načelo specifičnosti:** Pri urejanju vzorcev trenutno ne uporabljamo združevalnih metakategorij. Primer je skupina SM-3 (Tabela 7), v kateri so zbrani vzorci z nepreglašenimi oblikami, pri katerih se v deblu izpušča polglasnik (*meter*, *posel*, *kamen* ipd.). Soglasnik, ki se posledično pojavi v pregibnem delu oblike (*r*, *l*, *n*), glede na dano metodologijo opredeljuje in ločuje vzorec od drugih. Generalizirana oblika (te je mogoče vpeljati naknadno) bi vključevala metakategorijo za nabor soglasnikov in posamezne vzorce združila v enega.¹³
- g) **Načelo enovitosti:** Kadar je izluščeni vzorec sestavljen iz več posameznih identificiranih vzorcev, ga označimo s posebno oznako, ki nakazuje, da je v nadaljnjem procesiranju besedo treba umestiti k vsem

¹¹ Z določeno mero prizanesljivosti, npr. primeri *uho*, *oko*, *igo* so umeščeni med vzorce, ne izjeme, vsaj dokler ne izvedemo preverbe zastopanosti vzorca v referenčnem korpusu. Preverba podatkov v korpusu bo pomagala jasneje ločiti kategoriji vzorcev in izjem – če se bo njuno ločevanje za pripravo končnega rezultata sploh potrdilo kot smiselno oz. potrebno.

¹² V isti vzorec kot *mezeg* bi npr. spadalo lastno ime *Drozeg*, ki v Sloleks ni uvrščeno (v korpusu Gigafida se pojavi v dveh konkordancah).

¹³ 26 (pod)vzorcev v skupini SM-3 je na tak način mogoče strniti v 2.

ustrezajočim vzorcem (več v poglavju 3.3).

- h) **Načelo omejene variantnosti:** Variantnost beležimo, kadar se izbira pojavlja na ravni posameznih oblik, pri čemer so vse vrste tovrstnih pojavitev jasno opredeljene (Tabela 6). V redkih primerih, kjer je posebnosti pri pregibanju veliko, obenem pa se pojavlja sum, da podatki o variantnosti v leksikonu niso dosledno pripisani ali ažurni, so primeri označeni za nadaljnje analize in niso dodani med vzorce (npr. *drog, zid, voz, vol*).
- i) **Načelo ločenih vnosov:** Kadar se dvojnica v leksikonu pojavlja na ravni imenovalnika ednine (npr. *penal/penale*), se predlaga leksikonski vnos ločenih lem.
- j) **Načelo standardnosti:** Variantnosti, ki so posledica nestandardnih oblik (npr. *paparacom/paparacem*), pri pripravi vzorcev zanemarimo. Nestandardne variante beležimo, da jih bo v nadaljevanju mogoče sistematično preveriti v korpusnem gradivu (gl. tudi poglavje 3.3).

3.3 Standardne variante ter dvojni vzorci

Kadar se določena lema glede na Sloleks pregiba po več vzorcih, se po trenutni metodologiji podatki za vse vzorce izpišejo združeno. Pri analizi takšne primere ročno označimo z namenom, da bodo v naslednjem koraku ustrezno strojno prerazvrščeni in bo posledično tovrstnim leмам pripisanih po več ločenih vzorcev. Analiza je pokazala, da se dvojni vzorci pojavljajo pri naslednjih skupinah podatkov: (a) samostalniki, ki se lahko (glede na pomen) pregibajo po paradigmi za živo ali neživo (npr. *tip, nosilec, dvojček; Anton, Diego*); (b) samostalniki, ki se lahko pregibajo s podaljševanjem osnove ali brez (npr. *glas[ov]i, svet[ov]i, mand[e]lj[n]a, okvir[j]a, premier[j]a*); (c) samostalniki, kjer se v zapisu upošteva preglas ali ne (npr. *radiom/radiem*); (č) samostalniki, kjer se lahko polglasnik izpušča ali ne (npr. *meseca/mesca*); (d) samostalniki, ki se lahko pregibajo po paradigmah za različne spole (npr. *DNK-ja* ali *DNK-Ø, ledvica, skripta*); (e) priimki, ki izkazujejo sklanjatev za moški in ženski spol (*Potočnik – Potočnika* vs. *Potočnik – Potočnik*); (f) lastna imena, ki so lahko

različnih vrst, npr. ime ali priimek, osebno ali zemljepisno ime (npr. *Miro*, *Chelsea*, *Pearl*); (g) primeri, kjer lahko pride do različnega podaljševanja osnove (npr. *Arne* – *Arnet*a/*Arne*ja). Pri analizi samostalnikov, ki se lahko pregibajo po različnih vzorcih, so se razkrile tudi določene leksikonske nedoslednosti, h katerim se vračamo v poglavju 5.2.

Oblikovne variante pri pripravi vzorcev beležimo kot (neobvezni) dodatni del v imenu katerega koli obstoječega vzorca. Tako se npr. vzorec za tip *gospodje* loči od vzorca za tip *predsednik* po dodani oznaki za varianto V1 (SM-1(ž) vs. SM-1(ž)-V1, gl. Tabelo 7). Identificirani nabor variant navaja Tabela 6. Tudi na ravni beleženja variant se kažejo številna mesta leksikona, pri katerih bi bilo mogoče podatke urediti in poenotiti, predvsem pa je nujno vključiti preverbo obstoja v korpusnih podatkih in ločiti jezikovnosistemsko utemeljene možnosti od tistih, ki so prisotne v sodobni jezikovni rabi.¹⁴

Varianta	Opis	Primer
Moški spol V1	Imenovalnik množine: -(ov)i -je	<i>gospodi/gospodje</i>
Moški spol V2	Rodilnik ednine, pri samostalnikih, ki izražajo živost, tudi tožilnik ednine: -a -u	<i>mira/miru</i>
Moški spol V3	Mestnik množine, v določenih primerih tudi dvojine: -eh -(ov)ih	<i>gostih/gosteh</i>
Moški spol V4	Orodnik množine: -(ov)i -mi	<i>mostovi/mostmi</i>
Ženski spol V1	Rodilnik dvojine in množine: -Ø -a	<i>vod/voda</i>

¹⁴ Podatki o arhaičnih in stilnih variantah so lahko dragoceni za določene naloge obdelave naravnega jezika in jih nikakor ne gre zanemariti, za naloge, ki so vezane na procesiranje sodobnega in splošnega (če je mogoče tako imenovati jezik, ki ga reprezentira pisni referenčni korpus) jezika pa lahko njihova vsebnost deluje kontraproduktivno. Veljalo bi torej posebej označevati dvojnice, ki v sodobni rabi nimajo več potrditve, obenem pa leksikonske informacije (ustrezno opremljeno z metapodatki) obogatiti tudi s podatki iz korpusov, kot sta IMP, korpus starejših slovenskih besedil (Erjavec 2015), in Janes, korpus računalniško posredovane komunikacije (Fišer in dr. 2016). Slednji je nepogrešljiv vir tudi za nadaljnjo obravnavo nestandardnih oblik.

Ženski spol V2	Rodilnik dvojine in množine: -ac -c	ovc/ovac
Ženski spol V3	Orodnik ednine: -ijo -jo	rebrjo/rebrijo
Srednji spol V1	mestnik dvojine in množine: -eh -ih	sencih/senceh

Tabela 6: Oblikoslovne variante, ki se pojavljajo v identificiranih vzorcih.

3.4 Nestandardne variante

Kot nestandardne dvojnice se v leksikonu beležijo pogostejše težave jezikovne rabe. Pri obravnavi samostalnikov je najti primere: (a) nestandardnega sklanjanja kratic brez vezaja (npr. *CDja*); (b) neustreznega ne/izpuščanja polglasnika na ravni oblike ali leme (npr. *filem*, *ansambl*, *vrteca*, *luknj*, *podlaht*); (c) neustreznega ne/podaljševanja osnove pri pregibanju (npr. *filterja*); (č) neustreznega ne/preglaševanja (npr. *paparacom*); (d) neustreznega tvorjenja osnovne oblike po analogiji s pregibnimi oblikami (npr. *bukva*); in (e) primere regionalnih oblik (npr. *v Prekmurji*). K možnim izboljšavam beleženja nestandardnih oblik se vračamo v poglavju 5.2.

4 REZULTATI

Nabor vzorcev navajava v tabelarni obliki, ločeno za vse tri spole. Občnoimenski in lastnoimenski podatki so predstavljeni skupaj glede na vzorec, ločuje jih podpičje. Trenutno so pri lastnoimenskih samostalnikih moškega spola ločeno navedeni vzorci, ki pokrivajo priimke, in sicer zato, ker se slednji v leksikonu vedno pojavljajo tudi z vzorcem za sklanjanje z ničto končnico za ženski spol. Vzorci so razvrščeni po skupinah, ki so kratko opisane. Za vsak vzorec je navedena (trenutna) kratka koda ter opredelitev, ali je vzorec v leksikonu opredeljen v celoti ali le delno (npr. za ednino ali množino). Sledi navedba števila lem, ki jih vzorec v leksikonu pokriva,¹⁵ ter strojno pridobljenih tipskih

¹⁵ Podatki o številčni zastopanosti se bodo v končni različici povečali na račun prerazvrščenih samostalnikov. Z dopolnjevanjem skupin in popravljanjem nedoslednosti v korpusu se lahko v končni različici spremeni tudi izbira tipskega primera (npr. *konj* namesto redkega *prakonj*).

primerov. V prispevku ni prostora za navajanje celotnih vzorcev (kakor so denimo prikazani v Tabeli 3), vendar je do vseh podrobnosti mogoče dostopati s pomočjo spletne različice leksikona (<http://www.slovenscina.eu/sloleks>).

Koda	Celovitost	Št. lem (O; L)	Tipski primer(i)
1. skupina: Osnovni nepreglašeni vzorci za neživo (n) in živo (ž); posebej sta vzorca za leme na -o in -e . Variante razlaga Tabela 6.			
SM-1(n)	celotna	4.878; 2	čas; Windows
SM-1(n)-V1	ednina	523; 444	promet; Maribor
SM-1(n)-V2	celotna	1	ud
	celotna	4	mir
	ednina	6	sram
SM-1(n)-V2, V3	celotna	2	nos
SM-1(n)-V3, V4	celotna	1	kol
SM-1(ž)	celotna	2.528; 1.122; 434	predsednik; Potočnik; Janez
SM-1(ž)-V1	celotna	32	gospod
SM-1(ž)-V1, V3	celotna	1	gost
SM-1	množina ¹⁶	11; 25	hemoroidi; Helsinki
SM-1o(n)	celotna	98; 2	evro; Yugo
	ednina	18; 42	vaterpolo; Nato
SM-1o(ž)	celotna	25; 125; 75	dodo; Branko; Šukalo
SM-1e(n)	celotna	9	polfinale
	ednina	1; 7	pasodoble; Google
SM-1e(ž)	celotna	1; 21; 8	kamikaze; Stone; Mike

Izjeme: SM-1(ž)-otrok, SM-1(ž)-človek, SM-1(n)-las.

2. skupina: Osnovni preglašeni vzorec za neživo (n) in živo (ž); posebej je naveden vzorec za leme na -o.¹⁷

SM-2(ž)	celotna	853; 790; 53	prijatelj; Majdič; Franc
SM-2(n)	celotna	640	razvoj
	ednina	124; 69	hokej; Kranj
SM-2	množina	5; 34	tisoči; Radenci
SM-2o(ž)	celotna	7	Franjo

¹⁶ Množinske oblike ne izkazujejo razlik na ravni živosti, zato jih po tem kriteriju strojno ni mogoče ločevati.

¹⁷ Različica za leme na -e ni izpričana, čeprav jo je mogoče predvideti za imena tipa *Djordje* (651 pojavitev v korpusu Gigafida). Kot rečeno, bo luščenje podatkov iz korpusa omogočilo identifikacijo in dopolnitev manjkajočih vzorcev.

SM-20(n)	celotna	7	<i>pončo</i>
----------	---------	---	--------------

Izjeme: SM-2(ž)-*prakonj*, SM-2(ž)-*mož*.

3. skupina: Nepreglašeni vzorec za neživo (n) in živo (ž), pri katerem se izpusti polglasnik. Deljeni so glede na soglasnik, ki se zato pojavi v spremenljivem delu.

SM-3k(n)	celotna ednina	829 6; 5	<i>odstotek</i> <i>nameček; Podčetrtek</i>
SM-3k(ž)	celotna	232; 113; 15	<i>deček; Lipovšek; Božiček</i>
SM-3m(n)	celotna ednina	264 207; 1	<i>sejem</i> <i>turizem; Videm</i>
SM-3m(ž)	celotna	1; 1	<i>mikroorganizem; Erazem</i>
SM-3r(n)	celotna ednina	151 9; 6	<i>meter</i> <i>koper; Koper</i>
SM-3r(ž)	celotna	41; 26	<i>minister; Bešter</i>
SM-3ar(ž)	celotna	2	<i>Aleksandar</i>
SM-3ar(n)	ednina	2	<i>Zadar</i>
SM-3l(n)	celotna ednina	41 7	<i>posel</i> <i>Basel</i>
SM-3l(ž)	celotna	7; 7; 2	<i>osel; Rupel; Pavel</i>
SM-3ll(ž)	celotna	2	<i>Rusell</i>
SM-3n(n)	celotna ednina	29 2; 14	<i>kamen</i> <i>česen; München</i>
SM-3n(ž)	celotna	3; 23; 2	<i>oven; Verboten; Domen</i>
SM-3g(n)	celotna ednina	1 1	<i>mozeg</i> <i>bezeg</i>
SM-3g(ž)	celotna	1	<i>mezeg</i>
SM-3t(n)	celotna ednina	1 1	<i>hrbet</i> <i>ocet</i>
SM-3t(ž)	celotna	1	<i>valpet</i>
SM-3s(ž)	celotna ednina	1 1	<i>pes</i> <i>oves</i>

Izjeme: *SM-3r(n)-veter; SM-3r(n)-blagor.*

4. skupina: Preglašeni vzorec za neživo (n) in živo (ž), pri kateri se izpusti polglasnik. Deljeni so glede na soglasnik, ki se zato pojavi v spremenljivem delu.

SM-4c(n)	celotna ednina	405 24; 18	<i>marec</i> <i>svinec; Gradec</i>
SM-4c(n)-V3	celotna	1	<i>konec</i>
SM-4c(ž)	celotna	1.906; 98; 65	<i>igralec; Mavec; Avstrijec</i>
SM-4ac(ž)	celotna ednina	9 2	<i>Badovinac</i> <i>Karlovac</i>
SM-4lj(n)	celotna ednina	52 3; 7	<i>čevelj</i> <i>žajbelj; Bruselj</i>
SM-4lj(ž)	celotna	10; 43	<i>rabelj; Avbelj</i>
SM-4nj(n)	celotna ednina	12 1	<i>ogenj</i> <i>Sovodenj</i>
SM-4nj(ž)	celotna	3	<i>suženj</i>
SM-4š(n)	ednina	1	<i>Mengeš</i>

5. skupina: Vzorec za sklanjanje z uporabo vezaja, pri čemer se uporabljajo preglašene (-ja) in nepreglašene (-a) končnice.

SM-5ja(n)	celotna ednina	62 66; 357	<i>m</i> <i>DDV; BMW</i>
SM-5a(n)	ednina	4	<i>GSM</i> (tudi po SM-5ja(n)-ednina)

6. skupina: Vzorec za sklanjanje z ničtimi končnicami.

SM-6	celotna ednina 'Somei' 'Sometn'	3 15; 131 53 1	<i>mio</i> <i>foto; New</i> <i>EUR</i> <i>poštev</i>
------	---------------------------------------	-------------------------	---

7. skupina: Vzorec za leme na -a ali -ja, ki je enak ženskim vzorcem, ali pregibanje po preglašanih in nepreglašanih vzorcih za moški spol.¹⁸

SM-7ja(ž) ali SM-9ja(ž)	celotna	18; 7; 3	<i>zborovodja; Burja; Mitja</i>
-------------------------	---------	----------	---------------------------------

¹⁸ Vzorce, načeloma enake ženskim, beležimo tudi kot vzorce za moški spol. Enako ustrezna možnost bi bila pripis vzorcev za ženski spol ali nenazadnje oblikovanje poimenovanj tako, da spol samostalnika (ali tudi besedna vrsta, gl. SM-10(ž)) ni ločevalna značilnost. S tem bi zmanjšali število vzorcev, tudi denimo pri beleženju sklanjanja z ničtimi končnicami. Odločitev trenutno še ni jasna, mora pa biti optimalna z vidika strukturiranja podatkov v bazi.

SM-7a(ž) ali SM-1a(ž)	celotna	15; 54; 9	<i>panda; Slana; Miha</i>
SM-7a(ž) ali SM-2a(ž)	celotna	14; 32; 9	<i>kuža; Franca; Matija</i>

8. skupina: Vzorec za neživo (n) in živo (ž), ki v dvojini in množini izraža podaljšavo z *-ova-*. Variante razlaga Tabela 6.

SM-8(n)	celotna	6	<i>sok</i>
SM-8(n)-V2	celotna	11	<i>strah</i>
SM-8(n)-V2, V3, V4	celotna	1	<i>most</i>
SM-8(ž)	celotna	2	<i>bog</i>
SM-8(ž)-V2	celotna	1	<i>tat</i>

9. skupina: Vzorec za neživo (n) in živo (ž), ki podajšuje osnovo z *-j*, *-t* ali *-n*. Variante razlaga Tabela 6.

SM-9j(n)	celotna ednina	501 40; 65	<i>denar</i> <i>humor; Tivoli</i>
SM-9j(ž)	celotna	1.152; 507; 125	<i>direktor; Brezigar; Igor</i>
SM-9t(n)	celotna	1	<i>kofe</i>
SM-9t(ž)	celotna	6; 80; 23	<i>pezde; Blagne; Jože</i>
SM-9t(ž)-V1	celotna	2	<i>oče</i>
SM-9n(n)	celotna	3	<i>buhtelj</i>

Izjeme: SM-9-dan₁ (*dan - dneva*) in SM-9-dan₂ (Somer: *dan - dne*).

10. skupina: Vzorec, podoben pridevniškemu.

SM-10(ž)	celotna	5; 14	<i>moški; Cetinski</i>
----------	---------	-------	------------------------

Tabela 7: Skupine vzorcev za samostalnike moškega spola.

Koda	Celovitost	Št. lem (O; L)	Tipski primer(i)
1. skupina: Osnovni vzorec za samostalnike ženskega spola, ki se končajo na -a. Ločeni so primeri, kjer se lema konča na zev. Variante razlaga Tabela 6.			
SZ-1	celotna ednina množina	12.092; 458 865; 492 58; 181	<i>država; Amerika nafta; Slovenija finance; Jesenice</i>
SZ-1-V1	celotna	9	<i>voda</i>
SZ-1j	celotna	15; 23	<i>alinea; Maria</i>
Izjeme: <i>SZ-1-gospa; SZ-1-Golte</i>			
2. skupina: Osnovni vzorec za samostalnike ženskega spola, ki se ne končajo na -a.			
SZ-2	celotna ednina množina	5.202 22; 5 4	<i>možnost last; Podpeč obresti</i>
3. skupina: Vzorec za leme, ki se končajo na -ev.			
SZ-3	celotna ednina množina	819 1 2	<i>odločitev Lokev Ponikve</i>
4. skupina: Vzorec za samostalnike ženskega spola, ki se ne končajo na -a in imajo v množini v določenih sklonih v spremenljivem delu oblike -e-.			
SZ-4	celotna množina	95 2; 2	<i>stran sani; Ravni</i>
Izjeme: <i>SZ-4-kri; SZ-4-Žiri</i>			
5. skupina: Vzorec za samostalnike ženskega spola na -a, kjer se v rodilniku dvojine in množine vriva <i>e</i> ali <i>i</i> , redko tudi <i>a</i> .			
SZ-5r	celotna množina	64; 6 2; 1	<i>igra; Petra citre; Pekre</i>
SZ-5r-V1	celotna	1	<i>sestra</i>
SZ-5nj	celotna množina	53 4	<i>izkušnja Bitnje</i>
SZ-5lj	celotna množina	34 2; 13	<i>kaplja grablje; Trbovlje</i>
SZ-5lj-V1	celotna	1	<i>zemlja</i>
SZ-5l	celotna množina	33 3; 1	<i>megla orgle; Murgle</i>
SZ-5l-V1	celotna	2	<i>metla</i>
SZ-5m	celotna	28	<i>tekma</i>

SZ-5n	celotna množina	24; 2 3	<i>opna; Vesna Ravne</i>
SZ-5v	celotna	15	<i>spužva</i>
SZ-5j	celotna množina	5; 5 2; 4	<i>ladja; Katja škarje; Nazarje</i>
SZ-5c-v2	celotna	1	<i>ovca</i>
SZ-5k-v2	celotna množina	1 1	<i>deska nečke</i>

Izjeme: *SZ-5r-mati, SZ-5r-hči*

6. skupina: Vzorec za sklanjanje z ničtimi končnicami.

SZ-6	celotna ednina množina	7; 121 5 1; 1	<i>lady; Jennifer madame OI; ZDA</i>
------	------------------------	---------------------	--------------------------------------

7. skupina: Vzorec za samostalnike ženskega spola, ki se ne končajo na -a in vsebujejo izpustljiv polglasnik. V dv. in mn. so končnice z -i- (*bolezni*ma).

SZ-7-en	celotna	12	<i>bolezen</i>
SZ-7-enj	celotna	2	<i>povodenj</i>
SZ-7-el	celotna množina	3 5	<i>misel jasli</i>
SZ-7-em	celotna	1	<i>pesem</i>

8. skupina: Vzorec za samostalnike ženskega spola, ki se ne končajo na -a in vsebujejo izpustljiv polglasnik. V dv. in mn. so končnice z -e- (*ravnema*).

SZ-8-en-v3	celotna množina	2 2	<i>raven Ravni</i>
SZ-8-an-v3	celotna	1	<i>ravan</i>
SZ-8-er-v3	celotna	2	<i>reber</i>
SZ-8-et-v3	celotna	2	<i>lahet</i>

9. skupina: Vzorec, podoben pridevniškemu.

SZ-9	celotna	2	<i>častita</i>
------	---------	---	----------------

Tabela 8: Skupine vzorcev za samostalnike ženskega spola.

Koda	Celovitost	Št. lem (O; L)	Tipski primer(i)
1. skupina: Osnovni vzorec za preglašene samostalnike.			
SS-1	celotna ednina	5-740 197; 41	življenje zdravje; Celje
SS-1-V1	množina celotna	6 1	vratca sence
2. skupina: Osnovni vzorec za nepreglašene samostalnike.			
SS-2	celotna ednina množina	418 469; 17 15; 7	delo mleko; Kosovo vrata; Selca
Izjeme: SS-2-Rova.			
3. skupina: Vzorec za preglašene samostalnike, kjer se v rodilniku dv. in mn. vriva -i- ali -e-.			
SS-3i	celotna	564	podjetje
SS-3e	celotna	5	ozemlje
4. skupina: Vzorec za nepreglašene samostalnike, kjer se v rodilniku dv. in mn. vriva -e-.			
SS-4v	celotna	443	ministrstvo
SS-4l	celotna	21	geslo
SS-4n	celotna	10	okno
SS-4r	celotna množina	7 2	jutro jetra
SS-4m	celotna	2	pismo
Izjeme: SS-4l-tla; SS-4n-dno.			
5. skupina: Vzorec, ki vsebuje podaljševanje osnove s -t-, -n- ali -s-.			
SS-5t	celotna	16	dekle
SS-5n	celotna	11	ime
SS-5s	celotna	2	oje
6. skupina: Vzorec za samostalnike na -o, ki imajo podaljšavo z -es-.			
SS-6	celotna	9	telo
SS-6h	celotna	1	uho
SS-6k	celotna	1	oko

SS-6g	celotna	1	<i>igo</i>
Izjeme: SS-6-črevo.			
7. skupina: Vzorec za sklanjanje z ničtimi končnicami.			
SS-7	'Sosei' in 'Soset'	3	<i>dopoldne</i>
8. skupina: Vzorec, podoben pridevniškemu.			
SS-8o	celotna ednina	9 72	<i>valentinovo</i> <i>Laško</i>
SS-8e	ednina	5	<i>Trebnje</i>

Tabela 9: Skupine vzorcev za samostalnike srednjega spola.

5 VREDNOST PODATKOV ZA NADALJNI RAZVOJ VIROV

5.1 Program za pridobivanje vzorcev

Rezultati kažejo, da program za strojno pridobivanje vzorcev (pogl. 2.2) ponuja dobra izhodišča za nadaljnjo analizo. Možne so izboljšave, ki bodo optimizirale delo za preostale besedne vrste. Trenutno se denimo v podatkih skupaj izpisujejo enakopisni samostalniki različnih spolov (npr. *prst*, *čelo*, *bit*, *tv*), kar je že pri luščenju mogoče obravnavati ločeno, saj gre za ločene leksikonske enote. Kot drugo, težave so na ravni izpisa spremenljivega dela dvojnic: kadar sta varianti pri različnih besedah v leksikonu navedeni v različnem vrstnem redu (npr. *pandov* / *pand* vs. *nadvojov* / *nadvojovov*), program vzorca obravnava kot različna, kar bi bilo mogoče popraviti z dodatnim korakom, ki bi preverjal oz. zanemaril tovrstna zaporedja. Dodati bi bilo mogoče tudi predrazvrščanje delnih vzorcev pod celovite, kot tudi že izhodiščno umeščanje besed v dvojne vzorce, čeprav je pri slednjih primerih koristno, da se v prvem koraku izpisujejo ločeno in analizirajo ročno. Brez dvoma pa je strojno prerazvrščanje nujno razviti za drugi korak obravnave.

5.2 Leksikon Sloleks

Kot je bilo omenjeno na več mestih dosedanje razprave, rezultati pričujočega dela niso zanimivi samo za dodajanje novih leksikonskih informacij (vzorcev),

ampak tudi za urejanje obstoječe vsebine leksikona. Razvrščanje samostalnikov v vzorce in skupine namreč izpostavi primere, v katerih se pojavljajo neregularnosti. Kot stranski rezultat torej dobimo seznam lem, ki potrebujejo pregled. Skupine težav, ki jih razkriva analiza, so naslednje: (a) v trenutni različici Sloleksa je možnost podaljševanja osnov mestoma beležena nedosledno, npr. za razliko od vzorca *glas*, kjer se v dvojini in množini lahko pojavlja podaljšava *-ov-*, imajo *pas*, *sin*, *top* v vzorcu beleženo to možnost samo v dvojini (*pasa/pasova* vs. *pasovi*), *val* in *trak* pa samo v množini (*vala* vs. *valom/valovom*); (b) variantnost je včasih neidentificirana, npr. samostalnik *megla* nima pripisanih končniško naglašениh variant v rodilniku dvojine in množine (beležena je oblika *megel*, ne pa tudi *megla*); (c) kot omenjeno (pogl. 3.3) so težave z nedosledno beleženimi in v rabi neizpričanimi starinskimi ali stilno zaznamovanimi variantami, npr. beleženje dvojnice na *-eh* v mestniku množine moškega spola (*drogeh*, *zideh*, *noseh*), dvojnice na *-mi* v orodniku množine moškega spola (npr. *mostmi*) ali vrivanja samoglasnika *-a-* v rodilniku dvojine in množine ženskega spola (*desak*); (č) mestoma nedosledno je beleženje celovitosti pregibalnih možnosti, npr. ime *Karmen* ima v leksikonu samo edninski del vzorca, čeprav tip *Jennifer* prinaša celotni vzorec); (d) pri več podvzorcih, ki so omejeni na posamezno število, bi bilo mogoče predvideti tudi rabo drugih oblik (npr. tip *hokej*, *Slovenija*, *mami* imajo beležene samo oblike za ednino); (e) v Sloleksu se pojavlja navajanje variantnosti v imenovalniku ednine, ki bi zahtevalo ločene leksikonske vnose (npr. *penal/penale*); (e) sicer redko se pojavljajo napake pri beleženju kategorije živosti samostalnikov moškega spola (npr. *adagio* je umeščen v vzorec za živo); (f) redkejši so tudi lapsusi na ravni vpisa oblik, npr. pri samostalniku *počasnost* se pojavlja med oblikami *prepočasnost*, zaradi česar program izpiše neustrezen vzorec; (g) in nenazadnje se v podatkih pojavljajo določene problematične besednovrstne umestitve, npr. besede *jesti* med samostalnike srednjega spola¹⁹

¹⁹ Umestitev sledi Slovenski slovnici (Toporišič 2004: 301), ki samostalniško rabljene nedoločnike omenja pri 3. srednji sklanjatvi, primer *dobro jesti in piti*. Besednovrstno gre takšne primere v leksikonu obravnavati kot glagole.

ter (h) vsaj v podatkih, ki so na voljo v vmesniku, primeri nestandardno črkovanih lem, pri katerih pa nestandardnost ni označena, npr. *jeterca*, *pluča*.

Nekatere od naštetih težav se pojavljajo sporadično in jih je treba obravnavati ročno, dober delež pa je mogoče urediti sistemsko. Preverba pojavljanja oblik v različnih številih in ne/uporabljenih podaljšav sta denimo nalogi, ki ju je mogoče v korpusu preveriti strojno za celoten nabor ustrežajočih lem. Predvideno je, da bo podatke, pridobljene s strojnimi luščenji za določene prepoznane probleme, treba natančneje analizirati, npr. rabo redkejših dvojnic, ki sovpadajo z v jeziku pogosto rabljenimi oblikami (npr. rodilnik množine *vodá*, ki je v zapisu prekriven z osnovno obliko *vôda*).²⁰

Podatki o nestandardnih oblikah oz. tipičnih odstopih od trenutnega jezikovnega standarda na ravni oblikoslovja so dragoceni za uporabniško skupnost (Arhar Holdt in dr. 2013), vendar so trenutno v leksikonu beleženi zelo sporadično. V nadaljevanju je obstoj nestandardnih dvojnic treba preveriti sistematično pri vseh lemah, ki se pregibajo po določenem vzorcu. V leksikonu navedene nestandardne oblike vzamemo za izhodišče preverbe. Primer sta nestandardni obliki *bukva*, *bukvo* namesto standardne *bukev*; tovrstna nestandardnost je v leksikonu beležena pri dveh samostalniki, mogoče pa jo je iskati za vse samostalnike, ki se sklanjajo po vzorcu SZ-3 (*odločitev*). Kot omenjeno, je za preverbo nestandardne morfologije smiselno vključiti korpus računalniško posredovane komunikacije Janes (Fišer in dr. 2016). V literaturi (Dobrovoljc in dr. 2015b: 100) je tudi že bilo izpostavljeno, da bi natančnejša kategorizacija (nestandardne in standardne) variantnosti omogočila naprednejše povezovanje leksikona z drugimi viri, npr. Slogovnim priročnikom (Krek in dr. 2013). Slednji je zasnovan na osnovi tipičnih uporabniških jezikovnih zadreg in kot tak ponuja informacije in rešitve, komplementarne leksikonskim podatkom, kot tudi slovničnemu opisu, ki se mu posvečamo v

²⁰ Pri analizah je treba imeti v mislih, da oblikoskladenjsko označevanje korpusa Gigafida temelji na podatkih leksikona Sloleks in posledično odraža zgoraj navedene pomanjkljivosti.

nadaljevanju.

5.3 Slovnični opis

Pridobljene podatke je mogoče uporabiti za dopolnitev obstoječega slovničnega opisa oz. razmislek, kako zasnovati slovnični opis, da bo tovrstne izsledke lahko izčrpno in informativno vključeval. Pristop iz gradiva je dragocen, ker prinaša podatke o pogostosti in s tem tipičnosti pomembnih razlikovalnih elementov med vzorci, na osnovi pogostosti izbrane tipske primere, izhodiščno pa zato, ker temelji na avtentičnih podatkih o sodobnem slovenskem jeziku v rabi. Vprašanje prihodnjega slovničnega opisa, ki je seveda kompleksno in si v slovenskem prostoru brez dvoma zasluži več pozornosti, presega domet prispevka; vseeno pa se zdi na tem mestu smiselno izpostaviti nekaj primerov za boljšo predstavo, kako naprej.

Podatke, ki smo jih z opisano metodo pridobili za samostalnike srednjega spola, primerjamo s podatki, ki so na voljo v Slovenski slovnici (Toporišič 2004: 297–301). Slovnično poglavje, ki obsega samo štiri strani in pol, je na vrhnjem nivoju strukturirano po sklanjatvah. Največ informacij je na voljo za prvo srednjo sklanjatev (vzorec *mesto*), začeni z (a) naborom končnic, sledijo: (b) po odstavkih opisane premene osnove (daljšanje osnove s *-t-*, *-n-* in *-s-*; kakovostne premene naglasov na *-e-* in *-o-* v osnovi besede, npr. *srêbru/srébru*; vrivanje *-e-* oziroma *-i-* v rod. dv. in mn. pri določenih primerih); (c) po odstavkih opisane premene končnic (v im. ed. ničta končnica pri besedah, ki podaljšujejo osnovo; preglas za *c*, *č*, *j*, *š*, *ž*; končnica *-eh* namesto *-ih* v mest. dv. oz. mn. (npr. *drveh*); množinski končnici *-emi*, *-mi* (npr. *drvmi*); množinski varianti *blaga-blagovi*) in na koncu (č) naglasi (najprej jakostni po naglasnih tipih, nato pa še tonemski po naglasnih tipih in akutiranih oz. cirkumflektiranih osnovah). V naslednjem podpoglavju o 2. srednji sklanjatvi izvemo, da je ni, 3. sklanjatev z ničto končnico je z nekaj primeri omenjena kot redka (*vremja*), 4. srednja sklanjatev (*Krško*) prinaša končnice za ednino in opombo, da sta množina in dvojina redki, vendar ne nemogoči.

Na drugi strani strojno luščenje podatke loči v osem skupin, v katerih je daleč najpogostejše izpričan tip *življenje*, torej vzorec s preglašenimi končnicami, ki ga slovnica ne navaja eksplicitno. Tip *delo* oz. *mesto*, ki je edini primer, pri katerem slovnica navaja celotni nabor končnic, je od tipa *življenje* glede na število vsebovanih lem skoraj 14-krat redkejši. Preglaševanje se prvič omeni šele med premenami končnic (in niti ne na prvem mestu, prehiti ga pregibanje besed s podaljševanjem osnove, ki so v seštevku od tipa *življenje* redkejše približno 198-krat). Čeprav je v jezikoslovnem smislu preglaševanje mogoče razumeti kot sekundaren pojav, se torej kaže potreba, da slovnični opis – še zlasti, če je slednji pripravljen tudi za jezikovnodidaktične namene – izhaja iz podatkov o sinhroni jezikovni rabi in vsebine predstavlja na način, da so tipične in pogoste značilnosti postavljene v ospredje. Pogostost je pomembna tudi pri drugih vzorcih. Npr. v tretji in četrti najpogostejši vzorec po strojnem luščenju umeščamo primere, kjer se v rodilniku dv. in mn. vrivata samoglasnika -e- ali -i-. Ti vzorci so v slovnici omenjeni pod premenami osnove, vendar na izjemno nepregleden in nesistematičen način (ibid: 298):

Če se osnova samostalnikov končuje na nezvočnik – zvočnik ali na *rj*, *uj*, se v rod. mn./dv. pred (drugi) zvočnik vriva polglasnik, pred *j* pa *i*: *povesmo* – *povesem*, *kraljestvo* – *kraljestev*, *stêgno* – *stêgen*, *jétra* – *jéter*, *sédlo* – *sêdel* proti *obzídje* – *obzídij*, *morje* – *morij*, *nedrje*, *nedrij*, *gorovje* – *gorovij*. Skupna imena na -*je* in s koncem podstave na *n* (*osten-je*), sklop *nj* prav tako razbijajo z *i*: *osténij*. Kadar občutka za tvorjenost ni, tudi ni premene, prim. *korenj*. Tudi večina besed na -*lje* ima v rod. mn. premeno -*lij*: *nasélje* – *nasélj*. – Pri besedah *dnò* in *tlà* se v rod. mn. in dv. vriva *a*: *dán* (poleg običajnejšega *dnòv/dnôv*) in *tál*. Prim. str. 58.

Prednost celovite formalizirane obravnave oblikoslovja je torej urejenost (četudi na prvi pogled razdrobljenih) rezultatov, na osnovi katere je mogoče tudi slovnični opis pripraviti urejeno, ločeno po prepoznanih značilnostih in podprto s sodobnim gradivom. Že uporaba v sodobnem jeziku pogostih zgledov olajša razumevanje obravnavanih slovničnih pojavov. Tako bi preprosteje

zapisali,²¹ da se samostalniki srednjega spola na -o pregibajo drugače, če pred o-jem stoji kombinacija nezvočnika in zvočnika *v, l, n, r* ali *m*; npr. samostalniki *ministrstvo*, *geslo*, *okno*, *jutro*, *pismo*. Pri teh samostalnikih se v rodilniku dvojine in množine med nezvočnik in zvočnik vrine -e- (npr. *geslo* – *gesle*l in ne *geslo* – *gesl*, za razliko od običajnega pregibanja *delo* – *del*, *mesto* – *mest*). Posebnosti pregibanja samostalnikov, ki se končajo na -je, lahko nato opis obravnava ločeno, ker se v resnici obnašajo precej drugače. Prav tako se lahko ločeno obravnavajo specifike naglaševanja.

V primeru, da bodo novi slovnični opisi zasnovani za digitalno obliko, kar je zaželeno oz. pričakovano, je mogoče v besedilo dodati povezave na celotne paradigme za obravnavane zglede, kakor tudi na sezname vseh samostalnikov, ki se pregibajo na v razdelku opisani način. Podatkovna povezljivost na eni in kvantiteta na drugi strani lahko pomembno olajšata razumevanje slovničnih pojavov, sploh za jezikovnodidaktične namene. V smislu povezljivosti je kot zadnjo prednost predstavljene strojne obravnave mogoče izpostaviti sopostavitve lastnoimenskih samostalnikov ob občnoimenske. Vprašanja pregibanja lastnih imen se v Slovenski slovnici sicer pojavljajo, vendar ne sistematično, in velik del vprašanj, zlasti o pregibanju tujih lastnih imen, se prepušča v obravnavo pravopisnim priročnikom. Z vidika uporabnika je zaželeno, da slovnični opis pokrije tudi lastnoimenske podatke, splošnejšo slovnično obravnavo pa je mogoče povezati s problemskim pristopom, kot ga predvideva že omenjeni koncept Slogovnega priročnika (Krek in dr. 2013).

6 SKLEP IN NADALJNJE DELO

Relativno preprost pristop k strojni obravnavi oblikoslovnih vzorcev, kot so ga napovedali (Dobrovoljc in dr. 2015b) v sklopu priprav na novi slovar sodobne slovenščine, se po prvi implementaciji izkazuje za plodnega in učinkovitega,

²¹ Nikakor ni namen članka ponuditi alternativni slovnični opis za katero koli od prepoznanih jezikovnih značilnosti, za slednje je potreben celovitejši razmislek in več jezikoslovnih analiz gradiva kot izhodišče oblikovanim trditvam. Na tem mestu želiva le ponazoriti, kako lahko izgradivna urejenost rezultatov pripomore k urejenosti opisa.

pod pogojem, da mu sledi ročna jezikoslovna analiza, podprta z razumevanjem trenutnega ustroja leksikona Sloleks. V nadaljevanju dela bodo samostalniški vzorci prerazvrščeni in leksikonsko pridobljeni podatki posodobljeni. Z nadgrajenim programom bodo izluščene, nato pa ročno urejene paradigme za ostale besedne vrste. Sledilo bo dopolnjevanje gradiva s podatki iz korpusa oz. korpusov; v prvem koraku je v načrtu uporaba korpusa Gigafida 2.0 (Krek in dr. 2016), postopek je seveda mogoče uporabiti tudi na drugih korpusih, pri čemer bo posebna pozornost namenjena težavam na ravni nestandardnih in redkih arhaičnih oz. stilnih variant. Predvideno je, da bodo za ta del potrebne dodatne jezikoslovne analize, ki bodo opredelile metodologijo luščenja in de facto pojavnost redkih oblik. Po dopolnitvi s korpusnimi podatki bo pripravljen končni nabor vzorcev, ki ga bomo vpisali v leksikonsko bazo, strokovni javnosti pa bo dokumentirani in strukturirani seznam na voljo tudi na repozitoriju Clarin.si. Zadnji korak je dopolnitev leksikonskega vmesnika, ki bo omogočil, da s klikom na izpisano kodo vzorca uporabnik dostopa do zbranega nabora vseh ustrezajočih lem. Od tam je mogoče vzpostaviti tudi povezave na vire, ki lahko določene oblikoskladenjske specifikke natančneje obravnavajo.

ZAHVALA

Predstavljeno znanstvenoraziskovalno delo je rezultat projekta 'Nova slovnica sodobne standardne slovenščine: viri in metode' (šifra ARRS: J6-8256), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

LITERATURA

Arhar, Š. (2009): Učni korpus SSJ in leksikon besednih oblik za slovenščino.

Jezik in slovstvo 54 (3–4): 43–56.

