

MATEJ METERC

ANALIZA FRAZEOLOŠKE VARIANTNOSTI ZA SLOVARSKI PRIKAZ V ESSKJ-JU IN SPP-JU

COBISS: 1.01

[HTTPS://DOI.ORG/10.3986/JZ.25.2.2](https://doi.org/10.3986/jz.25.2.2)

Prispevek predstavlja teoretična in metodološka izhodišča za prikaz frazeoloških in paremioloških variant v Slovarju slovenskega knjižnega jezika (tretja izdaja), ki je na portalu Fran dostopen pod imenom eSSKJ, ter v Slovarju pregovorov in sorodnih paremioloških izrazov (SPP), katerega prvi prirastek bo objavljen leta 2020 na istem portalu. Najprej bomo podali razloge za natančno obravnavo variant tako v splošnem kot v specializiranem slovarju. Nato bomo opisali različne metode korpusne analize in vključevanja anketno pridobljenih podatkov vanjo.

Ključne besede: slovaropisje, frazeologija, frazem, paremiološka enota, pregovor

Analysis of Phraseological Variants for Their Lexicographic Representation in eSSKJ and SPP

This article presents a theoretical and methodological basis for lexicographic presentation of phraseological and paremiological variants in the third edition of *Slovar slovenskega knjižnega jezika* (Standard Slovenian Dictionary, eSSKJ), which has been published on the Fran portal, and in *Slovar pregovorov in sorodnih paremioloških izrazov* (Dictionary of Slovenian Proverbs and Similar Paremiological Expressions, SPP), the first volume of which will be published in 2020 on the same lexicographic portal. The need for accurate lexicographic presentation of variants in both dictionaries is then discussed. This is followed by a description of various methods of corpus analysis and combining them with data from sociolinguistic surveys.

Keywords: lexicography, phraseology, phraseme, paremiological unit, proverb

1 SMOTRI NATANČNE OBRAVNAVE FRAZEOLOŠKIH IN PAREMIOLOŠKIH VARIANT V DVEH RAZLIČNIH SLOVARJIH

Za hkratni prikaz obravnave variantnosti frazeoloških in paremioloških enot (v nadaljevanju FE in PE) v dveh različnih slovarjih smo se odločili, ker je v obeh primerih nujna natančna gradivska (zlasti korpusna) analiza. Poleg tega sta to prva slovarska priročnika, ki sistematično izboljšujeta nezadovoljivo stanje slovenske frazeografije in paremiografije; medtem ko imamo v teoriji jasno opredeljeno tipologijo glasoslovnih, oblikoslovnih, leksikalnih, skladenjskih in pravopisnih variant FE in PE (Kržišnik 1988: 156) in nadaljnje členitve (Gantar 2007: 229), pa so nam v do sedaj objavljenih splošnih in specializiranih slovarjih ter zbirkah na voljo le

Prispevek je nastal v okviru projekta Z6-9378, ki ga financira ARRS v okviru programa P60038.

naključno izpostavljene variante.¹ Pri eSSKJ upoštevamo opozorilo, da »nestrokovna presoja med frazeološkimi variantami, ki jih potrjuje gradivo, in njihova nehierarhična in zgolj informativna slovarska razporeditev za uporabnika ne bosta obvestilni« (Jakop 2014). Za slovenščino do eSSKJ-ja nismo imeli slovarja, ki bi ponujal skrbno empirično preverjen podatek o variantah posamezne FE in PE.

2 FRAZEOLOŠKA IN PAREMIOLOŠKA VARIANTNOST V OKVIRU SPLOŠNEGA RAZLAGALNEGA SLOVARJA eSSKJ

Nekaj besed namenimo obravnavi variant v prvi in drugi izdaji SSKJ-ja. Frazeološkega gradiva je v njem veliko, paremiološkega pa dokaj malo – za določanje slovenskega paremiološkega optimuma (Meterc 2017: 49) smo iz celotnega slovarja izpisali 599 PE. Več kot 270 se jih je uvrstilo v paremiološki minimum in so danes poznane več kot 79 % govorcev slovenščine. To lahko pripišemo dejstvu, da so avtorji SSKJ-ja izhajali iz analize množice aktualnih besedil. Izkazalo se je tudi, da je bilo v SSKJ vključene tudi precej manj poznane paremiologije – četrtnina PE v tem slovarju je poznana manj kot polovici slovenskih govorcev (Meterc 2017: 71), pri čemer smo upoštevali tudi anketirance, rojene v zadnjih desetletjih. SSKJ je k variantam pristopal nesistematično (Kržišnik 1988: 156–158). Delno iz tega razloga, delno pa verjetno tudi zaradi sprememb v samem frazeološkem fondu se mnoge razlikujejo od danes najbolj pogoste (Jesenšek – Ulčnik 2014: 284; Meterc 2017: 216). Ko smo v jezikovnem korpusu Gigafida preverili 300 najbolj poznanih PE (Meterc 2017: 219), se je izkazalo, da moramo kar tretjini izmed njih poleg variante iz SSKJ-ja dopisati varianto, ki danes prevladuje v rabi. Nekatere variante PE, ki so bile v SSKJ-ju predstavljene kot osnovne (in edine), najverjetneje niso bile reprezentativne niti v času nastanka slovarja, npr. *Nasprotja se privlačujejo* ali *Čas je zlato* namesto *Nasprotja se privlačijo* in *Čas je denar*.

