

## **PRIMERJALNI FRAZEMI V RABI IN SLOVARJU<sup>1</sup>**

**IZVORNI ZNANSTVENI RAD**

**UDK 811.163.6'373.7'374**

[HTTPS://WWW.DOI.ORG/10.17234/9789533790121.14](https://www.doi.org/10.17234/9789533790121.14)

Namen prispevka je razširitev razmišljanja v članku *Zoonimi v primerjalnih frazemih* (Kržišnik 2014), vendar z obrnjeno hierarhijo: zanimali nas bodo vsi primerjalni frazemi (PF), v članku navajani korpusni podatki o rabi zoonimov v njih pa bodo primerjalno izhodišče za dodatno preverjanje rabe v novejših in zvrstno različnih slovenskih besedilnih korpusih: Gigafida v2.0 (1990–2018), slWaC v2.1 (2014) in Janes v1.0 (zajem besedil 2015/2016). Nastajajoči eSSKJ, ki je rastoči slovar (od l. 2016) – »s popolnoma novim celovitim opisom besedja slovenskega knjižnega jezika od druge polovice 20. stoletja do danes« (www.fran.si) – v konceptualnih izhodiščih kot frekvenčni prag za uvrstitev (enobesednega) leksema v slovar navaja 100 pojavitev v (zvrstno) različnih korpusnih virih, za frazeme pa kot najnižji prag velja 5 pojavitev. Zanimalo nas bo, kaj to pomeni za PF.

**Ključne besede:** besedilni korpus; frazeologija; slovar; primerjalni frazemi

### **1. VZROK, NAMEN IN POTEK RAZISKAVE**

Kot nekorpusna jezikoslovka se lotevam ugotavljanja razmerja med rabo in slovarjem oz. uslovarjanjem primerjalnih frazemov (dalje PF) s pomočjo podatkov, pridobljenih iz korpusov slovenskega jezika, čeprav je jasno, da bi bilo to nalogo najbolje in najhitreje mogoče opraviti z vpogledom v slovar kolokacij. V slovenščini celo obstaja korpusni slovar z naslovom *Kolokacijski slovar sodobne slovenščine* (Kolokacije 1.0), vendar je uporabniku za zdaj dostopen le v obliki, ki je rezultat avtomatskega luščenja podatkov, prim. na strani <https://www.cjvt.si/raziskovalno-delo/projekti-cjvt/kolokacijski-slovar-sodobne-slovenscine>. Hkrati na Inštitutu za slovenski jezik SAZU od l. 2016 nastaja spletni rastoči slovar eSSKJ, »nov/i/ temeljn/i/ informativno-normativn/i/ razlagaln/i/ slovar slovenskega knjižnega jezika«, in sicer »na podlagi analize sodobnega jezikovnega gradiva« (Predgovor). V tem primeru gre v metodološkem smislu za nekorpusni slovar. Toda v prvi polovici 21. stol.

---

1 Ta članek je bil napisan kot voščilo dragi kolegici in prijateljici Željki Fink ob njeni okrogli obletnici, napisan pa je bil tudi kot premislek o eni največjih strukturnih skupin v frazeologiji in vključitvi njenih enot v splošni in frazeološki slovar – kar je pravzaprav prav tako posvečeno prof. dr. Ž. Fink in njenemu bogatemu slovarskemu opusu.

si je težko zamisliti, da bi slovar nastajal brez upoštevanja stanja v elektronskem besedilnem korpusu (še bolje, v več in različno profiliranih korpusih).<sup>2</sup>

Ne glede na metodološki pristop, lahko ugotovimo, da v nobenem od obeh navedenih nastajajočih slovarjev še niso obdelane iztočnice primerjalnih veznikov *kot*, *kakor* in *ko*, ki so identificirajoča sestavina PF v slovenščini. Zato je namen tega prispevka pregledati, kaj o rabi teh frazemov dobimo v slovenskih besedilnih korpusih *Gigafida 2.0* (1.333.360.653 pojavnic; dalje GF), *Janes* (252.904.238, dalje J) in *sIWaC* (895.903.321, dalje sIW), ki so različno obsežni,<sup>3</sup> a hkrati tudi jezikovno- in besedilnozvrstno različni: najtradicionalnejša, načeloma brez spletnih besedil in deduplicirana je GF; J vsebuje besedila blogov, spletnih forumov, komentarjev na novice in tvitov in je tako najboljši približek govorjenih besedil, sIW je korpus slovenskih spletnih besedil in je torej najboljši pokazatelj aktualne splošne rabe jezika. Prav pred kratkim je bil zgrajen korpus *metaFida* (4.463.244.126 p., dalje mF), ki združuje javno dostopne slovenske korpuse (ne le GF, J in sIW).<sup>4</sup> Iz tega skušamo dalje razmišljati, koliko in kako lahko dobljene podatke uporabimo pri delu s slovarjem, splošnim in/ali frazeološkim.