Arhar, Š. in Holozan, P. (2009): Leksikalna podatkovna zbirka ASES (Amebisov skupni elektronski slovar). V V. Mikolič (ur.): *Jezikovni korpusi v medkulturni komunikaciji*: 30–51. Koper: Univerza na Primorskem, Znanstveno-raziskovalno središče, Založba Annales:

Zgodovinsko društvo za južno Primorsko.

- Arhar Holdt, Š., Dobrovoljc, K. in Popič, D. (2013): Reprezentacija standardnega in nestandardnega v virih SSJ. V A. Žele (ur.): *Družbena funkcijskost jezika (vidiki, merila, opredelitve)*: 19–27. Ljubljana: Znanstvena založba Filozofske fakultete.
- Čibej, J., Arhar Holdt, Š., Erjavec, T. in Fišer, D. (2016): Razvoj učne množice za izboljšano označevanje spletnih besedil. V T. Erjavec in D. Fišer (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 40–46. Ljubljana: Znanstvena založba Filozofske fakultete.
- Dobrovoljc, K. (2015): Oblikoslovne informacije v sodobnih slovarskih priročnikih. V V. Gorjanc in dr. (ur.): *Slovar sodobne slovenščine: problemi in rešitve*: 64–79. Ljubljana: Znanstvena založba Filozofske fakultete.
- Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T. in Romih, M. (2015a): *Morphological lexicon Sloleks 1.2*. Ljubljana: Slovenian Language Resource Repository CLARIN.SI, 2015. <http://hdl.handle.net/11356/1039>
- Dobrovoljc, K., Krek, S. in Erjavec, T. (2015b): Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V V. Gorjanc in dr. (ur.): *Slovar sodobne slovenščine: problemi in rešitve*: 80–105. Ljubljana: Znanstvena založba Filozofske fakultete.
- Erjavec, T. in Krek, S. (2008): Oblikoskladenjske specifikacije in označeni korpusi JOS. V T. Erjavec in J. Žganec Gros (ur.): *Zbornik Šeste konference Jezikovne tehnologije: zbornik 11. mednarodne multikonference Informacijska družba - IS 2008*: 49–53. Ljubljana: Institut Jožef Stefan.
- Erjavec, T., Holozan, P., Krek, S., Pivec, M., Rigač, S., Rozman, S. in Velušček, A. (2008): *Specifikacije za leksikon besednih oblik – projekt Sporazumevanje v slovenskem jeziku, kazalnik 3*. Kamnik. Dostopno

- prek: <http://projekt.slovenscina.eu/Vsebine/Sl/Kazalniki/K3.aspx> (2. 12. 2018).
- Erjavec, T. (2015): The IMP historical Slovene language resources. *Language resources and evaluation*, 49 (3): 753–775.
- Fišer, D., Erjavec, T. in Ljubešić, N. (2016): JANES vo.4: korpus slovenskih spletnih uporabniških vsebin. V: D. Fišer (ur.). *Računalniško posredovana komunikacija, Slovenščina 2.0*, 4 (2): 67–994. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Gorjanc, V., Gantar, P., Kosem, I. in Krek, S., ur. (2015): *Slovar sodobne slovenščine: problemi in rešitve*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Gorjanc, V. (2017): *Nije rečnik za seljaka*. Beograd: Biblioteka XX vek.
- Grčar, M., Krek, S. in Dobrovoljc, K. (2012): Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. V T. Erjavec in J. Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije: zbornik 15. mednarodne multikonference Informacijska družba*: 89–94. Ljubljana: Institut Jožef Stefan.
- Krek, S., Dobrovoljc, H., Dobrovoljc K. in Popič, D. (2013): Online style guide for Slovene as a language resources hub. V I. Kosem in dr. (ur.): *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of eLex 2013 Conference*: 379–391. Ljubljana: Trojina, Institute for Applied Slovene Studies; Tallinn: Eesti Keele Instituut.
- Krek, S., Gantar, P., Arhar Holdt, Š. in Gorjanc, V. (2016): Nadgradnja korpusov Gigafida, Kres, ccGigafida in ccKres. V T. Erjavec in D. Fišer (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 200–202. Ljubljana: Znanstvena založba Filozofske fakultete.
- Ljubešić, N., Erjavec, T. in Fišer, D. (2014): Standardizing tweets with

- character-level machine translation. V A. Gelbukh (ur.): *Computational linguistics and intelligent text processing*: 164–175. Heidelberg [etc.]: Springer.
- Ljubešić, N., Fišer, D., Erjavec, T., Čibej, J., Marko, D., Pollak, S, in Škrjanec, I. (2015): Predicting the level of text standardness in user-generated content. *10th International Conference on Recent Advances in Natural Language Processing: Proceedings of RANLP 2015*: 371–378. Hissar, Bulgaria.
- Ljubešić, N. in Erjavec, T. (2016): Corpus vs. Lexicon Supervision in Morphosyntactic Tagging: The Case of Slovene. *Proceedings of Language Resources and Evaluation Conference (LREC) 2016*: 1527–1531. Portorož, Slovenia.
- Logar, N., Grčar, M., Brakuš, 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; Ljubljana: Fakulteta za družbene vede.
- Pollak, S. in Božinovski, B. (2014): Luščenje borzne terminologije. V T. Erjavec in J. Žganec Gros (ur.): *Jezikovne tehnologije: zbornik 17. mednarodne multikonference Informacijska družba*: 114–119. Ljubljana: Institut Jožef Stefan.
- Rejc, R. (2017): Generiranje slovenskih besednih oblik s pomočjo strojnega učenja [diplonsko delo]. Dostopno prek: <https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=slv&id=91151> (2. 12. 2018).
- Scherrer, Y. in Erjavec, T. (2016): Modernising historical Slovene words. *Natural language engineering*, 22 (6): 881–905.
- Slovar slovenskega knjižnega jezika* (2., dopolnjena in deloma prenovljena izd., elektronska objava, 2014). Ljubljana: SAZU in Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. Dostopno prek: www.fran.si (oktober

2018).

Stritar, M. in Dobrovoljc, K. (2013): Korpusi na poti v šole: jezikovnotehnoško izpopolnjevanje učiteljev. *Slovenščina 2.0*, 1 (1): 181–194.

Toporišič, J. (2004): *Slovenska slovnica*. Maribor: Obzorja.

Vintar, Š. (2015): Terminologija v spletnih forumih. V D. Fišer (ur.): *Zbornik konference Slovenščina na spletu in v novih medijih*: 69–74. Ljubljana: Znanstvena založba Filozofske fakultete.

MORPHOLOGICAL PATTERNS IN THE SLOLEKS LEXICON OF SLOVENE: AN INITIAL SET OF PATTERNS FOR NOUNS

The paper presents the first step to expanding the Sloleks lexicon of Slovene with morphological patterns, starting with nouns. In the first phase, the patterns were extracted automatically from the lexicon based on a selection of differentiating characteristics (morphosyntactic tags and variable word parts). This was followed by a manual categorization during which we (a) separated patterns that are either systemic or based on actual language use from examples extracted because of noise attributable to either the extraction method or inconsistencies in Sloleks; (b) arranged patterns into groups based on their content and relatedness; (c) analyzed and more clearly defined form variability, with both standard and non-standard word forms; (d) propose future steps for the further development of the extraction method and lexicon upgrades. The result is a set of formalized morphological patterns for (common and proper) nouns containing 10 groups (64 patterns) for masculine nouns, 9 groups (29 patterns) for feminine nouns and 8 groups (20 patterns) for neuter nouns. The preparation of the set of formalized patterns also resulted in numerous suggestions on how to upgrade the lexicon, while a machine-focused view of morphological flexion offers opportunities to improve the current grammatical description of Slovene. As part of our future work, we intend to expand the set of patterns with other parts of speech and corpus-based material. The final categorization of patterns will be included in the Sloleks lexicon, and the patterns will also be published on the CLARIN.SI repository in a machine-readable format.

Keywords: Sloleks, word form lexicon, morphological patterns, noun, Slovene

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/>



FORMULAIČNOST V SLOVENSKEM JEZIKU

Kaja DOBROVOLJC

Center za jezikovne vire in tehnologije, Institut "Jožef Stefan"

Dobrovoljc, K. (2018): Formulaičnost v slovenskem jeziku. Slovenščina 2.0, 6 (2): 67–95.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.67-95>.

Številne empirične raziskave povezovanja besed v jeziku razkrivajo, da je jezik prepreden z večbesednimi vzorci, ki vsaj na neki točki jezikovne rabe delujejo kot nerazstavljiva celota. To potrjujejo tudi dosedanje raziskave večbesednih enot v slovenščini, ki se običajno osredotočajo predvsem na strukturno in/ali pomensko zaključene enote, kot so kolokacije, stalne besedne zveze ali frazemi. Da bi omogočili bolj celovit vpogled v obseg in naravo formulaičnega jezika v slovenščini, ki bi ga bilo smiselno vključevati tudi v bodoče leksikalne in slovnične opise, v članku raziščemo koncept formulaičnih besednih nizov, tj. neprekinjenih nizov dveh ali več besed z izstopajočo pogostostjo pojavljanja, ne glede na njihovo strukturno, pomensko ali zaznavno celovitost. Predstavimo postopek identifikacije formulaičnih besednih nizov v referenčnih korpusih pisne in govorjene slovenščine, jih kvantitativno in kvalitativno analiziramo ter primerjamo. Velik delež identificiranih formulaičnih nizov v obeh korpusih dokazuje določeno stopnjo formulaičnosti, še zlasti v govorjeni slovenščini. Čeprav v obeh oblikah jezikovne rabe prevladujejo predvsem strukturno nezaključeni nizi funkcijskih besed, ugotavljamo več pomembnih razlik v skladenjskih in funkcijskih lastnostih najpogostejših nizov pisne in govorjene slovenščine, ki potrjujejo vpliv specifičnih sporazumevalnih okoliščin in ciljev na razvoj in obseg formulaičnega jezika.

Ključne besede: formulaičnost, besedni nizi, večbesedne enote, pisni jezik, govorjeni jezik

1 UVOD

V starejših jezikoslovnih teorijah je dolgo časa prevladovalo prepričanje, da govorci v svoj mentalni leksikon shranjujejo posamične besede, ki jih take tudi priključijo iz spomina in nato po kognitivno ločenem sistemu slovničnih pravil

združujejo v višje skladenjske enote, kot so besedne zveze, stavki, povedi in besedila. Da kombinatorika povezovanja besed v jeziku ni povsem poljubna, so v svojih razpravah opozarjali že posamezni jezikoslovci (Bolinger 1976; Coulmas 1979; Firth 1957; Pawley, Syder 1983), nove metodološke možnosti, kakršne so se s tehnološkim napredkom vzpostavile v zadnjih dveh desetletjih, pa so povzročile pravi razmah empiričnih – zlasti korpusnojezikoslovnih in psiholingvističnih – raziskav različnih vidikov besedne povezovalnosti, ki dokazujejo, da govorci v svoj spomin shranjujejo precej večje in kompleksnejše enote besedišča kot zgolj posamične besede.

Za sicer izredno heterogeno področje raziskav, ki se na kakršenkoli način ukvarjajo s pojasnjevanjem vzorcev besednega povezovanja v jeziku (za izčrpen pregled glej Wray 2013), se v zadnjem času vse bolj uveljavlja krovno poimenovanje t. i. raziskav formulaičnega jezika (angl. *formulaic language*), ki v svoji najsplošnejši opredelitvi zajema »kombinacije besed, ki vsaj v nekem vidiku niso popolnoma predvidljive, denimo zaradi svojega pomena, ki je bolj ali manj nenapovedljiv na podlagi pomena posamičnih besed, zaradi svoje funkcije, ki jo lahko doseže samo zveza kot celota, ali pa zaradi posebnosti v zgradbi, kot sta netipična morfologija ali besedni red« (prav tam: 317).

Kljub tej široki opredelitvi, preko katere se formulaični jezik vzpostavlja kot nadpomenka frazeologije, idiomatike in drugih sorodnih disciplin, pa so razvoj področja najbolj zaznamovala predvsem spoznanja raziskav, ki se osredotočajo na analizo statistično izstopajočih, formulaičnih besednih nizov,¹ ne glede na njihovo pomensko, strukturno ali zaznavno celovitost, kakršno v ospredje postavljajo tradicionalne frazeološke discipline. Izhajajo namreč iz predpostavke, da je prav nadpovprečna pogostost rabe tista, ki razkriva njihovo

¹ V skladu z multidisciplinarno naravo raziskovalnega področja poleg najpogostejših izrazov formulaični niz (angl. *formulaic sequence*) ali leksikalni skup (angl. *lexical bundle*) za prekrivne ali sorodne jezikovne enote pojavljajo še številni drugi izrazi, npr. *chunk*, *cluster*, *formula*, *prefabricated routine*, *prefab*, *ready-made utterance*, *lexical phrase*, *institutionalized utterance*, *sentence frame*, *lexical frame*, *lexicalized sentence stems*, (*language*, *lexical*, *syntactic*, *prefabricated*) *pattern*, *speech formula*, *conversational routine*, *recurrent string* itd.

posebno vlogo v usvajanju, razumevanju in tvorjenju jezika (Biber 2009; Cortes 2015; Wray 2002).

To hipotezo posredno ali neposredno potrjujejo številne raziskave, ki z različnimi metodološkimi pristopi ugotavljajo, da formulaične besedne kombinacije govorniki procesirajo hitreje in na drugačen način kot poljubne kombinacije besed (Conklin, Schmitt 2012; Lin 2010; Tremblay in dr. 2011). Po nekaterih ocenah naj bi delež vnaprej pripravljenih besednih kombinacij predstavljal kar med tretjino in polovico človeškega diskurza (Conklin, Schmitt 2012: 2–3), pri čemer opredelitev razmerja med sprotno tvorjenim in vnaprej pripravljenim diskurzom (Erman, Warren 2000; Sinclair 1991) ostaja eno ključnih odprtih vprašanj področja (Wray 2012).

Ni torej presenetljivo, da koncept formulaičnosti pronica tudi v nove generacije slovnčnih teorij, ki z zavračanjem tradicionalnega ločevanje jezika na sistem pravil (slovnico) na eni strani in enot pomena (leksikon) na drugi večbesedne enote iz obrobja jezikovnega opisa postavljajo v njegovo središče, kot so denimo leksikogramatika (angl. *lexicogrammar*, Halliday 1985), slovnica globinskih sklonov (*case grammar*, Fillmore 1968), shemska semantika (*frame semantics*, Fillmore 1982), konstrukcijska slovnica (*construction grammar*; Goldberg 2006), slovnica vzorcev (*pattern grammar*; Hunston, Francis 2000), teorija leksikalnega proženja (*lexical priming*; Hoey 2005) ali teorija konvencij in invencij (*theory of norms and exploitations*; Hanks 2013), če jih naštejemo le nekaj. To velja tudi za slovnčni opis ožjega področja formulaičnih besednih nizov, ki so denimo obravnavani v samostojnem poglavju Longmanove korpusne slovnice za angleščino (Biber in dr. 1999).

Čeprav se s porastom korpusnih raziskav večbesednih enot, kot so analize stalnih besednih zvez (Gantar 2007; Gantar 2015), oblikoskladenjskih nizov (Arhar Holdt 2011), kolokacij (Kosem in dr. 2013) ali pomenskovezljivostnih vzorcev (Krek in dr. 2016), ti trendi kažejo tudi v slovenskem prostoru, koncept formulaičnosti v slovenščini z izjemo razprav o posameznih funkcijskih

skupinah formulaičnih izrazov, kot so pragmatični frazemi (Jakop 2006), komunikacijski ali sporočanjejski frazemi (Stramljič Breznik 2001) ali diskurznofunkcijski nizi (Dobrovoljc 2018a), doslej še ni bil podrobneje raziskan. Cilj pričujoče raziskave je zapolniti to vrzel ter s sistematično analizo obsega in narave formulaičnih besednih nizov v slovenščini ponuditi empirično izhodišče za nadaljnje obravnave tega kompleksnega jezikovnega pojava, tudi v kontekstu načrtovanja novega slovnicega opisa slovenščine.

2 METODOLOGIJA

Glede na splošno neraziskanost koncepta formulaičnosti v slovenskem jezikoslovju in aktualnost tega raziskovalnega področja v mednarodnem prostoru smo s pričujočo raziskavo želeli odgovoriti na naslednja raziskovalna vprašanja:

- a) Ali je slovenščina formulaična?
- b) Kako formulaična je pisna slovenščina v primerjavi z govorno?
- c) Kateri so najpogostejši formulaični izrazi v pisni in govorni slovenščini?

Za ta namen smo v skladu z izbranim popolnim korpusnim pristopom (Biber 2009) v referenčnih korpusih pisne in govorne slovenščine (razdelek 2.1) z metodo prilagojenega štetja (razdelek 2.2) identificirali formulaične besedne nize in najpogostejše med njimi jezikoslovno ovrednotili (razdelek 2.3).

2.1 Gradivo

Uravnoteženi korpus sodobne pisne slovenščine Kres (Logar Berginc in dr. 2012) vsebuje približno 100 milijonov besed, ki predstavljajo besedilno uravnoteženo podmnožico krovnega korpusa Gigafida v obsegu milijarde besed, in sicer korpus Kres sestavljajo leposlovna besedila (17 %), priročniki in druga stvarna besedila (18 %), časopisi (20 %), revije (20 %), spletna besedila novinarskih portalov, podjetij in ustanov (20 %) ter drugi tipi besedil, kot so prepisi sej državnega zbora ali podnapisi (5 %). V raziskavi smo uporabili

različico 1.0, ki vsebuje besedila, nastala med letoma 1990 in 2011, in je za brskanje dostopna na uradni spletni strani korpusa.²

Referenčni korpus govorjene slovenščine Gos (Verdonik in Zwitter Vitez 2011) je največja prostodostopna zbirka posnetkov in transkripcij spontanega govora v slovenskem jeziku. Vsebuje transkripcije približno 120 ur posnetkov (1 milijon besed) spontanega oz. nepripravljenega govora v različnih vsakodnevnih sporazumevalnih situacijah, ki so uravnotežene glede na demografske lastnosti govorcev (spol, starost, regija, izobrazba), prenosnik (TV, radio, telefon, osebni stik) in vrsto govornega dogodka (javni informativni in izobraževalni 34 %, javni razvedrilni 20 %, nejavni nezasebni 15 % ter nejasni zasebni govor 29 %). Korpus Gos vsebuje ročni prepis govora v dveh različicah, pogovornem in standardiziranem zapisu, pri čemer se v tej raziskavi osredotočamo zgolj na standardizirano raven, ki nevtralizira narečno, zvrstno ali drugače pogojene izgovorne posebnosti slovenščine. V raziskavi smo uporabili različico 1.4, kakršna je dostopna preko prostodostopnega konkordančnika noSketchEngine³ (Erjavec 2013) in se od prvotne različice, dostopne preko specializiranega konkordančnika na uradni spletni strani,⁴ ki omogoča tudi hkratno poslušanje izvornih posnetkov, razlikuje v tem, da vsebuje hevristične ročne popravke nekaterih najbolj problematičnih kategorij zapisa, lematizacije in/ali oblikoskladenjskih lastnosti (Verdonik 2014).

2.2 Luščenje

Za identifikacijo najpogostejših formulaičnih nizov v obeh besedilnih korpusih smo uporabili orodje, razvito za luščenje besednih nizov (n-gramov) v korpusih slovenskega jezika (Dobrovoljc 2018b), ki glede na različne nastavljive parametre izdela seznam besednih nizov poljubnega tipa, razvrščenih po pogostosti. Med različnimi vrstami frekvenčnih seznamov, ki jih orodje

² www.korpus-kres.net

³ https://www.clarin.si/noske/run.cgi/corp_info?corpname=gos

⁴ www.korpus-gos.net

omogoča, smo za luščenje formulaičnih besednih nizov po vzoru sorodnih raziskav (Adolphs, Carter 2013; Buerki 2016) uporabili t. i. prilagojeni frekvenčni seznam (O'Donnell 2010). Ta pri štetju n-gramov upošteva medsebojno vsebovanost nizov različnih dolžin in pojavitev krajšega niza (npr. *kljub temu*) šteje le, kadar se ta ne pojavi kot del daljšega relevantnega niza (npr. *kljub temu da*). V primerjavi z običajnim štetjem nizov, pri katerem so daljši nizi vedno enako ali manj pogosti od vsebovanih podnizov, ta metoda omogoča realnejšo primerjavo pogostosti rabe nizov različnih dolžin (kot je denimo ugotovitev, da govorci zvezo *kljub temu da* uporabljajo pogosteje kot zvezo *kljub temu*).

Za oba korpusa smo izdelali prilagojeni frekvenčni seznam neprekinjenih nizov normaliziranih besednih pojavnic dolžine do 5 besed, tj. pojavnic v zapisu z malimi črkami v korpusu Kres in pojavnic v standardiziranem zapisu v korpusu Gos.⁵ Kot merilo za status formulaičnosti niza smo izbrali frekvenčni prag vsaj 20 pojavitev na milijon pojavnic (tj. 1942 pojavitev v korpusu Kres in 20 pojavitev v korpusu Gos), skladno s pristopi sorodnih raziskav formulaičnosti, kjer ta številka običajno variira med 10 (Biber in dr. 1999; Cortes 2015; Simpson-Vlach, Ellis 2010;) in 40 pojavitev na milijon pojavnic (Biber in dr. 2004; Biber 2009). Da bi izničili vpliv nizov, specifičnih zgolj za posamične avtorje ali govorce, smo upoštevali tudi pogoj pojavljanja niza v vsaj 2 različnih besedilih oz. govornih dogodkih, kot to omogoča funkcija določanja minimalnega besedilnega praga.

2.3 Kategorizacija

Najpogostejše formulaične nize, identificirane v vsakem izmed korpusov, smo nato razvrstili glede na različne slovnične kategorije, relevantne za njihovo

⁵Pri štetju v nobenem korpusu nismo upoštevali morebitnih ločil med pojavnicami (združeno štetje za nize '*kljub , temu da*', '*kljub temu , da*' in '*kljub temu da*'), v korpusu Gos pa smo na enak način iz štetja izločili še druge vrste nebesednih pojavnic, kot so zapolnjeni premori (*eee, eem, mmm, nnn, mm, een, nn*) ter označevalci premorov (*[gap], [pause], ...*), intonacije (*!, ?*) in drugih zvočnih signalov (*[voice], [all:laughter], [audience:laughter], [incident], [speaker:laughter]*).

izhodiščno kvalitativno analizo ter nadaljnje metodološke in vsebinske raziskave tega jezikovnega pojava.

Z vidika leksikalne zgradbe so bili formulaični nizi razvrščeni na nize, ki vsebujejo zgolj predmetnopomenske besede, tj. glagole, samostalnike, pridevnike, prislove in okrajšave; nize, ki vsebujejo zgolj funkcijske besede, tj. predloge, veznike, členke, medmete, zaimke in oblike pomožnega glagola *biti*; ter nize, ki vsebujejo tako predmetnopomenske kot funkcijske besede.

Z vidika skladenjske zgradbe so bili formulaični nizi razvrščeni na strukturno zaključene nize, ki predstavljajo skladenjsko celovite strukture, ki jim je mogoče pripisati samostojno skladenjsko vlogo (npr. v *tem primeru*), in strukturno nezaključene nize, kot so nizi z manjkajočimi skladenjsko ali pomensko obveznimi elementi (npr. *da bi se*). Glede na problematičnost določanja podrobnejših skladenjskih lastnosti strukturno nezaključenih nizov je bil vsakemu nizu dodan še robusten opis besednovrstne sestavinske zgradbe po slovnичnem sistemu JOS⁶ (Erjavec, Krek 2008), npr. predlog-zaimek-samostalnik za niz *v tem primeru* ali veznik-glagol-zaimek za niz *da bi se*.⁷

Pri končnem razvrščanju nizov glede na pomen oz. pragmatično funkcijo v komunikaciji smo sledili eni vplivnejših taksonomij za funkcijsko razvrščanje t. i. leksikalnih skupov (Biber in dr. 2004) oziroma njeni nadgrajeni različici, razviti za potrebe opisa formulaičnih nizov v znanstveni angleščini (Simpson-Vlach, Ellis 2010), ki je bila izdelana na večjem obsegu gradiva in ustrežneje naslavlja specifične govornega jezika. Po tej taksonomiji so bili nizi razvrščeni na: nize za opisovanje predmetnosti (angl. *referential expressions*, npr. *to je*

⁶ <http://nl.ijs.si/jos/>

⁷ V formulaičnih besednih nizih se pogosto pojavljajo večfunkcijske besede z visoko stopnjo gramatikaliziranosti, kot so *tako*, *glede*, *zato*, *tudi*, *pa*, *ne*, *a* ali *kaj*, pri katerih določanje besedne vrste predstavlja poseben izziv in pogosto točko razhajanj med obstoječimi slovnичnimi opisi slovenščine. Kljub nekaterim že identificiranim možnostim za nadaljnji razvoj in poenotenje označevalnega sistema JOS (Krek 2010; Dobrovoljc 2018a), smo pri kategorizaciji sledili izvornim smernicam besednovrstne kategorizacije (Holozan in dr. 2008), na podlagi katerih je bil izdelan tudi ročno označeni učni korpus ssj500k (Krek in dr. 2018).

bilo, v skladu z, uradni list RS št.); nize za izražanje stališča oz. odnosa do sporočanega (angl. *stance expressions*, npr. *na nek način, se mi zdi, naj bi bil, ja ne vem*); in nize za strukturiranje oz. upravljanje diskurza (angl. *discourse-organizing expressions*, npr. *kar pomeni da, to se pravi, tako da je, ja ja ja*).

Na vseh ravneh slovničnega opisa smo v primeru dvoumnosti ali večfunkcijskosti na podlagi ročnega pregleda konkordanc nizu pripisali tisto leksikalno, skladenjsko oz. pomensko interpretacijo, ki izhaja iz njegove najpogostejše vloge v danem korpusu. Tipični primer je denimo besedni niz *a ne da*, ki se v korpusu Gos pojavlja v različnih kontekstih rabe, a najpogosteje kot niz diskurznega členka *a ne* in veznika *da*, ki ga lahko opišemo kot strukturno nezaključen niz prirednega veznika, členka in podrednega veznika s pragmatično funkcijo diskurznega upravljanja.

3 REZULTATI

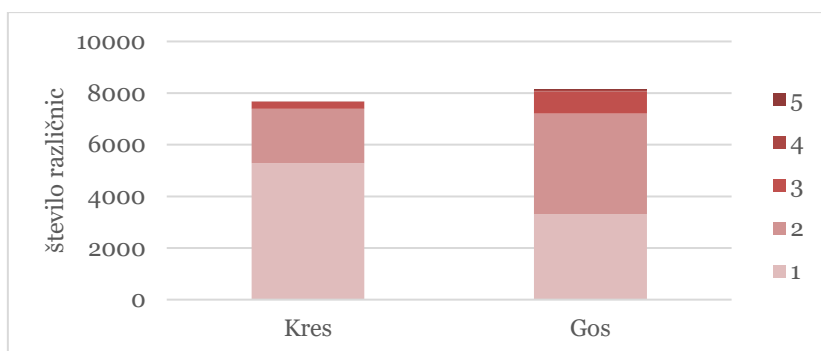
3.1 Delež formulaičnih besednih nizov

Tabela 1 povzema število izluščenih nizov dolžine 1 do 5 pojavnic z minimalno pogostostjo vsaj 20 pojavitev na milijon pojavnic v vsakem izmed korpusov, ki bi jih lahko opredelili tudi kot enote stalnega oz. pogostega besedišča v jeziku (statistično izstopajoči eno- ali večbesedni n-grami).

Št. besed	Kres		Gos	
	Različnic	Pojavnic	Različnic	Pojavnic
1	5.280	45.619.297	3.313	424.763
2	2.109	12.155.355	3.891	244.980
3	277	1.128.313	862	34.028
4	8	25.439	54	2.147
5	0	0	10	511
SKUPAJ	7.674	58.928.404	8.130	706.429

Tabela 1: Število identificiranih formulaičnih besednih nizov v korpusih Kres in Gos.

Primerjava skupnega števila vseh enot na Sliki 1 kaže, da se v obeh korpusih pojavlja približno podobno število različnih enot stalnega besedišča (7.674 v korpusu Kres oz. 8.130 v korpusu Gos), v obeh korpusih pa razmeroma visok delež vseh enot predstavljajo besedni nizi (31,2 % v korpusu Kres oz. 59,2 % v korpusu Gos), kar potrjuje formulaično naravo komunikacije v slovenščini. Še zlasti to velja za govorjeno slovenščino, kjer je število formulaičnih besednih nizov celo večje od števila posamičnih besed. To dokazuje, da je tudi v slovenščini govorjena raba izrazito bolj formulaična kot pisna (prim. npr. Biber in dr. 2004 za angleščino).

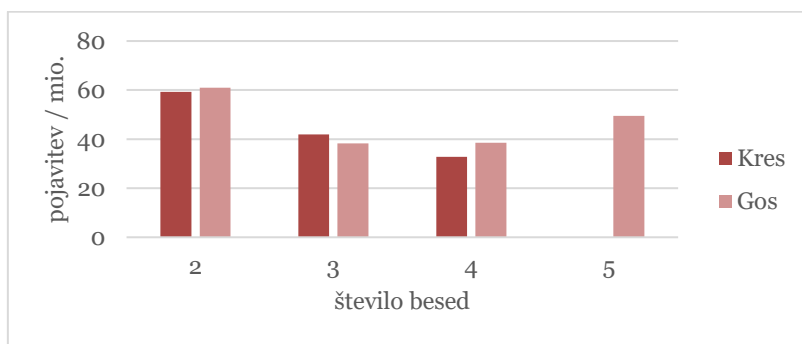


Slika 2: Število različnic na prilagojenem frekvenčnem seznamu korpusov Kres in Gos glede na število besed.

Podrobnejša analiza izluščenih formulaičnih nizov glede na dolžino nadalje

kaže, da med formulaičnimi nizi v slovenščini niso zgolj dvobesedne kombinacije, temveč tudi daljše enote, pri čemer je delež tri- ali večbesednih nizov še zlasti izrazit v govoru (11,9 % vseh besednih nizov v korpusu Kres oz. 19,2 % vseh besednih nizov v korpusu Gos). V korpusu pisne slovenščine Kres petbesednih formulaičnih nizov ni.

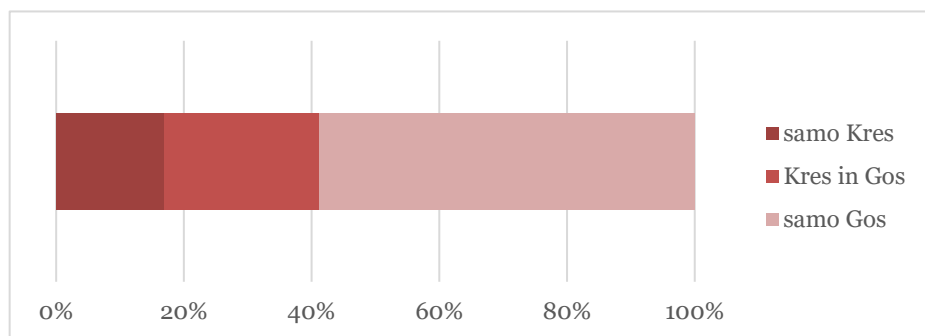
Kot kaže primerjava povprečne pogostosti nizov v vsakem izmed korpusov (Slika 2), se formulaični besedni nizi v povprečju pojavljajo s precej visoko in podobno pogostostjo (povprečno 57 pojavitev na milijon pojavnic v korpusu Kres oz. 56 pojavitev v korpusu Gos), tudi če primerjamo povprečno pogostost nizov posameznih dolžin. Medtem ko v pisni slovenščini pogostost rabe pada z dolžino niza, je v govorjeni slovenščini raba štiribesednih nizov povsem primerljiva z rabo tribesednih nizov. Nekoliko presenetljivo visoko povprečno pogostost petbesednih nizov lahko pojasnimo z dejstvom, da so med njimi predvsem nizi, ki so gradniki daljših besednih nizov (npr. del niza opornih signalov *ja ja ja ja ja* ali del telefonske številke *osem nič osem nič nič*), ki bi jim bila pogostost ob izbiri daljše izhodiščne dolžine iskanih nizov zaradi pojavljanja v daljših formulaičnih nizih ustrezno zmanjšana.



Slika 2: Povprečna relativna pogostost formulaičnih nizov različnih dolžin v korpusih Kres in Gos.

3.2 Prekrivnost formulaičnih besednih nizov

Ob dejstvu, da med obema oblikama jezikovne rabe oz. prenosnikoma prihaja do razlik v deležu formulaičnih nizov (Slika 1), se odpira vprašanje, v kolikšni meri so formulaični nizi, h katerim se govorci zatekajo v pisni in govorjeni rabi, sploh prekrivni. Rezultati, povzeti na Sliki 3, kažejo, da sta seznama zgolj delno prekrivna. V obeh korpusih se pojavlja 1.410 formulaičnih nizov, medtem ko je preostalih 984 (41 %) nizov korpusa Kres oz. 3.407 (71 %) nizov korpusa Gos unikatnih, na podlagi česar lahko predvidevamo, da se tako v pisni kot govorjeni jezikovni rabi glede na specifične komunikacijske okoliščine in cilje razvijajo specifične oblike formulaičnega jezika.



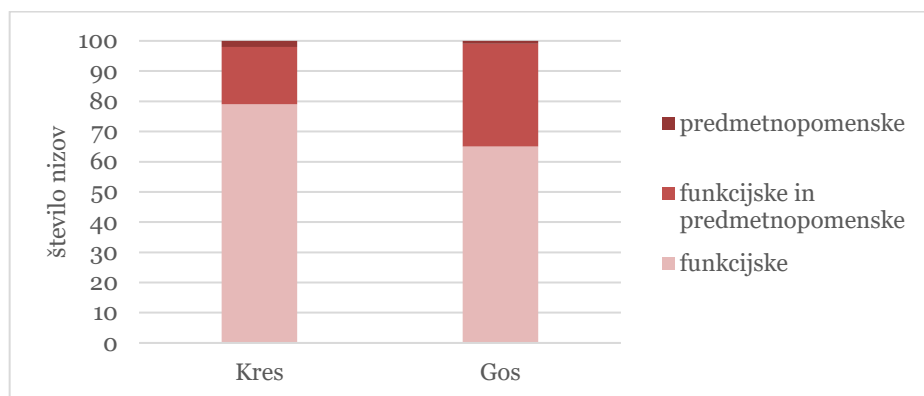
Slika 3: Prekrivnost formulaičnih nizov v korpusih Kres in Gos.

3.3 Analiza formulaičnih besednih nizov

Po kvantitativni analizi deleža in pogostosti rabe formulaičnih nizov v pisni in govorjeni slovenščini smo se v tretjem koraku raziskave osredotočili na kvalitativno analizo lastnosti najpogostejše rabljenih formulaičnih nizov v vsaki izmed oblik jezikovne rabe. Za ta namen smo analizirali 100 najpogostejših tri-ali večbesednih nizov v vsakem korpusu z vidika leksikalne zgradbe, skladenjske strukture ter pragmatične funkcije (razdelek 2.3). Njihov seznam je naveden v priloženih Tabelah 5 in 6.

3.3.1 FORMULAIČNI NIZI GLEDE NA LEKSIKALNO ZGRADBO

Kot povzema Slika 4 oz. ponazarjajo podatki v Tabelah 5 in 6, najpogostejše formulaične nize v slovenskem jeziku sestavljajo predvsem funkcijske oz. slovnične besede, kot so vezniki in členki (npr. *da, ne, pa*), zaimki (npr. *to, se, tem*), predlogi (npr. *v, na, s*) in oblike pomožnika *biti* (npr. *je, bi, bilo*), saj v obeh oblikah jezikovne rabe prevladujejo nizi z izključno funkcijskimi besedami (79 % v korpusu Kres oz. 65 % v korpusu Gos) oz. nizi, ki poleg predmetnopomenskih besed vsebujejo tudi funkcijske besede (19 % v Kres oz. 34 % v Gos). Zveze predmetnopomenskih besed so bistveno redkejše, saj se med najpogostejšimi formulaičnimi nizi pojavljajo zgolj trije tovrstni nizi (*d. o. o.* in *uradni list rs št.* v korpusu Kres in *leta dva tisoč* v korpusu Gos).

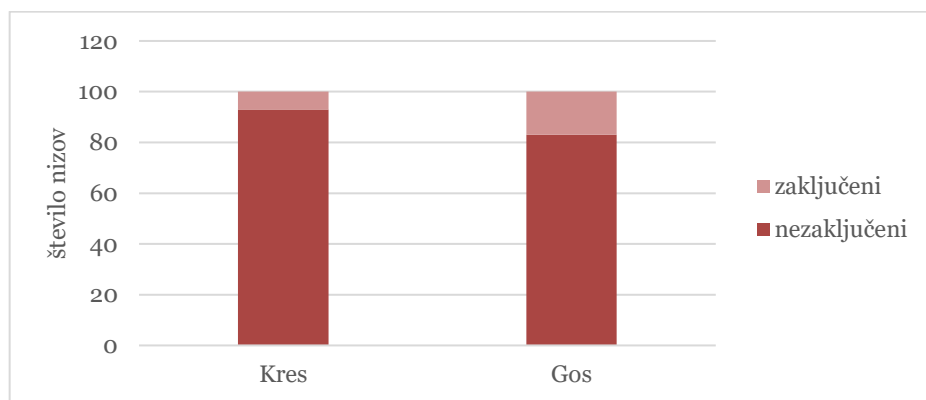


Slika 4: Vrsta besed v formulaičnih nizih pisne in govorjene slovenščine.

3.3.2 FORMULAIČNI NIZI GLEDE NA SKLADENJSKO ZGRADBO

Poleg specifične leksikalne zgradbe se formulaični nizi od drugih doslej obravnavanih tipov večbesednih enot v slovenščini ločijo tudi po svoji skladenjski zgradbi (Slika 4), saj med najpogostejšimi nizi prevladujejo predvsem skladenjsko nezaključeni nizi, kot so stavčni (npr. *da bi se, to pa je*)

ali besednozvezni fragmenti (npr. *člena tega zakona, uradni list rs št.*).⁸ Medtem ko se med skladenjsko zaključenimi nizi v korpusu Kres pojavljajo predvsem besedilnopovezovalne samostalniške besedne zveze (npr. *v tem primeru, na drugi strani, na ta način*), večinski delež strukturno zaključenih enot v korpusu Gos predstavlja predvsem kombinacije pragmatičnih oz. diskurznihi členkov (npr. *ja ja ja, ne tako da, ja ja saj*), s čimer lahko tudi pojasnimo nekoliko večji delež zaključenih struktur v primerjavi s pisnim korpusom.



Slika 5: Skladenjska zaključenost formulaičnih nizov pisne in govorjene slovenščine.

Velik delež strukturno nezaključenih formulaičnih nizov sovпада tudi z nadaljnjo analizo specifičnih skladenjskih struktur, saj v obeh korpusih prevladujejo predvsem stavčni fragmenti, ki vsebujejo kombinacije (delov) povedka in različnih slovničnih besed (npr. kazalnih zaimkov ali veznikov), bistveno redkejši pa so nizi, ki izhajajo iz besednih zvez drugih tipov, kot so samostalniške (npr. *v skladu z/s, v tem primeru, člena tega zakona*), prislovne (*in tako naprej*) ali medmetno-členkovne zveze (npr. *ne ne ne, mhm ja ja, ne*

⁸ Med strukturno nezaključene nize so bili umeščeni tudi pomensko nerazstavljivi nizi s skladenjsko odprtimi vezljivostnimi mesti, kot so *v skladu z, v zvezi z, za to da, ne glede na, to pomeni da* (Kres) ali *in tako naprej, in to je, zaradi tega ker* (Gos), ki sicer glede na svojo visoko stopnjo leksikaliziranosti predstavljajo vmesni pol med strukturno zaključenimi in nezaključenimi enotami.

tako da).

Kljub primerljivemu deležu nizov z glagolsko sestavino pa se oba korpusa pomembno razlikujeta v distribuciji posameznih tipov (Tabeli 2 in 3). Medtem ko v korpusu Gos prevladujejo predvsem kombinacije povedka in njegovih slovnično- in pomenskovezljivostnih določil (npr. *to je bilo, to je pa, se mi zdi*), v korpusu Kres izrazito izstopa delež fragmentov odvisnih stavkov (npr. *da bi se, ki ga je, ki se je*). Prav tako so v korpusu Kres pogostejši samostalniški besednozvezni nizi, v korpusu Gos pa nizi diskurznofunkcijskega besedišča.

Ta opažanja se skladajo s spoznanji sorodnih korpusnojezikoslovnih raziskav o skladenjskih (Dobrovoljc, Nivre 2016) in leksikalnih (Verdonik, Sepesy Maučec 2017) razlikah med obema oblikama sporazumevanja, obenem pa že nakazujejo tudi pragmatično dimenzijo razvoja in specializacije formulaičnega jezika, ki jo predstavimo v nadaljevanju.