Skok v kakovosti slovaropisne obravnave variant FE in PE prinaša eSSKJ z gradivsko (zlasti korpusno) potrditvijo vsake variante FE ali PE v sodobni knjižni slovenščini. Slovar prinaša celostno obravnavo FE in PE: »Frazeološkim enotam v novem razlagalnem slovarju namenjamo poseben razdelek, saj zaradi svoje strukturne, semantične in pragmatične specifikke zahtevajo celovito in enovito slovarsko obravnavo, zaradi uporabniških potreb pa tudi nazoren prikaz« (Gliha Komac idr. 2016: 26). Pri PE se korpusnim pristopom pridružujejo sociolingvistični s pomočjo ankete za slovenski paremiološki optimum (Meterc 2017): »Z izsledki te raziskave lažje pridobimo in presojamo podatke v besedilnem korpusu Gigafida. To je v primerjavi z besednozvezno ravno na paremiološki ravni bolj potrebno, ker korpusna orodja še niso dovolj zahtevna za daljše ustaljene enote, po drugi strani pa gre tudi za večjo stopnjo kombinatorike« (Gliha Komac

¹ Izjemo predstavlja spletni portal Sprichwort-Plattform, ki (v omejenem obsegu iztočnic) prinaša korpusno pridobljene podatke o variantnosti PE.

idr. 2016: 26). Mnogi pregovori so zelo variantni, kar dokazuje tudi Čermákova (2014: 210) raziskava, v kateri je s pomočjo jezikovnega korpusa ugotovil, da ima več kot 40 % od 241 analiziranih čeških pregovorov variante.

V eSSKJ-ju ne preverjamo samo FE in PE, ki so že bile zabeležene v SSKJ-ju ali Slovarju slovenskih frazemov (Keber 2012: 495, v nadaljevanju SSF), temveč poleg tega sistematično iščemo tudi do sedaj neuslovarjene, a za sodobno rabo značilne enote in njihove variante (Jakop – Meterc 2016: 55). Ker imajo nekatere FE in PE veliko število različno pogostih variant, je treba zamejiti, koliko variant in katere izmed njih bodo uporabniku sploh prikazane v splošnem slovarju. V eSSKJ-ju želimo uporabniku podati v knjižnem jeziku najbolj pogoste in s tem za uporabnika najbolj relevantne variante. Razporejene so po frekvenčnem merilu od najpogostejše navzdol, večinoma pa se omejujemo na prikaz največ treh variant (poleg osnovne). V slovarski bazi sicer beležimo tudi variante, ki ostanejo pod frekvenčnim pragom ali pa ga prestopijo, vendar jih po pogostnosti prehitijo dovolj drugih. Poudarjamo, da izbor osnovne slovarske oblike ne pomeni predpisovanja najbolj priporočljive variante. Gre samo za podatek o varianti, ki je najbolj značilna za knjižna besedila. Podrobna obravnava variant FE in PE je ključna, še preden se odločamo za izbor najbolj relevantnih variant za prikaz v slovarju. Od zadostne pogostnosti vsaj ene variante FE ali PE je namreč odvisna njena uvrstitev v slovar sploh. Za to, da bi se vsaj ena varianta – osnovna oblika, ki jo lahko razumemo tudi kot »ničto varianto« (Grzybek – Chlosta 1995: 72) – prebila v splošni slovar, uporabljamo frekvenčni prag približno petih prototipnih zgledov iste variante v korpusu Gigafida, v mejnih primerih pa je za nas pomembno tudi, ali nam obliko potrdi uravnoteženi korpus Kres. Pri določanju prototipnih zgledov se opiramo na Čermákova (2014: 115–117) merila in jih dopolnjujemo. Kot neprototipne najpogostejše izločamo množične rabe frazemov v okviru naslova del popularne kulture, očitne prepise iz zbirk, zglede v jezikoslovnih člankih ter podvajanja zaradi novinarskega navajanja iste izjave. Dodaten razlog za natančno obravnavo je tudi razmejevanje variant od številnih prenovitev, ki predstavljajo variantam vzporedno in prav tako z vidika jezikovnih ravnin zelo razčlenjeno pojavnost (Kržišnik 1987), ki pa v slovar ne sodi (Kržišnik 1988: 146; 2004: 203).² Le v redkih primerih iskanje razširimo na dodatne vire. Za primer navedimo FE *stric iz ozadja*, za katerega v Gigafidi z iskalnim pogojem [lemma="stric"][lemma="iz"][lemma="ozadje"] najdemo dva zгледа, v sLWaCu preko 1000, v deduplicirani Gigafidi 2.0 pa okoli 700 zgledov.

2 Za primer navedimo skrajno obliko prenovitve – frazeološko križanje enot *medvedja usluga* in *levji delež*. V korpusu Gigafida 2.0 najdemo z iskalnim pogojem [lemma="medvedji"][lemma="usluga"] presenetljivo število (38) zgledov križanja *levja usluga*, ki ga ne moremo šteti za varianto enega ali drugega izhodiščnega frazema. Zanimivo je, da se obratno križanje pojavi v istem korpusu zgolj v enem prototipnem zgledu: »Resnici na ljubo je računalnik precej glasnejši od predhodnika. Pa primerjava ni ravno poštena, saj že sama grafična karta prispeva medvedji delež hrupa.«

Enako merilo približno petih zgledov velja pri ESSKJ-ju za PE, čeprav se paremiološki fond od frazeološkega po pogostnosti izrazov dokaj razlikuje. Najpogostejše izmed najbolj znanih PE iz paremiološkega optimuma so se v korpusu FidaPLUS pojavile v največ 700 zgledih, tudi v Gigafidi pa redkokdaj presežejo 1000 zgledov, medtem ko besednozvezni frazemi dosegajo večkratnike teh števil. Zgled PE iz paremiološkega minimuma (ter optimuma) se pojavi na okoli 20.000 pojavnic v jezikovnem korpusu, kar se je potrdilo tako z analizo v nekdanjem FidaPLUS kakor tudi v veliko manjšem korpusu GOS (Meterc 2015a: 7). Vseeno pa že po nekaj letih dela v okviru ESSKJ-ja kaže, da se bo v ESSKJ kljub strogemu pragu prebilo precej več aktivno rabljene paremiologije kot v klasični SSKJ.