Primerjalno izhajamo iz že objavljene raziskave o PF, ki je bila gradivno omejena na PF z zoonimnimi sestavinami (dalje PFZ, prim. Kržišnik 2014), pri čemer smo – razen v točki 4 – analizo posplošili na vse PF, podatki v starejših korpusih pa so preverjeni in aktualizirani s podatki v zgoraj omenjenih novejših.

## 2. DISTRIBUCIJA RABE PRIMERJALNIH VEZNIKOV KAKOR – KOT – KO IN NJIHOVA VARIANTNOST

Razmerje med tremi primerjalnimi vezniki je v PF približno enakovrstno kot v slovenščini sploh – nezaznamovan je *kot*, zelo knjižni in (vsaj v frazeologiji) nekoliko starinski je *kakor*, medtem ko je primerjalni *ko* zaznamovan z govornostjo in pogovarjalnostjo (ne nujno neknjižno pogovornostjo). Ker v korpusu primerjalni in časovni veznik *ko* nista ločena, bomo o primerjalnem *ko* spregovorili samo v zvezi s PF, splošno kvantitativno razmerje med *kot* in *kakor* pa je naslednje (navedena količina velja na milijon pojavnic!):

Korpus	<i>kot</i>	<i>kakor</i>
GF	4.270,43/mio	204,28
J	3.407,50	132,46
sIW	4.326,46	204,48
mF	4.307,13	179,64

2 Iskanje gradivne potrditve na spletnih straneh seveda niso nadomestilo za po definiranih merilih urejeno korpusno besedilno gradivo.

3 Zato bomo ob primerjavi frekventnosti navajali količino pojavitev dane enote na milijon pojavnic.

4 O mF več na <https://docs.google.com/document/d/1Vn-T6REdU5pSGHCdCRwpJ4GzT3LuFN-csK5RYJOFXZrM/edit?usp=sharing>. Korpus je trenutno še v poskusni fazi, zato smo ga uporabili samo izjemoma, ko je sumarni rezultat povednejši od rezultatov o zvrstni različnosti.

2.1. Odstopanje od tega smo našli za oba zaznamovana veznika.

2.1.1. Z iskanjem *kakor*+desni kolokator naletimo vsaj na štiri PF, ki se rabijo samo ali pretežno s *kakor*.

Najprej sta to pragmatična frazema *kakor komu drago*, ki 'izraža govorčevo (zadržano) pritrdjevanje', in *kakor kdo ve in zna*, ki 'izraža govorčevo neprizadetost, nepripravljenost pomagati; privoščljivost, ukazovalnost' – zgled rabe za prvega »*Kdo bo poklical naše vozilo? Nočem, da bi še naprej norel po hiši. Naj poskrbijo zanj drugi. **Kakor vejo in znajo.***« in za drugega „[...] *toda moja navada je, vse sproti plačevati in te navade ne bom popustil.*« »**Kakor vam drago,**« pravi gostilničar in gre. – in dva pregovora, *kakor ti meni, tako jaz tebi* in *kakor si kdo postelje, tako leži / spi*. Z iskanjem [word="kakor"][tag="Z.\*"][word="drago"] dobimo za PF *kakor komu drago* v GF 152, J 40, sIW 130 in v mF 343 pojavitev.<sup>5</sup> Pri PF *kakor kdo ve in zna* lahko govorimo tudi o variantnosti *kakor / kot*, vendar je varianta s *kot* redkejša, prim. *kakor : kot* v GF 654 : 289, J 138 : 59, sIW 413 : 177.