Struktura	Št. nizov	Primer nizov
V G G	12	<i>ki je bil, da je bil, ki je bila, da je bila, pa je bil, ki so bili, da je bilo, pa je bila, pa je bilo, da bi bil, ko je bil</i>
V G Z	10	<i>da bi se, ki so se, ki so jih, da je to, ki so ga, ki so jo, da so se, pa so se, da bi ga, če bi se</i>
V Z G	9	<i>ki ga je, ki se je, da se je, ki jih je, ki jo je, pa se je, ko se je, da se bo, ki mu je</i>

Tabela 2: Tri najpogostejše skladenjske strukture formulaičnih nizov pisne slovenščine (G = glagol, V = veznik, Z = zaimek).

Struktura	Št. nizov	Primer nizov
Z Z G	6	<i>se mi zdi, to se pravi, to to je, kar se tiče, ja jaz sem, kaj se je</i>
V Z G	6	<i>in to je, da se je, saj to je, da se bo, ampak to je, pa to je</i>
Z G V	5	<i>to je pa, jaz mislim da, to pomeni da, jaz sem pa, se pravi da</i>

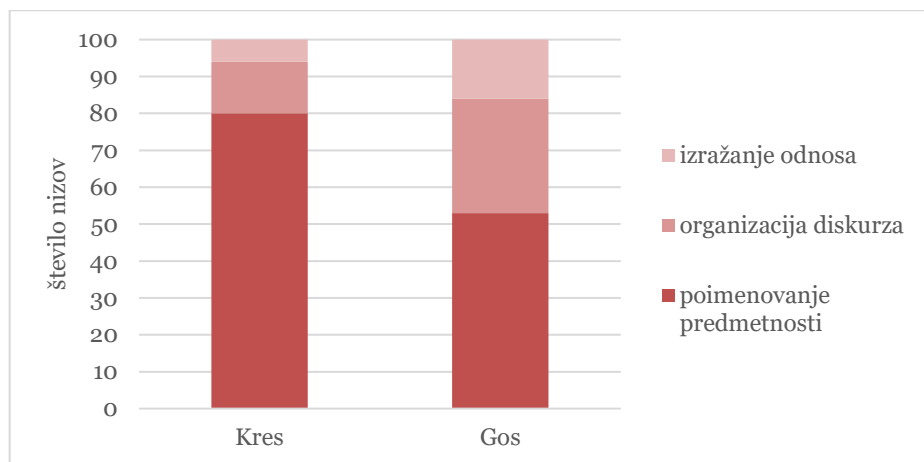
Tabela 3: Tri najpogostejše skladenjske strukture formulaičnih nizov govorne slovenščine (G = glagol, V = veznik, Z = zaimek).

3.3.3 FORMULAIČNI NIZI GLEDE NA FUNKCIJO

Končna analiza najpogostejših formulaičnih nizov glede na pragmatično funkcijo na Sliki 6 kaže, da v korpusu Kres med najpogostejšimi formulaičnimi nizi prevladujejo nizi za poimenovanje predmetnosti (80 %), sledijo nizi za izražanje odnosa, najmanj pa je nizov za strukturiranje diskurza. Čeprav enako velja tudi za govorni jezik, pa je med najpogostejšimi formulaičnimi nizi korpusa Gos izrazito več nizov za izražanje odnosa (16 %) in strukturiranje diskurza (31 %).

Pri tem velja opozoriti, da se pogostost teh dveh funkcij v diskurzu deloma odraža tudi v naboru predmetnopomenskih nizov, saj med njimi v skladu z izbrano izhodiščno tipologijo (Simpson-Vlach, Ellis 2010) najdemo tudi druge tipe nizov za izražanje nedoločnosti, kot so označevalci odprte propozicije (*in tako naprej, ali pa kaj*), ali nizov za vzpostavljanje koherence (*v tem primeru, na ta način*). Če bi tovrstne večfunkcijske nize namesto med nize za poimenovanje predmetnosti umestili med nize za izražanje odnosa oz.

upravljanje diskurza, bi bil delež predmetnopomenskih nizov v obeh korpusih, še zlasti pa v korpusu Gos, še precej manjši. Enako opozorilo po preudarnem tolmačenju pogostosti velja tudi za drugi dve funkcijski kategoriji, saj tudi ločnica med izražanjem odnosa in upravljanjem diskurza ni vedno jasno določljiva, zlasti v govoru (npr. *ja ja ja*).



Slika 6: Pragmatične funkcije formulaičnih nizov pisne in govorne slovenščine.

Korpusa pa se ne razlikujeta zgolj v deležih nizov posameznih funkcijskih skupin, temveč tudi v samem naboru nizov, ki razkrivajo specifike formulaičnega jezika v posamezni obliki jezikovne rabe (Tabela 4). Kot je sugerirala že primerjava nizov glede na skladenjsko zgradbo, se v pisni rabi med najpogostejšimi predmetnopomenskimi nizi pojavljajo predvsem fragmenti odvisnih stavkov, v govoru pa poleg teh tudi stavčni fragmenti s kazalnimi zaimki in sestavljenimi povedki, ki izpostavljajo vlogo poročanja in deiktčnosti v spontanem govoru. Podobno razliko vidimo tudi v primerjavi nizov za izražanje (epistemskega) odnosa, saj v korpusu Kres izstopajo nizi za relativizacijo dokaznosti oz. izvora informacije (*naj bi se, naj bi bil*), v korpusu Gos pa nizi za omilitev govorčevih subjektivnih sodb (npr. *jaz mislim da, pa ne vem, na neki način*).

Kategorija	Kres	Gos
Poimenovanje predmetnosti	<i>da bi se, ki ga je, ki se je, da se je, ki jih je, ki jo je, pa se je, ko se je, se je v, ki so se</i>	<i>da je to, in tako naprej, to je bilo, to je pa, to je to, da bi se, kaj je to, da je bilo, da se je, saj to je</i>
Izražanje odnosa	<i>naj bi se, pa naj bi, se mi zdi, mislim da je, naj bi bil, naj bi bila</i>	<i>jaz mislim da, se mi zdi, pa ne vem, ne vem kaj, mislim da je, ja ne vem, ne vem če, na neki način, se mi zdi da, ne vem kako</i>
Strukturiranje diskurza	<i>ne da bi, in s tem, ne glede na, tako da je, tako da se, in da je, na drugi strani, medtem ko je, to pomeni da, s tem da</i>	<i>ja ja ja, ja ja ja, ja ja ja ja, ja, ne ne ne, in to je, zaradi tega ker, ja to je, ne to je, to se pravi, to pa je</i>

Tabela 4: Deset najpogostejših formulaičnih nizov za poimenovanje predmetnosti, izražanje odnosa in strukturiranje diskurza v korpusih Kres in Gos.

4 DISKUSIJA IN ZAKLJUČEK

V članku smo predstavili pilotno raziskavo koncepta formulaičnosti v slovenskem jeziku na podlagi kvantitativne in kvalitativne analize najpogostejše ponavljajočih se besednih nizov v referenčnih korpusih pisne in govorjene slovenščine. Rezultati potrjujejo, da tudi slovenščina izkazuje določeno stopnjo formulaičnosti komunikacije, saj smo v obeh oblikah jezikovne rabe identificirali opazen delež besednih nizov z izstopajočo pogostostjo pojavljanja. Še zlasti to velja za govorjeno slovenščino, ki je izrazito bolj formulaična kot pisna, kar potrjuje pomen vnaprej pripravljenih besedilnih vzorcev v spontanem govoru, zlasti z vidika kognitivnega razbremenjevanja pritiskov tvorjenja v realnem času in vzdrževanja tekočnosti govora (Nattinger,

DeCarrico, 1993; Pawley, Syder 1983).

Nadaljnja analiza slovničnih lastnosti najpogostejših formulaičnih nizov razkriva, da se v obeh oblikah jezikovne rabe glede na specifične sporazumevalne okoliščine in cilje razvijajo specifične oblike formulaičnega jezika. Medtem ko v pisni slovenščini izstopajo formulaični nizi za pojasnjevanje oz. specificiranje, kot so stavčne matrike za uvajanje odvisnikov (npr. *npr. da bi se, ki ga je, ki se je*) in samostalniške zveze za izražanje predmetnih razmerij (npr. *v skladu z, v zvezi z, v tem primeru*), v govorjeni slovenščini izstopajo predvsem deiktični nizi za poročanje oz. pripovedovanje (npr. *to je bilo, to je pa, in to je*). Prav tako v govorjeni rabi izstopa delež metabesedilnih nizov za organizacijo diskurza in uravnavanje medosebnih odnosov, zlasti nizov, s katerimi govorec omiljuje in relativizira neposrednost svojih sporočil (npr. *jaz mislim da, se mi zdi, pa ne vem*). Te ugotovitve se skladajo s širokim in raznolikim naborom doslej izpostavljenih funkcij formulaičnega jezika, od nedvoumnega in jasnega posredovanja informacij do družbene solidarnosti (Schmitt, Carter 2004; Wray 2002), in potrjujejo eno temeljnih hipotez področja, da za vsako ponavljajočo komunikacijsko potrebo običajno obstaja konvencionalizirana jezikovna raba, ki jo izpolnjuje (Alali, Schmitt 2012).

Ne glede na samo opredelitev formulaičnega jezika, pri kateri odprto ostaja vprašanje razmerij med formulaičnostjo, pogostostjo in zaznavno celovitostjo (Wray 2012), pa je pomembna ugotovitev naše raziskave dejstvo, da najpogostejši besedni nizi v slovenščini izkazujejo specifične lastnosti v primerjavi z drugimi vrstami doslej preučevanih večbesednih enot, kot sta vsebovanost slovničnega besedišča in strukturna nezaključenost. Čeprav so bili tovrstni nizi doslej ravno zaradi svoje strukturno-pomenske razstavljivosti in nezaključenosti potisnjeni ob rob jezikoslovnega preučevanja, njihova izstopajoča pogostost rabe poleg razmisleka o njihovi vlogi v jeziku in metodoloških vidikov njihovega prepoznavanja (Dobrovoljc 2017) zahteva tudi razpravo o smotrnosti in načinu njihovega vključevanja v jezikovne opise.

V prvi vrsti so relevantni zlasti leksikalni opisi, namenjeni rojenim in tujim govorcem slovenščine. Sorodne raziskave namreč kažejo, da je raba formulaičnih nizov eden ključnih pokazateljev jezikovne kompetence (Boers in dr. 2006), zaradi česar se formulaični nizi danes vse bolj načrtno vključujejo v priročnike za poučevanje jezika (Martinez, Schmitt 2012) oz. njegovih podzvrsti (Simpson-Vlach, Ellis 2010). Čeprav so nekateri izmed nizov, identificiranih v tej raziskavi, že bili vključeni v Slovar slovenskega knjižnega jezika, njihova obravnava ni enotna, saj so bodisi izpostavljeni kot stalna zveza, ki se povezuje z enim izmed pomenov (npr. *in tako naprej, to se pravi, kar se tiče, v skladu z/s*) bodisi se brez pomenskega pojasnila pojavljajo zgolj v enem izmed zgledov (npr. *na neki način, v tem primeru, glede na to da, v zvezi s/z*). Ob posodobitvi leksikalnega opisa sodobne slovenščine bi tako veljalo razmisliti o doslednejšem pristopu k obravnavi (slovarsko relevantnih) formulaičnih nizov, vključno z natančnejšo opredelitvijo glede na druge tipe večbesednih leksikalnih enot, ter predvsem posodobitvi njihovega nabora na podlagi analize sodobnega gradiva, saj se že na našem razmeroma majhnem vzorcu najpogostejših nizov pojavlja kar nekaj stalnih nizov, ki jih SSKJ ne beleži (npr. *zaradi tega ker, s tem da, v tem času, se pravi da, to je to, ali pa kaj*).

Na podoben način pa rezultati naše raziskave, ki opozarjajo na rutinskost določenih besednih kombinacij, preizprašujejo tudi zasnovo obstoječih in bodočih slovničnih opisov slovenščine. Čeprav je teza, da jezika ni mogoče opisovati z razmejevanjem besedišča in skladnje, skozi korpusne in druge raziskave večbesednih enot v slovenskem prostoru uveljavljena že dolgo, pa izziv ostaja, kako v teorijo jezika in njegov opis vključiti tudi strukturno in pomensko radikalno razbremenjen koncept formulaičnosti. V tem kontekstu pričujoča raziskava predstavlja zgolj enega izmed korakov k sistematičnemu opisu formulaičnega jezika v slovenščini, ki bi ga bilo treba v prihodnje – ob temeljitejšem teoretskem preizpraševanju izhodiščnega koncepta in opredelitve – nadgraditi z nadaljnjimi raziskavami metod prepoznavanja in opisovanja formulaičnih besednih kombinacij na eni strani kot tudi z

raziskavami kognitivnih mehanizmov njihovega usvajanja, razumevanja in tvorjenja na drugi.

ZAHVALA

Znanstveno-raziskovalno 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.

LITERATURA

- Adolphs, S., in Carter, R. (2013): *Spoken Corpus Linguistics: From Monomodal to Multimodal*. London, New York: Routledge.
- Alali, F. A. in Schmitt, N. (2012): Teaching Formulaic Sequences: The Same as or Different From Teaching Single Words? *TESOL Journal*, 3(2): 153–180.
- Arhar Holdt, Š. (2011): *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Biber, D. (2009): A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14 (3): 275–311.
- Biber, D., Conrad, S., in Cortes, V. (2004): If you look at ...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, 25(3): 371–405.
- Biber, D., S. Johansson, S. Conrad in Finnegan, E. (1999): *Longman grammar of spoken and written English*. Harlow: Longman.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., in Demecheleer, M. (2006): Formulaic sequences and perceived oral proficiency: putting a Lexical

- Approach to the test. *Language Teaching Research*, 10 (3): 245–261.
- Bolinger, D. (1976): Meaning and memory. *Forum Linguisticum*, 1 (1): 1–14.
- Buerki, A. (2016): Formulaic sequences: a drop in the ocean of constructions or something more significant? *European Journal of English Studies*, 20(1): 15–34.
- Conklin, K., in Schmitt, N. (2012): The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, 32: 45–61.
- Cortes, V. (2015): Situating lexical bundles in the formulaic language spectrum. V V. Cortes in E. Csomay (ur.): *Corpus-based Research in Applied Linguistics: Studies in Honor of Doug Biber*: 197–216. Amsterdam_ John Benjamins Publishing.
- Coulmas, F. (1979): On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics*, 3 (3–4): 239–266.
- Dobrovoljc (2018a): *Leksikalne prvine govornega jezika v uporabniških spletnih vebinah: primer večbesednih diskurznihi označevalcev*. Doktorska disertacija. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Dobrovoljc, K. (2017): Multi-word discourse markers and their corpus-driven identification: The case of MWDM extraction from the reference corpus of spoken Slovene. *International Journal of Corpus Linguistics*, 22 (4): 551–582
- Dobrovoljc, K. (2018b): Frekvenčni sezname n-gramov v korpusih slovenskega jezika. V D. Fišer in A. Pančur (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 47–53. Ljubljana: Znanstvena založba Filozofske fakultete.
- Dobrovoljc, K., in Nivre, J. (2016): The Universal Dependencies Treebank of Spoken Slovenian. *Proceedings of the Tenth International Conference*

- on Language Resources and Evaluation* (LREC 2016): 1566–1573. Slovenija.
- Erjavec, T. (2013): Korpusi in konkordančniki na strežniku nl.ijs.si. *Slovenščina 2.0*, 1 (1): 24–49.
- Erjavec, T., in Krek, S. (2008): Oblikoskladenjske specifikacije in označeni korpusi JOS. V T. Erjavec in J. Žganec Gros (ur.): *Zbornik Šeste Konference Jezikovne Tehnologije*: 49–53.
- Erman, B., in Warren, B. (2000): The idiom principle and the open choice principle. *Text*, 20 (1): 29–62.
- Fillmore, C. J. (1968): The case for case. V E. Bach in R. T. Harms (ur.): *Universals in Linguistic Theory*: 1–88. New York: Holt, Rinehart and Winston.
- Fillmore, C. J. (1982): Frame semantics. V *Linguistics in the Morning Calm, Selected Papers from SICOL-1981*: 111–137. Seul: Hanshin Publishing Company.
- Firth, J. (1957): *Papers in linguistics 1934-1951*. Oxford: Oxford University Press.
- Gantar, P. (2007): *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC SAZU.
- Gantar, P. (2015): *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Goldberg, A. E. (2006): *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Halliday, M. A. K. (1985): *An introduction to functional grammar*. London: Edward Arnold.
- Hanks, P. (2013): *Lexical analysis: norms and exploitations*. Cambridge MA: MIT Press.

- Hoey, M. (2005): *Lexical priming: a new theory of words in language*. London: Routledge.
- Hunston, S., in Francis, G. (2000): *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins Publishing.
- Jakop, N. (2006): *Pragmatična frazeologija*. Ljubljana: Založba ZRC SAZU.
- Kosem, I., Gantar, P., in Krek, S. (2013): Avtomatizacija leksikografskih postopkov. *Slovenščina 2.0*, 1 (2): 139–164.
- Krek, S. (2010): *Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovníc*. Doktorska disertacija. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Krek, S., Gantar, P., Dobrovoljc, K., in Škrjanec, I. (2016): Označevanje udeleženskih vlog v učnem korpusu za slovenščino. V T. Erjavec in D. Fišer (ur.): *Zbornik konference Jezikovne tehnologije in digitalna humanistika*: 106–110. Ljubljana: Znanstvena založba Filozofske fakultete v Ljubljani.
- Lin, P. M. S. (2010): The phonology of formulaic sequences: a review. V D. Wood (ur.): *Perspectives on formulaic language: acquisition and communication*: 174–193. London: Continuum.
- 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; Založba FDV.
- Martinez, R., in Schmitt, N. (2012): A Phrasal Expressions List. *Applied Linguistics*, 33 (3): 299–320.
- Nattinger, J. R., in DeCarrico, J. S. (1992): *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

- O'Donnell, M. B. (2010): The adjusted frequency list: A method to produce cluster-sensitive frequency lists. *ICAME Journal*, 35: 135–170.
- Pawley, A., in Syder, F. H. (1983): Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. V J. C. Richards in R.W. Schmidt (ur.): *Language and Communication*: 191–226. London: Longman.
- Schmitt, N. in Carter, R. (2004): Formulaic sequences in action: an introduction. V N. Schmitt (ur.): *Formulaic sequences: acquisition, processing and use*: 1–22. Amsterdam: John Benjamins Publishing.
- Simpson-Vlach, R., in Ellis, N. C. (2010): An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31 (4): 487–512.
- Sinclair, J. (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja* (2014). Dostopno prek: www.fran.si (15. 11. 2018).
- Stramljič Breznik, I. (2001): Komunikacijski ali sporočanje frazemi. *Jezik in slovstvo*, 46 (5): 191–200.
- Tremblay, A., Derwing, B., Libbern, G. in Westbury, C. (2011): Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61: 569–613.
- Verdonik, D. (2014): Vprašanja zapisovanja govora v govornem korpusu Gos. V T. Erjavec in J. Žganec Gros (ur.): *Jezikovne tehnologije: Zbornik 17. mednarodne multikonference Informacijska družba*: 151–156.
- Verdonik, D., in Sepesy Maučec, M. (2017): A speech corpus as a source of lexical information. *International journal of lexicography*, 30 (2): 143–166.
- Verdonik, D., in Zwitter Vitez, A. (2011): *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Wray, A. (2002): *Formulaic Language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2012): What (do we think) we know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32: 231–254.

Wray, A. (2013): Formulaic Language. *Language Teaching*, 46 (3): 316–334.

PRILOGA

1	da bi se	26	ki so ga	51	da se ne	76	v tem času
2	ki ga je	27	ki je bila	52	na drugi strani	77	je v tem
3	ki se je	28	se mu je	53	kot da bi	78	mislim da je
4	da se je	29	da je v	54	da gre za	79	kar pomeni da
5	ki jih je	30	je da se	55	uradni list rs št.	80	o tem da
6	ki jo je	31	da je bila	56	da je bilo	81	ne glede na to
7	pa se je	32	v zvezi s	57	je bila v	82	glede na to da
8	ko se je	33	ne bi bilo	58	da bi lahko	83	ki mu je
9	se je v	34	v tem primeru	59	pa je tudi	84	so se v
10	ki so se	35	se je na	60	v tem da	85	da bi bil
11	ki je bil	36	člena tega zakona	61	medtem ko je	86	iz prejšnjega odstavka
12	je da je	37	za to da	62	pa je da	87	v kateri je
13	v skladu z	38	ki so jo	63	se ji je	88	naj bi bil
14	ki so jih	39	naj bi se	64	pa so se	89	zato ker je
15	da se bo	40	pa je bil	65	to pomeni da	90	ko je bil
16	ki naj bi	41	da ne bi	66	pa naj bi	91	se je da
17	ki je v	42	ne glede na	67	da bi ga	92	to je bilo
18	da je bil	43	je bil v	68	pa je bila	93	v katerem je
19	ne da bi	44	da so se	69	se mi zdi	94	na to da
20	da je to	45	tako da je	70	to da je	95	naj bi bila
21	se mi je	46	tako da se	71	pa je bilo	96	na ta način
22	v skladu s	47	je bilo v	72	da bi bilo	97	da ne bo
23	in s tem	48	ki so v	73	s tem da	98	in da se
24	v zvezi z	49	ki so bili	74	je še vedno	99	v primerjavi z
25	d. o. o.	50	in da je	75	če bi se	100	pa je v

Tabela 5: 100 najpogostejših tri- in večbesednih formulaičnih nizov v korpusu Kres.

1	ja ja ja	26	na neki način	51	je rekla da	76	ja jaz sem
2	ja ja ja ja	27	jaz ne vem	52	ne bi bilo	77	ampak to je
3	ja ja ja ja ja	28	se mi zdi da	53	to je zdaj	78	a je to
4	da je to	29	je v bistvu	54	pol je pa	79	ne ne ne ne
5	ne ne ne	30	ne vem kako	55	ja kaj pa	80	ali pa kaj
6	in tako naprej	31	ne tako da	56	s tem da	81	da je v
7	to je bilo	32	da je bilo	57	to je tako	82	kaj se je
8	to je pa	33	mhm mhm mhm	58	na ta način	83	leta dva tisoč
9	jaz mislim da	34	da se je	59	v tem primeru	84	ja to pa
10	se mi zdi	35	saj to je	60	to je bil	85	se pravi da
11	pa ne vem	36	to je to je	61	tako da se	86	da je bil
12	in to je	37	da ne bo	62	je bilo pa	87	da se ne
13	ne vem kaj	38	je bilo to	63	je pa to	88	je rekel da
14	to je to	39	to je res	64	da bi bilo	89	tako kot je
15	mislim da je	40	glede na to da	65	moram reči da	90	je to to
16	ja to je	41	ja ja saj	66	v tem trenutku	91	veš kako je
17	zaradi tega ker	42	to pomeni da	67	mhm ja ja	92	in da je
18	ne to je	43	zato ker je	68	je bilo tako	93	je pa tudi
19	da bi se	44	jaz sem pa	69	za to da	94	kaj je bilo
20	ja ne vem	45	da se bo	70	to je bila	95	pa je to
21	to se pravi	46	kaj pa je	71	kar se tiče	96	pa to je
22	to pa je	47	je bilo v	72	da bi bilo	97	da ne bo
23	kaj je to	48	ki so v	73	s tem da	98	in da se
24	ne vem če	49	ki so bili	74	je še vedno	99	v primerjavi z
25	tako da je	50	in da je	75	če bi se	100	pa je v

Tabela 6: 100 najpogostejših tri- in večbesednih formulaičnih nizov v korpusu Gos.

FORMULAICITY IN SLOVENE

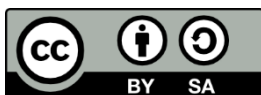
The multitude of usage-based investigations into different types of lexical patterns has revealed the abundance of multi-word combinations that function as a single unit in some aspect of language use. This has also been confirmed by previous research on multi-word units in Slovenian, which has mainly focused on structurally and semantically complete units, such as collocations, multi-word lexical units or idioms. To propose a different perspective on the formulaic nature of Slovenian language, the research presented in this article focuses on the so-called formulaic sequences, i.e. continuous sequences of two or more words with a high frequency of usage, regardless of their structural, semantic or psychological salience. We adopt a corpus-driven approach to the identification of formulaic sequences by extracting the most commonly recurring strings in the reference corpora of written and spoken Slovenian. The two resulting lists of formulaic sequences are analyzed and compared in terms of lexical composition, syntactic structure and pragmatic function. The large number of identified formulaic sequences in both corpora confirms a certain degree of formulaicity in Slovenian, in particular in spoken registers, where the number of formulaic sequences surpasses that of individual words. In both corpora, most formulaic sequences consist of structurally incomplete sequences of grammatical words, which demonstrates their unique characteristics in comparison to other types of multi-word expressions. Nevertheless, the lists of formulaic sequences in written and spoken Slovenian only partially overlap, with 41% of all written sequences and 71% of spoken sequences being unique to each mode. A subsequent in-depth comparison of the syntactic patterns and pragmatic functions confirms that the extent and the nature of formulaic language in Slovenian depends on the specific communicative circumstances and goals of each mode. While written formulaic sequences indicate the explanatory nature of written communication, spoken sequences illustrate the importance of speaker mitigation, discourse structuring and deixis.

Keywords: formulaic language, lexical bundles, multi-word expressions, written language, spoken language

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/>



IZPELJAVA SAMOSTALNIKOV IZ GLAGOLOV: KORPUSNA PRODUKTIVNOST IN SLOVNIČNI OPIS

Nataša LOGAR

Fakulteta za družbene vede Univerze v Ljubljani
Ljubljana, Slovenija

Logar, N. (2018): Izpeljava samostalnikov iz glagolov: korpusna produktivnost in slovnični opis. Slovenščina 2.0, 6 (2): 96–126.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.96-126>.

V prispevku predstavljamo razumevanje besedotvorne produktivnosti, kakršnega je na korpusni metodologiji konec 80. in v začetku 90. let 20. stoletja utemeljil Harald Baayen. Baayenov način izračunavanja produktivnosti *P* smo preizkusili na desetih slovenskih priponskih obrazilih, ki tvorijo samostalniške navadne izpeljanke iz glagolov, in sicer izpeljanke moškega spola s pomenom človeškega vršilca dejanja. Šlo je za obrazila *-vec*, *-telj*, *-ež*, *-uh*, *-avh*, *-ar*, *-lec*, *-ač*, *-aj* in *-ant*. Polavtomatski postopek pridobivanja podatkov smo izvedli na korpusu pisne slovenščine Kres ter med drugim ugotovili, da sta število tvorjenk, katerih del je določeno obrazilo, in produktivnost *P* tega istega obrazila lahko (zelo) različna. Npr.: največ tvorjenk izmed naštetih je tvorilo obrazilo *-lec*, medtem ko je bil *P* najvišji pri obrazilu *-ež*. Na podlagi podatkov iz Kresa smo nato prikazali še seznam obravnavanih obrazil skupaj s tvorjenkami, tj. seznam, kakršen bi lahko bil del nove slovnice slovenščine, ter ga na kratko komentirali, v zaključni razpravi pa smo se nato opredelili še do nekaterih prednosti in slabosti prikazanega pristopa.

Ključne besede: besedotvorje, produktivnost, H. Baayen, priponsko obrazilo, tvorjenka, slovnica, korpus

1 UVOD

Ustvarjanje besed z različnimi priponami je staro, kolikor daleč nazaj sledimo razvoj jezika. Prvotne pripone so večinoma omrtnile in se ohranile samo kot okamnine, besed ne določajo več ne po pomenu ne po obliki. Nastajale pa so nove pripone,

nekatero z zelo jasnim pomenom, druge z manj določenim, zato so nekatere v jeziku zelo tvorne in razširjene, druge manj (Bajec in dr. 1956: 70–71).

Navedena misel iz prve izdaje Slovnice štirih dobro povzema problematiko, ki ji raziskovalci besedotvorja kateregakoli jezika posvečajo precej pozornosti, namreč tematiko besedotvorne – pa sicer tudi širše, morfološke – produktivnosti (Baayen, Lieber 1991: 801; Bauer 2005: 315). Avtorji razprave o besedotvorni produktivnosti pogosto začenjajo z opozorilom o nenatančni rabi tega izraza in njegovi neenotni definiciji (npr. Aronoff 1976: 35; Lieber 1992: 1–9; Bauer 2005: 315, 330; Scherer 2015: 1781, 1782), večinsko pa se strinjajo, da gre za pojav z lastnostmi kontinuuma (Aronoff, Anshen 1998: 242–243). Na enem skrajnem koncu tega kontinuuma so obrazila (če za potrebe prispevka ostanemo le pri teh), ki jih pri tvorbi novih besed ne uporabljamo več – so "omrtvela", kot so se izrazili Bajec, Kolarič in Rupel; na skrajnem drugem koncu so obrazila, ki v tvorbeni postopek vstopajo zelo pogosto in so del zelo številnih tvorjenk – so torej "zelo tvorna in razširjena", medtem ko so vmes obrazila, ki niso niti zelo pogosta niti zelo redka (kot bi se po Likertu izrazili družboslovni metodologi). Pri tem je ključen izziv, na katerega so skušali odgovoriti nekateri, predvsem pa Baayen (1992; več gl. v nadaljevanju), kako to "kontinuumsko" lastnost izmeriti.

Tudi v slovenskem jezikoslovju je običajno, da raziskovalci besedotvornih pojavov svoje ugotovitve o besedotvornih morfemih, besedotvornih vrstah ali pa tvorjenkah iz njih interpretirajo z vidika njihove večje oz. manjše produktivnosti. Prim. nekaj takih ocen (poud. N. L.):

Nasploh lahko rečemo, da najdemo glagolske zloženke tipa *listopad* že v starocerkvenoslovanščini, v večji ali manjši meri pa so uporabljene v vseh slovanskih jezikih. Vsaj za slovenščino velja, da **tvorbena v glavnem niso več žive**, zlasti pa se ne spreminja obseg skladenjskopodstavnih glagolov (Vidovič Muha 1988: 53).

Tudi posnemovalni medmet *tik tak* je **produktiven** za glagol *tiktakati* in iz njega nastale tvorjenke *tiktakanje*, *tiktakajoč* in računalniški žargonizem *tiktakalo*

'processor' (Stramljič Breznik 2013: 104).

Sklenemo lahko, da je sklapljanje v slovenščini **manj produktiven** /.../, vendar pa enostaven in učinkovit način tvorjenja stilno zaznamovanih besed (Logar 2005: 191).

V zvezi s pomenotvornimi postopki velja omeniti tudi ugotovitev, da so tvorjenke, pri katerih se je zgodil pomenski prenos, **tvorbeno manj produktivne** od tistih, pri katerih pomenskega prenosa ni (Kern 2017: 228).

Ko smo deloma že videli, se pri takih in podobnih interpretacijah poleg izraza produktivnost (oz. njegove sopomenke *tvornost*) uporabljajo še sorodne besede, kot so živost, razširjenost, aktualnost in pogostost. To je še dodatno opazno v naslednjem odstavku z naslovom *Živost besedotvornih vrst in sredstev* v Toporišič (2000: 160–161; poud. N. L.):

Besedotvorne vrste in sredstva **niso vsi enako živi in pogosti**. Od besedotvornih vrst je **najbolj pogosta** izpeljava, pri glagolih tudi sestava, nato pridejo zloženke in na koncu sklopi. Vse vrste pa **so žive**. To se ne da reči za vsa besedotvorna sredstva, npr. za vsa priponska obrazila. **Popolnoma mrtvo** je npr. priponsko obrazilo *-sm*, kakor ga imamo v besedi *pesem* 'to, kar se poje'; tudi *-ezen/-azen* je **zelo redko**. Prav tako **netvorno** je npr. priponsko obrazilo *-t*, kot ga imamo v samostalniki tipa *pove* 'to, kar se pove'. **Zelo tvorno** pa je npr. priponsko obrazilo *-lec* (*bralec*, *čistilec*); z njim **je izpeljanih veliko besed**, in še vedno jih delamo. **Tvornost ponskih obrazil** lahko **zamre** samo za kak pomenski tip, za druge pomene pa je **še zmeraj zelo živa**; tako je s priponskima obraziloma *-(i)ca* in *-ka*, ko zaznamujeta ženski par moškemu, npr. *Smrekarica* /.../, namesto česar se danes, posebno v meščanskih krogih, oficialno rabijo posamostaljene pridevniške besede s pripono *-ova/-eva*, tj. *Smrekarjeva*.

V prispevku se bomo naprej odzvali na obe zgornji zadregi in skušali odgovoriti na vprašanja,

a) kaj je besedotvorna produktivnost (in kako se – vsaj v ožjem pomenu – loči od živosti, razširjenosti, aktualnosti in pogostosti oz. kaj ima z njimi skupnega) ter

b) kako besedotvorno produktivnost izmeriti (torej kako pri vsakem besedotvornem pojavu natančneje določiti njegovo mesto med "zelo tvornimi" in "omrtvelimi").

V nadaljevanju bomo na podlagi izračuna produktivnosti izbranih desetih priponskih obrazil ter na podlagi drugih podatkov o njih prikazali še,

c) kakšen bi lahko bil del besedotvornega poglavja v novi slovnici slovenščine, če bi k njegovi pripravi pristopili korpusno.

Pri tem smo za izhodišče izbrali Toporišičevo *Slovensko slovnico* (2000) in dve tuji slovnici (obe korpusni): Cvrček (2010): *Mluvnice současné češtiny 1, Jak se píše a jak se mluví*, ter Biber in dr. (1999): *Longman grammar of spoken and written English*.

2 BESEDOTVORNA PRODUKTIVNOST: DEFINICIJA IN IZRAČUN

Bauer (2005; gl. pregled tudi v Scherer 2015) obrat k resnejšemu proučevanju morfološke produktivnosti pripisuje Chomskyjevemu prispevku *Remarks on nominalization* iz leta 1970, čeprav zametke prepozna že v zgodnejših delih, najbolj izrazito pri Schultinku (1961)¹ in Zimmerju (1964).² Tako Bauer (prav tam: 318–321) kot Baayen in Lieber (1991) pa prvi poskus kvantitativnega formaliziranja produktivnosti pripisujejo Aronoffovemu delu *Word formation in generative grammar* (1976).

Aronoff v svoji knjigi poglavje z naslovom *Productivity* (35–45) začinja z ugotovitvijo, da prav v produktivnosti korenini naslednje "čudno in trdovratno dejstvo: čeprav so v morfologiji možne mnoge stvari, so nekatere bolj možne kot druge" (prav tam: 35). Na primeru angleških samostalniških priponskih

¹ Schultink, Henk (1961): Produktiviteit als morfologisch fenomeen. *Forum der Letteren*, (2), 110–125.

² Zimmer, Karl E. (1964): *Affixal negation in English and other languages. Supplement to Word* 20.

obrazil -ness in -ity, ki (oz. ko) se pripenjata na pridevnike, ki se končajo na -ous (npr. *fabulousness* : *fabulosity*), je nato dokazoval, da dotedanje početje raziskovalcev, tj. ugotavljanje produktivnosti s preprostim štetjem besed z določenim obrazilom – po logiki katerega je bolj produktivno tisto obrazilo, s katerim je tvorjenih več besed – ni skladno z intuitivnim občutkom o večji oz. manjši produktivnosti (v smislu: *perspectiveness* in *perceptivity* sta sicer obe možni, ampak druga se sliši "'čudno' ali 'izumetničeno'" (37)), hkrati pa je poudaril, da tako štetje tudi "ni fer" (36). In sicer že zato ne, ker ne upošteva morfoloških omejitev, npr.: obrazili -ment in -ion v angleščini obe tvorita samostalnike iz glagolov, vendar pa se -ion razvršča samo na latinične glagole, zato bo njegova produktivnost v smislu zgolj štetja besed, ki jih tvori, vedno manjša od produktivnosti obrazila -ment (o tem gl. tudi Aronoff in Anshen 1998: 243–244). Aronoffov predlog za izračun indeksa produktivnosti je bil zato naslednji: prešteti je treba število besed, za katere menimo, da bi iz njih z določenim obrazilom lahko nastala tvorjenka, nato preštejemo število dejanskih tvorjenk, ki so nastale na ta način, ter na koncu izračunamo še razmerje med obema vsotama. Indeks produktivnost bi bil v tem primeru torej razmerje med možnimi tvorjenkami z določenim obrazilom in dejanskimi tvorjenkami iz njih.

Avtor se je sicer že sam zavedal, da ima tudi njegov način izračunavanja produktivnosti pasti (npr. to, da številne besede niso uslovarjene, torej jih pri računanju produktivnosti nimamo na seznamu), dokončno pa sta njegov predlog zavrnila Baayen in Lieber (1991), ki sta se problematike lotila s korpusno metodologijo.³ Baayen in Lieber sta pri tem izhajala iz naslednje

³ V Aronoff, Anshen (1998) je razvidno, da je tudi Aronoff pozneje v celoti sprejel Baayenov način merjenja morfološke produktivnosti. Treba je še dodati, da gre kljub dvojnemu avtorstvu razprave Baayen in Lieber (1991) metodologijo pripisati Baayenu, ki je na to temo

(opomba se nadaljuje na naslednji strani)

Schultinkove (1961) definicije produktivnosti (nav. po Baayen, Lieber 1991: 808): *morfološka produktivnost je zmožnost jezikovnih govorcev, da nenamerno tvorijo številne besede, ki se jih načeloma niti ne da dokončno prešteti*. Pri tej definiciji sta se jima zdeli ključni: (a) nenamernost postopka ("če je besedotvorni postopek zares produktiven, potem bodo tvorjenke, ki nastanejo v njem, povsem neopazne" (prav tam: 808)) in (b) omejena zmožnost preštetja ("zelo produktivni besedotvorni postopki bodo načeloma dali neskončno število tvorjenk, medtem ko bodo neproduktivni besedotvorni postopki dali le določeno število tvorjenk, torej število, ki se ga da prešteti" (prav tam)). Njuna formula za izračun morfološke produktivnosti, v katero sta skušala zajeti oboje, se tako glasi:

$$P = \frac{n_1}{N},$$

pri čemer je n_1 število različnih tvorjenk z določenim obrazilom, ki se v vzorcu (korpusu) pojavijo natančno enkrat (hapax legomena), in N skupno število pojavitev vseh tvorjenk s tem obrazilom.⁴ " P je /torej/ ocena verjetnosti, da bomo naleteli na novo, doslej neopazeno tvorjenko, če je velikost vzorca opazovanih tipov tvorjenk enaka N " (prav tam: 809–810). Kar hkrati pomeni, da lahko napovemo tudi, po kateri stopnji bi število novih tvorjenk tega tipa naraščalo, če bi se odločili korpus povečati (prav tam: 811), seveda pa je to naraščanje tangentno, kar pomeni, da P kaže tudi, kako hitro se bo določen način tvorjenja v vzorcu izčrpal. Če je P velik, lahko sklepamo, da je število tvorjenk opazovanega tipa, ki so ostale zunaj vzorca, še veliko; če je P majhen, pa smo v vzorec zajeli skoraj vse tvorjenke tega tipa vsaj enkrat. V prvem

leta 1989 doktoriral (Vrije Universiteit, Amsterdam), prim. tudi Lieber (1992: 2, 4), Baayen (1992) in poznejše avtorje, ki izrecno pišejo le o Baayenovem merjenju produktivnosti (npr. Bauer 2005; Schreier 2015). Sicer pa je Baayen s korpusnim pristopom izrazito drugačen od predhodnikov (tudi Aronoffa, kot smo videli) po tem, da so ti produktivnost merili na osnovi slovarskih geslovníkov, ker boljšega vira do takrat pač še ni bilo (k njemu se je pozneje na primeru slovarja *The Oxford English Dictionary* pri proučevanju diahronih morfoloških postopkov sicer spet vrnil Plag (1999)).

⁴ Baayenovih formul je v resnici več, a kot je razvidno tudi iz Bauer (2005), je ta ključna.

primeru gre za produktiven besedotvorni postopek, za katerega je značilno veliko, morda celo neskončno število možnih tvorjenk; v zadnjem primeru pa gre za neproduktiven postopek, pri katerem nastane majhno in seveda končno število tvorjenk (prav tam: 813). Velik *P* posledično dobimo, če je število enkratnih tvorjenk zelo visoko ali če je v vzorcu sicer veliko različnih tvorjenk, a nimajo izrazito visoke pogostosti (prav tam: 815).

Na Baayenov način izračunavanja besedotvorne produktivnosti so se najbolj izrecno odzvali van Marle (1992) ter Frauenfelder in Schreuder (1991). Van Marle je ob izraziti pohvali Baayenu, da je v središče raziskovanja problematike s korpusnim pristopom postavil performanco (prav tam: 152), kritično opozoril na to, da Baayenova formula ni izračun zmožnosti določenih tvorbenih postopkov, da "dajo" določeno število tvorjenk, temveč je kvečjemu izračun stopnje, do katere lahko realno pričakujemo, da bodo določeni besedotvorni postopki podlaga za nove tvorjenke (prav tam: 152). Frauenfelder in Schreuder pa sta bila do Baayenove formule kritična z vidika psiholingvistike: osrednja pomanjkljivost, ki sta jo našla, je bilo Baayenovo neupoštevanje tega, da govorec ob tvorbi novih besed procesira tudi fonološke značilnosti ter semantično transparentnost podstavne ter nove besede, medtem ko Baayenova formula, utemeljena na korpusni pogostosti, upošteva samo hitrost tvorčevega spominskega dostopa do tam že obstoječih jezikovnih podatkov. Na obe kritiki je Baayen obsežno odgovoril v Baayen (1993) in tam ponovno zagovarjal svoj način izračunavanja produktivnosti na podlagi podatkov iz rabe (torej pogostosti), pri čemer zgornja formula po njegovem mnenju ustrezno odgovarja na vprašanje – kot ga je povzel Bauer (2005: 326), kolikšen del tvorjenk, ki so nastale po določenem postopku, je hapaxov. Z odgovorom na to vprašanje pa dobimo tudi podatek, ki ga lahko primerjamo z istovrstnimi podatki drugih besedotvornih pojavov in posledično ločimo tiste pojave, ki so v določenem korpusu bolj produktivni (imajo več hapaxov na skupno število pojavitev vseh tvorjenk z obravnavanim obrazilom), od tistih, ki so manj produktivni.