2.1 Paremiološka variantnost v luči razlogov za zasnovo paremiološkega slovarja SPP

Potreba po specializiranem paremiološkem slovarju izhaja že iz dejstva, da pri mnogih PE, ki so v sodobni rabi izpričane z besedilnimi zgledi ali pa anketnimi izsledki, nobena izmed variant ne dosega zgoraj opisanega praga za splošni slovar. Paremiologija je namreč znatno širša od svojega jedra najbolj pogostih enot, ki bodo uvrščene v splošni razlagalni slovar. Paremiološki fond ni vezan le na pisana knjižna (zlasti publicistična) besedila, ki jih pokriva en sam referenčni korpus. Anketa o poznavanju PE za slovenski paremiološki optimum potrjuje, da je veliko enot, ki so očitno veliko pogostejše rabljene v ustnem kot pisnem sporazumevanju (Meterc 2017: 186). Dober primer za to je *Mrzle roke, vroče srce*, ki se v Gigafidi 2.0 pojavi natanko enkrat, s čimer se v splošni slovar ne prebije, sodeč po anketi pa ga aktivno uporablja 21,2 %, pasivno pa pozna še 28,2 % vprašanih – skupaj skoraj polovica izmed 509 anketirancev. Sodobno rabo manj poznanih in pogostih PE preverjamo v okviru projekta Aktualna raba pregovornih izrazov slovenskega jezika: korpusni in slovaropisni vidik, v okviru katerega nastaja SPP.

Zaradi vsega naštetega mora biti prag za uvrstitev posamezne variante PE v paremiološki slovar nižji kot v primeru splošnega slovarja. Vseeno pa moramo tudi za uvrstitev v SPP s pomočjo širšega nabora virov (zlasti jezikovnih korpusov) ter anketnih pristopov potrditi ustaljenost vsake variante vsaj z nekaj prototipnimi zgledi rabe. Razložimo, glede na kakšna merila bomo razvrščali PE in katere zahteve mora izpolnjevati PE, da jo uvrstimo v slovar. Glede na merilo aktualne rabe ter merilo pomenske določljivosti s sobesedilom bomo razvrščali 200 PE: 50 iz Bojčeve zbirke (1987) z veliko zastaranega gradiva, 50 iz zbirke Marije Makarovič (1975) s potrjeno rabo v 70. letih 20. stoletja ter dveh naborov po 50 PE, ki sta nastala s pomočjo vprašalnika, ki ga bomo opisali v nadaljevanju (iz dodatnih navedkov anketirancev ter izmed PE z manj kot 30-odstotno poznanostjo). Zavedamo se tudi raziskovalnega potenciala zastaranih ter aktualnih, vendar do te mere redkih PE, da jih je v danem trenutku težko opremiti s korektno razlago pomena, temveč kvečjemu z rekonstrukcijami pomenov. Razvrščanje PE po navedenih merilih zato ni razmejevanje enot, ki se jim je vredno posvečati, od teh, ki se jim ni.

Gre za razvrščanje glede na načine, na katere bi se jim bilo treba v prihodnosti posvetiti bodisi z diahronega bodisi s sinhronega vidika. Iz nabora PE, ki jim bomo lahko dokazali aktualno rabo in tudi določili pomen, bo nastal prvi prirastek SPP-ja v okviru slovaropisnega portala Fran.

3 METODE ANALIZE FRAZEOLŠKE IN PAREMIOLŠKE VARIANTNOSTI ZA SLOVAROPISNE POTREBE

V eSSKJ-ju se posvečamo razmejitvi polja večbesednih leksikalnih enot (zlasti glede na pomenske premike in frazeologizacijo) na pogoste kolokacije (v okviru iztočnice) ter stalne besedne zveze in FE ter PE (v razdelkih Stalne zveze in Frazeologija), obenem pa se zavedamo, da gre vendarle za jezikovni kontinuum z mnogimi mejnimi primeri. Zavedanje o kontinuumu se v zadnjih desetletjih krepi tudi pri frazeologih (Steyer 2015: 281). Naštete tipe izrazov proučujemo s pomočjo sorodnih korpusnih metod. Jezikovni korpusi so frazeologom omogočili podrobne in dovolj sistematične analize variantnosti (Jesenšek – Ulčnik 2014: 282–285; Čermak 2014: 210–212; Đurčo 2014: 13; Đurčo – Hein – Steyer 2015: 152). Čim večje upoštevanje različnih pojavnih oblik – tako variant kot različno pogostih in priložnostnih pretvorb in prenovitev (Gantar 2007: 230) – s kombiniranjem različnih iskalnih pogojev je bilo nujno že za določanje pogostnosti posamezne PE za izdelavo paremiološkega optimuma (Meterc 2017: 74). V okviru eSSKJ-ja se v raziskavah variant FE in PE opiramo predvsem na Gigafido, pomožno pa tudi na druge korpusne, v analizah paremiološke variantnosti za prvi prirastek SPP-ja pa še na korpusa Janes in slWaC in dodatne vire (npr. dLib in Wikivir). Koristna orodja predstavljajo ankete. Pri obravnavi PE si v primeru obeh slovarjev v prvi vrsti pomagamo s sociolingvistično raziskavo poznanosti slovenskih PE med govorce vseh starostnih skupin ter z vseh koncev slovenskega govornega območja in izseljenstva, ki smo jo podrobneje že predstavljali (Meterc 2015b; 2017). V poglavju 3.2 bomo prikazali, kako izsledke ankete vključujemo v korpusno analizo variantnosti.