Pri pregovornih PF se varianta s *kot* pojavlja izjemoma (v mF dvakrat!) v primeru, ko primerjalni veznik ni na začetku, npr. *Tako kot ti meni, jaz tebi*.<sup>6</sup> Ker so pregovorni frazemi lahko in najpogosteje tudi so samostojne povedi, hkrati pa se kot samostojna poved lahko rabijo tudi stavčni pragmatični frazemi, sklepamo, da se varianta s *kakor* kot nezaznamovana ohranja na začetku. Koliko te vrste distribucija veznika *kakor* (ali vsaj njegova večja pogostnost v prvem, *kakor* v drugem delu zloženega stavka) velja tudi zunaj frazeologije, nismo preverjali, bi pa bilo smiselno.

Seveda je mogoče ugovarjati, češ da je PF *kakor komu drago* izcitatni frazem (Shakespeareova komedija *Kakor vam drago*, v angl. *As you like it*), zaradi tega tudi zaznamovan s knjižnostjo in da se zato drži *kakor*. Tudi pri pregovorih bi načeloma lahko govorili o kaki arhaizaciji v smislu časa nastanka in tradicionalnosti pregovora, a težko bi jim na splošno odvzemali mesto v govorjenem jeziku, vsaj tistim najpogosteje rabljenim. Zanesljivo pa je *kakor veš in znaš* pragmatični frazem, ki je – kot pragmatični frazemi pogosto – značilen ravno za govorjenje.

2.1.2. Medtem ko je veznik *kakor* v SSKJ<sup>2</sup> še razmeroma pogosto naveden v PF,<sup>7</sup> se *ko* najde praviloma le v nekaj PF v okviru iztočnice *ko*, pod II. in s kvalifikatorjem *pog[ovorno]*, tudi ne kot edina ali osnovna varianta PF, saj je npr. *star ko zemlja* v iztočnici *zemlja* naveden v obliki *star kot zemlja*. Za občutek stanja v sodobnih besedilih navedimo, da je v korpusih razmerje *ko : kot* v tem PF naslednje: GF 9 : 64, J 21 : 32 in v sIW 33 (vendar je od tega 10-krat naslov) : 78, in ugotovimo lahko, da zlasti stanje v Janesu potrjuje zvrstno zaznamovanost variante s *ko* na govorjenost.

Preizkus variantnost *ko / kot* v okviru skladijskega vzorca *Prid+ko / kot+Sam* v vseh treh korpusih je pokazal, da so upoštevanja vredni PF še *zdrav ko dren*, *drag ko žafran*, pa tudi *zdrav ko riba*, *lačen ko volk*, *zvit ko presta*, *pijan ko krava* (zanimivo, da ne bolj tradicionalni

5 Za *kot* jih z istim iskanjem dobimo v GF 6, J 1, sIW 4, mF 12.

6 V mF se enkrat pojavi varianta *kot* tudi na začetku (*kot ti meni, tako jaz tebi*) – kot vir je naveden Janes (blogi), kar ni zanemarljiv podatek.

7 Enako tudi v SSF, iztočnice *ko* pa SSF nima, tako da je težko presoditi, koliko je te variante v njem.

*pijan ko mavra*, ki je v slW tudi s kot rabljen le 14-krat), *suh ko prekla* in *seveda* (neknjižno pogovorni) *hladen ko špricer*. Razen zadnjega bi jih težko označili kot neknjižno pogovorne PF.

**2.1.2.1.** Posebno pozornost zahtevajo t. i. intenzifikatorji, npr. *ko(t) strela*, *ko(t) vrag*, *ko(t) hudič*. Prim. razmerja:<sup>8</sup>

	<i>kot strela : ko strela</i>	<i>kot vrag : ko vrag</i>	<i>kot hudič : ko hudič</i>
GF	464 : 20 <sup>9</sup>	105 : 21	111 : 66
J	116 : 74	30 : 31	206 : 210
slW	340 : 52	74 : 31	396 : 297

Za zdaj variante s *ko* ni ne v SSKJ<sup>2</sup> ne v SSF, a navedeni rezultati kažejo, da bi zaslužila uvrstitev ne le v frazeološki, temveč tudi splošni slovar, še posebej, ker kaže, da ustaljenost variante *ko* pri intenzifikatorjih ni nujno socialnozvrstno omejena.