Baayenova metodologija je bila v nadaljnjih letih širše sprejeta (prim. npr. že omenjena Aronoff, Anshen 1998; Bauer 2005; Säily 2011; Booij 2012), zato jo bomo uporabili tudi v tukajšnji analizi, skupaj z njim (in poznejšimi avtorji, ki so mu sledili) pa prevzemamo tudi zgoraj že navedeno Schultinkovo definicijo produktivnost, ki se v natančnem prevodu iz nizozemščine glasi:⁵

morfološka produktivnost je zmožnost jezikovnih govorcev, da s pomočjo morfološkega postopka, ki temelji na oblikovnem in pomenskem ujemanju nekaterih njim znanih besed, nenamerno tvorijo številne nadaljnje besede, ki se jih načeloma niti ne da dokončno prešteti.

S tem obravnavani pojem za potrebe tega prispevka terminološko ožimo na kvantitativni korpusni pristop in ga ločimo od sorodnih izrazov oz. ga z njimi povezujemo zgolj na naslednji način: *pogostost* upoštevamo tako, kot jo prikazuje Baayenova formula; *razširjenost* razumemo kot lastnost, ki izhaja iz korpusno ugotovljene prisotnosti obrazil (če seveda obravnavamo ta) po različnih jezikovnih zvrsteh, besedilnih vrstah, področjih rabe, tematikah ipd. (angl. *domain*, o tem gl. v Plag 1999; tudi Baayen 2009: 19–23); medtem ko *živost* razumemo kot sinonimno poimenovanje za bodisi ničto vrednost produktivnosti *P* (t. i. *mrtvo* morfemsko obrazilo) bodisi vsakršno drugo vrednost *P* (*živo* morfemsko obrazilo).

3 BESEDOTVORNA PRODUKTIVNOST: ANALIZA KORPUSNIH PODATKOV

Analizo smo izvedli v skladu z Baayen in Lieber (1992; za nemščino gl. npr. Evert, Lüdeling (2001), za poljščino Górski (2013: 70–74)), in sicer na korpusu Kres (Logar Berginc in dr. 2012: 77–97). Kres vsebuje 100 milijonov besed oz. dobrih 120 milijonov pojavnic iz besedil, ki so izšla med letoma 1990 in 2011. V celotni obseg korpusa 35 % besed prinašajo knjige (od tega 17 % leposlovje, 18

⁵ Za prevod se zahvaljujem lekt. dr. Aniti Srebnik s Filozofske fakultete v Ljubljani.

% stvarna besedila), 40 % periodični tisk (20 % časopisi in 20 % revije), 20 % internetna besedila (visoko brani novičarski portali ter predstavitvene spletne strani podjetij in ustanov), preostalih 5 % pa zapisi sej državnega zbora, podnapisi in postprodukcijska besedila RTV Slovenija.

Priponska obrazila, ki smo jih za izračun produktivnosti naključno izbrali iz Toporišič (2000: 161–163), so bila naslednja: *-vec*, *-telj*, *-ež*, *-uh*, *-avh*, *-ar*, *-lec*, *-ač*, *-aj* in *-ant*. Vsa se razvrščajo na glagolsko podstavo in tvorijo samostalnik moškega spola s pomenom človeškega vršilca dejanja.

Do korpusa smo dostopali v orodju Sketch Engine (Kilgarriff in dr. 2004), ki ga za slovenščino podpira Center za jezikovne vire in tehnologije Univerze v Ljubljani.⁶ Sezname smo pridobili z iskalnimi pogoji

```
[lemma=".+vec" &tag="Som.*"],
```

```
[lemma=".+telj" &tag="Som.*"] itd.,
```

nato pa smo na podlagi ogleda konkordanc ročno izločili najprej netvorjene besede, dalje pa še tvorjenke, ki niso iz glagola ali nimajo pomena človeškega vršilca dejanja, zloženske in druge neizpeljavne besedotvorne vrste ter zatipkane besede in napačno označena lastna imena. Pri manjšem delu tvorjenk smo sprejeli še naslednji dve odločitvi:

a) Kjer sta bili možni dve poti nastanka tvorjenke, tj. bodisi iz glagola (*žvižgač* < *žvižgati* < *žvižg*) bodisi iz samostalnika (*žvižgač* < *žvižg* < *žvižgati*), smo dosledno upoštevali prvo možnost in smo torej tvorjenko na naš seznam uvrstili.

b) Izmed tvorjenk, ki so v konkordancah izkazovale tako poimenovanje človeškega kot nečloveškega (npr. *nakladač*, *rezkar*, *pomivalec*, *upravljavec*, *nosilec*), smo na seznam uvrstili vse tiste, ki so lastnost človeškega izkazovale

⁶ <https://www.cjvt.si/>

vsaj enkrat. Pri tem pojavitev s kategorijo nečloveško nismo odšteli – če bi namreč hoteli biti v takem odštevanju dosledni, bi morali natančno ročno pregledati prav vse konkordance iskanih tvorjenk, to pa bi nam vzelo izredno veliko časa (vseh konkordanc iskanih tvorjenk je bilo namreč skoraj 365.000, med njimi so nekatere, pri katerih je že hiter ogled konkordanc potrdil, da imajo pomen tako človeškega kot nečloveškega, zelo pogoste, npr. že omenjeni *nosilec* ima 5.546 zadetkov, *povzročitelj* jih ima 1.542, *upravljavec* 1.407 itd.). Ocenili smo, da ima prisotnost kategorije nečloveško v tem obsegu na končne rezultate te študije le zanemarljiv vpliv.

Tabela 1 prikazuje številske podatke opisanega polavtomatskega postopka in končni izračun produktivnosti P po Baayenovi formuli.

Priponsko obrazilo	Število pojavitev	Število različnic	Število iskanih tvorjenk med različnicami	Število vseh pojavitev iskanih tvorjenk	Število hapaxov med iskanimi tvorjenkami	P
<i>-vec</i>	85.549	2.153	50	35.916	17	0,00047
<i>-telj</i>	105.609	461	168	101.201	41	0,00041
<i>-ež</i>	58.324	926	14	1.092	5	0,0046
<i>-uh</i>	38.462	210	15	5.175	2	0,00039
<i>-avh</i>	28	11	0	0	0	0
<i>-ar</i>	262.210	4.004	77	19.746	14	0,00071
<i>-lec</i>	239.456	1.948	662	179.718	213	0,0012
<i>-ač</i>	22.882	652	142	12.879	34	0,0026
<i>-aj</i>	130.632	1.053	3	1.251	0	0

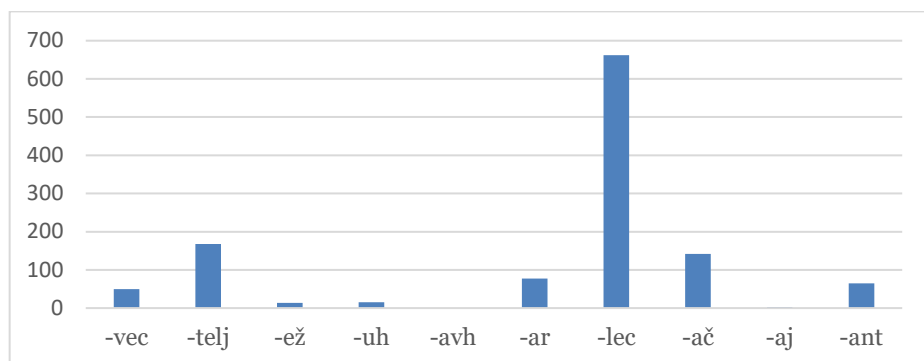
-ant	34.971	812	65	7.981	10	0,0013
------	--------	-----	----	-------	----	--------

Tabela 1: Število pojavitev samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja z 10 izbranimi priponskimi obrazili v korpusu Kres ter njihova produktivnost *P*.

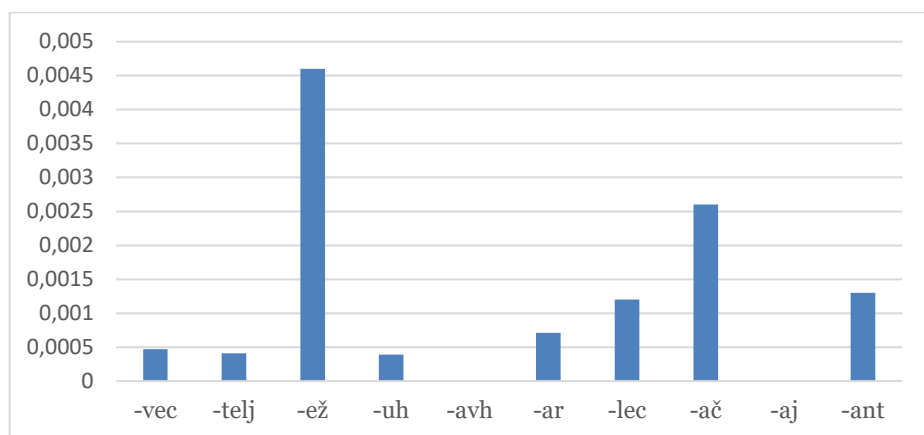
V Tabeli 1 je razvidno, da je bilo v Kresu daleč največ različnih besed z izbranim pomenom tvorjenih z obrazilom *-lec* (gl. 4. stolpec in Sliko 1), npr. *igralec*, *prebivalec*, *obiskovalec*, *bralec*, *gledalec*, *izvajalec*; sledili sta obrazili *-telj* in *-ač* (*prijatelj*, *učitelj*, *vlagatelj*, *pisatelj*, *voditelj*, *ljubitelj*; *navijač*, *dirkač*, *potapljač*, *tekač*, *berač*, *kovač* itd.), nato pa *-ant* in *-vec* (*diplomant*, *reprezentant*, *maturant*, *muzikant*, *projektant*, *emigrant*; *delavec*, *pevec*, *volivec*, *upravljavec*, *razstavljavec*, *pivec*) ter *-uh* in *-ež* (*varuh*, *potepuh*, *ovaduh*, *požeruh*, *oderuh*, *stremuh*; *vedež*, *klatež*, *hlinež*, *burkež*, *rogovilež*, *kurbež*). Obrazilo *-aj* v iskanem pomenu v Kresu izkazujejo le 3 tvorjenke (*strežaj*, *čuvaj*, *točaj*), z obrazilom *-avh* pa v Kresu ni nobene iskane tvorjenke. Tudi sicer so razlike v njihovi pogostosti zelo velike (*točaj* ima npr. 146 pojavitev, *igralec* pa kar 20.608).

Ugotovljeno obenem še ne pomeni, da je *-lec* v izbranem vzorcu tudi najbolj produktivno obrazilo, tj. najbolj produktivno v pomenu *P*. Če smo natančni, smo potrdili le naslednje: za *-lec* je značilno, da se na glagolsko podstavo razvršča z malo omejitvami (je pomensko nevtralen in fonetično široko družljiv, ker se pripenja na končni samoglasnik glagolske osnove *-a-* ali *-i-* (Toporišič 2000: 163)), ugodno pa je tudi, da ima slovenščina zelo veliko zelo pogostih glagolov, ki nase pač lahko sprejmejo to obrazilo. Taka spodbudna podlaga seveda prinese veliko število tvorjenk, vendar pa – če pogledamo z druge strani – to še ne pomeni, da bo obrazilo *-lec* tudi najbolj verjetna prva izbira pri tvorbi naslednje nove slovenske besede (in vsake nadaljnje) s pomenom človeškega vršilca dejanja moškega spola. Razmerje med številom hapaxov (6. stolpec Tabele 1) sicer na prvi pogled med izbranimi obrazili kaže približno enaka

razmerja kot število različnih tvorjenk, vendar pa izračun vrednosti P , torej produktivnosti po Baayenovi formuli (7. stolpec v Tabeli 1), izpostavlja druge "dobitnike medalj" (Slika 2).



Slika 1: Število različnih samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja z 10 izbranimi priponskimi obrazili v korpusu Kres.

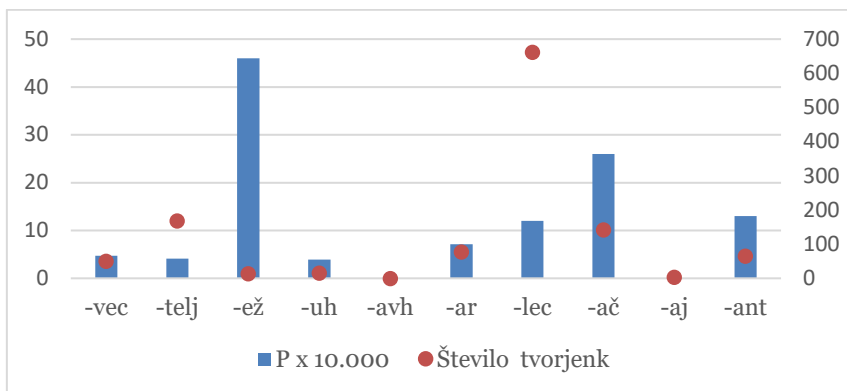


Slika 2: Produktivnost P 10 izbranih priponskih obrazil samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja v korpusu Kres.

Slika 2 kaže, da je v izbranem vzorcu najbolj produktivno obrazilo *-ež*; med le 14 tvorjenkami s skupnim razmeroma nizkim številom vseh pojavitev (1.092) je namreč v Kresu kar 5 hapaxov, ki se končajo na *-ež*: *prdež*, *pofuklež*, *pizdež*,

buljež in *štrbunkež*. Dalje smo izluščili kar 34 hapaxov z obrazilom *-ač*: *čofotač*, *zavirač*, *zapravljač*, *stokač*, *savnač*, *ravskač* itd., kar to obrazilo po produktivnosti *P* uvršča na drugo mesto, sledi pa obrazilo *-ant* (npr. *rehabilitant*, *nategant*, *malverzant*, *mahinant*, *konzultant*, *kapitulant*). Obrazilo *-lec* je tako šele na 4. mestu s sicer v absolutnem številu velikim številom hapaxov (213, npr. *zalagalec*, *terjalec*, *prebujevalec*, *zajtrkovelec*, *točkovelec*, *sesljalec*), na 5. mestu je *-ar* (*savnar*, *plužar*, *pizdar*, *linčar*, *driblar*, *balincar*), nato pa s skoraj enakimi vrednostmi sledijo še *-vec*, *-telj* in *-uh* (*zaklinjavec*, *pospravljavec*, *hlastavec*, *premagavec*, *izjavljavec*, *belivec*; *žalitelj*, *pesnitelj*, *mrzitelj*, *obdavčitelj*, *opremitelj*, *prilastitelj*; *zaležuh*, *oprezuh*). Poleg obrazila *-avh*, za katerega nismo našli nobene tvorjenke z obravnavanim pomenom, nobenega hapaxa ni imelo še obrazilo *-aj*.

Razmerja med produktivnostjo *P* in številom različnih tvorjenk s posameznimi obrazili, ki smo jih zgoraj obravnavali ločeno, spodaj združeno prikazujemo še na Sliki 3.



Slika 3: Razmerja med produktivnostjo *P* in številom različnih tvorjenk z 10 izbranimi priponskimi obrazili samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja v korpusu Kres.

Na Sliki 3 je hitro opazno, da so razlike v razmerjih največje pri obrazilih *-ež* (v smislu veliko večje produktivnosti *P* kot števila različnih tvorjenk) ter *-ač* in

-ant (zopet v smislu večje produktivnosti) – v drugi smeri (torej kot precej manjša produktivnost v primerjavi s številom različnih tvorjenk) pa pri -lec in -telj. Če torej za kazalnik produktivnosti vzamemo razmerje med številom hapaxov, ki imajo določeno obrazilo, in številom pojavitev vseh tvorjenk, ki imajo to obrazilo, potem je izmed 10 izbranih obrazil, ki v slovenščini iz glagolov tvorijo samostalnike moškega spola s pomenom vršilca dejanja, najbolj produktivno – ponavljamo – obrazilo -ež. Ali še drugače: sodeč po deležu hapaxov med vsemi tvorjenkami z izbranimi obrazili, je najbolj verjetno, da bo imela naslednja nova tvorjenka iz glagola in z obravnavanim pomenom obrazilo -ež, sledil mu bo -ač, nato -ant, -lec itd.

Na zgornje dejstvo seveda lahko vplivajo tudi značilnosti podstavnih glagolov (glagolska pripona, fonetične značilnosti, naglas). Hiter pregled je pokazal, da se vsa obravnavana obrazila razvrščajo na besedotvorne podstave iz glagolov na -a-ti in -i-ti (lahko tudi -eva-/-ova-ti in -e-ti), izjema je le obrazilo -ant. To obrazilo je edino, ki se izmed 10 v našem naboru pripenja na besedotvorno podstavo, ki je nastala iz prevzetih glagolov na -ira-ti, npr. *emigrant* < *emigrirati*, *projektant* < *projektirati*, *simulant* < *simulirati* (pravzaprav je sama tvorjenka največkrat pogostejša od postavnega glagola ali pa ga celo edina dokazuje, npr. *vizitant* < **vizitirati*, čeprav slednje velja tudi za tvorjenke (sploh hapaxe) z drugimi obrazili, npr. *popackovalec* < **popackovati*). V nekaj primerih se je -ant sicer razvrstil še na besedotvorno podstavo iz glagolov na -a-ti oz. -ava-ti (*prevarant*, *zabušant*, *zafrkant*, *zajebant* in *nategant*). Pri vseh drugih obrazilih pa, kot rečeno, v zvezi z glagolsko pripono podstavnega glagola ni posebnosti. Tudi fonetičnih omejitev na morfemskem šivu pri izbranih obrazilih ni, z izjemo obrazil -lec in -vec, pri katerih je drugi v prednosti, če se *l* ali *lj* pojavi že v besedotvorni podstavi (npr. *pripravljavec* : *pripravljalec*). Z izjemami v mislih lahko torej sklepamo, da je primerjava izračunane produktivnost med izbranimi obrazili relevantna.

4 SLOVNIČNI OPIS

Zdi se, da je produktivnost v širšem pomenu besede že v samem izhodišču tudi v slovničnih opisih slovenskega besedotvorja. Tako npr. Breznik v uvodu v ta del svoje slovnice ugotavlja: "Večina slovenskih besed je nastala po izpeljavi" (1934: 158), tej ugotovitvi pa nato neposredno sledi prav obravnava omenjene besedotvorne vrste. Tudi sicer se slovnična poglavja o besedotvorju v slovenščini praviloma začenjajo z izpeljavo (gl. npr. Janežič, Sket 1900: 124–157; Breznik 1934: 158–179; Toporišič 2000: 161–232; Herrity 2000: 347–362), zaključujejo pa z besedotvorno vrsto, po kateri nastane najmanj tvorjenk – v zadnji Slovenski slovnici (Toporišič 2000) je to pri tvorbi samostalnikov, pridevnikov in prislovov sklapljanje.

4.1 Seznam obrazil

Slovnične opise besedotvornih značilnosti večinoma sestavljajo sezname obrazil skupaj z zgledi, kakršne za slovenščino, češčino in angleščino na primeru priponskih obrazil za tvorbo samostalnikov prikazujejo spodnje slike. Ti sezname niso osnovani na vseobsegajočnosti in izčrpnosti, temveč zopet – tako je mogoče sklepati – prav na lastnosti, o kateri razmišljamo tu: produktivnost – oz. če smo natančni: avtorji v odstavkih pred temi sezname ali neposredno za njimi pišejo, kot smo deloma že videli, o veliki oz. manjši tvornosti, produktivnosti, pogostosti, tipičnosti in živosti (Toporišič 2000: 160–161; Cvrček 2010: 81–124; Biber in dr. 1999: 320–325).

Priponsko obrazilo	Zgled
-ø	<i>prêrok, spàk, pék (zelo zelo redko);</i>
-a	<i>shúga;</i>
-ja	<i>vódja;</i>
-c	<i>ígrc;</i>
-ec [əc]	<i>bórec, hválec, govórec, jézdec, písec, vídec, jédec, lóvec, védec, strélec, jókec;</i>
-(á)lec [əc]	<i>brálec, dvigálec, malikoválec, sesálec, godálec;</i>
-(í)lec [əc]	<i>branílec, gasílec, pletílec, drobílec, gnetílec, volílec;</i>

-í nec [əc]	<i>mrgolínec;</i>
-unec [əc]	<i>begúnec;</i>
-vec [əc]	<i>pívec, brívec, pévec, mlévec, klávec;</i>

Slika 4: Priponska obrazila za izpeljavo samostalnikov s pomenom vršilca dejanja moškega spola in zgledi v Toporišič (2000: 161–162), del.

Priponsko obrazilo	Zgled
-a	<i>bambula, něha, osma, chvála, čiča;</i>
-ba	<i>honba;</i>
-oba	<i>staroba;</i>
-tba	<i>kletba;</i>
-da	<i>strejda, pravda, obejda;</i>
-ajda	<i>žákajda;</i>
-nda	<i>šeptanda;</i>
-anda	<i>fešanda;</i>
-uha	<i>ostruha</i>
-echa	<i>macecha</i>

Slika 5: Priponska obrazila za izpeljavo samostalnikov in zgledi v Cvrček (2010: 93), del.

Priponsko obrazilo	Pomen(i)	Zgled
-age	'collection of N'	<i>baggage, leafage, plumage</i>
	'action/result of V'	<i>breakage, haulage, wastage</i>
	'cost of N/V-ing'	<i>brokerage, haulage, postage</i>
	'measure in N-s'	<i>acreage, mileage, tonnage</i>
	'place for N'	<i>hermitage, orphanage, vicarage</i>
-al	'action/instance	
	of V-ing'	<i>arrival, burial, denial</i>
-an, -ian	'person who lives in N'	<i>American, Estonian, Korean</i>
	'language of N'	<i>Estonian, Korean</i>
	'person associated with N'	<i>Darwinian, Victorian</i>

-ance, -ence	'action or state of V-ing'	<i>adherence, assistance, resemblance</i>
	'state of being A'	<i>abundance, dependence, ignorance</i>
-ant, -ent	'person who V-s'	<i>assistant, consultant, student</i>
	'something used for V-ing'	<i>coolant, defoliant, intoxicant</i>
-cy	'state or quality of being A/N'	<i>accuracy, adequacy, infancy</i>
-dom	'state of being A/N'	<i>boredom, freedom, stardom</i>
-ee	'person who has been or is to be V-ed'	<i>employee, evacuee, trainee</i>
	'person to whom something has been or is to be V-ed'	<i>assignee, licensee</i>
	'person who V-s or has V-ed'	<i>escapee, retiree, standee</i>
	'person who is A'	<i>absentee, devotee</i>
	'person who V-s'	<i>advertiser, driver, governor</i>
-er, -or	'something used for V-ing'	<i>computer, filler, silencer</i>
	'person concerned with N'	<i>astronomer, footballer, geographer</i>
	'person living in N'	<i>cottager, Londoner, New Yorker</i>
	'action/instance of V-ing'	<i>bribery, robbery, thievery</i>

Slika 6: "Nekatera značilna /angl. *common*/" priponska obrazila za izpeljavo samostalnikov, njihov pomen in zgledi v Biber in dr. (1999: 321), del.

V novi slovnici slovenščine bi bilo treba take prikaze pripraviti na korpusnih podatkih, pri čemer predlagamo, da bi veljalo naslednje:

a) pri izboru obrazil in obrazilnih morfemov se upošteva njihova produktivnost, izračunana po Baayenovi (ali kateri drugi, v prihodnje nastali) formuli, kar pomeni, da se obrazila in obrazilni morfemi z ničelno ali zelo nizko vrednostjo *P* (meja je dogovorna) izpustijo;

b) pri zaporedju obrazil in obrazilnih morfemov se upošteva število različnih tvorjenk, katerih del so (in ne npr. abeceda);

c) pri zgledih se upoštevata število in besedilna razpršenost korpusnih pojavitev tvorjenk, pri čemer enkratnih tvorjenk ali tvorjenk z nizkim številom pojavitev (meja je zopet dogovorna) med zgledi ne navajamo.

V skladu s povedanim bi bila slovnična slika tu obravnavanih priponskih obrazil skupaj z zgledi naslednja (Slika 7):

Priponsko obrazilo	Zgled
-lec	<i>igralec, prebivalec, obiskovalec, bralec, gledalec, izvajalec, raziskovalec, proizvajalec, nosilec, zmagovalec</i>
-telj	<i>prijatelj, učitelj, vlagatelj, pisatelj, voditelj, ljubitelj, predlagatelj, skladatelj, ravnatelj, ustanovitelj</i>
-ač	<i>navijač, dirkač, potapljač, tekač, berač, kovač, krojač, veslač, jahač, nosač</i>
-ar	<i>slikar, vladar, natakár, kuhar, smučar, stražar, šolar, ropar, zidar, romar</i>
-ant	<i>diplomant, reprezentant, maturant, muzikant, projektant, emigrant, komendant, protestant, prevarant, migrant</i>
-vec	<i>delavec, pevec, volivec, upravljavec, razstavljavec, pivec, brivec, razpravljavec, klavec, pripravljavec</i>
-uh	<i>varuh, potepuh, ovaduh, požeruh, oderuh</i>
-ež	<i>vedež, klatež, hlinež, burkež, rogovilež</i>

Slika 7: Izbrana priponska obrazila za izpeljavo samostalnikov s pomenom vršilca dejanja moškega spola in zgledi – prikaz na podlagi podatkov iz korpusa Kres.

S Slike 7 smo izpustili obrazili *-avh* in *-aj*, ki v Kresu za iskani tip tvorjenk

izkazujeta ničelno produktivnost *P* (gl. Tabela 1); obrazila smo razvrstili v skladu z različnim številom tvorjenk, katerih del so (največ jih ima *-lec*, sledi *-telj* itd.); zgledi pa so navedeni po korpusni absolutni pogostosti, pri čemer smo s številom zgledov (10 : 5 – števili sta arbitrarni) posredno nakazali še, kateri dve obrazili sta del znatno manjšega nabora tvorjenk, ki so obenem tudi manj pogoste.⁷

Hkrati naj opozorimo, da smo za razliko od Toporišičeve slovnice naglas obrazil zaenkrat zanemarili (prim. npr. *slik-ár* : *kúh-ar*), ker menimo, da bi bile tu nujno potrebne predhodne sodobne fonetične in fonološke potrditve (Toporišič npr. loči med *-ánt*, npr. *diplománt*, in *-ànt*, npr. *zabušànt* – prim. Jurgec (2011), ki ugotavlja, da je treba ločiti med samoglasnikom *a* in (novim) srednjim nizkim samoglasnikom *Λ*, ki se razlikujeta po formantnih frekvencah, torej kvaliteti, ne po kvantiteti (dolžini)). Kljub zgledu v Biber in dr. (1999) pa smo se odločili, da ne vključimo informacije o pomenu obrazil, ker je ta okvirno vsebovan že v poimenovanju celotne skupine: človeški vršilec dejanja.

4.2 Komentar seznama obrazil

Avtorji slovnice seznane obrazil komentirajo predvsem z vidika omejitev njihove pojavnosti in razvrščanja ter modifikacij. Kot npr.:

Priponsko obrazilo *-ec* krni glagolsko pripono in se sklaplja s soglasniškim izglasjem podstave (*govor-iti* + *-ec* > *govorec*) (Toporišič 2000: 163).

Razlog, zaradi katerega so predpomska obrazila manj produktivna kot priponska, je verjetno v tem, da je večina prvih grškega izvora, medtem ko so skoraj vsa priponska obrazila, naštetá v 4.8.1.2, po izvoru romanska ali germanska (Biber in dr. 1999: 324).

Pri tvorbi besed iz samostalnikov lahko nastanejo podstave, ki se na besedotvorni

⁷ V eventualni elektronski izdaji so seveda možni še daljši sezname tako obrazil kot zgledov ter prekop v korpusne pojavitve, pa seveda tudi v druge spletno dostopne jezikovne vire.

meji zaključujejo s podvojenimi glasovi (-ss-). Podvojeni -ss- je na takem mestu predvidljiv, tako priponsko obrazilo v skladu s tradicijo razumemo kot obrazilo -ský: / *Brandýs(ø)* – podstava: *Brandýs* > *brandýský*; priponsko obrazilo -sk(ý) (Cvrček 2010: 87).

V slovnici avtorjev Biber in dr. (1999) seznamu obrazil, katerega del smo prikazali na Sliki 6, neposredno sledi še poglavje z naslovom *Pogostost značilnih priponskih obrazil za izpeljavo samostalnikov*. Poglavje začenja trditev: "Distribucija izpeljanih/⁸/ samostalnikov, kot tudi samostalnikov nasploh, je po zvrsteh /angl. registers/ zelo različna" (prav tam: 322). V nadaljevanju avtorji navajajo ugotovitve, kakršni sta npr. naslednji dve: (a) pogovorna zvrst ima daleč najmanj izpeljanih samostalnikov in pa (b) z eno izjemo (-ness pri leposlovju) relativna pogostost izpeljanih samostalnikov od pogovorne zvrsti prek leposlovja in časopisja do akademskih pisnih besedil hitro raste (prav tam). Komentar dopolnjuje graf, v katerem je prikazano število tvorjenk s štirimi najbolj značilnimi priponskimi obrazili (-tion, -ity, -ism in -ness), ki tvorijo abstraktne samostalnike, in iz katerega je npr. razvidno, da je tvorjenk z obrazilom -tion v leposlovju trikrat toliko kot v pogovorni zvrsti, nadalje v časopisju 9-krat toliko, v akademskih besedilih pa kar 22-krat toliko.

Tudi zadnja Slovenska slovnica povsem brez zvrstnih komentarjev ni, kot smo videli že zgoraj v Uvodu v navedku iz Toporišič (2004: 160–161) v zvezi s tvorjenko *Smrekarjeva* (: *Smrekarica*), ki da se rabi v "meščanskih krogih, oficialno" (prav tam: 161), še pogostejše pa so v Toporišičevi slovnici stilne (ekspresivne) opombe tipa: priponska obrazila z *u* (npr. *smrduh*) so slabšalna, podobno se slabšalnost "drži ustnične prvine tudi v priponskih obrazilih -avs, -avt ipd." (prav tam: 164).

Oboji – pojavno-razvrstilni in zvrstno-stilni – komentarji obrazil so za opis besedotvornih pojavov pomembni, zato bi jih bilo smiselno ohraniti tudi v novi

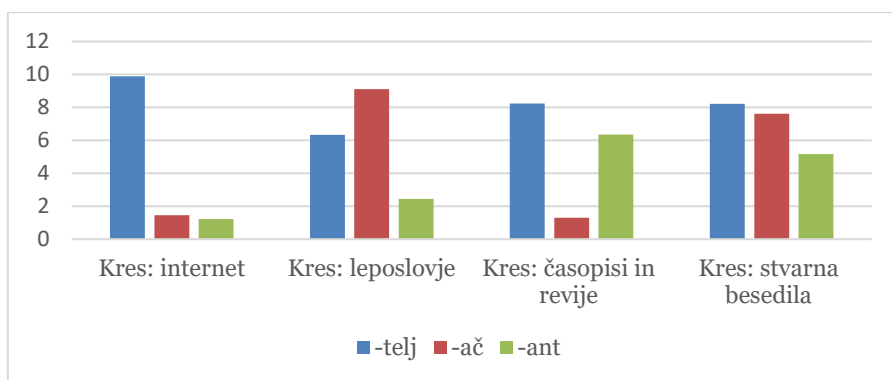
⁸ Angl. termin *derived nouns* sicer vključuje tudi sestavljenke.

slovnici, pri čemer bi morali avtorji pri zvrstno-stilnih komentarjih natančno upoštevati značilnosti besedilne in tematske zgradbe uporabljenega korpusa (oz. korpusov).

Zgolj za ponazoritev grafičnega elementa, omenjenega v zvezi s slovnico Biber in dr. (1999), prikazujemo še podkorpusno razporeditev treh naših obrazil, in sicer obrazil *-telj*, *-ač* in *-ant* (Tabela 2 in Slika 8).

Priponsko obrazilo	Kres: internet	Kres: leposlovje	Kres: časopisi in revije	Kres: stvarna besedila
<i>-telj</i>	9,88	6,34	8,24	8,22
<i>-ač</i>	1,45	9,12	1,29	7,62
<i>-ant</i>	1,23	2,45	6,36	5,16

Tabela 2: Število pojavitev samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja s 3 izbranimi priponskimi obrazili v delih korpusa Kres (na 100.000 pojavnic).



Slika 8: Število pojavitev samostalniških tvorjenk iz glagolov, moškega spola in s pomenom človeškega vršilca dejanja s 3 izbranimi priponskimi obrazili v delih korpusa Kres (na 100.000 pojavnic).

Iz Tabele 2 in s Slike 8 je razvidno, da so tvorjenke z obrazilom *-telj* z iskanim pomenom najbolj pogoste v internetnem delu Kresa, najmanj pa v leposlovju, skoraj enak delež imajo v periodiki in priročnikih; vendar pa tu razlike niso zelo velike. Večje so pri tvorjenkah z obrazilom *-ač*, ki jih je v internetnih besedilih ter v časopisih in revijah izrazito manj kot v leposlovju, pa tudi v stvarnih besedilih so precej pogoste. Tvorjenke z obrazilom *-ant* so najbolj pogoste v revijah in časopisih, sledijo stvarna besedila, malo in še manj pa jih je v leposlovju ter v internetnem delu Kresa.

Nasplošno naj še dodamo, da je take prikaze na podlagi korpusnih podatkov mogoče pripraviti za številne in zelo različne besedotvorne pojave, zato bi bilo treba razmisliti o tem, kako njihov prikaz v temeljnem jezikovnem viru, slovnici, omejiti na najbolj izstopajoče in povedne; vsekakor pa bi bilo treba pri tem poleg tvorjenk iz predvidljivih tvorbeno-pretvorbenih postopkov upoštevati tudi tvorjenke, ki niso sistemske (Logar 2006; Voršič 2013). Odprto puščamo tudi vprašanje širše zgradbe celotnega slovnicega besedotvornega poglavja, zdi pa se, da bi tudi prikaz besedotvornih vrst nasploh lahko sledil korpusnim podatkom.

5 RAZPRAVA

Uporaba velikih korpusov ima v analizi vsakršnih jezikovnih pojavov mnoge prednosti (npr. Leech 1992; Tognini - Bonelli 1993; Čermák 1995/2005; Biber in dr. 1998; Teubert 1999/2005; Kennedy 1999), v analizi besedotvorne ravnine jezika pa hkrati tudi eno pomembno slabost: velik obseg ročnega dela. Bogato oblikoslovje v slovenščini dopolnjuje prav tako pestra besedotvorna morfološkost s sedmimi jezikovnosistemskimi besedotvornimi vrstami (Vidovič Muha 1988) in močno stopenjskostjo (Kern 2017). Poleg tega ima slovenščina glasovno ničto priponsko obrazilo (tudi pripono ter medponsko obrazilo in medpono) ter obrazilne morfeme, ki so v deležu, ki ga ne moremo zanemariti, še končniško homonimni. Naš izbor obrazil in tip tvorjenk, ki smo ga analizirali, je bil v tem smislu razmeroma nezahteven: omejili smo se na

neničta obrazila, ki niso končniško homonimna (kot je npr. prav obrazilo *-ø* ali pa obrazilo *-a*), in se ustavili pri številki 10 (Toporišičev seznam je veliko daljši: vsebuje kar 54 morfemov (2000: 161–163)),⁹ iskali smo samostalniške tvorjenke (ne pa npr. prislovnih, ki bi prinesle še izločanje napačno oblikoskladenjsko označenih pridevnikov (prim. Grčar in dr. 2012)). Poleg tega smo izbrali izpeljavno besedotvorno vrsto, zanemarili pa besedotvornomorfemsko zahtevnejše (vsaj kar se avtomatskega pridobivanja podatkov tiče) zlaganje, pa tudi sestavljanje (sploh glagolsko). Kot že omenjeno, ne smemo pozabiti na napake pri korpusnem označevanju: te sicer niso predstavljale težav v primerih, ko so napačne "tvorjenke" prišle na seznam (takrat smo jih pač ročno izločili), zelo verjetno pa obstaja v korpusu še kakšna tvorjenka, ki se je zaradi napake v označitvi izognila priklicu (sploh če je šlo za hapaxe, za katere smo videli, da so pri izračunu produktivnosti ključni). Več ročnega dela obenem pomeni tudi večji vpliv človeške napake (brez dvoma tudi v tem prispevku). Opozoriti je treba tudi na to, da smo tu ravnali po načelu delnega korpusnega pristopa (Gorjanc 2005: 24), medtem ko bi bil za končni slovnični opis ustrenejši popolni korpusni pristop s kolikor se da predprocesiranimi podatki.¹⁰

Baayenove formule za izračun morfološke produktivnosti še danes veljajo za sicer ne idealne, ob natančni interpretaciji pa vendarle znanstveno sprejemljive. Pri tem ne smemo spregledati avtorjevega opozorila (Baayen, Lieber 1998: 820; tudi Bauer 2005: 325), da produktivnost *P* velja samo za določen korpus, kar pomeni, da *P*-ji, izračunani na podatkih iz različnih

⁹ V lasten zagovor lahko povemo, da so se tudi vsi raziskovalci, na katere smo se z literaturo opirali pri obravnavi besedotvorne produktivnosti, omejili le na nekaj obrazil (od 2 do pribl. 10).

¹⁰ Možnost večjega avtomatskega predprocesiranja podatkov za izračun morfološke produktivnosti sta na primeru nemščine testirala Evert in Lüdeling (2001) ter pri tem ugotovila, da sistema, ki bi dal za ta namen dovolj kakovostne podatke brez ročnega pregleda, še ni. Za slovenščino si je mogoče obetati vsaj delno tovrstno pomoč z zbirkami podatkov, ki bodo nastale pri projektu *Nova slovnica sodobne standardne slovenščine: viri in metode* (ARRS, 2017–2019, <http://slovnica.ijs.si/>).

korpusov, med sabo niso primerljivi. Vpliv korpusne zgradbe na to meritev je torej izrazit.

Baayen na več mestih utemeljuje, da so njegove formule precej skladne z jeziko(slov)no intuicijo (Baayen, Lieber 1991: 801, 808, 811, 817; Baayen 1992: 11; Baayen 1993: 81). Ker gre za subjektivno lastnost, deloma priučeno na preteklih načinih interpretiranja raziskovalnih ugotovitev (pri jezikoslovcih) bodisi temelječo na jezikovnem občutku (pri teh in vseh drugih govorcih jezika), se je z avtorjem mogoče tako strinjati kot mu nasprotovati. V slovenskem prostoru vrednosti P kot "matematično izražene jezikoslovne predstave morfološke produktivnosti" (Baayen, Lieber 1991: 811) do zdaj še nismo preizkusili, zato je na neki način tudi nismo navajeni, a vrednost zgoraj prikazanega je vsekakor v objektivnosti, ki deluje prepričljivo. Vseeno pa – če se nadalje še enkrat navežemo na zgornjo Sliko 7 – vrednosti P tu nismo izrazito postavili v središče slovničnega opisa. Ocenili smo, da je pri prikazu obrazil še vedno bolje izhajati iz števila in besedilne razpršenosti korpusno izkazanih tvorjenk, ker je ta količinskost za opis tehtnejša (in priznajmo: tudi intuitivnejša). Pri tem smo skladni z Biber in dr. (1999), ki pod naslovom *Produktivnost značilnih samostalniških izpeljank* (prav tam: 323) naslovni termin definirajo kot skupno število različnih tvorjenk, ki imajo določeno obrazilo, in posledično med drugim npr. povzemajo, da je "obrazilo *-tion* daleč najbolj produktivno tako v smislu skupnega števila samostalnikov, ki jih tvori, kot v skupini relativno redkih tvorb (tj. izpeljanih samostalnikov, ki se pojavijo manj kot desetkrat na milijon besed)" (prav tam). Kljub temu pa se produktivnost P zdi upoštevanja vredna pri izbiri prikazanih obrazil (natančneje: opustitvi neproduktivnih) in pri njihovi zvrstno-stilni interpretaciji (sploh ko gre za sopomenska obrazila). Je pa – ponavljamo – izredno pomembno, kakšen je korpus, iz katerega za take izračune pridobivamo podatke.

Kljub temu da smo se tu ukvarjali le s priponskimi obrazili, velja ob koncu razprave omeniti še to, da bi Baayenu lahko sledili tudi pri izračunu

produktivnosti posameznih besedotvornih vrst. Čeprav bi tak podvig zahteval ročni pregled še veliko večje količine podatkov, bi bil zanimiv (in morda v marsičem presenetljiv).