3.1 Korpusna analiza frazeološke variantnosti za eSSKJ

Za primer vzemimo FE, ki je bila v prvi in drugi izdaji SSKJ-ja pri sestavinah *hodi*, *mačka*, *vrel* in *kaša* uslovarjena samo v obliki *hodi kakor mačka okrog vrele kaše* s pomenom 'ne upa se lotiti jedra problema', v SSF-ju pa z enakim pomenom v osnovni obliki *hoditi kot mačka okrog vrele kaše* ter v variantah s sestavinama *maček* in *vrel* ali pa tudi z izpustom sestavine (*hoditi okrog vrele kaše*). V zgleidih, ki jih navaja SSF, najdemo tudi druge variante (npr. s sestavinama *sukati* ter *okoliti*). Natančna korpusna analiza, ki za SSF ni bila opravljena, pokaže veliko več variant tega frazema, predvsem pa se izriše pogostnostno razmerje med njimi, kar predstavlja dragocen podatek, na podlagi katerega lahko izberemo zgolj najbolj

relevantne variante za prikaz v splošnem slovarju. Iskanja te FE se lahko lotimo s shemo iskalnih pogojev s po dvema sestavinama, npr. [lemma="mačka"][]{}{0,5} [lemma="kaša"] in izključevanjem sestavin s pomočjo negativnega filtra (npr. z obliko *vrel*). S takim iskanjem dobimo najprej niz 193 konkordanc, nato pa le šest konkordanc, med katerimi so prosta besedna zveza (prvi primer), dve pretvorbi (drugi in tretji primer) ter skupno tri oblike s pridevnikom *vroč* (od tretjega do vključno šestega primera):

Odprl sem konzervo s hrano za mačke in stresel kašo v skodelico.

»Vse je to vrela mačka okoli kaše,« sem rekel.

Če se ne oziram preveč na besedne vragolije v slogu hoje mačke okrog vroče kaše [...]

[...] res ni razloga, da hodiva kot mačka okoli vroče kaše [...]

Nikar se ne motovelite kot mačka okoli vroče kaše. Hitro pridite z besedo na plan.

In ne dobiš NIČ! Vse se štuli okoli tebe ko mački okol vroče kaše.

Oblike s pridevnikom *vroč* se ne prebijejo preko praga, saj se vsaka pojavi zgolj enkrat. Nadalje bi za potrebe specializiranega slovarja lahko preverili, ali gre za redke variante ali zgolj prenovitve. Podobno se nam z uporabo pozitivnega filtra z lemo *sukati* izkaže, da tudi obliki *sukati se kot mačka okoli vrela kaše* in *sukati se kot mačka okrog vrela kaše* – vsaka s po tremi zgledi – ne dosežeta praga. Ker smo naleteli na več pojavitev iste oblike, je to vredno zabeležiti v slovarsko bazo, saj obstaja večja verjetnost, da gre za (redkejšo) variante FE. Z natančnim pregledom konkordanc po še več kombinacijah iskalnih pogojev ugotovimo, da so za slovarski prikaz najbolj relevantne naslednje variante:

virteti se kot mačka okoli vrela kaše

hoditi kot mačka okoli vrela kaše

virteti se kot mačka okrog vrela kaše

hoditi kot maček okrog vrela kaše

virteti se kot maček okrog vrela kaše

hoditi kot maček okoli vrela kaše

Pod slovarskim pragom ostanejo vse variante s sestavinami *sukati*, *vroč* in *kakor*, pa tudi mnoge druge.

Navedimo še en očitni primer nesistematičnega dela z variantami FE. Enote *dahniti/izreči (usodni) da* v SSKJ-ju ne najdemo, SSF pa jo navaja kot *dahniti (veliki) da*, kar je v nasprotju z današnjo rabo, ki kaže veliko variant s pridevnikom *usoden*. Najbolj uspešno je iskanje z iskalnim pogojem [word="da"][]{}{0,5} [lemma="dahniti"][]{}{0,5} [word="da"]. Z naslednjim iskalnim pogojem se lahko prepričamo, da je varianta s sestavino *veliki* redka: [lemma="velik"][]{}{0,5} [lemma="dahniti"][]{}{0,5} [lemma="velik"]. Najde-

mo vsega štiri zglede, izmed katerih sta dve varianti z zaimkom *svoj* (*dahniti svoj veliki da*). Slovaropisec si v analizi lahko pomaga z lastnim jezikovnim čutom; po njem smo se v obravnavanem primeru najprej odločili za domnevno najbolj pogosti sestavini *da* in *dahniti*, v prejšnjem primeru pa za sestavini *mačka* in *kaša*. Bistveno je, da si z intuicijo v korpusni analizi variantnosti pomagamo, vendar se ne zanašamo zgolj nanjo (Jesenšek – Ulčnik 2014: 282), zato je v konkretnem primeru nujno preveriti npr. tudi druge glagole govorenja (npr. *izreči*) in členek *ja*. Že začetni iskalni pogoji nam potrdijo, da je poleg variante s sestavino *usoden* med najbolj relevantnimi varianta *dahniti da* brez pridevniške sestavine. Vseeno pa je potreben natančen pregled nizov konkordanc, ki jih dobimo že po prvem navedenem iskalnem pogoju s sestavinama *da* in *dahniti*, saj vsebujejo tudi naslednje oblike, pri katerih gre za manj pogoste variante ali pa za prenovitve:

A sta le dva mladoporočenca v farni cerkvi dahnila slovesni da.

Gaspari je pozneje kot guverner Banke Slovenije dahnil končni da za mestni nakup banke.

[...] tudi nadzorni svet Slovenske razvojne družbe dahnil težko pričakovani da za pobot SRD z Mesno industrijo Pomurka [...]

Na podoben način kot sestavinske (vključno s fakultativnimi sestavinami) iščemo tudi pravopisne in oblikoslovne variante. Skladenjske variante uspešno iščemo z iskalnimi pogoji, kakršen je v primeru prislovne FE *dan in noč* oziroma *noč in dan* [word="noč"]{0,2}[word="dan"]{0,2}[word="noč"]. Z iskanjem po lemah bi v primeru pomensko tako sorodnih sestavin našli veliko preveč konkordanc (ok. 13.000). Skladenjsko variantnost lahko preverimo tudi z enostavnima iskalnima pogojema [word="noč"][word="in"][word="dan"] (2919 konkordanc) ter [word="dan"][word="in"][word="noč"] (2166 konkordanc).