**2.1.3.** Če povzamemo: razmerje med *kot* in *kakor* v PF je na splošno enako kot sicer v slovenščini, zato v slovarju zadošča samo navajanje z nezaznamovanim *kot*, dvojnost je lahko omenjena v uvodnem delu, kjer so podana splošna navodila za »branje« slovarskih podatkov. Izjema pri varianti *ko* so intenzifikatorji, prav tako je vredno preveriti morebitno širšo omejenost stalnosti za veznik *kakor*.

**2.2.** Fakultativnost primerjalnega veznika ni nič posebej novega v skupini PF. Primerjava rezultatov v korpusih z zabeleženostjo fakultativnosti veznika v slovarjih kaže nekatere novosti, ki bi se lahko pokazale kot tendence, vredne pozornosti.

**2.2.1.** Mogoče je najti variante, ki so že v obeh za frazeologijo najpomembnejših slovarjih, SSKJ<sup>2</sup> in SSF, npr. *je (kot) na dlani*, *stresti (kot) iz rokava*. Dalje take, ki jih SSKJ<sup>2</sup> še nima, so pa v SSF, npr. *biti (kot) pljunek v morje*, *(kot) po tekočem traku* ali *biti / sedeti (kot) na trnih*.

Korpusno gradivo pokaže PF, ki (še) niso uslovarjene variante, tako npr. *kaj gre za med* kot PF z variantama *kaj gre / se prodaja kot za med*: GF 44/26 pojavitev, J 10/2, slW 14/7. Še več je samostalniških frazemov, ki imajo pretežno ali izključno sintaktično funkcijo imenskega dela predikata, npr. *dvorezni meč*, *podaljšana roka koga / česa*, se pa pojavljajo v obliki *kot dvorezni meč*, *kot podaljšana roka*. Ker je pojavitev, tudi ob *biti* kar nekaj, nekatere težko zavrnemo kot le posledico glagolske valence, kar zagotovo je v *počutiti se kot dvorezni meč*, ne pa tudi v primeru *Obstajajo takšni stili vedenja, ki lahko organizaciji koristijo ali pa jo omejujejo*. **Mnogi so kot dvorezni meč**, s katerim je treba ravnati previdno. Vsekakor je vredno premisleka, ali gre res za variantno enoto ali le sintaktično (valenčno) povzročeno rabo, primerov za zadnje je seveda več (npr. *kupovati kot mačka v žaklju*).

8 Pravi intenzifikator se načeloma ne prilagaja vezljivostnim zahtevam, zato smo upoštevali le obliko *ko(t) hudič*, prim. razliko med *bati se koga kot hudič* in *bati se koga kot hudiča*.

9 Odštete so nestalne zveze in drugi PF, npr. *ko strela z jasnega* ali *bati se ko hudič križa* itd. Konkordance so bile pregledane ročno, zato so manjša odstopanja možna.

**2.2.2.** Med PF, ki v slovenskih slovarjih še niso našli svojega mesta in za katere še težko z gotovostjo rečemo, v katero smer bodo šli (PF <-> ne-PF?), sodita varianti PF *imeti spomin kot zlata ribica* in *imeti spomin zlate ribice*. O njem je bilo govora že v Kržišnik (2014), in sicer kot edinem primeru PF s potencialnim statusom samostalniškega frazema. Stanje sedem let kasneje in ob bogatejšem gradivu je naslednje:

	spomin kot zlata ribica	spomin zlate ribice
GF 2.0	3 (vsi <i>imeti</i> )	21 (5 <i>imeti</i> ) <sup>10</sup>
J	28 (24 <i>imeti</i> )	206 (88 <i>imeti</i> )
slW	19 (15 <i>imeti</i> )	101 (48 <i>imeti</i> )
mF	53 (41 <i>imeti</i> )	335 (151 <i>imeti</i> )

Iz navedenega lahko razberemo, da je varianta ne-PF še vedno v povprečju 6-krat pogostejša. Za razliko od vseh drugih navedenih primerov variantnosti je pri tem nenavadno, da je težko določiti, v kateri smeri je od samega začetka potekal razvoj (frazem ni star, niti v SSKJ<sup>2</sup> niti v SSF ni vključen). V najbolj povzemalni mF (združeni korpusi) dobimo za iskanje obeh variant naslednje podatke:

- iskanje *spomin zlate ribice*:<sup>11</sup>

2003	2005	2006	2008	2009	2011	2012	2013	2014	2015	2016	2017	2018
1	2	1	1	3	6	3	36	80	44	37	18	4