6 SKLEP

Obravnava besedotvornih prvin slovenščine ima močno tradicijo in je tudi v sodobnosti eno živahnejših raziskovalnih področij (prim. npr. Michelizza 2008; Vidovič Muha 2009; Gložančev in dr. 2009; Stramljič Breznik 2013; Stramljič Breznik 2016; Kern 2014; Štumberger 2015; Žele 2016; Sicherl, Žele 2018). Veliki korpusi lahko tovrstne študije vsekakor podatkovno okrepijo – raziskovalce pa spodbudijo tudi k pripravi novih teoretičnih izhodišč –, kar pa še ne pomeni, da olajšajo analitični del takih raziskav. A to je lahko hkrati tako slabost kot prednost; namreč v domeni, ki je bila v času prve izdaje zadnje Slovenske slovnice (Toporišič 1976) še povsem jezikoslovna, lahko zdaj poznavalcem morfemske zgradbe jezika pomagajo tudi računalničarji, metodologi in matematiki.

Kot je bilo v prispevku razvidno, je pri korpusnem pristopu k besedotvorju malo zadreg le pri naboru zgledov. Če prvi del analize podatkov nekako še zmorejo strokovnjaki za morfematiko, pa bi bilo treba k pripravi komentarja obrazil, zajetih v slovnični opis, povabiti tudi poznavalce stilistike, besedilne zvrstnosti, normativistike in etimologije. Pri prihodnjem slovničnem prikazu besedotvornega dela slovenščine tako brez timskega dela ne bo šlo – vsaj če se odgovorno zavežemo nalogi, da delamo v vseh pogledih sodoben jezikovni opis (in predpis).

ZAHVALA

Članek je nastal v okviru programskega financiranja ARRS. Raziskovalni program št. P6-0215 je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

LITERATURA

- Aronoff, M. (1976): *Word formation in generative grammar*. Cambridge: MIT Press.
- Aronoff, M., in Anshen, F. (1998): Morphology and the lexicon: Lexicalization and productivity. V A. Spencer in A. M. Zwicky (ur.): *The handbook of morphology*: 237–246. Oxford: Blackwell Publishers.
- Baayen, H. (1993): On frequency, transparency and productivity. V G. Booij in J. van Marle J. (ur.): *Yearbook of Morphology 1992*: 181–208. Dordrecht: Springer.
- Baayen, H. (1994): Productivity in language production. *Language and cognitive processes*, 9 (3): 447–469.
- Baayen, H. (2009): Corpus linguistics in morphology: morphological productivity. V A. Lüdeling in M. Kytö (ur.): *Corpus linguistics: An international handbook*: 900–919. Berlin; New York: Mouton de Gruyter.
- Baayen, H., in Lieber, R. (1991): Productivity and English derivation: A corpus-based study. *Linguistics*, 29 (5): 801–843.
- Bajec, A., Kolarič, R., in Rupel, M. (1956): *Slovenska slovnica*. Ljubljana: Državna založba Slovenije.
- Bauer, L. (2001): *Morphological productivity*. Cambridge: Cambridge University Press.
- Bauer, L. (2005): Productivity: Theories. V P. Štekauer in R. Lieber (ur.): *Handbook of word-formation*: 315–334. Dordrecht: Springer.
- Biber, D., Conrad, S., in Reppen, R. (1998): *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., in Finegan, E. (1999): *Longman grammar of spoken and written English*. Essex: Pearson

Education.

Booij, G. (2012): *The grammar of words: An introduction to linguistic morphology*. Oxford: Oxford University Press.

Breznik, A. (1934): *Slovenska slovnica za srednje šole*. Celje: Družba sv. Mohorja.

Center za jezikovne vire in tehnologije, Univerza v Ljubljani. Dostopno prek: <https://www.cjvt.si/> (22. junij 2018).

Cvrček, V. (2010): *Mluvnice současné češtiny: 1, Jak se píše a jak se mluví*. Praga: Karolinum.

Čermák, F. (1995/2005): Jezikovni korpus: sredstvo in vir spoznanj. V V. Gorjanc in S. Krek (ur.): *Študije o korpusnem jezikoslovju*: 137–171. Ljubljana: Krtina./*Slovo a slovesnost*, 56: 119–140.

Evert, S., in Lüdeling, A. (2001): Measuring morphological productivity: Is automatic preprocessing sufficient?. V P. Rayson idr. (ur.): *Proceedings of the Corpus Linguistics 2001 Conference*: 167–175. Lancaster: UCREL.

Frauenfelder, U. H., in Schreuder, R. (1992): Constraining psycholinguistic models of morphological processing and representation: The role of productivity. V G. Booij in J. van Marle (ur.): *Yearbook of Morphology 1991*: 165–183. Dordrecht: Springer.

Gložančev, A., Jakopin, P., Michelizza, M., Uršič, L., in Žele, A., ur. (2009): *Novejša slovenska leksika: v povezavi s spletnimi jezikovnimi viri*. Ljubljana: Založba ZRC, ZRC SAZU.

Gorjanc, V. (2005): *Uvod v korpusno jezikoslovje*. Domžale: Založba Izolit.

Górski, R. (2013): Polish derived expressive adjectives, a corpus-based study. V M. Polczyńska, Ł. Piotr Pakuła in D. Jaworska (ur.): *Young linguists' insights: Taking interdisciplinary approaches to the core*: 65–80.

Varšava: Wydział Anglistyki UAM.

Grčar, M., Krek, S., in Dobrovoljc, K. (2012): Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. V T. Erjavec in J. Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*: 89–94. Ljubljana: Institut "Jožef Stefan".

Herrity, P. (2016): *Slovene: A comprehensive grammar*. London; New York: Routledge.

Janežič, A., in Sket, J. (1900): *Slovenska slovnica*. Celovec: Družba sv. Mohorja.

Jurgec, P. (2011): Slovenščina ima 9 samoglasnikov. *Slavistična revija*, 59 (3): 243–268.

Kennedy, G. (1999): *An introduction to corpus linguistics*. London; New York: Longman.

Kern, B. (2014): Zagadnienia słowotwórcze w wybranych podręcznikach do nauczania języków słoweńskiego i polskiego jako obcych. V E. Pałuszyńska (ur.): *Glottodydaktyka – media – komunikacja, Negocjowanie znaczeń*: 211–220. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Kern, B. (2017): *Stopenjsko besedotvorje*. Ljubljana: Založba ZRC, ZRC SAZU.

Kilgariff, A., Rychlý, P., Smrz, P., in Tugwell, D. (2004): The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*: 105–116. Lorient: Université de Bretagne-Sud.

Leech, G. (1992): Corpora and theories of linguistic performance. V J. Svartvik (ur.): *Directions in corpus linguistics*: 105–122. Berlin: Mouton de Gruyter.

Lieber, R. (1992): *Deconstructing morphology: Word formation in syntactic theory*. Chicago; London: The University of Chicago Press.

- Logar, N. (2005): Besedotvorni sklopi. *Slavistična revija*, 53 (2): 171–192.
- Logar, N. (2006): Stilno zaznamovane nove tvorjenke – tipologija. *Slavistična revija*, 54 (pos. številka): 87–101.
- Michelizza, M. (2008): Nove tvorjenke v spletnih besedilih (primer Wikipedije). V M. Košuta (ur.): *Slovenščina med kulturami*: 328–338. Celovec; Ljubljana: Slavistično društvo Slovenije.
- Nova slovnica sodobne standardne slovenščine: viri in metode*. Dostopno prek: <http://slovnica.ijs.si/> (22. junij 2018).
- Plag, I. (1999): *Morphological productivity: Structural constraints in English derivation*. Berlin; New York: Mouton de Gruyter.
- Säily, T. (2011): Variation in morphological productivity in the BNC: Sociolinguistic and methodological considerations. *Corpus linguistics and linguistic theory*, 7 (1): 119–141.
- Scherer, C. (2015): Change in productivity. V P. O. Müller idr. (ur.): *Word-formation: An international handbook of the languages of Europe*: 1781–1793. Berlin; Boston: De Gruyter Mouton.
- Sicherl, E., in Žele, A. (2018): Prekrivanke v slovenščini z vidika vpliva angleškega jezika. *Jezik in slovstvo*, 63 (1): 75–88.
- Stramljič Breznik, I. (2013): Besedotvorna produktivnost posnemovalnih medmetov za zvoke naprav v slovenščini. *Slavia Meridionalis*, 13: 97–107.
- Stramljič Breznik, I. (2016): Slovene. V P. O. Müller (ur.): *Word-formation: An international handbook of the languages of Europe*: 2979–2998. Berlin; Boston: De Gruyter Mouton.
- Stramljič Breznik, I., in Voršič, I. (2011): Word-formational productivity of the Slovene language in the case of sports neologisms. *Linguistica*, 51: 23–38.

- Štumberger, S. (2015): Besedotvorje novejšje slovenske leksike: medponskoobrazilne zloženke. V D. Zuljan Kumar in H. Dobrovoljc (ur.): *Zbornik prispevkov s simpozija 2013*: 155–163. Nova Gorica: Založba Univerze.
- Teubert, W. (1999/2005): Korpusno jezikoslovje in leksikografija. V V. Gorjanc in S. Krek (ur.): *Študije o korpusnem jezikoslovju*: 103–136. Ljubljana: Krtina./Korpuslinguistik und Lexikographie. *Deutsche Sprache*, 4.
- Tognini - Bonelli, E. (1993): *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins.
- Toporišič, J. (2000): *Slovenska slovnica*. Maribor: Založba Obzorja.
- Van Marle, J. (1992): The relationship between morphological productivity and frequency: A comment on Baayen's performance-oriented conception of morphological productivity. V G. E. Booij in J. van Marle (ur.): *Yearbook of Morphology 1991*: 151–163. Dordrecht: Springer.
- Vidovič Muha, A. (1988): *Slovensko skladenjsko besedotvorje ob primerih zloženk*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Vidovič Muha, A. (2009): Skladenjska interpretacija glagolskih predponskih obrazil – vprašanje propozicije. *Slavistična revija*, 57 (2): 251–261.
- Voršič, I. (2013): *Sistemska in nesistemska leksikalna tvorba v novejšem besedju slovenskega jezika: Doktorska disertacija*. Maribor.
- Žele, A. (2016): Aktualni jezikovni načini izražanja v slovenščini: sklapljanje kot naravni in aktualni odraz nepretrganosti govora v narečnem in knjižnem jeziku. *Annales*, 26 (4): 709–714.

NOUN DERIVATIVES FROM VERBS: CORPUS WORD-FORMATION PRODUCTIVITY AND GRAMMATICAL DESCRIPTION

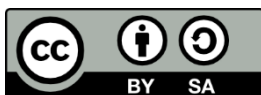
This paper discusses the understanding of word-formation productivity, as was founded by Harald Baayen at the end of the 1980s and in the early 1990s. Baayen's method of calculating productivity P is based on corpus linguistics methodology, therefore by using data from the Kres corpus, a corpus of written Slovene, we tested the method on the following ten Slovene suffixes, which all form noun derivatives from verbs (gender: masculine, meaning: human agent): *-vec*, *-telj*, *-ež*, *-uh*, *-avh*, *-ar*, *-lec*, *-ač*, *-aj*, and *-ant*. Among other findings, one was indicative: the number of derivatives with a certain suffix, and the productivity P of that same suffix could be (quite) diverse. For example: the suffix that had the largest number of derivatives was *-lec*, while the suffix *-ež* had the highest P score. On the basis of the data from the corpus, we then presented a list of the ten aforementioned suffixes along with example derivatives, i.e. the list, that could be part of a new grammar of Slovene. We also briefly commented on it and then, in the final discussion, pointed out some advantages and disadvantages of the approach presented.

Keywords: word-formation, productivity, H. Baayen, suffix, derivative, grammar, corpus

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/>



OSEBNI ZAIMKI V FUNKCIJI OSEBKA: PRAGMATIČNI VIDIKI SLOVNIČNE KATEGORIJE

Agnes PISANSKI PETERLIN

Filozofska fakulteta Univerze v Ljubljani

Tamara MIKOLIČ JUŽNIČ

Filozofska fakulteta Univerze v Ljubljani

Pisanski Peterlin, A., Mikolič Južnič, T. (2018): Osebni zaimki v funkciji osebk: pragmatični vidiki slovnične kategorije. Slovenščina 2.0, 6 (2): 127–153.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.127-153>.

S korpusno analizo je mogoče pridobiti relevantne podatke o slovničnih pojavih, na podlagi katerih se lahko oblikujejo poglobljeni slovnični opisi. Korpusna metodologija je posebej uporabna za slovnične raziskave, ki se osredotočajo na pragmatiko in variacije v rabi izbranih slovničnih elementov. Namen pričujoče raziskave je analizirati rabo osebnih zaimkov v funkciji osebk v slovenščini s pragmatičnega vidika in tako pridobiti nove podatke za bolj podroben opis rabe osebnega zaimka kot osebk v slovenščini. Slovenščina je jezik z opustljivim osebkom, kar pomeni, da osebk stavka v njej ni treba eksplicitno izraziti. Raziskava temelji na korpusni analizi korpusov Kres in Gos; narejena je bila s pomočjo spletne programske aplikacije za analizo besedil NoSketch Engine. Rezultati pokažejo precejšnjo razliko v pogostosti rabe osebnih zaimkov kot osebkov med govornim in pisnim diskurzom. Rezultati prav tako pokažejo, da število osebnih zaimkov upada z naraščanjem formalnosti. Na podlagi funkcijske analize so izraženi osebni zaimki v najširšem smislu klasificirani v kategoriji razdvoumljanja in poudarjanja. V okviru razdvoumljanja na rabo osebnih zaimkov kot osebkov vplivajo oblikoslovni, skladijski in diskurzni dejavniki, in tudi prenosnik. V kategoriji poudarjanja je prenosnik posebej relevanten dejavnik. Analiza pokaže, da je tematika osebnega zaimka kot osebk v slovenščini mnogo kompleksnejša, kot bi lahko sklepali iz dosedanjih opisov. Izsledki raziskave odpirajo nekatera zanimiva vprašanja glede rabe osebnega zaimka kot osebk v slovenščini v zvezi z dialektalno specifikom in v hibridnih

internetnih žanrih, prav tako pa tudi vprašanja o vplivu lektorskih posegov in o vplivu usvajanja pismenosti.

Ključne besede: osebni zaimek kot osebek, jezik z opustljivim osebkom, korpusna analiza

1 UVOD

Uporaba korpusnega pristopa za raziskave slovničnih pojavov predstavlja odmik od tradicionalnih raziskav slovnične problematike, ki so vir za slovnične opise jezikovnega sistema (gl. tudi Krek 2013). Kot opozarja Susan Conrad (2010: 227) tradicionalne raziskave slovničnih pojavov, ki ne izhajajo iz analize rabe jezika (ki jo je mogoče opazovati na korpusnem gradivu), navadno temeljijo na razlikovanju med slovnično pravilnimi ali nepravilnimi primeri. Conrad (ibid.) poudarja, da je tak tradicionalni pristop, osnovan na dihotomiji ustrezno-neustrezno, uporaben za opise nekaterih slovničnih elementov, ne pa vseh: pri nekaterih slovničnih kategorijah je mogoče opazovati več variacij glede na različne sociolingvistične in funkcijske spremenljivke. Conrad (2010: 228) tako kot prispevek korpusnega jezikoslovja k opisom slovnice navaja zlasti to, da korpusni pristop omogoča proučevanje velikih zbirk besedil, v katerih je mogoče opazovati tipične in netipične vzorce v različnih kontekstih. Lahko bi torej povzeli, da je korpusni pristop posebej uporaben za slovnične raziskave, ki se osredotočajo na pragmatiko in variacije v rabi izbranih slovničnih elementov, seveda pa je treba vedno upoštevati tudi sestavo in vrsto korpusa.

Primer slovničnega problema, ki ga ni mogoče opisati z dihotomijo ustrezno-neustrezno, v slovenskem jeziku predstavlja raba osebnega zaimka v funkciji osebk. V slovenščini kot v mnogih drugih jezikih osebk glavnega stavka ni treba eksplicitno izraziti (prim. tudi Golden 1996: 171, Toporišič 2000: 307); osebk torej ni treba še dodatno zaimkovno opredeljevati, ker je v večini primerov izražen v končnici osebnih glagolskih oblik. Jeziki z opustljivim osebkom (izraz je povzet po Golden 1996: 261) so bili podrobno analizirani v sklopu tvorbeno-pretvorbenega jezikoslovja (prim. npr. Jaeggli & Safir 1989),

raziskave tega pojava pa segajo prek okvirov teoretičnega jezikoslovja: razlikovanje med jeziki z opustljivim in obveznim osebkom je posebej zanimivo z vidika stika med jeziki, npr. usvajanja tujega jezika (npr. Tsimpli in sod. 2004, Sorace in Filiaci 2006), dvojezičnosti (npr. Paradis in Navarro 2003, Serratrice 2007) in prevajanja (npr. Mauranen 2005, Baroni in Bernardini 2006).

Prav študije medjezikovnih razlik in medsebojnih vplivov jezikov so pokazale, da je tudi med jeziki z opustljivim osebkom mogoče razlikovati različne podtipe (prim. tudi klasifikacijo, ki jo navajata Roberts in Holmberg 2010: 5–13) v okviru katere se slovenščina, podobno kot npr. italijanščina, španščina in sodobna grščina, uvršča v kategorijo jezikov s konsistentno opustljivim osebkom, ki so zaenkrat najbolj raziskana kategorija jezikov z opustljivimi osebki (Roberts in Holmberg 2010: 6). Vendar so podrobnejše analize pokazale, da je celo znotraj posameznih podtipov jezikov mogoče opazovati razlike, znotraj posameznih jezikov pa dialektalne mikrovariacije (prim. npr. Sorace in Filiaci 2006). Za slovenščino je primerjava slovenskih in italijanskih zaimkov v funkciji osebkov (Pisanski Peterlin in Mikolič Južnič 2018) pokazala velike razlike v pogostosti rabe med jezikoma, kar je precej presenetljivo, saj jezika sicer sodita v isti podtip jezikov z opustljivim osebkom. Na podlagi izsledkov omenjene raziskave je mogoče za slovenščino sklepati o zelo specifični rabi osebnih zaimkov v funkciji osebkov, ki je v primerjavi z nekaterimi drugimi jeziki z opustljivim osebkom, bolj omejena.

V opisu rabe osebnega zaimka kot osebkov v slovenščini v *Slovenski slovnici* (Toporišič 2000: 307) je navedeno, da se v slovenščini osebni zaimek v imenovalniku uporablja v specifičnih situacijah: v odgovoru na vprašanja po glagolski osebi (npr. kot odgovor na vprašanje, ki se začne s *kdo*), če je glagolska oseba eden izmed delov priredno sestavljenega dela povedi (npr. *vi in jaz*), če ima glagolska oseba še kakšna določila (Toporišič *ibid.*) navaja primer *Jaz, knez modruški ...*), kadar glagolska oseba iz povedka in sobesedila ni razvidna (npr. pogojnik v ednini), če hočemo podati ozračje domačnosti (navedenih je nekaj primerov iz neformalnega govora) in če se hoče vzpostaviti kontrast

(Toporišič (ibid.) navaja primer *Jaz grem na delo* in argumentira, da je *jaz* v takem stavku smiselno le v primeru, kadar želimo »povedati, da pa za druge ne ve, kam gredo«). O močni tendenci opuščanja osebnega zaimka kot osebkov v knjižnem jeziku priča tudi opažanje, ki je navedeno v Koletnik (2008: 39), in sicer, da se v prekmurskem narečju pojavlja osebni zaimek, kjer je v knjižni slovenščini pričakovan »ničti zaimek«. Podobno bi lahko tudi izpostavili možnost nadomeščanja osebnega zaimka s kazalnim ali z zaimkom *sam* (prim. Pisanski Peterlin in Mikolič Južnič 2018: 242). K temu velja dodati, da se osebni zaimki v imenovalniku v slovenščini navadno lahko nanašajo le na žive osebe: posebnih raziskav na temo rabe zaimkov v imenovalniku za nežive osebe sicer še ni bilo, zdi pa se, da je takšna raba redka in zaznamovana.

Namen pričujoče raziskave je analizirati rabo osebnih zaimkov v funkciji osebkov v slovenščini s pragmatičnega vidika in tako nadgraditi obstoječe slovnične opise. Na podlagi korpusne analize korpusov Kres (<http://www.korpus-kres.net/>) in Gos (<http://www.korpus-gos.net/>) so osebni zaimki v funkciji osebkov razčlenjeni z žanrskega in diskurzno-pragmatičnega vidika.

2 IZRAŽANJE OSEBKA KOT DISKURZNO-PRAGMATIČNO SREDSTVO

Referenca kot oblika slovnične kohezije se izraža skozi kazalne in osebne zaimke (za podrobnejši pregled koncepta za slovenščino, prim. Korošec (1981) in Gorjanc (1999)). Gorjanc (1999: 141) poudari, da je pogoj za referenčnost v besedilu prisotnost nanašalnice in naveznika/napovednika, ki znotraj besedila vzpostavljata koreferenčno besedilno vez. Serratrice (2007: 1058–9) izpostavi dejstvo, da je funkcija jezikovnih elementov, ki vzpostavljajo referenčno kohezijo, označevanje informacije o nekem elementu v smislu razlikovanja med znanim in novim.

Uporaba osebnih zaimkov kot kohezijskih sredstev se med jeziki razlikuje. V jezikih z obveznim osebkom se razlikovanje med novimi in starimi informacijami odraža skozi pronominalizacijo (Paradis and Navarro 2003: 375); v teh jezikih se nove informacije vpeljejo s samostalniško besedno zvezo,

stare informacije pa se izrazijo z zaimkom. Po drugi strani pa se v jezikih z opustljivim osebkom razlikovanje med novimi in starimi informacijami odraža skozi izražen ali opuščen osebek. V teh jezikih se osebni zaimke kot osebek pojavlja ob tematskih premikih, da bi se izognili referenčni dvoumnosti, in na mestih, kjer uvajajo kontrast ali poudarek (prim. Paradis and Navarro 2003; Rothman 2009). Serratrice (2007) v svoji študiji referenčne kohezije v pripovedih angleško-italijansko dvojezičnih otrok poudari razliko med vzdrževanjem in ponovno vpeljavo nanašalnice: v jezikih z opustljivim osebkom osebek ni eksplicitno izražen, kadar se v dveh zaporednih stavkih ohranja ista tema, kadar pa pride do tematskega premika, se uporabi osebni zaimke kot osebek.

Vendar pa nekatere druge raziskave jezikov z opustljivim osebkom kažejo, da poleg opisanih diskurzno-pragmatičnih dejavnikov na odločitev za opustitev osebkov vpliva lahko še vrsta drugih dejavnikov. V literaturi se tako kot drugi potencialni dejavniki, ki vplivajo na rabo osebnih zaimkov, omenjajo žanr oziroma vrsta diskurza v širšem smislu (Travis (2007) ugotavlja občutne razlike med monološko pripovedjo in dialoškim pogovorom), referent slovnične osebe (Posio (2012) pokaže, da je v evropski španščini in portugalščini mogoče opazovati različno pogosto rabo osebnega zaimka kot osebkov med inkluzivno in ekskluzivno prvo osebo ednine) in dialektalna specifika (Paradis and Navarro (2003: 375) navajata izrazito specifiko karibske španščine v primerjavi z drugimi španskimi dialekti). V tem kontekstu velja omeniti, da tudi za slovenščino Koletnik (2008: 39) opozori na dialektalno specifiko rabe osebnega zaimka kot osebkov pri upovedovanju prekmurskega narečja, kjer prihaja do odstopanja od knjižne norme, pri čemer v prispevku narava knjižne norme ni nadalje pojasnjena. Očitno je torej, da je raba osebkov v jezikih z opustljivim osebkom kompleksno vprašanje in zahteva večplastni opis. Zato predstavlja vsak poskus, da bi se pri opisu omejili na en sam dejavnik, poenostavitev, ki je v nekaterih primerih morda uporabna, v raziskavi, kot je pričujoča, katere namen je nadgraditi obstoječe razumevanje tega pojava v slovenščini, pa seveda

ni.

Na podlagi pregleda literature se postavlja vprašanje, ali je s korpusno analizo dejanske rabe osebnega zaimka v slovenščini mogoče pridobiti natančnejše podatke o pragmatičnih vidikih izražanja in opuščanja osebkov v slovenščini. Konkretno želiva v pričujoči študiji odgovoriti na naslednji raziskovalni vprašanji:

- 1) Ali je pogostost rabe osebnih zaimkov kot osebkov pogojena z besedilno zvrstjo?
- 2) Kateri dejavniki vplivajo na rabo osebnih zaimkov kot osebkov?

3 KORPUS IN METODA

3.1 Korpus

V analizi bila sta uporabljena korpusa Kres in Gos, ki sta nastala v okviru projekta *Sporazumevanje v slovenskem jeziku*; za več podrobnosti o korpusu Kres gl. npr. Logar in sod. (2012), za več podrobnosti o korpusu Gos gl. Verdonik in Zwitter Vitez (2011), za več podrobnosti o samem projektu pa projektni spletni portal (www.slovenščina.eu). Korpus Kres je uravnoteženi pisni korpus slovenskega jezika, ki po podatkih z omenjenega portala obsega 99.831.145 besed v besedilih, ki so izšla med letoma 1990 in 2011¹ (prim. tudi Erjavec in Logar Berginc 2012), korpus Gos pa je govorni korpus slovenskega jezika, ki obsega 1.035.101 besedo (prim. tudi Verdonik in sod. 2013).²

Za korpusno analizo je uporabljen korpus Gos v celoti. Korpus Gos omogoča iskanje po štirih sklopih oziroma podkorpusih (Verdonik in sod. 2013), ki so ločeni po besedilnovrstnih kriterijih, in sicer: javni informativno-izobraževalni govor (informativni medijski diskurz ter pedagoški diskurz v osebnem stiku), javni razvedrilni (medijski diskurz), nejavni nezasebni (telefonski in osebni

¹ Omeniti velja, da je velik delež besedil v korpusu Kres lektoriran, saj gre za objavljena literarna, neliterarna in medijska besedila.

² Podatki o obsegu korpusov so bili veljavni v času, ko je bila narejena analiza (julij 2016).

pogovori v uradih, trgovinah itd.) in nejavni zasebni (telefonski in osebni pogovori v družini in med prijatelji).

Korpus Kres za korpusno analizo ni bil uporabljen povsem v celoti, čeprav je bila analizirana velika večina korpusa (v skupnem obsegu 85.993.551 besed), kar predstavlja več kot 85 % celotnega korpusa Kres. Za primerjavo med besedilnimi zvrstmi je bilo namreč smiselno zagotoviti čim večjo žanrsko in registrsko homogenost posameznih zvrsti. Tako so bili uporabljeni naslednji sklopi besedil iz korpusa Kres:³ leposlovje, revije, časopisi, stvarna besedila brez podkategorije »drugo« in besedil Zavoda za šolstvo (za zagotovitev čim večje homogenosti) in internetna besedila brez transkriptov iz Državnega zbora (zaradi hibridne narave besedil), ne pa tudi kategorija »drugo«, ki je bila preveč heterogena za smiselno uporabo. V pričujoči raziskavi so tako rezultati analize korpusa Kres združeni v naslednjih podkorpusih, glede na besedilno zvrst: leposlovje, periodika (obsega sklopa časopisi in revije), strokovna besedila (obsega poljudnoznanstvena in strokovna neliterarna besedila v sklopu stvarnih besedil) in internetna besedila (obsega novičarske portale s komentarji in spletne strani podjetij).

3.2 Metoda

V prvi fazi analize so bile s pomočjo spletne programske aplikacije NoSketch Engine (<http://nl.ijs.si/noske/index.html>) identificirane vse pojavitve osebnih zaimkov v imenovalniku v vseh podkorpusih, zajetih v analizi. Iskalni parametri so obsegali besedno vrsto »osebni zaimek« in sklon »imenovalnik«,⁴ zato je bilo pri večini zaimkov šuma, torej primerov, ko ni šlo za osebne zaimke v imenovalniku, izjemno malo.

³ Zaradi doslednosti opisa metode dela, preciznosti rezultatov in omogočanja ponovljivosti se pri kategorizaciji korpusa Kres ne sklicujeva na teoretično klasifikacijo v Logar in sod. (2012), temveč na kategorizacijo, ki izhaja iz strani spletnega konkordančnika korpusa Kres (<http://www.korpus-kres.net/>).

⁴ Zaradi narave korpusnega iskanja, ki izhaja iz oblikoskladenjske označenosti uporabljenih korpusov, je bilo med zadetki nekaj primerov imenovalnikov, ki niso bili uporabljeni v vlogi osebk, temveč npr. povedkovnega določila, ki so bili iz kvalitativne analize izločeni.

Nekoliko več šuma se je pojavilo pri zaimkih *ona*, *ono* in *oni* zaradi homonimije s kazalnim zaimkom *oni*. Pri zaimkih *ona* in *oni* je bilo na podlagi pregleda 900 primerov ocenjeno, da je šuma še vedno malo (v nobenem podkorpusu ne več kot 10 %, v nekaterih tudi občutno manj), zato so pri teh dveh zaimkih navedene številke vseh zadetkov. Za zaimek *ono* pa so bili vsi rezultati prečiščeni ročno, saj je bilo skupno število zadetkov zelo nizko.

Za osebne zaimke *ti*, *mi*, *me* in *ve* je prvo iskanje pokazalo, da se med rezultati pojavlja veliko šuma, ki je posledica homonimij s kazalnim zaimkom *ta*, dajalnikom in tožilnikom osebnega zaimka *jaz* in tretjo osebo ednine glagola *vedeti*, zato je bilo za te zaimke opravljeno natančnejše iskanje, in sicer na podlagi iskalnih kriterijev besedna vrsta »osebni zaimek«, sklon »imenovalnik«, bližina glagola v ustreznem številu in osebi (oddaljenost do 5 besed) in dodatno pri *ti* izločitev zadetkov, ki se nahajajo neposredno pred samostalnikom (in lahko tudi pridevnikom in samostalnikom) v imenovalniku (za izločitev kazalnega zaimka).

Rezultati so bili nato izraženi v obliki števila pojavitev na 1.000 besed, kar je omogočalo natančnejšo primerjavo pogostosti med posameznimi vrstami diskurza.⁵

V drugi fazi analize so bili ročno pregledani naključno vzorčeni sezname zadetkov za vsak zaimek posebej, in sicer po 50 zadetkov za vsak sklop besedil ali manj, kjer je bilo skupno število zadetkov v sklopu manjše od 50. Skupno je to pomenilo pregled 450 zadetkov za vsak zaimek, oz. nekoliko manj v primerih, kjer je bilo število zadetkov manjše od 50. Vsak primer je bil klasificiran glede na rabo. Osnovna delitev na kategoriji *razdvoumljanje* in *poudarek* je izhajala iz literature (prim. razdelek 2) in zlasti iz prejšnje raziskave rabe osebnih

⁵ V pričujočem prispevku je termin diskurz rabljen v smislu ravni jezika nad ravniyo povedi in ne le v smislu govornega jezika. Termin »diskurz« je namreč na področju diskurzno-pragmatičnih študij zaimkov v jezikih z opustljivim osebkom dobro uveljavljen in ga ni mogoče preprosto nadomestiti, saj omogoča osredotočanje na vprašanje jezika v rabi (za natančnejši opis kompleksnosti termina diskurz v uporabnem jezikoslovju prim. npr. Hyland in Paltridge (2011: 1)).

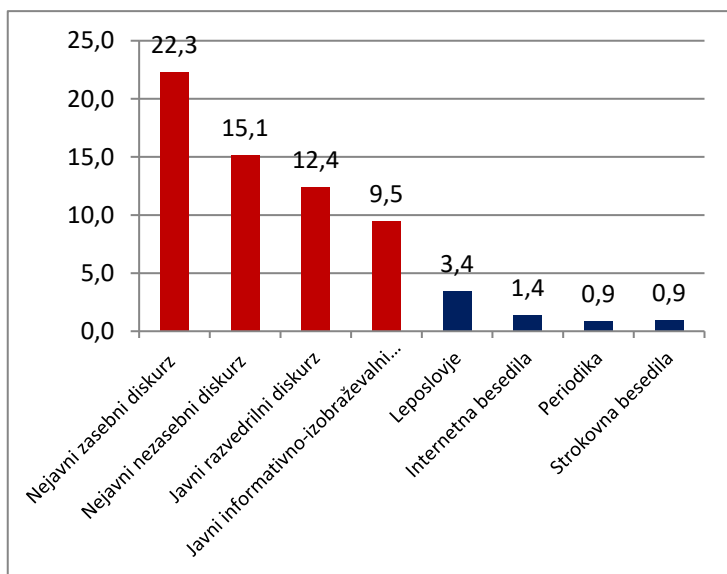
zaimkov za slovenščino in italijanščino (prim. Pisanski Peterlin in Mikolič Južnič 2018), kjer se je delitev izkazala za učinkovito; dodatne podkategorije so izhajale iz korpusne analize.

4 REZULTATI IN DISKUSIJA

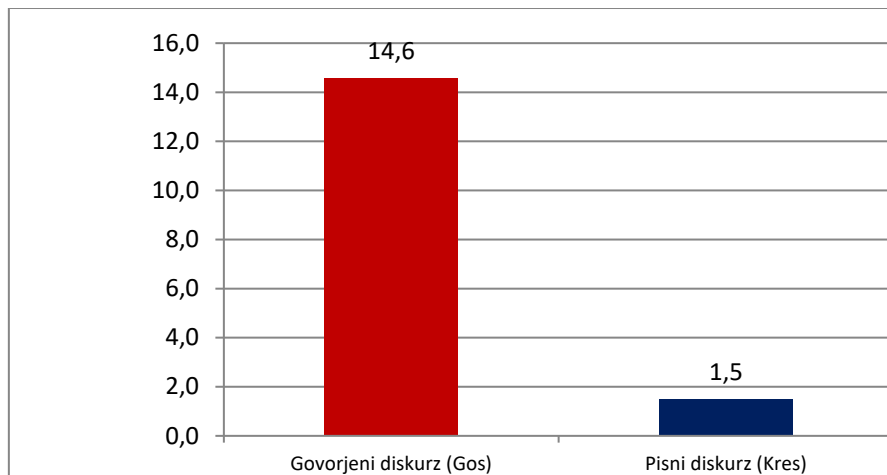
V nadaljevanju so ločeno predstavljeni in razčlenjeni kvantitativni rezultati (razdelek 4.1) in kvalitativni izsledki analize (razdelek 4.2)

4.1 Pogostost rabe osebnih zaimkov kot osebkov

Povprečna pogostost pojavitev osebnih zaimkov kot osebkov v celotnem analiziranem gradivu je 1,7 pojavitve na 1000 besed. Na Sliki 1 so v obliki stolpčnega grafa predstavljeni podatki o pogostosti rabe osebnih zaimkov kot osebkov na 1000 besed v analiziranih podkorpusih besedil korpusov Gos in Kres glede na besedilno zvrst.



Slika 1: Pogostost rabe osebnih zaimkov kot osebkov v posameznih podkorpusih: število pojavitev na 1000 besed.



Slika 2: Primerjava pogostosti rabe osebnih zaimkov kot osebkov v govorjenem in pisnem diskurzu: število pojavitev na 1000 besed.

Primerjava podatkov presenetljivo jasno pokaže, da se raba osebnega zaimka kot osebkov v slovenščini v različnih zvrsteh močno razlikuje. Če kot kriterij za razvrstitev zvrsti upoštevamo formalnost,⁶ je za govorjeni diskurz smiselno privzeti, da je nejavni zasebni diskurz (npr. pogovor med prijatelji ali družinskim članki) nekoliko manj formalen od nejavnega nezasebnega diskurza (npr. pogovor v uradu), pa tudi da je javni razvedrilni diskurz, ki se pojavlja v medijih (npr. razvedrilna oddaja), sicer bolj formalen od nejavnega diskurza, a manj formalen od javnega informativnega diskurza (npr. informativna oddaja). Podobno je mogoče za pisni diskurz privzeti, da je v leposlovju (npr. dialog) in internetnih besedilih (npr. spletni komentar) mogoče pričakovati več elementov, ki so tipični za spontani govor in manj formalni, kot v stvarnih

⁶ Termin formalnost je v prispevku rabljen v uveljavljenem jezikoslovnem smislu, povzemava pa še slovarsko definicijo, ki jo navaja Crystal (1992: 142), kjer je formalnost opisana kot »dimenzija družbenega vedenja«, pri kateri se stopnja regulacije odraža skozi različne jezikovne lastnosti. Visoka stopnja formalnosti jezika pomeni zelo skrbno organiziran diskurz, pogosto tudi kompleksno skladnjo in besedišče, ki v veliki meri sledi standardnemu jeziku. Visoka stopnja neformalnosti se v jeziku odraža skozi ohlapno strukturiranje, številne pogovorne izraze in pogosto odstopa od standarda (npr. skozi rabo slenga, dialektalnih izrazov, neologizmov itd.).

(opomba se nadaljuje na naslednji strani)

besedilih in v periodiki. Čeprav za gradivo, uporabljeno v pričujoči analizi, velja, da so zvrsti, ki sestavljajo oba korpusa, razmeroma heterogene, je prav tako smiselno privzeti, da je govorjeni diskurz na splošno manj formalen od pisnega:⁷ za večji del analiziranega govorjenega diskurza je mogoče ugotavljati, da je tvorjen bolj ali manj spontano in vsebuje več neformalnih elementov, za večino analiziranega pisnega diskurza pa velja, da gre za besedila, ki so načrtovana, revidirana in objavljena. Pomembno odstopanje od te na prvi pogled jasne delitve predstavljata javni informativni diskurz, ki je večinoma v naprej pripravljen, revidiran in bran, in internetni komentarji, ki so večinoma zelo spontano tvorjeni in se približujejo spontanemu govoru. A ker podkorpus javni informativno-izobraževalni govor vsebuje tudi pedagoški govor v osebнем stiku, ki je tudi spontano tvorjen, internetna besedila pa tudi medijske vsebine z novičarskih portalov, ki so načrtovana in revidirana, ni mogoče trditi, da ta dva podkorpusa v celoti odstopata od ostalih govorjenih in pisnih besedil v smislu formalnosti.

Če torej sprejmemo tezo, da je mogoče podkorpuse razvrstiti glede na formalnost, je iz Slike 1 razvidno, da število osebnih zaimkov v analiziranem gradivu izrazito upada z naraščanjem formalnosti posamezne vrste diskurza. V govorjenem diskurzu (korpusu Gos) je tako raba osebnih zaimkov skoraj desetkrat pogostejša kot v pisnem (korpusu Kres): razmerje je 14,6 pojavitve proti 1,5 pojavitve na 1000 besed (prim. Sliko 2). Podatki na Sliki 1 jasno pokažejo, da je v govorjenih vrstah diskurza mogoče opazovati razlike med nejavnim in javnim diskurzom, pri čemer je v nejavnem diskurzu raba osebnih zaimkov pogostejša, in sicer je najpogostejša v zasebnem diskurzu

⁷ Jasno je, da so razlike med govorjenim in pisnim jezikom kompleksne in ne zadevajo le stopnje formalnosti, temveč celo vrsto dejavnikov, od neposrednega stika med tvorcem in naslovnikom, neverbalnih elementov, ki so prisotni v komunikaciji (gestikulacija in mimika proti grafičnim elementom), stopnje dialoščnosti oziroma monološkosti, spontanosti, vnaprejšnje priprave, prostorskega in časovnega stika, in še bi lahko naštevali. Nekateri od omenjenih dejavnikov gotovo lahko do neke mere vplivajo na rabo osebnih zaimkov kot osebkov in so v analizi tudi obravnavani (neposredni nagovor v dialogu, netekočnost v spontanem govoru).

(22,3 pojavitve na 1000 besed), torej v neformalni, spontani, sproščeni osebni ali telefonski komunikaciji med sorodniki, prijatelji in znanci, nekoliko manj pogosta pa je v polformalnih situacijah nezasebnega diskurza (15,1 pojavitve na 1000 besed), torej v osebni ali telefonski komunikaciji v trgovini, v uradu, v pisarni (za natančnejši opis vrst diskurza v Gosu prim. Verdonik in Zwitter Vitez 2011). V javnem diskurzu – v veliki meri gre za medijski diskurz, pri katerem je stopnja formalnosti višja – je raba osebnih zaimkov kot osebkov v primerjavi z nejavnim diskurzom nekoliko manj pogosta: znova je mogoče zaznati razlike med razvedrilnim diskurzom (12,4 pojavitve na 1000 besed), v katerem je osebnih zaimkov kot osebkov več, v primerjavi z bolj formalnim informativno-izobraževalnim diskurzom, v katerem je zaimkov manj (9,5 pojavitve na 1000 besed).