Pozorni moramo biti na prepletanje sestavinske in skladske variantnosti, ki pogosto privede do skrajnih primerov variantnosti, kakršne bomo prikazali na primeru PE v naslednjem poglavju. V navezavi na enobesedno iztočnico *mačka* je bilo za eSSKJ treba poleg sestavinskih variant primerjalne FE *gledati se kot pes in mačka*, *biti kot pes in mačka* ter *prepirati se kot pes in mačka*, ki so, sodeč po korpusnem gradivu, najbolj reprezentativne, preveriti tudi skladske variante z zamenjavo sestavin *pes* in *mačka* z iskalnim pogojem [lemma="mačka"][word="in"][lemma="pes"] (790 konkordanc) s pozitivnim filtrom glagolov *gledati* (2 konkordanci brez FE), *biti* (275 konkordanc) in *prepirati* (brez konkordanc). Polnopomenski sestavini smo iskali z lemama zaradi hkratnega preverjanja edninskih, dvojninskih in množinskih oblik, pri katerih pa v tej FE ne gre za oblikoslovne variante. Našli smo zgolj štiri zglede za varianto *biti kot mačka in pes*. V uravnoteženem korpusu Kres potrditve za to nismo našli, tako da je v eSSKJ nismo uvrstili. Podobno smo iskali v primeru variant naslednje FE, ki so se z dovolj visoko pogostnostjo uvrstile v slovar:

mešati jabolka in hruške
 mešati hruške in jabolka
 pomešati jabolka in hruške
 pomešati hruške in jabolka
 primerjati jabolka in hruške
 primerjati hruške in jabolka
 seštevati jabolka in hruške
 seštevati hruške in jabolka

V eSSKJ se ni uvrstila pogovorna varianta *mešati hruške in jabolke*. Poleg tega, da smo med konkordancami našli precej neprototipnih pojavitev te FE (naslova oddaje), so med preostalimi nekaj več kot desetimi zgledi skoraj izključno zgledi z dveh internetnih forumov, kar ne ustreza pogoju razpršenosti v različnih tipih besedil, poleg tega pa je sobesedilo mnogih navedenih zgledov pogovorno.

3.2 Korpusni in sociolingvistični pristopi k paremiološki variantnosti v primeru eSSKJ-ja in SPP-ja

Pri PE lahko uporabimo podobno metodo iskanja s kombinacijo dveh sestavin ter dodajanjem pozitivnega in negativnega filtra kot pri FE. Tako smo poiskali najbolj pogoste variante PE, ki opisuje situacijo, ko se kdo burno odzove na dejanje ali izjavo, zlasti ob omembi, izpostavljanju njegovih pomanjkljivosti, napak, slabosti. V prvem stolpcu navajamo (od najpogostejše do najmanj pogoste) variante, ki smo jih uvrstili v slovar, v drugem stolpcu oblike, ki so prav tako segle preko praga, vendar jih v slovarju zaradi prostorske omejitve ne bomo prikazovali, v preostalih stolpcih pa te, ki so pod pragom, ter take, ki jih v Gigafidi nismo zasledili:

Ko mački stopiš na rep, zacvili. (20 zgledov)
 Če mački stopiš na rep, zacvili. (19)
 Če mački stopiš na rep, zamijavka. (7)
 Če mačku stopiš na rep, zacvili. (6)

Če mački na rep stopiš, zacvili. (6)
 Ko mački stopiš na rep, zamijavka. (5)
 Če mačku na rep stopiš, zacvili. (5)
 Ko mačku stopiš na rep, zacvili. (4)
 Ko mački na rep stopiš, zacvili. (4)
 Ko mačku na rep stopiš, zacvili. (3)

Če mački na rep stopiš, zamijavka. (1)
 Če mačku na rep stopiš, zamijavka. (1)

Ko mačku na rep stopiš, zamijavka. (0)
 Ko mački na rep stopiš, zamijavka. (0)

Pri izbiri med variantama z enakim številom zgledov (6) smo se odločili za varianto s sestavino *maček*, saj ocenjujemo, da je zaradi povezave s še eno enobesedno slovarsko iztočnico pomembneje prikazati dodatno sestavinsko varianto kot zgolj še eno izmed skladenjskih.

Glede iskanja s kombinacijami leksikalnih sestavin omenimo, da se sestavine iz skupnih tematskih polj pogosto pojavljajo v več PE hkrati. Tak je na primer par *sejati* in *žeti*, s katerim najdemo *Kdor seje veter, bo žel vihar, Kar kdo seje, to bo tudi žel* in *Eden seje, drugi žanje*. Pomenska bližina sestavin nam tako v korpusni analizi včasih koristi, včasih pa nam je zaradi preobsežnih seznamov konkordanc v napoto in se raje zatečemo k sestavinam, ki so pomensko bolj oddaljene ali pa pričakujemo celo, da bomo njihove sopojavitve našli skorajda samo v okviru PE (Đurčo – Hein – Steyer 2015: 148).

Plodno je dopolnjevanje zgoraj opisanega iskanja s tako imenovanim iskanjem PE »po temi in remi« (Đurčo 2014: 13), ki pride v poštev pri dvodelnih PE, kakršne so enote s konstrukcijskimi vzorci *Kdor X, ta Y, Če ni X, ni Y, Kakor X, tako Y* in podobno. Za primer vzemimo PE *Kjer je volja, tam je pot* iz skupine 50 PE, ki smo jih izbrali iz množice dodatnih navedkov v že omenjeni anketi (Meterc 2015b; 2017) za slovaropisno obravnavo v prvem prirastku SPP-ja. Če v Gigafidi iščemo po temi (*kjer je volja*), najdemo oblike *Kjer je volja, je tudi pot, Kjer je volja, tam je pot* in *Kjer je volja, tam je tudi pot*, katerih ustaljenost bo v nadaljevanju treba preveriti. Iskanje po remah (npr. po *tam je pot*) iste PE pa nam nakazuje, da bi bilo dobro preveriti tudi druge oblike, npr. *Tam, kjer je volja, tam je pot, Kjer je želja, tam je pot* ter *Kjer je moč, tam je pot*. Iskanje nadaljujemo po posameznih novih odkritih polovicah PE.

Raba dodatnih korpusov, ki pokrivata sfero spletnih besedil (slWaC) in besedil z družbenih omrežij (Janes), močno dopolni paremiološko sliko. Zanima nas na primer, ali kateri izmed treh korpusov v rabi potrdi pogovorno varianto *Ne čohaj se, kjer te ne srbi*, ki jo navaja SSKJ ob glagolu *čohati*, pri čemer gre za enega redkih navedkov variante kake PE v tem slovarju sploh. Če uporabimo remo (*kjer te ne srbi*), najdemo poleg variant s *praskati* bodisi z uporabo enostavnega iskalnega pogoja bodisi s filtriranjem glagolov v okolici te zveze v treh korpusih naslednje oblike:

Kaj pa se čehljaš, kjer te ne srbi. (Gigafida)

Če ne razumeš prekmurščine se ne šrabaj tam kjer te ne srbi. (Gigafida)

[...] in bi bilo veliko bolje, da se ne vpletaš tam, kjer te ne srbi. (Gigafida)

[...] vendar po starem prepričanju, ne čohaj, kjer te ne srbi, vsi modro molčijo. (Gigafida)

Ne usajaj se, kjer te ne srbi. (slWaC)

Mal se pomiri in ne nabijaj tam, kjer te ne srbi. (slWaC)

Povsem druga stvar je tvoje vtikanje, kjer te ne srbi. (Janes)

Ne čohaj se tam kjer te ne srbi Idi nazaj [...]. (Janes)

Kaže, da sestavina *čohati* v rabi je prisotna, vendar smo zaenkrat našli le dve obliki pregovora, ki se pojavita po enkrat; možno pa je, da bomo z obravnavo več virov ali anketiranjem lahko potrdili ustaljenost katere izmed variant z njo in jo uvrstili v SPP.

Poleg večjega števila korpusov je v primeru SPP-ja treba upoštevati tudi bogate anketne podatke, ki jih do določene mere vključimo tudi v analize za ESSKJ. Do junija 2019 se je v anketo o poznanosti 918 PE, prepisanih iz SSKJ-ja ter frazeološkega slovarja v petih jezikih (Pavlica 1960), vključilo okoli 2400 govorcev, od katerih jih je 509 vprašalnik izpolnilo v celoti. Posamezna PE je zaradi naključnega vrstnega reda prikaza vsakemu govorniku deležna približno enakega števila odgovorov, tako da je bila vsaka ocenjena približno 810-krat. Navedke variant preverjamo v korpusu, pri čemer se kažejo zanimiva razhajanja v rezultatih korpusnih in anketnih metod, ki smo jih skupaj z razlogi zanje že predstavili (Meterc 2018: 144). Opisali smo že primer skrajne variantnosti PE *Ako mu prst pokažeš, te bo za roko prijel* (Meterc 2017: 223–225) z več kot 60 navedki različnih oblik, pri čemer je med odgovori najbolj pogosta oblika *Ponudiš prst, vzame roko* s petimi zgledi, poleg nje pa sta bili več kot enkrat – obe le dvakrat – navedeni le še varianti *Ponudiš prst, pa vzame roko* in *Ponudiš prst, zagradi roko*. Še več variant smo našli v Gigafidi, zanimivo pa je, da smo ob poskusu določanja osnovne slovarske oblike ugotovili, da bi bila po Gigafidi ta drugačna (*Ponudiš prst, zagrabijo roko* s štirimi zgledi) od najpogostejšega anketnega navedka.

Prikažimo analizo še ene skrajno variantne PE za SPP s pomočjo treh korpusov in anketnih navedkov. Enota *Ko ima hudič mlade, jih ima veliko* je bila vključena v vprašalnik za paremiološki optimum, zato smo najprej izpisali vse navedke in tiste, ki so bili navedeni večkrat, prešteli. Nato jih je bilo treba poiskati v Gigafidi, sIWaCu in Janesu. Že teh oblik je bilo več kot 20. Nato smo preverili, katere kombinacije prvega in drugega dela PE manjkajo (takih je bilo dodatnih 20), in njihovo rabo preverili v Gigafidi in ostalih dveh korpusih. Nato smo v korpusih dodatno iskali še po zaenkrat najdenih polovicah ter si zapisovali oblike, ki so se pojavile več kot enkrat. Po pregledu v treh korpusih se je izkazalo, da merilu vsaj dveh potrditev ustreza kar okoli 25 variant. Za kar 45 izmed preostalih oblik smo našli po en zgled v korpusu ali anketni navedek, za 15 potencialnih oblik pa nismo našli potrditve. Ob tako velikem številu variant s potrjeno aktualno rabo bo treba razmisliti o inovativnem izbirnem prikazu variant spletnemu uporabniku SPP-ja.

Perspektivno raziskovalno področje se odpira s proučevanjem frazeoloških in paremioloških vzorcev (Đurčo 2014: 13; Đurčo – Hein – Steyer 2015: 88; Steyer 2015; Meterc 2016a; 2017: 101), kakršni so na primer *X na Y, X za predsednika, Kdor X, ta Y* in *X je lepa čednost*. Te raziskave lahko prinesejo paradigmatski obrat v frazeografiji in paremiografiji s povsem novim tipom slovarjev, v katerih bodo imeli vzorci ključno vlogo (Steyer 2015: 293). V okviru tega članka se bomo omejili na koristnost uporabe vzorcev FE in PE pri analizi njihove variantnosti. S splošnejšim iskalnim pogojem [word="kdor"][]{}{0,4}[word="ta"], v katerem zapolnitve X glede na besedne vrste in njihova mesta ne določimo (kar sicer

omogoča zelo koristna iskanja po podvzorcih), najdemo številne PE (prvih pet zgledov) in tudi neleksikalizirane zapolnitve tega vzorca (zadnji zgled):

Kot otroci smo govorili: kdor laže, ta krade. In kdor krade, bolhe ima.
 Iznajdljivost in vztrajnost ali – kdor išče, ta najde!
 [...] ker jim v očeh javnosti škodi ljudski rek »kdor molči, ta pritrdi".
 Po principu izreka »kdor hitro da, ta dvakrat da«.
 Rek, da kdor pije, ta plača, upošteva tudi zdrava gospodarska politika.
 [...] bolj kot kdor koli drug ves ta čas [...]