V pisnih vrstah diskurza je raba osebnih zaimkov kot osebkov izrazito redka; skupno je v analiziranih delih Kresa v povprečju 1,5 pojavitve osebnih zaimkov kot osebkov na 1000 besed (prim. Sliko 2). Kot je razvidno iz Slike 1, so med vsemi vrstami diskurza najpogostejši v literarnem diskurzu (3,4 pojavitve na 1000 besed), kar verjetno odraža dejstvo, da ta vsebuje veliko dialoga (saj gre primarno za pripovedno prozo), ki je poskus imitacije govorjenega diskurza, hkrati pa je v dialogu v primerjavi z monologom mnogo pogostejša funkcija nagovora. V drugih treh vrstah diskurza se osebni zaimki kot osebkovi pojavljajo izrazito redko. Morda je na prvi pogled nekoliko presenetljivo, da pri tem ne izstopajo internetna besedila (1,4 pojavitve na 1000 besed) s svojo hibridno naravo: verjetno je to mogoče pripisati dejstvu, da v korpusu Kres internetni diskurz vsebuje velik delež novic na novičarskih portalih, ki so po formalnosti podobne klasičnemu medijskemu diskurzu, pregled primerov pa pokaže, da se osebni zaimki kot osebkovi najbolj izrazito pojavljajo v neformalnih spletnih komentarjih. Zastavlja se torej zanimivo vprašanje, ali bi bilo mogoče izrazitejša odstopanja opazovati v bolj specializiranem korpusu računalniško posredovane komunikacije, kakršen je Janes (Fišer in dr. 2017). V tiskani periodiki in strokovnem diskurzu zelo redka raba osebnih zaimkov kot osebkov (v obeh

primerih je pogostost 0,9 pojavitve na 1000 besed) nikakor ne preseneča, glede na to, da gre za zelo formalni vrsti diskurza.

4.2 Funkcijska analiza rabe osebnih zaimkov kot osebkov

Glede na izrazito korelacijo med pogostostjo rabe osebnih zaimkov in formalnostjo besedilnih zvrsti, predstavljeno v 4.1., bi na prvi pogled lahko pričakovali, da se (ne)izražanje osebnega zaimka neposredno navezuje na žanrske konvencije in jo je mogoče z njimi preprosto pojasniti. Natančnejša funkcijska analiza pokaže, da gre za mnogo kompleksnejši pojav, na katerega vpliva preplet različnih dejavnikov. Neposredna, enoznačna navezava razlage na žanrske konvencije bi predstavljala poenostavitev realnosti, hkrati pa bi tak pristop predstavljal odmik od namena korpusne raziskave, ki je v prvi vrsti popisati dejansko stanje v jeziku.

Funkcijska analiza je potrdila, da je v skladu z literaturo (prim. npr. Rothman 2009, Pisanski Peterlin in Mikolič Južnič 2018) smiselno te elemente razdeliti na dve širši, osnovni funkciji, in sicer na *razdvoumljanje* in *poudarjanje*, v okviru vsake od njih pa je mogoče identificirati različne diskurzno-pragmatične dejavnike, ki vplivajo na rabo osebnih zaimkov kot osebkov. V nadaljevanju sta podrobneje opisani obe funkciji, znotraj vsake pa so predstavljeni posamezni diskurzno-pragmatični dejavniki z ilustrativnimi primeri rabe iz analiziranih korpusov Kres in Gos; za vsak primer je naveden tako korpus kot podkorpus. Na koncu so strnjena še nekatera druga opažanja v zvezi z rabo osebnih zaimkov kot osebkov.

1 Razdvoumljanje

O razdvoumljanju govorimo, kadar identiteta vršilca dejanja v slovnični funkciji osebkov ni jasna, zato mora biti ta eksplicitno izražen, da se referenčna nejasnost razdvoumi. V analiziranem korpusu se funkcija razdvoumljanje pojavlja iz različnih vzrokov, ki se navezujejo na različne ravni jezika.

1.1 Oblikoslovna raven: Osebek ni razviden iz glagolske oblike

V nekaterih primerih razlogi za uporabo osebnega zaimka kot osebka izvirajo iz povsem oblikoslovnih razlogov. V teh primerih je eksplicitno izražen osebek nujen, ker ni nedvoumno razviden iz glagolske oblike.

1.1.1 Elipsa glagola

Za razdvoumljanje je uporaba osebnega zaimka kot osebka nujna ob elipsi glagola. V primeru (1), ki je del športnega komentarja, je glagol (npr. *smučati* v smislu »ona pa smuča z napako«) opuščen, brez osebka bi bilo »pa z napako« nemogoče smiselno interpretirati.

(1) *menjavam eee smeri /// [1 premor] [prazen] /// ja tu nekoliko lažji prehod / **ona pa z napako** / zavrtelo jo je v samem vhodu v (GOS Javni razvedrilni diskurz)*

1.1.2 Oblikoslovna nejasnost

Prav tako je osebek nujen v stavkih, kjer prihaja do oblikoslovne nejasnosti, ker so glagolske oblike enake za različne osebe (npr. pogojnik ali druga in tretja oseba dvojine).

(2) *veš **on bi** nam lahko povedal **jaz bi** pa kakšno parcelo kupila (GOS Nejavni nezasebni diskurz)*

(3) *njenim vrhom božati dlan. "Vem, da nista **vedve** sprožili predaje, mislim, da celo Harold to ve (KRES Leposlovje)*

(4) *ob tem pa nihče ne omenja, da sta **onadva** v finalu zbrala čez trideset tisoč glasov in konkurentke pustila (KRES Periodika)*

1.2 Skladenjska raven

Uporaba osebnega zaimka kot osebka je v nekaterih primerih pogojena s skladenjskimi vzorci.

1.2.1 Priredni osebek

Kadar je osebek sestavljen iz priredno vezanih elementov, so ti elementi izraženi kot samostalniki ali zaimki.

(5) ***Ona** in brata in sestre so vsi že odrasli; mati.* (KRES Strokovna besedila)

1.2.2 S povedkovim določilom

Prav tako je uporaba osebnega zaimka nujna v stavkih s povedkovim določilom, v katerih je izražena identičnost med osebkom in povedkovim določilom, saj bi bili sicer takšni stavki nesmiselni.

(6) *če se ji tako zahoče. Ker **ona** je ona.* (KRES Periodika)

1.3 Diskurzna raven: Členitev po aktualnosti

V kontekstih, kjer se zamenja tema, je za smiselno interpretacijo nujen eksplicitno izražen osebek: kadar je referent novega osebkata znan že od prej, je osebni zaimek eden od elementov, ki lahko uspešno izrazi referenta. V primeru (7) je eksplicitno izražen osebek nujen za ustrezno posredovanje želene informacije (prim. hipotetični primer (7a), ki sicer nakazuje, da je prišlo do menjave teme, vendar pragmatično ni ustrezen).

(7) *Sva enakopravna. **On** odloča, kje bo kateri kuhal in vlagal.* (KRES Periodika)

(7a) *Sva enakopravna. Odloča, kje bo kateri kuhal in vlagal.*

1.3.1 Kontrast

Zelo očitna situacija, pri kateri se tema zamenja, je vzpostavljanje kontrasta.⁸ V takšnih situacijah, v katerih sta vedno sopostavljena vsaj dva referenta (lahko tudi več), je osebek zaradi jasnosti eksplicitno izražen. V hipotetičnem primeru

⁸ Zaradi jasnosti uporablja izraz »kontrast«, s katerim želiva opisati pragmatično funkcijo in nikakor ne nujno slovnično kategorijo protivnosti.

(8a) nastopi potencialna pomenska nejasnost v kombinaciji s spremembo pomena.

(8) Delam, kakor se znajdem. **Ona** pa vidi samo denar. (KRES Periodika)

(8a) Delam, kakor se znajdem. Pa vidi samo denar.

1.3.2 Primerjava

Podobno kot pri kontrastu sta tudi pri primerjavi nujno prisotna dva referenta, ki se ju med seboj primerja. Hipotetični primer brez izraženega osebka eliptičnega stavka (9a) ni mogoč, vendar ne zaradi elipse (prim. hipotetični primer (9b) brez elipse glagola, ki je še vedno pragmatično neuspešen). Primerjava namreč zahteva eksplicitno poimenovanje obeh primerjanih entitet, zato sodi v kategorijo razdvoumljanja, kjer je raba eksplicitnega osebka iz pragmatičnih razlogov obvezna, in ne poudarjanja, kjer je možna, ne pa nujna.

(9) Tudi **ona** je naša gostja, tako kot **vi**,« je (KRES Leposlovje)

(9a) Tudi **ona** je naša gostja, tako kot,« je

(9b) Tudi **ona** je naša gostja, tako kot ste,« je

1.3.3 Prehod iz reme v temo

V nekaterih primerih pri razdvoumljanju lahko opazujemo prehod iz reme v temo.

(10) slabo spoznam za to imamo dobro plačane ljudi in naj **oni** razmišljajo o tem. (KRES Periodika)

1.4 Raven prenosnika: Netekočnosti v spontanem govoru

Pri opisu razdvoumljanja je treba posebej opozoriti na to, da se v spontanem govoru iz očitnih razlogov (napačni začetki, popravljanja, ponavljanja, hkratni govor) pogosto pojavlja dodatna potreba po razdvoumljanju.

(11) odgovoriš na na tisti /// aha / ja ma saj sem ji /// **ona** je si rekla

da je poslala na konferenčni naslov a (GOS Nejavni nezasebni diskurz)

2 Poudarjanje

O poudarjanju govorimo, kadar je identiteta osebk a sicer nedvoumna in jasna, vendar se osebek vseeno izrazi z namenom, da se referent poudari.

(12) *Pride dan, ko bi komu rekel: **tudi jaz** sem samo človek, tudi jaz imam kup problemov.* (KRES Periodika)

Pogosto se poleg zaimka uporabi še kakšno drugo sredstvo za poudarjanje, kot *tudi* v primeru (12) in *prav* in *sam* v primerih (13) in (14):

(13) *Seveda ob tem zavestno pozablajo, da so bili prav oni v preteklih letih nekakšni vladni zgodovinarji, ki so služili* (KRES Periodika)

(14) *Vidim vaju, vidim vaju; **jaz sama** predstavljam celo množico, množico, Garcin, množico* (KRES Strokovna besedila)

2.1 Pragmatična smiselnost: Razlika med poudarjanjem in razdvoumljanjem

V tovrstnih primerih velja posebej izpostaviti dejstvo, da lahko seveda do poudarjanja pride tudi pri razdvoumljanju. Razlikovanje, ki ga uvajava v pričujočem prispevku, tako temelji na tem, da je pri razdvoumljanju osebek vedno obvezen, saj bi bil stavek sicer pragmatično nesmiseln, če pa gre le za poudarjanje, bi bilo osebek mogoče tudi opustiti, kot npr. v hipotetičnem primeru (14a).

(14) *kako naj pa pogledam tovarišici Lidiji v oči, če **ona** stanuje na boljšem kot mi, in bodo rekli,* (KRES Leposlovje)

(14a) *kako naj pa pogledam tovarišici Lidiji v oči, če stanuje na boljšem kot mi, in bodo rekli,*

Poudarjanje se pogosto uporablja v primerih, kjer se tema očitno ohranja.

(15) *so že prišli gasilci, prav hitri so bili, **oni** so ju pozneje rešili.* (KRES Periodika)

2.2 Spontani govor

2.2.1 Slovnica govornjenega jezika in netekočnosti

Kot pri razdvoumljanju velja tudi pri opisu podfunkcij poudarjanja izpostaviti dejstvo, da je raba v spontanem govoru specifična, saj je poudarjanje lahko izrazitejše, morda vezano celo na idiolekt posameznega govorca.

(16) *zapacano ampak njej je uspelo tisto namazati in kako je **ona** bila takrat ponosna [1 delno] nase a ne / in to to* (GOS Javni informativno-izobraževalni diskurz)

(17) *me zanima nekaj ... kaj si **ti** poslušal kaj je **ona** prebrala ? si figo poslušal /// ne /// ja /// **ti** si poslušal* (GOS Javni informativno-izobraževalni diskurz)

(18) *že videla da ne morem se pogovarjati z njo ker **ona** je bila ful pijana / potem je šla noter in ven* (GOS Javni informativno-izobraževalni diskurz)

2.2.2 Nagovor v dialogu

V spontanem govoru je poudarjanje pogosto pri nagovoru, zato je ta raba izrazito vezana na drugo osebo:

(19) ***ti** kaj vidva študirata* (GOS Nejavni zasebni diskurz)

(20) *»Fanta, **vidva** bi že zdavnaj morala biti v postelji,« je⁹*
(KRES Leposlovje)

V istem kontekstu se kot potencialna razlaga za rabo osebnih zaimkov kot osebkov za poudarjanje ponuja tudi možnost, da gre za dialektalno specifiko. Ugotavljanje vpliva dialektov presega okvire pričujoče raziskave, se pa postavlja

⁹ Navedeni primer je sicer iz Kresa, vendar iz dvogovora v literarnem besedilu, ki odraža posnemanje spontanega govora. Prav takšni primeri dokazujejo, da poenostavitve v smislu neposredne korelacije med monološkostjo proti dialoškosti in pisnim ter govornjenim diskurzom za pričujoče gradivo niso smiselne.

kot zanimiva možnost za nadaljnje študije na tem področju.

*(21) je rekla je prišla punčka domov cela žalostna zato ker **ona** ne sme gledati eee mislim da Sosedovi ali neke oddaje (GOS Javni informativno-izobraževalni diskurz)*

3 Druga opazanja

Primeri, v katerih se osebni zaimki navezujejo na nežive in kolektivne referente, so sicer redki, vendar se pojavljajo.

*(22) Kar pa se tiče bančnih računalnikov, so seveda tudi **oni** ravno tako ranljivi kot vsi drugi računalniški sistemi (KRES Leposlovje)*

*(23) vodstvo odgovorno za proces proti četverici in da ga je **ono** tudi insceniralo, za političen konstrukt, zgrajen na neresnicah (KRES Periodika)*

5 SKLEP

Namen raziskave je bil na podlagi korpusne analize razčleniti rabo osebnih zaimkov v funkciji osebka v slovenščini s pragmatičnega vidika in pridobiti dodatne podatke za podrobnejši slovnični opis tega pojava, ki je bil v dosedanjih slovničnih opisih, npr. v Toporišič (2000), obravnavan manj podrobno. V spontano tvorjenih slovenskih besedilih, ki jih tvorijo rojeni govorci, se raba zaimka morda redko problematizira, po drugi strani pa predstavlja vsaj v prevedenih besedilih izziv za prevajalca v slovenščino (prim. Pisanski Peterlin in Mikolič Južnič 2018). Analiza, predstavljena v pričujočem prispevku, je tako pokazala, da gre pri rabi osebnega zaimka kot osebka nedvomno za mnogo kompleksnejšo slovnično problematiko, kot bi lahko sklepali iz dosedanjih opisov. Pokazalo se je tudi, da je na podlagi podatkov, pridobljenih s korpusno analizo, mogoče izdelati podroben slovnični opis rabe osebnih zaimkov v funkciji osebka v slovenščini.

Pri analizi sva izhajali iz dveh raziskovalnih vprašanj. Prvo je bilo, ali je

pogostost rabe osebnega zaimka kot osebkaj pogojena z besedilno zvrstjo. Rezultati korpusne analize so pokazali, da je v analiziranem gradivu mogoče opazovati velike razlike v pogostosti rabe osebnega zaimka kot osebkaj med posameznimi sklopi besedil, pri čemer je raba v spontanem, zlasti dialoškem, govoru in manj formalnih besedilih izrazito pogostejša kot v pisnih besedilih in v bolj formalnih žanrih.

Drugo raziskovalno vprašanje je bilo, kateri dejavniki vplivajo na rabo osebnega zaimka kot osebkaj. Rezultati analize so identificirali vrsto dejavnikov, ki v različni meri vplivajo na rabo osebnega zaimka kot osebkaj. V najširšem smislu bi lahko rekli, da je tudi za slovenščino, podobno kot je že bilo predlagano za druge jezike s opustljivim osebkom (glej npr. argumentacijo v Rothman 2009), mogoče rabo zaimkov ugotavljati v funkcijah razdvoumljanja in poudarjanja. Podrobnejša funkcijska analiza pa je pokazala, da so pri razdvoumljanju v ospredju dejavniki, ki se navezujejo na uvajanje nove teme, različne vrste primerjav ali protistavljanja, priredno vezane osebkaj in oblikoslovne nejasnosti. Pri poudarjanju se osebni zaimek kot osebek pojavlja v kontekstih, v katerih se tema ohranja. Tako pri razdvoumljanju kot pri poudarjanju se osebni zaimek kot osebek pogosto pojavi v spontanem govoru, kar je posledica specifike dialoškosti (nagovor) in netekočnosti.

Pričujoča raziskava je potrdila, da je korpusni pristop pomemben za pridobivanje podatkov za slovnično-pragmatične opise. Hkrati rezultati, ki zelo jasno kažejo postopno upadanje pogostosti rabe osebnih zaimkov kot osebkov z naraščanjem formalnosti, potrjujejo, da je žanrska zasnova korpusov Kres in Gos za tovrstne slovnične opise nadvse primerna.

Rezultati odpirajo nekatera zanimiva vprašanja, ki bi lahko bila izhodišče za nadaljnje raziskave. Tako se postavlja vprašanje dialektalnega vpliva na rabo zaimkov v vlogi osebkaj, saj rezultati korpusa Gos identificirajo nekatere rabe, ki bi lahko bile specifične za posamezne dialekte, vendar bi za natančnejše sklepanje o tem potrebovali drugačen nabor podatkov.

Drugo vprašanje, ki bi ga veljalo razjasniti, je, kakšno je stanje na področju internetnega diskurza: rezultati pričujoče študije kažejo, da je osebni zaimek kot osebek v internetnih besedilih najpogostejše prisoten v spontano tvorjenih, neformalnih komentarjih. Če bi se tovrstne raziskave lotevali s korpusnega vidika, bi lahko predstavljala analiza specializiranega korpusa računalniško posredovane komunikacije, kakršen je Janes, zanimivo nadgradnjo pričujoče študije.

Tretje vprašanje, ki izhaja iz pridobljenih rezultatov, pa se navezuje na rabo osebnega zaimka kot osebka v slovenščini v primerjavi z drugimi jeziki z opustljivim osebkom. Dobljeni podatki potrjujejo, da so funkcije osebnega zaimka kot osebka v slovenščini funkcijsko podobne kot v nekaterih drugih jezikih z opustljivim osebkom (španščina, italijanščina), v katerih so bile tovrstne funkcije že podrobneje razčlenjene. Tako se postavlja vprašanje, čemu je torej mogoče pripisati velike razlike v pogostosti rabe, ki so bile zaznane v predhodni kontrastivni raziskavi (Pisanski Peterlin in Mikolič Južnič 2018). Nekaj dodatnih podatkov v zvezi s tem bi morda lahko pridobili tudi z neposredno primerjavo slovenščine s sorodnimi, slovanskimi jeziki, čeprav je na podlagi novejših klasifikacij jezikov z opustljivim osebkom mogoče ugotavljati, da sorodstvena razmerja med jeziki niso nujno ključni dejavnik za iskanje podobnosti med njimi pri vprašanju opustljivega osebka. Tako npr. Roberts in Holmberg (2010: 11) navajata, da je celo med evropsko in brazilsko portugalsčino mogoče opazovati velike razlike: medtem ko je evropska portugalsčina, enako kot slovenščina, jezik s konsistentno opustljivim osebkom, se brazilka portugalsčina, podobno kot ruščina, uvršča med jezike z delno opustljivim osebkom. Morda je odgovor v zvezi s specifikko slovenščine vsaj delno mogoče iskati v opažanju, da v slovenščini pogostost rabe osebnega zaimka kot osebka upada z naraščanjem formalnosti, vendar bi bilo za bolj podkrepljeno sklepanje o tem treba izvesti raziskavo o lektorskih posegih v osebne zaimke kot osebke; če bi bila raziskava te tematike korpusne narave, bi lahko bila narejena na korpusu Lektor. Drugo možnost za nadaljnje iskanje

odgovorov predstavlja raziskava opismenjevanja oziroma usvajanja pisnega standarda, ki bi pokazala, ali je mogoče opazovati vzporednice med dolžino šolanja in upadom rabe zaimkov v pisnih besedilih šolarjev, kar je mogoče raziskovati z različnimi metodami, vključno s korpusno (npr. na korpusu Šolar).

ZAHVALA

Članek je nastal v okviru programskega financiranja ARRS. Raziskovalna programa št. P6-0215 in P6-0218 je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

LITERATURA

- Baroni, M., in Bernardini, S. (2006): A new approach to the study of translationese: Machine-learning the difference between original and translated text. *Literary and Linguistic Computing*, 21 (3): 259–274.
- Chomsky, N. (1981/1988): *Lectures on Government and Binding: the Pisa Lecture. 5th Edition*. Dordrecht/Providence: Floris Publications.
- Conrad, S. (2010): What can a corpus tell us about grammar. V A. O’Keefe in M. Michael (ur.): *Routledge Handbook of Corpus Linguistics*: 227–240. London in New York: Routledge.
- Crystal, D. (1994): *An Encyclopedic Dictionary of Language and Languages*. London: Penguin.
- de Prada Pérez, A. (2009): *Subject Expression in Minorcan Spanish: Consequences of Contact with Catalan*. Doktorska disertacija. The Pennsylvania State University.
- Fišer, D., Erjavec, T. in Ljubešić, N. (2017): The compilation, processing and analysis of the Janes corpus of Slovene user-generated content. V C. R. Wigham in G. Ledegen (ur.): *Corpus de communication médiée par les réseaux: construction, structuration, analyse. Collection Humanités Numériques*. Paris: L’Harmattan.

- Golden, M. (2001): *O jeziku in jezikoslovju*. Ljubljana: Filozofska Fakulteta Univerze v Ljubljani, Oddelek za Primerjalno in splošno jezikoslovje.
- Gorjanc, V. (1999): Kohezivni vzorec matematičnih besedil. *Slavistična revija*, 47 (2): 139–159.
- Hyland, K., in Paltridge, B. (2011): Introduction. V K. Hyland in B. Paltridge (ur.): *The Continuum Companion to Discourse Analysis* 1–5. London in New York: Continuum.
- Jaeggli, O., in Safir, K. (ur.) (1989): *The Null Subject Parameter. (Studies in Natural Language and Linguistic Theory)*. Dordrecht: Kluwer.
- Koletnik, M. (2008): Prekmursko narečje v medijih. V M. Pezdirc Bartol (ur.): *Slovenski jezik, literatura, kultura in mediji. 44. seminar slovenskega jezika, literature in kulture* 36–42. Ljubljana: Filozofska fakulteta, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik.
- Korošec, T. (1981): Besediloslovna vprašanja slovenščine. *XVII. seminar slovenskega jezika, literature in kulture*: 173–186. Ljubljana: FF.
- Korpus Gos – spletni konkordančnik*. Dostopno prek: <http://www.korpus-gos.net/>. (Datum dostopa: 26. 12. 2017.)
- Korpus Kres – spletni konkordančnik*. Dostopno prek: <http://www.korpus-kres.net/>. (Datum dostopa: 26. 12. 2017.)
- Krek, S. (2013): Korpusne metode in njihov odsev v jezikoslovnih teorijah 20. stoletja. *Slovenščina 2.0.*, 1 (1): 4–23.
- 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.
- Mauranen, A. (2005): Contrasting languages and varieties with translational corpora *Languages in Contrast*, 5 (1): 73–92.

NoSketch Engine. Dostopno prek: <http://nl.ijs.si/noske/index.html>. (Datum dostopa 26. 12. 2017.)

Paradis, J., in Navarro, S. (2003): Subject realization and crosslinguistic interference in the bilingual acquisition of Spanish and English: What is the role of the input? *Journal of Child Language*, 30 (2): 371–393.

Perlmutter, D. (1968): *Deep and Surface Structure Constraints in Syntax*. Doktorska disertacija. Cambridge, MA: Massachusetts Institute of Technology.

Pisanski Peterlin, A., in Mikolič Južnič, T. (2018): Contrasting pronominal subjects: A cross-linguistic corpus study of English, Italian and Slovene. *Languages in Contrast*, 18 (2): 230–251.

Posio, P. (2012): Who are ‘we’ in spoken Peninsular person plural subjects. *Language Sciences*, 34 (3): 339–360.

Roberts, I., in Holmberg, A. (2009): Introduction: parameters in minimalist theory. V T. Biberauer, A. Holmberg, I. Roberts in M. Sheehan (ur.): *Parametric Variation: Null Subjects in Minimalist Theory*: 1–57. Cambridge: Cambridge University Press.

Rothman, J. (2009): Pragmatic deficits with syntactic consequences? L2 pronominal subjects and the syntax–pragmatics interface. *Journal of Pragmatics*, 41 (5): 951–973.

Serratrice, L. (2007): Referential cohesion in the narratives of bilingual English-Italian children and monolingual peers. *Journal of Pragmatics*, 39 (6): 1058–1087.

Sorace, A., in Filiaci, F. (2006): Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22 (3): 339–368.

Sporazumevanje v slovenskem jeziku – spletni portal. Dostopno prek: <http://www.slovenscina.eu/>. (Datum dostopa 26. 12. 2017.)

- Toporišič, J. (2004): *Slovenska slovnica*. 4. izdaja (1. izdaja 1976). Maribor: Obzorja.
- Travis, C. E. (2007): Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19 (2): 101–135.
- Tsimpli, I., Sorace, A., Heycock, C., in Filiaci, F. (2004): First language attrition and syntactic subjects: A study of Greek and Italian near-native speakers of English. *International Journal of Bilingualism*, 8 (3): 257–277.
- Verdonik, D., in Zwitter Vitez, A. (2011): *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Verdonik, D., Kosem, I., Zwitter Vitez, A., Krek, S., in Stabej, M. (2013): Compilation, transcription and usage of a reference speech corpus: the case of the Slovene corpus GOS. *Language Resources and Evaluation*, 47 (4): 1031–1048.

SUBJECT PERSONAL PRONOUNS IN SLOVENE: PRAGMATIC ASPECTS OF A GRAMMATICAL CATEGORY

Corpus analysis can yield relevant data on a range of grammatical features; this data can be used to produce in-depth descriptions of grammar. Corpus methodology is particularly useful for grammar research focusing on pragmatics and variation in the use of selected grammatical features. The present study aims to analyse the use of subject personal pronouns in Slovene from a pragmatic point of view in order to obtain new data for a more detailed description of the use of subject personal pronouns in Slovene. Slovene is a null-subject language; this means that it does not require the explicit expression of the subject of the clause. The study is based on corpus analysis of the Kres and Gos corpora; NoSketch Engine, Web-based text-analysis software, was used in the analysis. The results identify a considerable difference between the frequencies of subject personal pronouns in spoken discourse as opposed to written discourse. The results also show that the frequency of subject personal pronouns decreases with an increase in formality. Functional analysis reveals that overt subject personal pronouns can be classified broadly into two main categories, disambiguation and emphasis. In the category of disambiguation, the use of subject personal pronouns is influenced by morphological, syntactic and discursive factors, as well as the medium of communication. In the category of emphasis, the medium of communication is a particularly relevant factor. The analysis shows that the use of subject personal pronouns in Slovene is far more complex than previous descriptions suggest. The findings raise interesting questions regarding the use of subject personal pronouns in Slovene in terms of dialects and hybrid Internet genres, as well as questions relating to editorial interventions and the impact of literacy development.

Keywords: subject personal pronoun, null-subject language, corpus analysis

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/>



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 temelječih 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.), skladenjskega (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 skladenjskem 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 sopojavaitev 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 skladišne 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 skladenjskih 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.

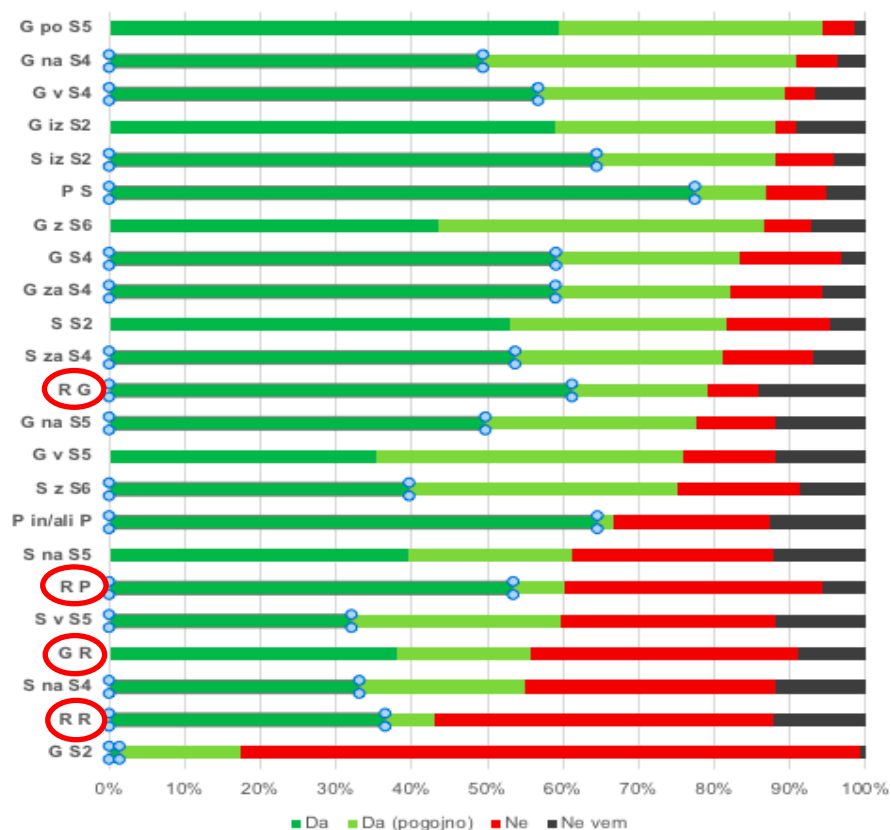
Označevalci jezikoslovci so še vedno izbirali med 3 krovnimi odgovori (DA, NE, NE VEM), a so jim bile ponujene podopcije:

- 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 zaimек 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 skladenjsko 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 skladenjskih 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, četrtrič (dvakrat poplaviti: *je mesto dvakrat poplavila Savinja*; četrtrič zbrati: *Srednja vas je že četrtrič 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, samostalni 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 osvetljuje 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] + samostalniček 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 PBZ0

Tipične kolokacije strukture RBZ PBZ0, 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, četrt, 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 pripravljeni*); 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 oslABLJENE 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, doslej, 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	tod, 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 pripravljeni]), 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 števniskosti, kot je kratnost ali zaporedje (zaporednostni prislovi), ki jih zaradi raznolike semantične relevantnosti ne moremo strukturno omejiti (*četrtrič 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 rodilnika v strukturi glagol + samostalnik v rodilniku), 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 podpori 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.
- Kilgariff, 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

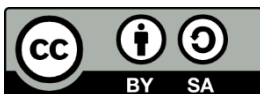
analyses, in particular a description of each grammatical relation from the perspective of its collocationality. In addition, bad collocation candidates that are the result of errors in morphosyntactic annotation will enable the improvement of sketch grammar and relatedly the quality of automatic extraction 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/>



SKLADENJSKE KONSTRUKCIJE MED PODREDJEM IN PRIREDJEM

Mojca SMOLEJ

Filozofska fakulteta Univerze v Ljubljani

Smolej, M. (2018): Skladenjske konstrukcije med podredjem in priredjem. Slovenščina 2.0, 6 (2): 186–205.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.186-205>.

V prispevku se najprej osredotočamo na različne tipe odvisniških stavkov in jih razdelimo v štiri osnovne skupine: v skupino odvisnikov, ki so del matičnega stavka, skupino odvisnikov, ki niso del jedrne propozicije matičnega stavka, skupino odvisnikov, ki so izhodiščno nestavčnočlenski, in skupino členkovnih odvisnikov. V nadaljevanju podrobno analiziramo zadnji dve skupini, saj sta bili do sedaj v slovenskem jezikoslovju skoraj popolnoma prezrti. Izhajajoč iz analize, predstavimo tudi možne nadaljnje teoretične obravnave odvisnikov.

Ključne besede: odvisniki, psevdohipotaksa, protivni odvisniki, členkovni odvisniki

1 UVOD

Stavki v obliki odvisnikov opravljajo številne skladenjske in pomenske vloge. Prav na osnovi njihovih primarnih skladenjskih in pomenskih oz. modifikacijskih vlog lahko zapišemo, da je množica različnih oblik in vlog odvisnikov široka, žal pa še ne popolnoma popisana. Nekateri odvisniki so tako znotraj slovenističnega jezikoslovja skoraj popolnoma prezrti. Dotaknili se bomo vzrokov za pomanjkljivo obravnavo odvisnikov, nato pa skušali podati pregled najpogostejše rabljenih oblik odvisnikov in njihovih skladenjsko-pomenskih vlog. Opozorili bomo tudi na nujnost vključitve skladenjskega večfunkcijskega razumevanja vezniških sredstev, ki ne sme biti prekrivno s pomensko ravnino.

Uvodoma bomo na kratko pregledali jezikoslovna dela, ki so med prvimi

opozarjala na omenjeno problematiko.

1.1 Priredja, izražena s podrednimi sredstvi

Leta 1968 je J. Toporišič v članku z naslovom *Priredni odnosi v slovenskem knjižnem jeziku* zapisal:

Čas, ki nam je bil na razpolago za raziskavo, ne dopušča podrobneje govoriti o priredjih, izraženih s podrednimi sredstvi. /.../ Zdi se nam, da je vsa priredna razmerja mogoče izraziti tudi s podrednimi sredstvi. Vendar bo treba vprašanje, ali imamo v teh primerih še opravka s priredji, in ne že s podredji, rešiti ob kaki drugi priliki /.../ (po Toporišič 1982: 142).

Tovrstne problematike, izražanje priredja s podrednimi sredstvi, so se dotaknili še A. Sovre, J. Gelb in B. Pogorelec. Prvi je v članku *Nepravi relativniki* (1965) opisal primere, ko se oziralni odvisniki rabijo v vlogi prirednih členov. A. Sovre nepravi relativnik imenuje sintaktični omnibus.¹

Če da zveza *tisti, ki* zadovoljiv smisel, je podredje zdravo, če ne, potem nekaj ni v redu. Primer: Za njima je prihitel miličnik, ki ju je vprašal, kaj nosita v aktovkah. /.../ Stavek ki ju je vprašal je torej le po slovnici odvisen od odnosnice miličnik, po svojem bistvu in notranji moči pa ne, zato je ponarejen, nepravi in napačen (Sovre 1965: 125).

Pseudohipotakso je J. Gelb obravnavala v članku *Podredje v vlogi priredja*. V njem popiše različne primere prirednih skladenjskih razmerij, ki so izražena s podrednimi vezniki.

Protivno razmerje se da izraziti z naslednjimi podrednimi vezniki: namesto da, medtem ko, čeprav, dasi, ko, če, da. /.../ Zanimiv je naslednji primer iz Finžgarja, ko se dekleti pogovarjata o svojem potencialnem možu: Jaz pa tebi rečem: Če ga je meni umoril Herod, ga bo tebi ustrelil cesar = Jaz pa tebi rečem: Meni ga je ustrelil Herod, tebi pa ga bo ustrelil /.../ (Gelb 1969: 137–138).

Na mejne strukture je B. Pogorelec (1963) opozarjala v svoji disertaciji *Veznik*

¹ »/O/mnibus zato, ker se relativnemu odvisniku rada po nemarnem oveša funkcija do malega vseh stavčnih zvez, prirednih nič manj nego podrednih.« (Sovre 1965: 125).

v slovenščini. Pri obravnavi veznikov *da*, *če*, *ko*, *kar* se je osredotočala na tiste zveze, ki so po strukturi podredne, po pomenu pa priredne.

Kot druga relativna vezna sredstva uvaja tudi *kar* stavke, ki po svojem pomenu niso relativni odvisniki. Največkrat to po pomenu sploh niso podredni stavki; v formalnem pogledu stavke sicer uvajajo relativna vezna sredstva, po pomenu pa so to različne priredne ali nerelativne podredne zveze. S prirednimi konstrukcijami (sindetičnimi ali asindetičnimi) lahko zamenjamo *kar* stavke v naslednjih primerih: /.../ Vera se je sklonila nad mizico, kar je zdaj prvič storila /.../ (Pogorelec 1963: 155).

Vsem štirim jezikoslovcem je skupno, da so pri analizi izhajali iz dihotomije skladenjskega razmerja in leksikalnimi oz. vezniškimi sredstvi, ki naj bi pripadali le eni izmed dveh skupin vezniških sredstev: priredni ali podredni. Skladenjsko in pomensko razmerje sta pri vseh močno prepleteni in drugo drugo pogojujeta, zato je pri vseh zaznati zadrego slovnične opredelitve nekaterih vezniških sredstev, ki so, kot je zapisala B. Pogorelec, po strukturi podredni, po pomenu pa priredni.

V nadaljevanju se bomo najprej osredotočili na osnovno razumevanje vloge odvisnikov in vrste odvisnikov ter izhajajoč iz podanih spoznanj, skušali podati možen pogled, možno rešitev prikazane »zadrega« navidezne neusklajenosti površinske in globinske, pomenske ravnine.

2 OSNOVNE SKUPINE ODVISNIKOV

Odvisniki se glede na svoje primarne funkcije in pomene med seboj razlikujejo. Razdelili bi jih lahko v štiri osnovne skupine:² a) odvisniki, ki so del matičnega stavka (gre za neke vrste propozicijo v propoziciji),³ b) odvisniki, ki niso del jedrne propozicije matičnega stavka (gre za družljivost), so pa izhodiščno (neobvezno) stavčnočlenski, c) odvisniki, ki so podredni le na strukturni ravni

² O prvih dveh skupinah je pisala tudi A. Žele (2016a).

³ O prilastkovem odvisniku kot propoziciji v propoziciji matičnega stavka je pisal tudi R. Cazinkić (2000/01 in 2004).

in so izhodiščno nestavčnočlenski (če izhajamo iz razumevanja strukturalistične slovnice) in č) odvisniki, ki imajo stavčno strukturo in so podredni le na strukturni ravni, na pomenski oz. modifikacijski ravni pa opravljajo vlogo modifikatorjev matičnega stavka in so po funkciji blizu členkom, zato jih imenujemo členkovni stavki.

Primeri⁴ posameznih skupin odvisnikov:

a) Odvisniki kot del matičnega stavka

- Kdor ne bo plačeval, ne bo mogel pridelovati hrane.
- Mislim, da v mojem primeru naziv in poklic nimata posebnih korelacij.
- Jasno je, da jim preveč godi posedanje ob oblastnem koritu.
- Ljudje, ki se radi gibljejo, živijo bolj zdravo.⁵
- Grem, kamor hočem.

Navedeni primeri se med seboj semantično in skladenjsko razlikujejo (prvi je osebkov odvisnik, drugi predmetni, tretji povedkovodoločilni, četrti prilastkov,⁶ peti pa prislovnodoločilni odvisnik), vendar pa jim je skupno to, da so na strukturni in pomenski ravni obvezni. Vsi opravljajo vlogo obveznega stavčnega člena. Vzporedno z vezljivostno kategorijo lahko zapišemo, da je med matičnim stavkom in navedenimi odvisniki izražena vezljivostna moč, kar nadalje pomeni, da vsi odvisniki opravljajo vlogo obveznih določil in so kot taki del matičnega stavka. Odvisnik kot celota predstavlja element propozicije matičnega stavka, zato lahko upravičeno govorimo o propoziciji znotraj

⁴ Vsi primeri so vzeti iz korpusa Gigafida, <http://www.gigafida.net/> (22. 12. 2017).

⁵ V skupini odvisnikov, ki so del matičnega stavka, so odvisniki dveh vrst. V bodoče bi jih bilo smiselneje obravnavati v dveh ločenih skupinah. Prvo vrsto predstavljajo osebkov odvisnik, predmetni odvisnik in prislovnodoločilni odvisniki kraja in časa. A. Žele (2016a: 302) jih je poimenovala kot prvostopenjske odvisnike. V drugo skupino sodijo razvijajoči odvisniki (A. Žele 2016a: 3012), kot so npr. prilastkovi odvisniki in povedkovoprilastkovi odvisniki. Za obe vrsti odvisnikov je značilna propozicija v propoziciji oz. medpropozicijska vezljivost, zaradi česar so bili vsi navedeni primeri uvrščeni v isto kategorijo oz. skupino.

⁶ Gl. tudi S. Kordić (1995).

propozicije. Tudi odvisnik ima namreč primarno svojo lastno propozicijsko zgradbo. Ker prva skupina ne predstavlja posebnih težav, še posebej ne na ravni slovnične sistematizacije (slovničnega opisa), nadalje pa tudi ne na ravni šolskega oz. didaktičnega prenosa, se bomo raje osredotočili na zadnji dve skupini. Predhodno pa se bomo na kratko ustavili še pri skupini odvisnikov, ki so izhodiščno neobvezno stavčnočlenski.

b) Odvisniki, ki niso del jedrne propozicije matičnega stavka⁷

- Čeprav je dodatni kisik pri višinski bolezni koristen, je edino pravo zdravilo hiter sestop.⁸
- Ko dobim priložnost, se moram nenehno dokazovati.
- Ker je kalifornijski standard sprejelo še dvanajst drugih zveznih držav, se kmalu obetajo nove tožbe.
- Če bomo perilo sušili v sušilnem stroju, mora biti v pralnem stroju ožeto z najmanj 800 vrtljaji na minuto.