Pomembno je, da (poleg prenovitev) s to metodo najdemo tudi številne variante, npr. *Kdor laže, ta tudi krade*, *Kdor išče, ta tudi najde* in *Kdor molči, ta pritrjuje*. Vseeno pa tudi ta iskalni pogoj ne zadošča – z njim npr. ne najdemo variant *Kdor išče, najde* in *Kdor hitro da, dvakrat da*, kar potrjuje, da je pri iskanju variant PE dobro kombinirati različne omenjene metode. Glede uporabnosti konstrukcijskih in tematsko-konstrukcijskih vzorcev smo ugotovili, da jih je v korpusni analizi dobro kombinirati z iskanjem po posameznih sestavinah:

Če je za pregovor značilna sestavinska variantnost ali pa je pogosto prenovljen, se to odraža na večjem številu zgledov z iskanjem po konstrukcijskih vzorcih. Če je več skladišne variantnosti, je to razvidno pri iskanju po polnopomenskih sestavinah. Če pri iskanju po prvem tipu zanemarimo določene pogoste skladišne variante in vrstne rede sestavin, to zaznamo pri iskanju po polnopomenskih sestavinah in posledično razširimo iskalne pogoje; obratno pa lahko prvi tip iskanja prinese podatke o sestavinski variantnosti, ki jo nato vključimo v iskalne pogoje po drugem tipu iskanja. (Meterc 2017: 85)

Tak način iskanja je koristen pri zelo variantnih PE v okviru eSSKJ-ja in SPP-ja in omogoča analizo celih skupin PE, ki so osnovane na istem vzorcu, ter tudi iskanje novih PE.

4 ZAKLJUČEK

Čeprav je v splošnem slovarju treba prikazati izbor najbolj relevantnih variant FE in PE, za specializirani paremiološki slovar pa čim več variant PE z dokazljivo aktivno rabo, so metode v teh dveh primerih podobne. Pristopi se razlikujejo zlasti v tem, da je analiza v primeru SPP-ja razširjena še na dva korpusa ter dodatne vire, pa tudi v tem, da sta v primeru specializiranega slovarja natančnost in sistematičnost dodatno izostreni pri manj poznanih in redko rabljenih PE, ter v težnji, da bi v specializiranem slovarju prikazali čim več v rabi potrjenih variant. V analizi variant s pomočjo jezikovnega korpusa se ne smemo zanašati samo na jezikovno intuicijo in na eno samo iskalno metodo, ker bi s tem lahko zajeli le omejen del variantnosti. Pomembno je kombinirati različne opisane korpusne metode, najbolj optimalno pa je v te postopke vključiti anketne izsledke. eSSKJ in SPP sta prva

slovarja, ki prinašata tako skrbno pridobljene podatke o bogati variantnosti slovenske frazeologije in paremiologije.

VIRI

- Bojc 1987** = Etbin Bojc, *Pregovori in reki na Slovenskem*, Ljubljana: DZS, 1987.
Fran: slovarji Inštituta za slovenski jezik *Frana Ramovša ZRC SAZU*, www.fran.si (8. 9. 2018).
Gigafida: *korpusna besedilna zbirka*, <http://www.gigafida.net> (25. 9. 2019).
Keber 2012 = Janez Keber, *Slovar slovenskih frazemov*, Ljubljana: Založba ZRC, ZRC SAZU, 2012.
Korpus slovenskega spleta slWaC v2.1, https://www.clarin.si/noske/run.cgi/corp_info?corpname=slwac&struct_attr_stats=1 (25. 9. 2019).
Korpus slovenskih spletnih uporabniških vsebin Janes v1.0, https://www.clarin.si/noske/run.cgi/corp_info?corpname=janes&struct_attr_stats=1 (25. 9. 2019).
Makarovič 1975 = Marija Makarovič, *Pregovori, življenjske resnice*, Ljubljana: ČZP Kmečki glas, 1975.
Pavlica 1960 = Josip Pavlica, *Frazeološki slovar v petih jezikih*, Ljubljana: Državna založba Slovenije, 1960.
Sprichwort-Plattform, <http://www.sprichwort-plattform.org/sp/Sprichwort-Plattform> (25. 9. 2019).

LITERATURA

- Čermák 2014** = František Čermák, *Proverbs: their lexical and semantic features*, Burlington: The University of Vermont, 2014.
Đurčo 2014 = Peter Đurčo, Paremiologija i korpusnaja lingvistika, *Vestnik novgorodskogo gosudarstvennogo universiteta: filologičeske nauki* 77 (2014), 13–17.
Đurčo – Hein – Steyer 2015 = Peter Đurčo – Katrin Hain – Kathrin Steyer, *Sprichwörter im Gebrauch*, Trnava: Univerzita Sv. Cyrila a Metoda v Trnave, 2015.
Gantar 2007 = Polona Gantar, *Stalne besedne zveze v slovenščini: korpusni pristop*, Ljubljana: Založba ZRC, ZRC SAZU, 2007.
Gliha Komac idr. 2016 = Nataša Gliha Komac – Nataša Jakop – Janoš Ježovnik – Simona Klemenčič – Domen Krvina – Nina Ledinek – Matej Meterc – Mija Michelizza – Tanja Mirtič – Andrej Perdih – Špela Petric – Marko Snoj – Andreja Žele, Novi slovar slovenskega knjižnega jezika – predstavitev temeljnih konceptualnih izhodišč, v: *Zbornik prispevkov s simpozija 2015*, ur. Franc Marušič – Petra Mišmaš – Rok Žaucer, Nova Gorica: Založba Univerze v Novi Gorici, 2016 (Škrabčevi dnevi 9), 16–33.
Grzybek – Chlosta 1995 = Peter Grzybek – Christoph Chlosta, Empirical and Folkloristic Paremiology: Two to Quarrel or to Tango?, *Proverbium* 12 (1995), 67–85.
Jakop 2014 = Nataša Jakop, Frazeologija v sodobnem slovarju slovenskega jezika, v: *Novi slovar za 21. stoletje: e-zbornik s Posveta o novem slovarju slovenskega jezika na Ministrstvu za kulturo, 12. februar 2014*, ur. Irena Grahek – Simona Bergoč, Ljubljana: Ministrstvo za kulturo, 2014, <http://www.dlib.si/details/URN:NBN:SI:DOC-UANHTHWG> (25. 9. 2019).
Jakop – Meterc 2016 = Nataša Jakop – Matej Meterc, Lexikografske spracovanje frazeoloških variantov v novom slovníku slovenského spisovného jazyka, v: *Akademický slovník současné češtiny a software pro jeho tvorbu aneb Slovníky a jejich uživatelé v 21. století: sborník abstraktů z workshopu*, ur. Michaela Lišková – Veronika Vodrážková – Zuzana Džengeová, Praha: Ústav pro jazyk český AV ČR, 2016, 55–56.
Jesenšek – Ulčnik 2014 = Vida Jesenšek – Natalija Ulčnik, Spletni frazeološko-paremiološki portal: redakcijska vprašanja ob slovenskem jezikovnem gradivu, v: *Več glav več ve: frazeologija in paremiologija v slovarju in vsakdanji rabi*, ur. Vida Jesenšek – Saša Babič, Maribor: Oddelek za germanistiko Filozofske fakultete Univerze v Mariboru – Ljubljana: Inštitut za slovensko narodopisje ZRC SAZU, 2014, 80–98.

- Kržišnik 1987** = Erika Kržišnik, Prenovitev kot inovacijski postopek, *Slava* 1.1 (1987), 49–56.
- Kržišnik 1988** = Erika Kržišnik, Frazeološko gradivo v Slovarju slovenskega knjižnega jezika, *Slava* 2.2 (1988), 143–162.
- Kržišnik 2004** = Erika Kržišnik, Poskusni zvezek slovenskega frazeološkega slovarja, *Slavistična revija* 52.2 (2004), 199–208.
- Meterc 2015a** = Matej Meterc, Enote slovenskega paremiološkega minimuma v govornem korpusu GOS, *Slavistična revija* 63.1 (2015), 1–16.
- Meterc 2015b** = Matej Meterc, Online questionnaire providing information on most well-known and well-understood proverbs in Slovene language, v: *Phraseologie im Wörterbuch und Korpus = Phraseology in Dictionaries and Corpora*, ur. Vida Jesenšek – Peter Grzybek, Maribor: Univerza v Mariboru, 2015 (Zora 97), 205–216.
- Meterc 2016** = Matej Meterc, Tematsko-konstruktivski vzorci nastanka in prenovitev stavčnih frazemov, *Slavistična revija* 64.2 (2016), 125–138.
- Meterc 2017** = Matej Meterc, *Paremiološki optimum: najbolj poznani in pogosti pregovori ter sorodne paremije v slovenščini*, Ljubljana: Založba ZRC, ZRC SAZU (Zbirka Lingua Slovenica 10), 2017.
- Meterc 2018** = Matej Meterc, Psiholingvistična presoja rezultatov sociolingvistične raziskave poznavanja paremioloških ničtih variant, v: *Slavofraz 2016: Phraseologie und (naive) Psychologie*, ur. Agnieszka Będkowska-Kopczyk – Heinrich Pfandl, Hamburg: Verlag Dr. Kovač, 2018, 141–155.
- Steyer 2015** = Kathrin Steyer, Usuelle Wortverbindungen: zentrale Muster des Sprachgebrauchs aus korpusanalytischer Sicht, *International Journal of Lexicography* 28.2 (2015), 260–265.

SUMMARY

Analysis of Phraseological Variants for Their Lexicographic Representation in eSSKJ and SPP

The third edition of *Slovar slovenskega knjižnega jezika* (Standard Slovenian Dictionary, eSSKJ) and *Slovar pregovorov in sorodnih paremioloških izrazov* (Dictionary of Proverbs and Similar Paremiological Units, SPP) are the first Slovenian dictionaries with a lexicographic presentation of phraseological and paremiological variants based on modern methodological approaches. These approaches include precise analysis of different (orthographic, morphological, lexical, syntactic, and other) variants using language corpora combined with specialized surveys of Slovenian speakers. First, the need for an accurate lexicographic presentation of variants in both dictionaries is explained along with the conceptual differences between eSSKJ and SPP. eSSKJ provides its users with the most frequent variants of phraseological and paremiological units. The main reference corpus (Gigafida) is used to check whether all phraseological and paremiological units and their variants meet these requirements. SPP—a growing online paremiological dictionary that will be published in 2020—will include many more paremiological units and their variants. It will present every unit and each of its variants if their active usage in modern Slovenian can be confirmed using three large language corpora (Gigafida, the Slovenian web corpus slWaC, and Janes, a corpus containing tweets, forum posts, blog texts, and comments) and additional sociolinguistic questionnaires. Various specialized search procedures with CQL (Corpus Query Language) are presented in detail in the article. A critical comparison between corpus data and the results of a paremiological survey in which more than five hundred language users took part is proposed as the most efficient tool for acquiring reliable data for the dictionary.