Navedeni primeri se od predhodne skupine ločijo po tem, da na ravni strukture oz. glede na matični stavek niso obvezni. Vzporedno z vezljivostno kategorijo jih lahko opredelimo kot družljive, kar pomeni, da propozicija odvisnika ni vključena v propozicijo matičnega stavka, sta pa obe propoziciji povezani na ravni modifikacije pomenske podstave matičnega stavka. Navedeni odvisniki (prvi je dopustni, drugi časovni, tretji vzročni, četrti pa pogojni) opravljajo vlogo neobveznih stavčnih členov, kar pomeni, da za samo strukturo matičnega stavka niso nujni, so pa obvezni na ravni smisla oz. na ravni ilokucije in posledično perlokucije. Obe propoziciji, tako matičnega stavka kot odvisnika,

⁷ Izraz jedrna propozicija je uporabljen zaradi razlikovanja med propozicijo matičnega stavka in propozicijo odvisnika, ki je hkrati tudi del propozicije matičnega stavka. Matični stavek tako zaobjema lastno, jedrno propozicijo, katere del je tudi propozicija obveznega stavčnočlenskega odvisnega stavka.

⁸ Vsi primeri so vzeti iz korpusa Gigafida.

stojita ena poleg druge, soobstajata in sta, kot je bilo že omenjeno, družljivi.

c) Izhodiščno nestavčnočlenski odvisniki

- Medtem ko prvo drži, pa zadnja navedba zahteva zgodovinsko pojasnilo.⁹
- Vozim se v napačno smer, proti vzhodu, namesto da bi dirkal proti zahodu, proti Čilu.
- To je obstajalo vseskozi, samo da je bilo zaradi ukvarjanja z drugimi temami manj opazno.
- Če si ti brezposeln, sem jaz super nadobremenjen!

V nasprotju s predhodno skupino, kjer so navedeni odvisniki, ki jih večina slovníc (npr. Slovenska slovnica 1956, Slovenska slovnica 2000), osnovnošolskih in srednješolskih učbenikov za slovenski jezik obravnava in po vsej verjetnosti zaradi enostavnejšega razumevanja predstavlja kar skupaj z obveznimi stavčnočlenskimi odvisniki, je tretja skupina zapostavljena, saj na skladenjskofunkcijski in pomenski ravni izstopa in se vede samosvoje. Kot je bilo navedeno že v uvodu, se je na tovrstne odvisnike opozarjalo¹⁰ zelo redko. Njihova posebnost je v tem, da so po strukturi vsi podredni, po pomenu pa ne sodijo v nobeno izmed kategorij, ki jih npr. predvideva Slovenska slovnica (2000: 432–436). Prav vse navedene odvisnike lahko namreč pretvorimo v protivna priredna razmerja, s čimer je delno pojasnjena njihova posebnost. V

⁹ Pri izboru primerov, ki so vsi vzeti iz korpusa Gigafida, sta upoštevani diplomski nalogi: D. Gabrovšek (2016) in M. Kos (2016).

¹⁰ Npr. M. Smolej 2011: 78; 2015: 127–137.

Na tovrstno problematiko je opozarjal tudi K. Ahačič (2017: 122):

»Pri določanju odvisnikov in priredij se ne smemo nikoli zanesti samo na značilni veznik, ampak moramo vedno upoštevati tudi pomen in možnost rabe v prenesenem pomenu. Poglejmo si težak primer iz sodobnega jezika: Med odbojgarskimi legionarji smo s pomočjo Odbojgarske zveze Slovenije v letu 1993 našli le brata Urnaut, medtem ko je spisek za 2005 kar obsežen. Tu veznik medtem ko ne izraža časovnosti: po njem se vsebinsko ne moremo vprašati kdaj? Ipd., pomen pa ne ustreza nobenemu od odvisnikov, ki smo jih spoznali. Pomensko razmerje med odvisnikom in glavnim stavkom je namreč enako kot pri protivnem priredju z veznikom pa.«

priredno izražena razmerja lahko pretvorimo tudi primere druge skupine: npr. *Dodatni kisik pri višinski bolezni je koristen, vendar je edino pravo zdravilo hiter sestop; Kalifornijski standard je sprejelo še dvanajst drugih zveznih držav, zato se kmalu obetajo nove tožbe* itd., vendar nobena izmed navedenih zvez ni tudi izhodiščno protivna oz. sklepalna, vse namreč pravo protivnost (ali sklepalnost, vezalnost, pojasnjevalnost) zadobijo šele po pretvorbi. Odvisniki tretje skupine so protivni že izhodiščno. Natančneje se bomo na pričujočo skupino osredotočili v naslednjem poglavju.

č) Členkovni odvisniki

- Ne maram zanj, da boš vedel.
- Če ti rečem, da ne smem!
- Nihče v tem filmu ne pije, kadi ali preklinja, kaj šele, da bi užival kake substance.

Osnovna značilnost odvisnih stavkov zadnje skupine je ta, da je njihova vloga izražanje modifikacije in da kot taki niso vključeni v propozicijo matičnega stavka, kar jih povezuje z drugo in tretjo skupino odvisnikov. Od predhodnih dveh skupin pa se ločijo po tem, da je v okviru modifikacije njihova primarna vloga krepitev ali slabitev gotovostne naklonskosti, zato jih lahko imenujemo tudi členkovni stavki. Izhodiščno so nestavčnočlenski, na ravni skladenjskega razmerja med matičnim stavkom in členkovnim stavkom pa lahko govorimo o psevdohipotaksi.

3 PROTIVNI ODVISNIKI

Osredotočili se bomo na odvisnike tretje skupine oz. na odvisnike, ki so že izhodiščno po tradicionalni slovnici obravnavani kot nestavčnočlenski. Po funkciji so zelo blizu neobvezno stavčnočlenskimi odvisnikom. S propozicijo matičnega stavka so povezani na ravni modifikacije, zato propozicija odvisnega stavka ni del propozicije matičnega stavka, sta pa obe propoziciji na sporočilni oz. smiselni ali ilokucijski ravni zagotovo neločljivo povezani in se dopolnjujeta.

Druži jih tudi podredno-priredna pretvorba, le da neobvezno stavčnočlenski odvisniki primarno izražajo eno izmed predvidenih in opisanih pomenskih razmerij (npr. časovno, načinovno, posledično, pogojno, vzročno, namerno), ki jih slovnica večinoma pripisuje podrednim vezniškim zvezam oz. podrednim odvisnikom.¹¹ Odvisniki tretje skupine ne izražajo nobenega izmed predvidenih pomenskih razmerij, kar je bil po vsej verjetnosti tudi eden izmed vzrokov za skoraj popolno izključitev iz slovnične obravnave. Eden izmed vzrokov je zagotovo tudi ta, da se je v slovenski slovnici na odvisnike največkrat gledalo kot na stavčne ustreznike stavčnih členov.

V tej knjigi se pri razporejanju odvisnikov držimo načela njihove stavčnočlenske vloge, obdelani so torej po vrsti odvisniki za osebek, predmet, povedkovo določilo, povedkov prilastek, prislovna določila in za prilastek. (Slovenska slovnica 2000: 638)

Razumevanje odvisnikov kot stavčnih členov se je preneslo tudi v osnovne in srednje šole, žal pa se ni nikoli osredotočalo oz. vidneje opozarjalo na odvisnike, ki svojih ustreznikov med predvidenimi stavčnimi členi nimajo.¹² Vsekakor pa je treba poudariti, da tudi Slovenska slovnica ta problem nakazuje, ga pa ne razreši in mu ne nameni obširnejše analize.

Izvezalni odvisnik je stavčno prislovno določilo izvzemanja h glavnemu (ali nadrednemu) stavku. Pretvorljiv je v protivno priredje, npr. Prišli so vsi, le da ne tudi prostovoljno. → Prišli so vsi, vendar ne prostovoljno. Še zgledi: Kar pojdi, samo da se ne prehladiš. – Tu ne bi hotel živeti, razen ko je pomlad. Ta odvisnik bi bilo bolje imenovati protivni. (ibid.: 643)¹³

¹¹ Npr. Slovenska slovnica 2000: 432444.

¹² Problematično je tudi izhajanje iz strukturnega opisa. Smiselneje bi bilo izhajati iz pomena. Na ta način bi se izognili ne vključevanju posameznih konstrukcij, ki, strukturno gledano, ne sodijo v nobeno izmed predvidenih kategorij: ne podredno ne priredno.

¹³ Na tem mestu se zastavlja vprašanje smiselnosti uvrstitve po pomenu različnih odvisnikov v isto skupino (izvezalni odvisniki oz. protivni odvisniki). Če sta primera *Prišli so vsi, le da ne tudi prostovoljno*; *Kar pojdi, samo da se ne prehladiš* brez pomenskih posegov v glavni stavek res pretvorljiva v protivno zvezo (*Prišli so vsi, vendar ne prostovoljno*; *Kar pojdi, vendar se ne prehladiš*), pa primer *Tu ne bi hotel živeti, razen ko je pomlad, zagotovo ni*, kar

Podobno je obravnavan tudi odvisnik, ki ga uvaja protivno vezniško sredstvo *medtem ko*.

Protivna priredja /.../ izražajo: a) Kontrast ali razliko: Ti boš mlinar, jaz pa ti bom nosila hrano. /.../ Pretvorba v podredje: Medtem ko boš ti mlinar, ti bom jaz nosila hrano: to je protivno podredje. (ibid.: 651)

Podobnih opredelitev je v Slovenski slovnici še več, vendar so obrobne in ne predvidevajo nadaljnjih obravnav in posledično sistematičnih rešitev, ki so nujne, če ne zaradi drugega, zaradi poučevanja slovenščine kot prvega in kot tujega jezika.

Če se vrnemo na obravnavanje odvisnikov kot ustreznikov (nestavčnih) stavčnih členov, je razvidna pomanjkljivost, ki je posledica tega, da med (nestavčnimi) stavčnimi členi ni stavčnega člena protivnosti oz. prislovnega določila protivnosti. Prekrivnost nestavčnih prislovnih določil (ibid.: 619–627) s stavčnimi prislovnimi določili (ibid.: 640–645) je namreč skoraj¹⁴ popolna (nestavčna prislovna določila so: prislovno določilo kraja, časa, pravega načina, primere, sredstva in orodja, mere, posledice, ozira, vzroka, namena, pogoja in dopustitve; prislovnodoločilni odvisniki pa so prav tako: odvisnik prostora, časa, načina, primere, posledice, sredstva, ozira, vzroka, namena, pogoja in dopustitve). Odmik k popolni prekrivnosti in usmeritev k nujnosti razširitve prislovnodoločilnih odvisnikov je uvedba izvzemačnega odvisnika (ibid.: 643).

Nujnost razširitve in popolnejšega pregleda pomenske in obenem funkcijske vrednosti (nestavčnih in stavčnih) členov v stavku, predvsem pa prislovnih določil, je posredno nakazana, kot je bilo že večkrat omenjeno, v Slovenski

nadalje nakazuje, da bi ga bilo potrebno obravnavati posebej oz. da ne izraža protivnosti, pač pa pravo izvzemačnost. Poleg tega je odvisnik izvzemačnosti nemalokrat možno pretvoriti tudi v nestavčno izvzemačno zvezo oz. je pomen izvzemačnosti lahko izražen tako stavčno kot nestavčno (kot stavčni člen, ki ga Slovenska slovnica (2000: 619627) ne predvideva.

¹⁴ Slovenska slovnica (2000: 625) je uvedla tudi (nestavčno) prislovno določilo vršilca glagolskega dejanja.

slovnici tudi pri navedbi prislovnih določil lastnosti.¹⁵

Za osvetlitev predhodno napisanega bodo navedeni trije primeri, s katerimi bomo skušali podati tudi eno izmed možnih rešitev obravnave izhodiščno nestavčnočlenskih (protivnih) odvisnikov in njihovih skladenjsko-pomenskih vlog.

- *Medtem ko* Vasalisa spi, njena punčka postori zahtevana opravila.¹⁶
- *Medtem ko* se v razvitem svetu zavestno, čeprav simbolično odrečemo nekaterim jedem, da bi se osredotočili na duhovno misel in jo okrepili, šestina človeštva ali ena milijarda trpi lakoto, pri čemer ji je takšno stanje vsiljeno.
- Po štirih mesecih ima javna zdravstvena blagajna "zgolj" za okoli 13 milijonov evrov primanjkljaja, *medtem ko* ga je bilo v trimesečju že 24 milijonov in je grozilo povečanje na prek 100 milijonov evrov do konca leta.

V vseh treh primerih je uporabljeno vezniško sredstvo *medtem ko*, ki izraža časovno pomensko razmerje, le da je to v drugem, posebej pa še v tretjem primeru zakrito. Prvi primer je grajen iz matičnega stavka in odvisnika, katerega propozicija ni del propozicije matičnega stavka. Med njima je vzpostavljeno podredno skladenjsko razmerje in pravo časovno pomensko razmerje. Drugi in tretji primer sta po konstrukciji popolnoma enaka prvemu, se pravi, da sta prav tako grajena iz matičnega stavka in neobvezno stavčnočlenskega odvisnika, se pa od njega razlikujeta glede na vzpostavljeno pomensko razmerje. *Medtem ko* je v prvem primeru izraženo časovno, je v drugem in tretjem poudarjeno predvsem protivno razmerje. V drugem primeru je poleg protivnosti vzpostavljeno tudi časovno pomensko razmerje, ki pa je

¹⁵ *Lastnostna prislovna določila so: (pravega) načina, sredstva, orodja, mere in primere ...* (Slovenska slovnica 2000: 622). V navedenem citatu je pomenljivo končno ločilo tropičje, s katerim je nakazano, da je množica prislovnih določil zagotovo širša, vendar je njen prikaz nepopoln.

¹⁶ Vsi primeri so vzeti iz korpusa Gigafida.

delno zakrito oz. postavljeno na drugo raven. V tretjem primeru je časovno razmerje popolnoma zakrito oz. neizraženo. Vzpostavljeno časovno ali/in protivno pomensko razmerje je/sta odvisno/odvisna ne le od vezniškega oz. konektorskega sredstva, pač pa celotne konstrukcije, zato bi v analizo morali vključiti tudi vsa druga leksikalna in slovnična sredstva.

Vsekakor pa lahko, izhajajoč le iz danih treh primerov, zaključimo a) da bi bilo nujno korpusno izluščiti prav vsa vezniška oz. konektorska sredstva, b) natančno analizirati vse njihove skladenjske in pomenske vloge, pri čemer bi v analizo morali vključiti konstrukcijski pristop,¹⁷ in c) pri razvrščanju analiziranih konektorskih sredstev bi morali ločevati med skladenjskim in pomenskim razmerjem, kar nadalje pomeni, da bi morali razširiti oz. dopolniti/izpopolniti shemo pomenskih razmerij in posledično shemo strukturnih vlog odvisnikov. Kot je bilo že omenjeno, bi bilo pri analizi smiselno izhajati iz pomenske vloge, ki je lahko strukturno izražena na različne načine (npr. podredno ali priredno). Tako bi poudarili soodvisnost semantične ravni s skladenjsko, pri čemer bi bila primarno analizirana pomenska, sekundarno pa strukturna oz. oblikovna izraženost preučevanega pomena. Ob vsem tem bi bilo nujno v analizo vključiti tudi jezikovnostilistični vidik, ki bi nakazoval dvofunkcijskost oz. neke vrste homonimijo (dvojno ali celo trojno pomensko in ilokucijsko vrednost) iste strukture.¹⁸

Ločevanje¹⁹ skladenjskega in pomenskega razmerja pomeni, da bi se izognili

¹⁷ Konstrukcijski pristop pri analizi kateregakoli jezikovnega elementa (na katerikoli jezikovni ravni) vedno vključuje tako pomen kot obliko oz. strukturo. Izhodišče analize predstavlja funkcija, ki jo določen jezikovni element (npr. določena beseda, besedna zveza, stavek itd.) opravlja. Konstrukcijsko metodo oz. konstrukcijsko slovnico je med drugimi natančno opredelil (in zagovarjal) W. Croft v delu *Radical construction grammar* (2009). Konstrukcije so po njegovem osnovne, temeljne enote vsakega skladenjskega preučevanja in zaobjemajo tako obliko/strukturo kot pomen (2009: str. 14–62).

¹⁸ Zveza stavkov *Medtem ko Janez spi, France dela* lahko izraža zgolj časovno razmerje (istočasnost dveh dejanj), lahko pa zaradi drugačnih kontekstualnih danosti poudarja protivnost, v katero je sicer vključena tudi časovnost, le da je ta delno ali popolno zakrita (*Ata dela, mama pa spi.*).

¹⁹ Ločevanje je na tem mestu izpostavljeno le zaradi nujnosti presejanja tradicionalnega ločevanja med prirednimi vezniškimi sredstvi, ki so jim »predpisana« točno določena

tradicionalnemu ločevanju med prirednimi vezniškimi sredstvi, za katera so tipična točno določena pomenska razmerja (npr. vezalno, ločno, protivno itd.), in podrednimi vezniškimi sredstvi, za katera so značilna druga pomenska razmerja kot pri podrednih vezniških sredstvih (npr. načinovno, posledično, vzročno itd.).²⁰ Ločevanje med skladenjskimi in pomenskimi razmerji je izhodiščno seveda nujno, vendar s temeljnima poudarkoma, da lahko isto konektorsko sredstvo izraža različna pomenska razmerja in da lahko isto pomensko razmerje izražajo različna konektorska sredstva (npr. tako podredna kot priredna). To nadalje pomeni, da bi bilo, izhajajoč iz korpusne analize, nujno razširiti oz. dopolniti obravnavo skladenjskih in pomenskih vlog (stavčnih in nestavčnih) obveznih in neobveznih stavčnih členov in družljivih oz. priredno povezanih stavčnih konstrukcij.

4 ČLENKOVNI ODVISNIKI IN PSEVDOKONJUNKTORJI

Členkovni stavki so stavki, ki opravljajo vlogo modifikacije in so blizu ali enaki členkom. S pred ali za njimi stoječimi stavčnimi strukturami vstopajo v priredna ali podredna skladenjska razmerja, vendar le na formalno skladenjski ravni, na pomenski ravni ostaja to razmerje neizpolnjeno.

- Če kdaj, potem tačas koristi »možganska nevihta« z neobremenjenim tretjim. Če kdo, potem Perry: z izkušeno, matematično glavo /.../.²¹
- Sicer pa, če kaj vem, upokojitev Mance Košir pomeni samo to, da bo počenjala mnoge stvari.
- Znanost pa ne more zanikat boga, če še ne veš.

pomenska razmerja, in podrednimi vezniškimi sredstvi, za katera naj bi bila značilna druga pomenska razmerja kot pri podrednih vezniških sredstvih. Tradicionalno ločevanje kot izhodišče postavlja strukturo (podredje, priredje) oz. vezniško sredstvo, in ne pomen. Kot smo že poudarili, bi moral biti kot izhodišče analize postavljen pomen, nadaljnji korak analize pa bi bila preučitev oblikovne/strukturne izpeljave analiziranega pomena.

²⁰ V Slovenski slovnici lahko npr. preberemo:

Priredni veznik se načeloma ne rabi tudi v podredni zvezi, podredni pa načeloma ne tudi v priredni. (Slovenska slovnica 2000: 432)

²¹ Vsi primeri so vzeti iz korpusa Gigafida.

- Z nobeno kretnjo ali grimaso ni pokazal ne veselja ne jeze, kaj šele da bi preklinjal.
- Od kod veš? se je začudil Bolton. Res si as, če ti rečem.
- Preden greš na izpit, glej, da vse še enkrat ponoviš.

Osredotočili se bomo na prvi primer, in sicer na razmerje med podčrtanima stavkoma *če kdaj* in *če kdo* ter matičnima stavkoma, na katera se nanašata. Med njimi namreč ni vzpostavljeno pogojno ali časovno razmerje, tudi protivno ali dopustno ne. Pomenska interpretacija kaže na to, da je vloga stavka *če kdaj* blizu členkovni vlogi oz. vlogi modifikacije izražanja podkrepiteve. Njena osnovna vloga je torej krepitev gotovostne naklonskosti stavka, na katerega se nanaša (*če kdaj* → *zagotovo*). Stavek *če kdaj* je ekspresivna stavčna struktura oz. psevdohipotaktična struktura, ki vstopa v podredno razmerje z za njo stoječim stavkom le na formalni ravni, na pomenski ravni pa je to razmerje prazno, saj opravlja (le) vlogo modifikacijskega sredstva oz. členkovnega stavka. Členkovni stavek kot tak ni del propozicije matičnega stavka, prav tako ne opravlja stavčnočlenske vloge, je pa vsekakor del smiselno-sporočanjske ravni povedi, katere gotovostno naklonskost krepi. Podobno velja tudi za drugi, tretji in predzadnji primer. Nekoliko poseben je četrti primer, saj se razlikuje od preostalih. Če so členkovni stavki v preostalih primerih kot celota modifikatorji (to pomeni, da vsa leksikalna in slovnična sredstva, ki gradijo stavek, skupno opravljajo vlogo modifikacije) in zato pravi členkovni stavki, v četrtem primeru (*Z nobeno kretnjo ali grimaso ni pokazal ne veselja ne jeze, kaj šele da bi preklinjal.*) vlogo modifikatorja ne opravlja stavek kot celota, pač pa le določena leksikalna in slovnična sredstva, ki ga gradijo. Tu namreč vlogo izražanja modifikacije opravljajo (le) členkovna zveza *kaj šele*, veznik oz. uvajalec odvisnika *da*, struktura stavka (odvisnik) z izraženim pogojnim glagolskim naklonom. Napisano lahko prikažemo s shemo:

(par.)²² kaj šele – (jun.) da – (hipot.) struktura stavka – VF (kond.) → krepitev zanikanja

Tabela 1: Konstrukcijski prikaz sooblikovanja vloge krepitev zanikanja

V obravnavanem primeru torej ni pravega členkovnega stavka,²³ saj vlogo modifikacije, kot je bilo že omenjeno, opravljajo le določena leksikalna in slovnična sredstva. Kljub temu pa so si vsi predhodno navedeni primeri podobni, saj je pri vseh vzpostavljeno psevdohipotaktično razmerje.

Za konec se bomo na kratko osredotočili še na zadnji primer (*Preden greš na izpit, glej, da vse še enkrat ponoviš.*) V podčrtanem stavku vlogo modifikatorja skupaj opravljajo tako leksikalna kot slovnična sredstva, in sicer členek *glej*, veznik oz. uvajalec odvisnika *da* in povedni glagolski naklon. Njihovo modifikacijsko vrednost (podkrepitev trditve oz. izražanje nujnosti, da vršilec izvede dejanje) bi lahko izrazili/prikazali tudi s pomočjo drugih modalnih izrazov, npr. glagolom *morati* in členkom *nujno/obvezno*.

(par.) glej – (jun.) da – (hipot.) struktura stavka - VF (ind. prez ali fut.)²⁴ → morati (ind. prez ali fut.) – (par.) nujno/obvezno

Tabela 2: Konstrukcijski prikaz sooblikovanja modifikacijske vloge

Tudi tu je obravnavani stavek psevdohipotaktičen, saj je v resnici nadreden stavku (opravlja vlogo glavnega stavka), za katerim stoji: »Preden greš na izpit,

²² Par. = členek; jun. = vezniško sredstvo; hipot. = hipotaksa; kond. = pogojni naklon.

²³ Obravnavani primer smo v skupino členkovnih stavkov uvrstili zaradi izraženega psevdohipotaktičnega razmerja, čeprav bi bilo po vsej verjetnosti bolj smiselno, da bi ga uvrstili v skupino izhodiščno nestavčnočlenskih odvisnikov (gl. 2. poglavje pričujočega prispevka: skupina c med odvisniki), saj na pomenski ravni izraža stopnjevalno razmerje, ki ga Slovenska slovnica (npr. Toporišič 2000: 432433) pripisuje le prirednemu skladenjskemu razmerju. Uvrstili bi ga torej lahko med stopnjevalne odvisnike, ki jih, kot že napisano, Slovenska slovnica (2000: 638646) ne predvideva.

²⁴ Ind. = povedni naklon; prez. = sedanjik; fut. = prihodnjik.

moraš vse še enkrat ponoviti // glej, da vse še enkrat ponoviš.«²⁵

5 SKLEP IN NALOGE ZA NAPREJ

Po kratkem pregledu nekaterih najpogostejših oblik odvisnikov lahko kot sklep ponovno zapišemo, da bi bilo nujno korpusno izluščiti vse pojavnice odvisniških oblik in podati njihove skladenjsko-pomenske in pragmatično-sporočanjske vloge. Nujno bi bilo prav tako korpusno pregledati vsa vezniška oz. konektorska sredstva in analizirati njihove skladenjske, pomenske in širše besedilne vloge, pri čemer pa se skladenjskih razmerij ne bi smelo zamenjevati s pomenskimi oz. se določenega pomenskega razmerja ne bi smelo pripisovati izključno določenemu skladenjskemu razmerju, saj bi se lahko le na ta način izognili pomanjkljivi obravnavi odvisnikov in konektorskih sredstev oz. hotenemu ali nehotenemu prezrtju določenih odvisniških konstrukcij in z njimi povezanimi konektorskimi sredstvi. Temeljita analiza vseh oblik odvisnikov in vseh konektorskih sredstev bi nadalje lahko vodila do vzpostavitve novih teoretičnih dognanj.

Vzporedno z analizo večinoma pisnega jezika bi morali vse predhodno napisano preveriti tudi v govornih korpusih. Raven izražanja podrednih medstavčnih skladenjskih razmerij v govorjenem jeziku je zelo slabo raziskana, kar je velika pomanjkljivost, saj ravno poznavanje zakonitosti govorjenega jezika lahko olajša ali pojasni marsikatero težavo, ki se pojavlja v pisnem jeziku. Poleg tega so časovni, vzročni in načinovni odvisniki, katerih propozicije niso del

²⁵ Pri uvrstitvi primera *Preden greš na izpit, glej, da vse še enkrat ponoviš* med členkovne stavke se zastavlja kar nekaj vprašanj, ki deloma nakazujejo, da bi ga bilo morda smiselneje obravnavati kot zvezo glavnega stavka s časovnim in predmetnim odvisnikom. Prvo vprašanje je vezano na opredelitev same besede *glej* oz. zveze *glej da*. V pričujočem prispevku jo razumemo kot frazeologizirano členkovno zvezo s pozivno funkcijo (gl. npr. še SSKJ 2, kjer je beseda *glej* opredeljena kot medmet), čeprav je *glej* primarno glagolska beseda. Če bi besedo *glej* opredelili izključno kot glagol (*gledati* → *glej, glejta, glejte* itd.), se s tem odpira drugo vprašanje oz. pomislek o upravičenosti uvrstitve obravnavanega primera med členkovne stavke. Zvezo bi namreč lahko, kot je bilo uvodoma napisano, opredelili tudi kot zvezo glavnega stavka (*glej*) z družljivim časovnim odvisnikom (*preden greš na izpit*) in obveznim predmetnim odvisnikom (*da vse še enkrat ponoviš*).

propozicije matičnega stavka, tipični predvsem za pisni jezik, v govorjenem jeziku so ta razmerja izražena največkrat s prirednimi skladenjskimi sredstvi.²⁶ V izbranem korpusu oz. zbirki²⁷ spontano govorjenih narativnih besedil 34 govorcev,²⁸ ki je bila oblikovana prav zaradi analize skladenjskih značilnosti spontano govorjenega jezika, se potrjuje, da je med najpogostejše rabljenimi vezniškimi sredstvi v spontano govorjenem jeziku prav veznik *in*, ki povezuje tudi propozicije oz. stavke, ki so med seboj v posledičnem, vzročnem, načinovnem ali časovnem pomenskem razmerju. Vzporedna analiza izražanja medstavčnih skladenjskih in pomenskih razmerij je torej nujna ne le zaradi kontrastive in same jezikoslovne stroke, pač pa predvsem zaradi boljših izhodišč pri poučevanju slovenščine kot prvega in tujega/drugega jezika.

LITERATURA

Ahačič, K. (2017): *Slovnica na kvadrat. Slovenska slovnica za srednjo šolo*.

Ljubljana: Rokus Klett.

Bajec, A., Kolarič, R., in Rupel, M. (1956): *Slovenska slovnica*. Ljubljana: DZS.

Cazinkić, R. (2001): Kategorizacija in razvrstitev oziralnikov ki in kateri.

Slavistična revija, 49 (1–2): 55–73.

Cazinkić, R. (2004): Pojmovanje odvisnika in razmerja med nadrednim in odvisnim stavkom. *Jezikoslovni zapiski*, 10 (1): 43–58.

²⁶ Večkrat je bilo že potrjeno, da se je podredno izražanje časovnih, vzročnostnih in načinovnih razmerij uveljavilo s pisnim jezikom, sicer pa so ista razmerja v govorih navadno oz. praviloma priredno izražena. (Grepl 2011: 170 po Žele 2016: 88).

[U] govorenem jeziku se općenito ljudi jednostavnije izražavaju i koriste kraće rečenične konstrukcije pa se zato mnoge rečenice koje bi se u pisanom jeziku ostvarile kao nerestriktivne relativne rečenice ostvaraju u govorenom jeziku kao nezavisne rečenice /.../. (Kordić 1995: 6263)

²⁷ Skupni obseg posnetih besedil je 148.427 znakov brez presledka v zapisani obliki besedila. Besedila so v zapisani obliki objavljena v: Smolej, Mojca, 2012: *Besedilne vrste v spontanem govoru*.

²⁸ Vsi govorci so zaradi enotnosti in usklajenosti jezikovne analize na ravni pokrajinske oz. zvrstne opredelitve jezika iz Ljubljane ali njene okolice. Govorci, ki so tvorili vsak le po eno besedilo, se razlikujejo glede na spol (12 moških, 22 žensk), starost (najmlajši govorci so bili star 21 let, najstarejši pa 72) in vrsto izobrazbe.

- Croft, W. (2009): *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Gabrovšek, D. (2016): *Pseudohipotaksa: primer vezniških besed medtem ko, namesto da, kar*. Diplomsko delo, Filozofska fakulteta, Univerza v Ljubljani.
- Gelb, J. (1969): Podredje v vlogi priredja. *Jezik in slovstvo*, 5 (14): 136–139.
- Gigafida. Dostopno prek: <http://www.gigafida.net> (22. december 2017).
- Goldberg, A. E. (1995): *A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.
- Kordić, S. (1995): *Relativna rečenica*. Zagreb: Hrvatsko filološko društvo, Matica hrvatska.
- Kos, M. (2016): *Skladenjske konstrukcije med podredjem in priredjem*. Diplomsko delo, Filozofska fakulteta, Univerza v Ljubljani.
- Pogorelec, B. (1963): *Veznik v slovenščini*. Doktorska disertacija, Filozofska fakulteta, Univerza v Ljubljani.
- Slovar slovenskega knjižnega jezika*² (2014). Ljubljana: ZRC SAZU.
Dostopno prek: <https://fran.si/> (7. november 2018).
- Sovre, A. (1965): Nepravi relativniki. *Jezikovni pogovori*: 125–131. Ljubljana: Cankarjeva založba.
- Smolej, M. (2011): *Skladnja slovenskega knjižnega jezika. Izbrana poglavja z vajami*. Ljubljana: Študentska založba.
- Smolej, M. (2012): *Besedilne vrste v spontanem govoru*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Smolej, M. (2015): Particle clauses. *Slovenski jezik – Slovene Linguistic Studies*, 10: 127–137.
- Toporišič, J. (1982): *Nova slovenska skladnja*. Ljubljana: DZS.

Toporišič, J. (2000): *Slovenska slovnica*. Maribor: Obzorja.

Žele, A. (2016a): Tipologija odvisnikov v slovenščini: merila. E. Kržišnik (ur.): *Toporišičeva obdobja*: 299–306. Ljubljana: Znanstvena založba Filozofske fakultete.

Žele, A. (2016b): Odvisniki v slovenščini: vsebinski odvisniki in nepravi prislovnodoločilni odvisniki. *Slavistična revija*, 64 (2): 81–94.

Žele, A. (2017): Razmerje osebek proti osebkov odvisnik v slovenskih povedih. *Slavistična revija*, 65 (1): 81–97.

SYNTAX CONSTRUCTIONS BETWEEN HYPOTAXIS AND PARATAXIS

Sentences in the form of subordinate clauses perform numerous syntactic and semantic roles. The number of different forms and roles of subordinate clauses is large but unfortunately not yet fully described. Some subordinate clauses are almost completely ignored in the Slovenian linguistics. On the basis of primary functions, subordinate clauses are divided into four basic groups: a) subordinate clauses that are part of the main clause; b) subordinate clauses that are not part of the semantic knot of the main clause; c) subordinate clauses that are subordinate only on the structural level and are originally not part of the speech; d) subordinate clauses that are subordinate only on the structural level, but perform the role of modifiers of the main clause on the semantic level and are therefore called particle clauses. In this article, we then focus more precisely on subordinate clauses of the third group that do not express any of the foreseen semantic relations; and that is probably one of the reasons why they are almost completely excluded from grammatical treatment. Surely, one of the reasons is also that in the Slovenian grammar the subordinate clauses are often looked at as the clause equivalents of sentence clauses. The understanding of subordinate clauses as sentence clauses was transferred also to primary and high schools but sadly, there has never been a focus or strong emphasis on those subordinate clauses that do not have their own equivalents among expected sentence clauses. Similar are particle subordinate clauses which form coordinate or subordinate syntactic relationships with the pre or after positioned sentence structures only on the formal syntactic level and not on the semantic level. The above written is also one of the main reasons why the Slovenian grammar has not yet included them into its study. The analysis shows that it would be necessary to find all occurrences of subordinate forms in the corpus and define their new syntactic-semantic and pragmatic-communication roles. It would also be necessary to look at all connector means in the corpus and analyse their syntactic, semantic, and broader text roles, and by so doing the syntactic relations should not be confused with semantic ones. A thorough analysis of all forms of subordinate clauses and all connector means could further lead to new theoretical findings. Everything

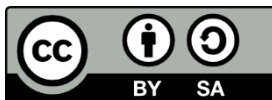
written so far should, along with the analysis of the majority of the written language, also be verified in spoken corpora.

Keywords: subordinate clauses, pseudohypotaxis, oppositional/contrastive subordinate clauses, particle clauses

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/>



BESEDNOVRSTNA KATEGORIZACIJA KOT SLOVNIČNI TEMELJ POMENSKEGA OPISA V SLOVARJU

Jerica SNOJ

Inštitut za slovenski jezik Frana Ramovša

Snoj, J. (2018): Besednovrstna kategorizacija kot slovnični temelj pomenkega opisa v slovarju. Slovenščina 2.0, 2018 (2): 206–226.

DOI: <http://dx.doi.org/10.4312/slo2.0.2018.2.206-226>.

V prispevku se obravnava besednovrstna kategorizacija s stališča navajanja besednovrstnih oznak v slovarskem priročniku za slovenski jezik na splošni ravni, veljavni ne glede na določeni slovarski koncept. Uvodoma je prikazana svojost besednovrstnega kategoriziranja v slovnični teoriji, čemur sledi ponazoritev vloge besednovrstnega določanja leksikalnih enot znotraj slovarskega pomenkega opisa. Ob primerih za povedkovnik, členek in izdeležniške tvorjenke je prikazan razvoj besednovrstnega kategoriziranja v doslejnjih slovenskih slovarjih in na osnovi tega so podane smernice za besednovrstno označevanje leksikalnih enot v prihodnjih slovenskih slovarskih priročnikih.

Ključne besede: jezikovni opis, pomen, slovnica

1 UVOD

Pomenska analiza, na osnovi katere nastane pomenki opis leksikalne enote v slovarju, vključuje besednovrstno kategorizacijo kot nujni organizacijski temelj celotnega slovarskega prikaza, tako v okviru posameznega slovarskega sestavka kot na ravni zgradbe celotnega slovarskega priročnika. Besedne vrste, ki so v slovnični znanosti definirane kot abstrakcija zlasti skladenjskih značilnosti besedja, so v slovaropisju kategorizacijsko sredstvo, ki omogoča enotno obravnavo določenih razredov leksikalnih enot, s čimer se v celovitem pomenskem opisu zagotavlja nujna usklajenost in enotnost. Besednovrstna

kategorizacija je v tej vlogi temeljnega pomena ne glede na to, katere vrste je gradivo, s katerim se pripravlja slovar; gradivo mora biti v vsakem primeru ustrezno pripravljeno, da je omogočena interpretativna pomenska analiza, v temelju vključujoča besednovrstno kategorizacijo. V postopku slovaropisne identifikacije posamezne leksikalne enote in nato v njenem pomenskem opisu se tako vsakokrat znova soočata slovnična definicija besedne vrste kot slovnične kategorije in celotni nabor pomenskih značilnosti te enote. Na izrazni ravni slovarskega pomenskega opisa je rezultat tega postopka razviden v obliki, kot jo določa pri vsakem slovarju slovarski koncept.

Besednovrstna kategorizacija slovenskega jezika je v zadnjem času deležna precejšnje pozornosti z različnih stališč.¹ Prispevek se tem obravnavam pridružuje tako, da izhaja iz neposrednih slovaropisnih izkušenj, pridobljenih pri pomenskih analizah in pri sestavljanju slovarskih sestavkov v slovenskih slovarjih.² Osredotoča se na vprašanja, kot jih zastavlja upoštevanje besednovrstne kategorizacije in praktično navajanje besednovrstnih oznak v predpostavljenem slovarskem priročniku na splošni ravni, veljavni ne glede na specifični slovarski koncept.³ V tem okviru se predpostavlja slovarski priročnik kot zaključeno leksikografsko delo, obstoječe v elektronski obliki ali v knjižni obliki, v obeh primerih pa na ravni zaključenega slovaropisnega izdelka z razvidno zasnovo in eksplikacijo pomenskega opisa, ki v svoji celovitosti omogoča uporabniku vpogled v pomensko razčlenjenost sodobnega slovenskega besedja.⁴ Zanima nas torej eksplicitno izražanje besednovrstne kategorialnosti, ki mora biti v kvalitetnem slovarskem priročniku za uporabnika razvidno in nedvoumno, če naj bo v podporo v slovarskem sestavku

¹ V ospredju so zlasti objave: Balažic Bulc 2015, Grošelj 2015, Žele 2015, Gantar 2015, Snoj 2015, Stramljič Breznik 2014.

² Ti slovarji so: *Slovar slovenskega knjižnega jezika* (prva izdaja), *Slovar Slovenskega pravopisa* (2001) in *Sinonimni slovar slovenskega jezika* (2016).

³ Razmerje med obravnavo besednovrstne kategorizacije v slovnici nasproti tisti v slovarju je sicer podrobneje obravnavano v Snoj 2015.

⁴ Iz navedenega naj bi bilo razvidno, da se v tukajšnji obravnavi ločujeta slovarski priročnik kot v samem sebi zaključen leksikografski izdelek in leksikalna baza z drugačno, specifično funkcijo. Prim. Gantar 2015: 116.

prikazani pomenski razčlenitvi.

Obravnava je uvedena s povzetkom o značaju besednovrstnega kategoriziranja v slovnični teoriji, temu sledi prikaz mesta besednovrstne kategorizacije znotraj koncepta leksikalnega pomena. V nadaljevanju je ob nekaterih primerih ponazorjeno, zakaj je v slovaropisju nujno dosledno podrejanje besednovrstne kategorizacije leksikalnemu pomenu kot celoviti jezikovnopomenski danosti.

2 BESEDNE VRSTE V SLOVNICI

V skladu s splošno predstavo je, da slovar prevzema že izdelane definicije besednih vrst iz slovnice in besedje za slovar razvršča in opremlja z besednovrstnimi oznakami glede na te definicije. V kolikšni meri ta predstava ustreza dejanskosti, je mogoče presojati tako, da se s tega vidika oceni obravnava besednih vrst v za slovenski jezik referenčni slovnici Jožeta Toporišiča (Toporišič 2004).

V slovnici je težišče obravnave besednih vrst v oblikovanju definicij za posamezne kategorije s poudarkom na njihovi medsebojni različnosti. Slovnica J. Toporišiča obravnava besedne vrste v okviru oblikoslovja, torej »tistega dela slovnice, ki slovarskim enotam določa vrstne, oblikovne in funkcijske značilnosti« (Toporišič 2004: 255). Besedne vrste so predstavljene »kot pojmi za množice besed z enakimi skladenjskimi vlogami in drugimi lastnostmi«, od katerih so nakazane »npr. tvorjenost, slovnične kategorije, konverzivnost ipd.«. Uvodni določitvi pojma »besedna vrsta« in navedbi devetih v slovenskem jeziku prepoznanih besednih vrst neposredno sledijo splošni opisi kategorij, s katerimi so posamezne besedne vrste razločevane: pregibnost, oblikoslovna vzorčenost, naglasna vzorčenost, morfematika; v ločeni skupini z naslovom »Inherentne kategorije sklonljivega« so prikazane kategorije spol, sklon, število, številskost in oseba. Splošni opis kategorij se zaključuje z razdelkom »Drugo«, ki napoveduje, da so pri razločevanju besednih vrst udeležene še druge kategorije, prikazane pri obravnavi posameznih besednih vrst (Toporišič 2004: 255–273). S tem pristopom je eksplicitno izraženo, v čem temelji

razmejevanje, ločevanje besednih vrst in pripravljena je pot v posamezni opis vsake od njih. Znotraj teh opisov (samostalniška beseda, pridevniška beseda, glagol itd.) so nato podrobneje obravnavane za vsako od besednih vrst relevantne razločevalne kategorije tako, da so obsežno in izčrpno prikazane njihove oblikoslovne in skladenjske značilnosti.

Slovenska slovnica slovarniku omogoča, da si izoblikuje razmeroma jasno predstavo o besednih vrstah kar zadeva oblikoslovne in skladenjske značilnosti posameznih kategorij.⁵ Povezanost besednovrstne kategorizacije z leksikalnim pomenom pa slovnica tega tipa pričakovano zgolj okvirno nakazuje, pri posameznih besednih vrstah v različni meri. Tako je npr. pri samostalniški besedi in pridevniški besedi prikaz besednovrstnih razločevalnih lastnosti razdeljen v obsežno predstavitev »skladenjskih lastnosti« in v kratek povzetek o »pomenskih lastnostih« (Toporišič 2004: 274, 318), toliko da je nakazano, kako je dana besedna vrsta kot kategorija določena z obojno razločevalnostjo.

Pomenskim lastnostim v ožjem smislu je v Slovenski slovnici J. Toporišiča še največja pozornost namenjena v tem, da se z njimi utemljujejo nekatere obsežne podrazvrstitve znotraj dane besedne vrste. Tako je npr. besedna vrsta členek prikazana celo z dvema podrazvrstitvama, ki temeljita na pomenskih lastnostih in sta še dalje členjeni v mnoge podrazvrstitve. Tako podrobna pomenska členitev naj bi bila prikladna za slovarnika, saj se s svojo obsežno razčlenjenostjo kaže kot popolni povzetek mogočih leksikalnih pomenov pri tej besedni vrsti. Če pa se ta obsežna pomenska razčlenitev primerja s samo kategorialno definicijo členska kot besedne vrste, postane jasno, da ima obsežna pomenska razčlenitev vlogo nujnega dopolnila ali celo nadomestka za skopo kategorialno definicijo, ki odraža dejstvo, da je členek kategorialno težje določljiv in da se nekateri členki »po vlogi približujejo veznikom, drugi

⁵ Izraz »slovenska slovnica« se v tem kontekstu nanaša pretežno na slovnično védenje o besednih vrstah, kot ga vsebuje *Slovenska slovnica* J. Toporišiča in kot je razviden v *Slovenskem pravopisu* 2001; kljub pomanjkljivostim in starosti je to najsodobnejši celoviti opis besednovrstnosti slovenskega jezika in zato trenutno edini, ki je uporaben v praktičnem slovaropisju.

prislovom«, poleg tega je okrog ena četrtnina členkov enakozvočna z vezniki in prislovi (Toporišič 2004: 445).

Ne glede na obstoječe definicije besednih vrst v slovnici ostaja raziskovanje povezave med besednovrstno kategorizacijo in leksikalnim pomenom dane leksikalne enote naloga slovaropisca.

3 SPLOŠNO O BESEDNOVRSTNI KATEGORIZACIJI V SLOVARSKEM POMENSKEM OPISU

Besednovrstna kategorizacija kot leksikalnopomenska danost je v jeziku zmeraj realno prisotna v pomenskorazločevalni vlogi; naloga slovaropisca je, da jo pri posamezni leksikalni enoti ustrezno identificira. Besednovrstne oznake kot slovaropisna dogovorna oblika navajanja besednovrstne kategorizacije so v slovarskem priročniku eden od običajnih, pričakovanih podatkov, s katerim je v slovarju identificirana vsaka iztočnica. Na najbolj splošni ravni besednovrstna oznaka izraža umestitev jezikovnega izraza v poimenovalni sistem danega jezika, saj v gospodarni obliki prikazuje številne lastnosti leksikalne enote: skladenjsko vlogo v stavčni povedi, pregibanjske značilnosti, tipski denotativni pomen. Uporabnik slovarja se pri povprečno zahtevni uporabi slovarja vsebine in organizacijske vloge teh oznak v celoti navadno niti ne zaveda. Toliko bolj pa se mora te vloge zavedati slovarnik, ki mora označevanje izpeljati tako, da so leksikalne enote določene besednovrstne kategorije prikazane enako, v skladu s pomensko vrednostjo dane kategorije.

Glede na to, da se besedne vrste pojmujejo kot slovnična danost, bi se lahko pričakovalo, da so dokončno definirane v slovnici in da je s tem že rešeno tudi vprašanje besednovrstnega določanja v slovaropisju. Pri neposrednem oblikovanju slovarskega sestavka v skladu z danim slovarskim konceptom in upoštevanje za slovar določeno besedilno gradivo se, nasprotno, potrjuje tipična razlika, tudi sicer obstoječa med pogledom slovnice na jezikovno resničnost in potrebami slovaropisja pri slovarskem prikazu te resničnosti. Slovnica je v zvezi z besednimi vrstami zavezana ugotavljanju in definiranju besednih vrst kot

kategorij znotraj slovničnega opisa danega jezika; v slovarskem priročniku pa naj bi bilo za vsako leksikalno enoto ali sestavino večbesedne leksikalne enote navedeno, v katero besednovrstno kategorijo se uvršča. V idealnem primeru bi slovnica imela predvidene in definirane besednovrstne kategorije, s katerimi bi bilo mogoče nedvoumno kategorizirati prav vse leksikalne enote in njihove sestavine. V slovaropisnih priročnikih se ta postopek prikazuje kot neproblematično opisovanje slovničnih lastnosti leksikalnih enot in njihovih sestavin; z navedbo besednovrstne oznake naj bi bila podana temeljna slovnica v zvezi z določeno leksikalno enoto skoraj v celoti.⁶ Besednovrstne oznake so neproblematične celo s stališča izbire, kaj v slovarju navajati iz leksikalne baze, saj so samoumevne (Attkins 2008: 219, 400). Dejansko pa prav v slovaropisju postane jasno, da jezikovna resničnost s svojo kompleksnostjo idealnosti slovničnega opisa v nekaterih primerih nasprotuje. V realizaciji slovarskega priročnika ima poleg tega besedo še slovarski koncept, ki postavlja svojske zahteve glede izraženosti besednovrstnega kategorizacijskega sistema.

Besednovrstno označevanje v slovarju dejansko vključuje soočenje enkratnih pomenskih lastnosti dane leksikalne enote s kategorizacijo besednih vrst, kot je dostopna v slovničnem opisu danega jezika. Slovnični opis slovenskega jezika sicer vključuje definicije posameznih besednih vrst, vendar to še ne zadošča za oblikovanje praktično uporabnega temelja, na katerem bi se gradila besednovrstna kategorizacija v slovarju. Uvrstitev leksikalne enote v določeno besednovrstno kategorijo ima ustrezno vsebino v pomenskem opisu leksikalne enote samo, če je besednovrstna kategorizacija sama po sebi (kot slovnična danost) pojmovana kot sestavina leksikalnega pomena.

V tem pogledu se v praktičnem slovaropisju kot primerno uporaben potrjuje strukturalni model leksikalnega pomena, vključujoč denotativni pomen in

⁶ Besednovrstna kategorizacija je v slovaropisnem priročniku lahko obravnavana na docela tehnični ravni, v okviru navodil za oblikovanje slovarskih razlag. Eno takih navodil je, da mora biti razlagalna beseda besednovrstno enakovredna razlagani besedi. Za posamezne besedne vrste se lahko predvidijo čisto določene tipske razlage ipd. Prim. Landau 2008: 138–144.

kategorialni pomen kot obvezna neločljiva dela leksikalnega pomena, vključno z izhodiščem, da ima vsaka leksikalna enota določljivi leksikalni pomen (Vidovič Muha 2013: 17–113). Kategorialni pomen, obstoječ v kategorialnih pomenskih lastnostih (npr. sklanjatev, števnost pri samostalniku, glagolski vid, vezljivost pri glagolu, stopnjevanje, določnost pri pridevniku itd.), je nosilec besednovrstne kategorizacije. Kategorialne pomenske sestavine, uresničevalke kategorialnih pomenskih lastnosti v pomenskih realizacijah leksikalnih enot (npr. števno+, števno-, moški/ženski/srednji spol, dovršnost/nedovršnost itd.), so besedilna konkretizacija besednovrstne kategorialnosti kot systemske danosti. Ob tako modeliranem leksikalnem pomenu je omogočen enotni pogled na vlogo besednovrstne kategorizacije pri raznovrstnih leksikalnih enotah, kar je v zvezi s pomenskim opisom v slovarju bistvenega pomena.⁷

4 RAZVOJNOST BESEDNOVRSTNEGA KATEGORIZIRANJA V SLOVENSKIH SLOVARJIH

Tako iz slovarskih prikazov kot iz načelne obravnave besednovrstne problematike je razvidno, da nekatera vprašanja povzročajo dvome in nesoglasja glede besednovrstne kategorizacije in sploh glede slovarskega prikaza vse od SSKJ dalje.⁸ Med v tem smislu problematične se uvrščajo: povedkovnik, členek (zlasti v razmejevanju nasproti prislovu), prvi del zloženk, vrstni pridevnik in področje konverznih tvorjenk nasploh. Problematika tega področja je v nadaljevanju ponazorjena s primeri izbranih leksikalnih enot. Navedbi leksikalne enote v ponazarjalnem sobesedilu, relevantnem za določanje besednovrstne pripadnosti,⁹ sledijo prikazi v *Slovarju slovenskega*

⁷ Obstoje tudi druge možnosti modeliranja leksikalnega pomena. Tukaj navedeni, povzet po *Slovenskem leksikalnem pomenoslovju*, je v slovenističnem slovaropisju pač na prvem mestu med drugim zato, ker je skladen z edinim celovitim pomenskim opisom slovenskega jezika v SSKJ.

⁸ Med drugim je to razvidno celo iz vzporednega branja obravnave besednih vrst pri leksikalni bazi v Gantar 2015 in opisom besednovrstne problematike v Černelič 1984. Kljub časovni razdalji so izpostavljena istovrstna vprašanja.

⁹ Za leksikalno enoto, navedeno v ponazorilu, se predpostavlja, da izpolnjuje pogoje za sprejem v predpostavljjeni slovar.

knjižnega jezika, v prvi izdaji (SSKJ1) in v drugi izdaji (SSKJ2), v *Slovarju Slovenskega pravopisa 2001* (SP 2001) in *Sinonimnem slovarju slovenskega jezika 2016* (SSSJ), iz katerih je razviden razvoj razmišljanja o besednovrstnosti v slovenističnem slovaropisju. Iz sprememb, kot so razvidne od slovarja do slovarja, je povzeta ocena, v kolikšni meri je dana slovarska rešitev v skladu z zahtevo, da je besednovrstna kategorizacija smiselno vgrajena v siceršnji pomenski opis.

4.2 Povedkovnik

*Ta kraj je zanj svet.

SSKJ1: **svét** -a -o prid. ... 5. ekspr., v povedni rabi ki ima za koga zelo veliko vrednost zaradi svoje povezanosti s čim zelo cenjenim, ljubljenim: ta kraj je zanj svet; zakon jim je bil svet in nedotakljiv; izgubil je vse, kar mu je bilo sveto

SP: **svét** -a -o ... poud. **svet** komu/čemu Ta stvar mu je ~a |ima zanj zelo veliko vrednost|; **svet** za koga/kaj Ta kraj je zanj ~ |ima zanj zelo veliko vrednost|

SSKJ2: = SSKJ1

SSSJ: **svét**⁴ -a -o povdk. 1. |izraža, da je kaj zaradi popolnosti, izjemnosti vredno najvišjega čaščenja, spoštovanja| Vsako življenje je sveto... SINONIMI: **božanski, božanstveni** 2. komu/za koga ekspr. |izraža, da ima kaj za koga zelo veliko vrednost zaradi povezanosti s čim zelo cenjenim, ljubljenim| Materni jezik je bil Slovencem v nekaterih časih svet SINONIMI: **svetinja, sveta stvar, sveto, paladij, relikvija**

Na osnovi pomenske analize je v SSKJ1 povedkovniški pomen pri 5. prikazan pri pridevniku v samostojni pomenski enoti, pomenskorazlikovalno opremljeni s slovničnim kvalifikatorskim pojasnilom »v povedni rabi«. Navedena je torej skladenjska značilnost, s katero se ta pomenska enota razlikuje od prejšnjih (od 1. do 4.).¹⁰ Pomen ima samostojno razlago, ki pa ne vključuje kategorialne

¹⁰ V skladu s sodobnim slovenističnim leksikološkim védenjem je povedkovniški tudi pomen 1. (*Bog je svet*), medtem ko nadaljnji pomeni vključujejo stalne besedne zveze, v katerih nastopa vrstni pridevnik, torej leksikalna enota sveti -a -o prid. : 1. *sveti Duh, sveta družina*, (opomba se nadaljuje na naslednji strani)

prilagoditve in je tipsko enaka razlagi lastnostnega pridevnika v prilastkovni funkciji. Vezljivost kot kategorialna pomenska lastnost (*za koga* in variantno *komu*) je prikazana na ravni slovarskega ponazorila, kot besedilna danost, in ne kot strukturna značilnost. Povedkovniški pomen kot leksikalni pomen je bil v SSKJ1 torej na osnovi gradivne analize ustrezno prepoznan in osamosvojen z navedbo skladenjske značilnosti »v povedni rabi«, izrazna raven slovaropisnega prikaza (formulacija razlage, (ne)navedba vezljivostnega določila) pa odraža značaj sočasnega pojmovanja predikativnosti v slovnici, pri katerem se predikativnost ni eksplicitno povezovala s specifičnostjo leksikalnega pomena.

V SP 2001 je obravnavani pomen prikazan z navedbo leksema v podiztočnici in eksplicitnim izpisom vezljivostnega določila, vendar brez eksplicitne navedbe, da se podiztočnica nanaša na novi, od pridevniške iztočnice besednovrstno različni povedkovniški leksem. Iz prikaza ni prav razvidno, ali podiztočnica ni morda mišljena samo kot izpostavitev besedilne rabe, sicer prikazane v slovarskem ponazorilu, ki sledi podiztočnici. Mehanično navajanje vezljivostne oznake v obeh variantah, za vezljivostno določilo s kategorialno pomensko sestavino človeško+ in človeško- (*komu/čemu, koga/kaj*), nadalje dokazuje, da slovarski prikaz ne temelji na upoštevanju leksikalnega pomena, saj bi upoštevalje pomen morala biti navedena samo vezljivost za človeško+. ¹¹ Glede na siceršnje navajanje povedkovnikov v tem slovarju ¹² postane neupoštevanje leksikalnega pomena še bolj razvidno, saj vse kaže, da izpridevniški povedkovniki (npr. **svét** -a -o *komu*) niso prepoznani kot samostojna

sveta brata Ciril in Metod; 1.// sveti kraji itd.; **3.** *sveti bik* itd.; **3.// svete knjige** itd.; **6.** *sveti mir*; **6.// sveta dolžnost**.

¹¹ Primere iz SP na tem mestu navajamo zgolj zaradi prikaza besednovrstne problematike. Odsotnost leksikografsko domišljenega koncepta pri Slovarju Slovenskega pravopisa sicer onemogoča kritično razmišljanje o leksikografskih rešitvah v tem slovarju, pri obravnavanem primeru npr. navedba dveh ponazoril, ki imata zapostavljeno isto razlago (»ima zanj veliko vrednost«).

¹² Prim. sestavke pri **blagor** (*Pravi blagor je, da...*), **bomba** (*Film je bil bomba*), **človeško** (*Človeško je biti žalosten*), **čudež** (*Čudež je bilo, da je ...*), **hudo** (*Hudo mi je bilo pri srcu*) itd.

besednovrstna kategorija, kljub enakovrednosti z izsamostalniškimi povedkovniki (**blagor**, **bomba**, **čudež**) in izprislovnimi povedkovniki (**človeško**, **hudo**). Pojmovanje povedkovnika v slovarskem delu Slovenskega pravopisa 2001 tako ostaja na pol poti. Pridevniški leksem dobi vezljivostno oznako (**svet** ⇒ **svet** *komu/čemu*, **svet** *za koga /za kaj*), s čimer se nakazuje neka strukturna povezava z glagolskimi besednovrstnimi lastnostmi, in nič več kot to. Povedkovniška pomenskost in besednovrstna razlikovalnost ostajata nepojasneni, celo manj pojasnjeni kot v SSKJ, kjer je povedkovniškost v povezavi z leksikalnim pomenom razmeroma sistematično in natančno v vseh primerih izražena s slovničnim kvalifikatorskim pojasnilom »v povedni rabi«. Neusklajenost slovarskih rešitev v slovarju SP (različni prikaz izsamostalniških in izpridevniških povedkovnikov) dodatno vnaša nejasnost glede pojmovanja povedkovnika.

Druga izdaja SSKJ na področju pojmovanja povedkovniške besednovrstnosti dejansko ne prinaša nobenih izpopolnitev v primerjavi s prvo izdajo, in to kljub siceršnjemu navajanju oznake povdk.¹³

V SSSJ je obravnavani povedkovniški pomen (*Ta kraj je zanj svet*) eden od dveh pomenov, navedenih v slovarskem sestavku povedkovniške iztočnice **svét⁴** -a -o povdk., ki je na četrtem mestu med homonimi: **svét¹** -á m, **svèt²** svéta m, **svét³** -a oo prid. (SINONIM *dober* v pomenu 'ki kaže pozitivne moralne lastnosti') in **svét⁴** -a -o povdk.¹⁴ Različna nabora sinonimov med drugim dokazujeta, da ima povedkovnik svét -a -o dva različna povedkovniška pomena. Slovarska razlaga obeh pomenov je oblikovana tako, da odraža kategorialno specifičnost povedkovniškega pomena ('izraža, da ...'). Pri drugem pomenu je v obliki zaimenske izpostavitve navedeno vezljivostno določilo v obeh ugotovljenih variantah (*svet komu*, *svet za koga*), kot je v celotnem

¹³ V uvodu k drugi izdaji SSKJ je sicer navedeno, da je obravnava povedkovnika usklajena s SP 2001, vendar slovarski sestavki tega ne potrjujejo.

¹⁴ Poleg navedenih homonimnih iztočnic je sicer v SSSJ še iztočnica **svéti** -a -o prid. (sestavina stalnih besednih zvez) in večbesedna iztočnica **vsí svéti** vseh svetih m mn. v pomenu 'praznik'.

slovarju določeno za prikaz kategorialne pomenske sestavine, uresničuječe vezljivost kot kategorialno pomensko lastnost glagola. Tako je površinsko izkazana tipska povezanost povedkovniškega besednovrstnega pomena z glagolskim pomenom in na ravni slovaropisnega prikaza upoštevano pomenoslovno dejstvo, da je povedkovnik imenski del zloženega glagola z lastnim leksikalnim pomenom (Snoj 2010: 94–151). V sinonimnem slovarju prikaz sinonimije, sinonimnega razmerja med leksemi že sam po sebi narekuje, da se besednovrstna določitev leksikalne enote upošteva dosledno znotraj leksikalnega pomena kot celovite pomenske danosti. Prikaz sinonimije, kot ga posreduje slovarski sestavek pri **svét**⁴ -a -o povdk. v SSSJ, prepričljivo kaže, da je prav ustrezno upoštevanje in eksplicitno izkazovanje besednovrstne kategorialnosti (v tem primeru povedkovniške) tisto, ki v slovarskem priročniku v temelju omogoča pregleden in verodostojen prikaz pomenskih danosti.

4.2 Členek

**Hoja je gotovo zdrava.* (← Hoja je zdrava. {O tem sem prepričan.}) **Jutri gotovo pridem.* (← Jutri pridem. {O tem ni potrebno dvomiti, odločitve ne bom spreminjal.})

SSKJ1: **gotòv** -óva -o **prid.** ... **gotóvo** 1. **prislov od gotov**: na ta način se da najbolj gotovo ugotoviti, kako se ptice selijo ... 2. nav. ekspr. **izraža a)** *prepričanost o čem*: takemu delu gotovo ne bo kos ... **b)** **v povedni rabi podkrepitev trditve**: gotovo je, da o tem ne bi smel govoriti ... **c)** *nepreklicnost česa*: gotovo bom naredil do jutri 3. **v medmetni rabi izraža a)** *nepreklicnost česa*: vrnem vam z obrestmi, gotovo ... **b)** *soglasje, pritrditev*: »tega ne bi smel reči.« »Gotovo, predaleč je šel« ...

SP: **gotóvo** **soglaš. člen.** To bo ~ držalo; Počakaj me, ~ pridem; Vrnem vam z obrestmi, ~

(**gotóvo**¹ **nač. prisl.** Glas mu je zvenel ~)

(**gotóvo**² **povdk.** ~ je, da bi o tem morali govoriti javno)

SSKJ2: = SSKJ1, le spremenjeni naglas gotôvo in sprememba pri **gotôvo 2.**: v **členkovni rabi** izraža a) *prepričanost o čem*: takemu delu gotovo ne bo ... (nadaljevanje enako kot v SSKJ1)

SSSJ:

(**gotôvo**¹ -ega **s** 'gotovina')

(**gotôvo**² **prisł.** 'samozavestno, odločno')

gotôvo³ **poudar. člen.** **1.** |poudarja prepričanost o čem| *Hoja je gotovo zdrava*
SINONIMI: **brez dvoma, brez oklevanja, definitivno** ... **2.** |poudarja nepreklicnost česa| *Jutri gotovo pridem* SINONIMI: **zagotovo, gvišno, vsekakor, za prmej** ... GLEJ ŠE pri dominantah: **res** (1. **soglašalni členek**: *Gotovo, to ni pošteno*, 2. **poudarjalni členek**: *Gotovo ne lažem*), **seveda** (**soglašalni členek**: *Me vzameš s sabo? Gotovo*)

(**gotôvo**⁴ **povdk.** |izraža, da je kdo prepričan o resničnosti trditve| *O tem bi morali razpravljati javno; to je gotovo* SINONIMI: **dejstvo, enkrat ena, fakt, faktum, nesporno, zagotovo**)

Iz navedenih slovarskih prikazov členka gotovo je razvidno, kako je besednovrstno določanje vključevalo določeni razvoj, ki prav v zvezi s to besedno vrsto še vedno nima zaključka v povsem jasnih definicijskih merilih (Balažič Bulc 2017). Za slovarski pomenski opis pa so bistvene prav enoumne, zanesljive besednovrstne določevalne lastnosti. Če za členek morda še ni ugotovljen popolni nabor besednovrstno razlikovalnih lastnosti, imajo v slovarskem pomenskem opisu toliko večji pomen tiste, ki so doslej nedvoumno ugotovljene. Pri členu je to njegova temeljna razlikovalnost v primerjavi s prislovom, razvidna v dejstvu, da ima prislov vselej določujočo vlogo za glagolski, pridevniški ali prislovni pomen, nikoli pa za samostalniški pomen; prislov odgovarja na vprašalnico, členek ne. Pomenska specifičnost členka, da ima sicer težko določljive kategorialne pomenske lastnosti in da je v mnogih primerih konveržno tvorjen iz prislova, ne opravičuje mešanja s prislovom in povedkovnikom v primerih, ko je prislovni in povedkovniški pomen nedvoumno prisoten.

Vzemimo primer: **Hudičevo** se moti, kdor tako misli (Žele 2015: 29). Po obstoječih slovarskih prikazih (SSKJ1, SSKJ2, SSSJ, SP) je **hudičevo** v tem primeru v vlogi prislova, saj nedvoumno določa glagol. V besedilnem gradivu je te vrste izrazitev s prislovom **hudičevo** na pretek, s tako rekoč vsemi vrstami glagolov in pridevnikov (*hudičevo se naprezati/uspevati* itd., *hudičevo mrzel/lep* itd.), tako da je neproblematičnih potrditev za prislov v tem primeru dovolj. V *Slovarju slovenskih členkov* je **hudičevo** v tej vlogi opredeljen kot členek, celo z navedeno zamenljivostjo z »zelo« in »prekleta« (Žele 2015: 29).¹⁵ Iz podobnih primerov v *Slovarju slovenskih členkov* je razvidno, da se razumevanje in definiranje besedne vrste kot kategorialnopomenske danosti meša z zaznavanjem konotativnega in pragmatičnega pomena, kot ga uresničuje **hudičevo** v navedenem primeru in kar dokazuje zapisana razlaga: »izraža močno čustveno prizadetost«. »Močna čustvena prizadetost« je udeležena v izbiri stilno močno označenega prislova **hudičevo** (namesto npr. stilno neoznačenega **zelo**), torej v izbiri določene izrazitvene variante, ki vključuje visoko stopnjo ekspresivnosti v konotativnem pomenu. Izbira izrazitve s konotativnim pomenom (ekspresivnost) je eno, kategorialni pomen izbranega leksema pa drugo. V tem primeru je **hudičevo** prislov, ki intenzivnostno določa glagolsko dejanje, izraženo z glagolom *motiti se*.

Podobno je v *Slovarju slovenskih členkov* kot členek obravnavan leks **hudič** v primerih: *Hudič je stati dve uri na mrazu* ali *Hudič je, če te nihče ne uboga* in *Hudič je vse skupaj* (Žele 2015: 29). Vsi trije primeri so uresničitve stanjskega metaforičnega povedkovnika **hudič** s pomenom 'izraža stanje, ki ga kdo doživlja kot neprijetno, težavno', za katerega je nevtralna izrazitev s povedkovnikom **hudó** in sinonimi **težko**, **bog pomagaj**, **groza**, **grozno**, **grozota**, **pokora**, **obupno**, **zafrknjeno** itd.

¹⁵ Niti navedba prislova zelo kot zamenljivostne variante k **hudičevo**, kar bi samo po sebi zahtevalo enako besednovrstno določitev glede na osnovna pravila usklajenega slovarskega prikaza, ni avtorice napeljalo k ugotovitvi napake.

4.3 Izdeležniške tvorjenke

**Usekani prst ga boli. – *Usekan je na meso. – *Usekan je nanjo.*

SSKJ1: **usékati** -am dov. ... **usékan** -a -o **1.** deležnik od usekati: usekana roka ga boli **2.** pog., ekspr. čudaški, neumen: lepa je res, vendar usekana **3.** pog., ekspr., navadno v povedni rabi, v zvezi z na ki čuti, ima veliko željo po tem, kar izraža določilo: usekan je na delo, ples; biti usekan na meso / usekana je nanj zelo ga ima rada

SP: **usékan** -a -o ... Vsak je po svoje ~ čudaški, omejen **usekan** na koga/kaj biti ~ ~ delo |zelo rad delati|; Čisto ~a je nanj |zelo ga ima rada|

SSKJ2 = SSKJ1

SSSJ: **usékan**¹ -a -o prid. ⇒ neumen (*usekan človek*)

usékan² -a -o povdk. ⇒ ljubiti (biti usekan na koga, biti usekan na kaj), čudaški (biti malo usekan, biti po svoje usekan)

**Rad je športno oblečen.*

SSKJ1: **obléči** oblečem**obléčen** -a -o: sem že oblečen in obut

SP: **obléči** oblečem dov. obléci -te/-íte; oblekel -kla, obléč/obléč, obléčen -a; obléčenje ...

SSKJ2 = SSKJ1

SSSJ: **obléčen** -a -o povdk. ⇒ nositi (*nositi civilno obleko = biti oblečen v civil ...*), oblačiti se (*športno se oblačiti = biti športno oblečen*)

Izdeležniške tvorjenke v doslejšnjih slovarskih prikazih narekujejo jasno ločevanje med (1) deležniškim pomenom, (2) povedkovnikom, (3) izdeležniškim lastnostnim pridevnikom in (4) izdeležniškim vrstnim pridevnikom. Razlikovanje je razvidno že v SSKJ1, npr. v zgoraj navedenem sestavku pri **usekati**, kjer je deležnik prikazan v podiztočničnem lastnem sestavku. Deležniški pomen je pri **1.** eksplicitno izkazan z razlago »deležnik od«. Pri tem pomenu je v primerjavi z izhodiščnim glagolom prisotna samo

kategorialnopomenska sprememba: dejanje $\Rightarrow$ posledica dejanja, ne da bi prišlo do spremembe v denotativnem pomenu. V levem prilastku je pri tem pomenu zmeraj prisotna določnost: **usekati prst $\Rightarrow$ *Prst je usekan $\Rightarrow$ *Usekani prst ga boli*. Pri **2.** je prikazan lastnostni pomen, in sicer s sinonimno razlago (*čudaški, neumen*), ki v slovarju najbolj prepričljivo potrjuje razmeroma težko ulovljivi in redko resnično lastnostni pomen izdeležniškega pridevnika. Pri **3.** je prikazan povedkovniški pomen, vezljivost je delno celo izpostavljena (*»v zvezi z na«*). SP tudi v tem primeru ostaja na pol poti. Iztočnica **usékan** -a -o je pridevniška, čeprav so vsa ponazorila povedkovniška. Vezljivost je izpostavljena, iz zapostavljene razlage je razvidno, da gre za povedkovniški pomen, in ne za pridevniški. SSKJ2 ne prinaša spremembe v primerjavi s SSKJ1. SSSJ ločuje pridevniško iztočnico, s katero izdeležniški lastnostni pridevnik **usekan** pošilja v sinonimni niz pri **neumen**, in povedkovniško iztočnico, iz katere je razvidno, da je usekan kot povedkovnik udeležen v izrazitvah, sinonimnih z ljubiti in biti čudaški.

Za primer izdeležniškega vrstnega pridevnika (4. kategorija izdeležniških tvorjenk) se lahko navede **obložèni** -a -o prid., kot je udeležen v večbesedni leksikalni enoti *obložena govedina*.

5 NEKATERE SMERNICE ZA BESEDNOVRSTNO KATEGORIZIRANJE V PRIHODNJIH SLOVENSKIH SLOVARJIH

5.1 Besednovrstna kategorialnost je v slovaropisju videna znotraj leksikalnega pomena kot celote, in sicer natanko v okviru kategorialnega pomena. Pri besednovrstnem kategoriziranju v slovarju je potrebno dosledno upoštevati, da poimenovalni sistem (slovar) in slovnica nista ločeni jezikovnosistemski danosti, pač pa različni področji enega kontinuuma. Model leksikalnega pomena, v katerem se pomen modelira kot hkratnost denotativnega, kategorialnega, konotativnega in pragmatičnega pomena, je pri utemeljevanju besednovrstne kategorizacije uporaben, ker lahko pripomore k doslednemu uvrščanju besednovrstne kategorizacije na določeno »mesto« v celovitem

pomenskem opisu.

5.2 Besednovrstna kategorizacija kot področje specifične slovaropisne problematike je enovito obvladljiva, če se rešuje v okviru jasne predstave, kaj je leksikalna enota in kaj leksikalni pomen, iz slovnice vzeti pojem besedna vrsta in tamkajšnja definicija različnih besednih vrst pa je ob tem pragmatično uporabljeno orodje za identifikacijo istovrstnih pomenskih danosti. Za slovarski pomenski opis je besedna vrsta kategorija, obstoječa kot nabor kategorialnih pomenskih lastnosti.¹⁶ Besednih vrst je toliko, kolikor je ugotovljenih različnih naborov kategorialnih pomenskih lastnosti, s katerimi se v danem jeziku medsebojno razlikujejo skladenjske vloge in hkrati s temi tipski denotativni pomen (npr. samostalniška besedna vrsta za poimenovanje stvari, oseb, pojmov; glagol za poimenovanje dogodkov itd.).¹⁷ V tukajšnjem prispevku je izpostavljen strukturalni model slovarskega pomena, ker je edini preizkušeni v doslejšnjem slovenističnem slovarstvu, čeprav ga lahko nadomesti kateri drugi. Pomembno je, da je slovaropiscu jasno, kaj mu znotraj izbranega pojmovanja, kaj je leksikalni pomen, pomeni besedna vrsta.

5.3 Določanje leksikalnih enot z besednovrstnimi oznakami v slovarskem priročniku je v vsakem primeru uvrščanje dane leksikalne enote v določeno besednovrstno kategorijo, tj. povezovanje, identificiranje določene leksikalne enote z besedno vrsto kot abstraktno danostjo, in to v vseh primerih, ne glede

¹⁶ Kako v tem okviru vzpostaviti ustrezno povezavo med leksikalno bazo in slovarskim priročnikom, v katerem se pričakuje eksplicitna pomensko utemeljena besednovrstna kategorizacija, je posebno vprašanje. Vsekakor je uporabnost leksikalne baze omejena, če v njej niso realno izkazane kategorialne pomenske lastnosti leksikalnih enot (če npr. lastnostni pridevniki niso jasno razlikovani od vrstnih pridevnikov, če metaforični konverzni izsamostalniški povedkovnik ni ločen od izhodiščnega samostalnika ali pa če je posamostaljenje evidentirano z brezobrazilno pridevniško obliko).

¹⁷ V tem okviru je potrebno razumeti tudi razvojnost slovenističnega nauka o besednih vrstah. V SSKJ ni členka, ker ob začetku tega dela, prvega razlagalnega slovarja slovenskega jezika, še niso mogle biti identificirane tipske pomenske značilnosti, s katerimi se členek razlikuje od prislova. Prav tako v SSKJ ni oznake povedkovnik, pač pa pogosto slovnično kvalifikatorsko pojasnilo »v povedni rabi«, ki z imenovanjem zgolj skladenjskega mesta nakazuje, da gre za posebni pomen dane leksikalne enote, ne da bi bili že ugotovljeni tipski pomeni povedkovnika kot besedne vrste (lastnostnodoločitveni pomen, stanjski pomen itd.). V Slovarju SP 2001 povedkovnik že nastopa kot posebna besednovrstna kategorija na osnovi skladenjske vloge, medtem ko tipski leksikalni pomen povedkovnika kot besedne vrste ni določen.

na to, ali se določitev nanaša na prvotno besednovrstno kategorialnost ali na konveržno pridobljeno. Zato razlikovanje med eksplicitno navedbo besednovrstne oznake pri dani leksikalni enoti (npr. navedbo roditeljske končnice in oznake za spol pri samostalniških leksikalnih enotah) in posrednim določanjem besednovrstnosti (npr. »kot samostalnik«, »kot pridevnik« ipd.) s stališča sinhronega pomenskega opisa ni smiselno. Prav tako je s stališča slovaropisja zanemarljivo, ali se določena kategorija interpretira kot »skladenjsko-pomenska kategorija« (Žele 2003) ali kot »besedna vrsta«. V slovarskem priročniku je pomembno, da je slovarniku jasno, katero kategorialnopomensko vsebino simbolizirajo uporabljene besednovrstne oznake.

5.4 Konverzne tvorjenke so samostojne leksikalne enote z lastno leksikalnopomensko vrednostjo, kar mora biti v slovarskem pomenskem opisu podprto z jasno besednovrstno določitvijo. Izkazovanje povezanosti, ki obstoji med leksikalnimi enotami zaradi konverzne tvorjenosti (sistem podiztočnic ipd.), je drugotno v primerjavi s tem, da morajo biti konverzne tvorjenke v slovarju pomensko opisane kot samostojne leksikalne enote.

5.5 Razločna prisotnost besednovrstnega označevanja in dosledno upoštevanje meril pri dodeljevanju besednovrstnih oznak v celotnem slovarju, od začetka do konca, zanesljivo veliko prispeva k uporabnosti slovarskega priročnika.

6 ZAKLJUČEK

Pojmovanje besednovrstne kategorizacije v slovarskem pomenskem opisu se podreja dejstvu, da je slovar kot jezikovni priročnik namenjen prikazu pomenskosti. Slovar je pri besednovrstnem kategoriziranju dolžen upoštevati spoznanja in definicijske zmožnosti slovnične znanosti, za slovenski jezik razmeroma obsežne in temeljite, in sicer v smislu čim bolj jasnega ločevanja med različnimi besednimi vrstami. Hkrati pa mora biti slovaropisje kritično do priložnostnega slovničarskega eksperimentiranja in v jezikoslovni teoriji nujnega teoretičnega relativiziranja definicijskih lastnosti posameznih

besednih vrst. Obstoječe védenje o besednih vrstah v slovenskem jeziku se bo v slovničarskem preučevanju v prihodnje še širilo z novimi pogledi in ugotovitvami. V slovaropisju pa je pomembno, da so merila za besednovrstno določanje leksikalnih enot razlikovalno utemeljena in v slovarski zasnovi jasno določena. Relativiziranje in mešanje meril, prisotno v drugi izdaji SSKJ (Ahlin 2014) in v Slovarju slovenskih členkov, v slovaropisju ni perspektivno.

V postopku praktične izdelave določenega slovarja o obliki in načinu izraženosti besednovrstne kategorizacije odloča zasnova slovarja, in sicer v okviru svoje temeljne vloge, da natančno predvidi vsebino in zgradbo načrtovanega slovarja in v sami sebi zagotovi konsistentnost posameznih slovarskih rešitev v njihovi medsebojni povezanosti. V zvezi z besednovrstno kategorizacijo je pri tem odločilna jezikoslovna izobraženost načrtovalca slovarja in njegova ambicioznost glede stopnje natančnosti v prikazu leksikalnopomenskih razlik med leksikalnimi enotami. Ob dovolj znižani natančnostni ravni v izkazovanju leksikalnopomenskih danosti se še dalje lahko ignorira doslejšnje ugotovitve v zvezi s pomensko razločevalnostjo v slovenskem jeziku: lahko se meša prislove in členke, lahko se ukine povedkovnik in pri roki je povratek v preživeto preteklost s pojmom »nesklonljivo«; v imenu »rabe« se lahko zanemari obrazilno izraženo pomensko razlikovalnost med vrstnimi pridevniki ter lastnostnimi pridevniki ipd. O tem, kako bodo že doslej prepoznane pomenskorazlikovalne danosti, med katerimi so besednovrstne kategorije temeljne, izkazane v prihodnjih slovarjih, odločata zgolj zmožnost in hotenje slovarnikov, koliko natančen in glede na jezikovno realnost verodostojen pomenski opis so sposobni udejanjiti.

LITERATURA

Ahlin, M., B. Lazar, Z. Praznik, in J. Snoj (2014): Slovar slovenskega knjižnega jezika. Druga, dopolnjena in deloma prenovljena izdaja. *Jezik in slovstvo* 59 (4). 121–127.

Atkins, B. T. S., in M. Rundell (2008): *The Oxford Guide to Practical*

Lexicography. Oxford: Oxford University Press.

Balažic Bulc, T. (2015): Členek v slovenskem jezikoslovju in slovarju. V V. Gorjanc in dr.: *Slovar sodobne slovenščine: problemi in rešitve*: 524–538. Ljubljana: Znanstvena založba Filozofske fakultete UL.

Černelič-Kozlevčar, I. (1988): Reševanje besednovrstnih vprašanj v Slovarju slovenskega knjižnega jezika. V B. Paternu in F. Jakopin (ur.): *Sodobni slovenski jezik, književnost in kultura*: 289–300. Ljubljana: Filozofska fakulteta.

Gantar, P. (2015): *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana: Znanstvena založba Filozofske fakultete UL.

Grošelj, R. (2015): Besedne vrste v slovenskem jeziku. V V. Gorjanc idr.: *Slovar sodobne slovenščine: problemi in rešitve*. 498–513. Ljubljana: Znanstvena založba Filozofske fakultete UL.

Landau, S. I. (1989): *Dictionaries. The Art and Craft of Lexicography*. Druga izdaja. Cambridge: Cambridge University Press.

SSKJ: *Slovar slovenskega knjižnega jezika* (1970–1991/spletna različica: 2000). Ljubljana: ZRC SAZU. Dostopno prek: <https://fran.si/>.

SSKJ2: *Slovar slovenskega knjižnega jezika* (2014). Druga, dopolnjena in deloma prenovljena izdaja. Dostopno prek: <https://fran.si/>.

SP: *Slovenski pravopis* (2001). Ljubljana: ZRC SAZU. Dostopno prek: <https://fran.si/>.

Snoj, J. (2010): *Metafora v leksikalnem sistemu*. Ljubljana: Založba ZRC, ZRC SAZU.

Snoj, J. (2015): Od definicije besedne vrste v slovnici do besednovrstne oznake v slovarju. *Slovenski jezik* 10: 163–175.

Stramljič Breznik, I. (2014): *Medmeti v slovenskem jeziku*. Maribor: Založba Pivec.

Toporišič, J. (2004): *Slovenska slovnica*. Maribor: Založba Obzorja.

Vidovič Muha, A. (2013): *Slovensko leksikalno pomenoslovje*. Druga, dopolnjena izdaja. Ljubljana: Znanstvena založba Filozofske fakultete.

Žele, A. (2003): Slovarska obravnava povedkovnika. *Jezik in slovnstvo* 48 (2). 3–15.

Žele, A. (2015): *Slovar slovenskih členkov*. Ljubljana: Založba ZRC, ZRC SAZU.

WORD-CLASS CATEGORIZATION AS A GRAMMATICAL FOUNDATION OF THE LEXICOGRAPHIC DESCRIPTION OF MEANING

The paper deals with word-class categorization from the standpoint of providing word-class tags in dictionaries of Slovene at a general level, applicable regardless of the lexicographic concept. We first demonstrate the individuality of word-class categorization in the theory of grammar, followed by the demonstration of the role of word-class categorization of lexical units within the description of meaning in a dictionary. Providing examples for the predicative, particle, and participial compounds, we demonstrate the development of word-class categorization in the existing Slovene dictionaries and provide guidelines for future dictionaries of Slovene.

Keywords: linguistic description, meaning, grammar

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/>