- iskanje *spomin kot zlata ribica*:

2004	2006	2007	2010	2011	2012	2013	2014	2015	2016	2017
1	2	1	2	1	1	4	11	3	5	3

Edini zanesljivi sklep, ki ga iz teh časovnih podatkov (slW jih nima, zato v tem iskanju ni zajet) lahko izpeljemo, je, da je v vseh letih pogosteje rabljena varianta *spomin zlate ribice*, in morda dodatno pragmatično spoznanje, da je bilo l. 2014 leto »slabega spomina« – obe varianti sta v tem letu najpogostejši.

Zanesljivo lahko o smeri prehoda PF v tudi variantni ne-PF govorimo pri primeru *kaj gre / teče kot po maslu*. Primerov rabe *kaj gre po maslu* očitno ni veliko, nanjo smo naleteli po naključju v *Kolokacijskem slovarju* pod lemo *maslo* skupaj s ponazarjalnimi zgledi, ki so ustrezali frazeološkemu pomenu, npr. *Daewoo je po lastnih trditvah na dobri poti, da se izkoplje iz največjih zadreg, po drugi strani pa nekatere številke kažejo, da vse ne gre po maslu*. Pregled po korpusih le z iskanjem po najpogostejši glagolski sestavini *iti* v položaju levostično pokaže naslednje: v Gf 18 (0,01), Janes 7 (0,03), slW 13 (0,01). Te vrste fakultativnost se bo, mogoče celo pri več PF, razkrila šele, ko bo dokončan slovar kolokacij.

<sup>10</sup> Številko v oklepaju beri: od navedenega skupnega števila jih je toliko z glagolsko sestavino *imeti*.

<sup>11</sup> Vse sestavine so iskane po lemi, razen *kot*, ki je iskan po besedni obliki (ker pač druge paradigmatske oblike veznik nima, homonimni Sam pa se v tej okolici ne pojavlja).

### 3. POMOČ KORPUSA PRI UGOTAVLJANJU DVO– IN TRIDELNOSTI TER DOLOČANJU VARIANTNOSTI PF

V tem delu bi bil dokončani slovar kolokacij najboljša pomoč, toda iskanje pod iztočnico (oz. lemo) *kot*, kjer smo skušali najti morebitne primere vezniškega *kot* (ki je enakopisnica Sam *kót*) smo z iskanjem po predlogih našli 10 PF (pri dvodelnih) oz. njihovih predložnozveznih delih (pri tridelnih), in sicer *kot + iz filma, iz pravljice, iz sanj, iz škafo, iz škatlice, iz topa, na dlani, v pravljici, za stavo, za šalo*. Ta način iskanja smo zaradi skoposti podatkov opustili in se dela lotili z iskanjem po skladijskih vzorcih, ki so se oblikovali pri analizi PFZ (Kržišnik 2014), že pred tem pa so navedeni tudi v Fink-Arsovski (2002: 12–20). V članku o PFZ smo ugotovili, da glagolski PF predstavljajo 40 % in pridevniški PF dodatnih 20 % vseh PF. Na podlagi tega smo sestavili skladijske pogoje, s katerimi smo zveze iskali po korpusih, npr. glagolske *Glag+kot+Sam* (občno ime) = [tag="G.\*"] [word="kot"] [tag="So.\*"];<sup>12</sup>

*Glag+kot+predložni Sam* (občn.) = [tag="G.\*"] [word="kot"] [tag="D.\*"] [tag="So.\*"]; *Glag+kot+deležnik* = [tag="G.\*"] [word="kot"] [tag="Pd.\*"] itd. Shranili in pregledali smo 1000 najpogostejših pojavitev zvez, ki smo jih dobili, ko smo vse konkordance s pomočjo urejanja po frekvencah spravili v (lematizirane) frazeološke in nefrazeološke zveze. Tako smo po skladijskem vzorcu *Glag+kot+Predl+Sam* v GF med 1000 najpogostejšimi poleg že zgoraj iz *Kolokacijskega slovarja* navedenih našli še (navajamo le predložnozvezne dele): *kot + iz svinca, iz rokava, iz puške, na krožniku, na pladnju, na punčico [svojega očesa], na trnih, po jajcih, po loju, po maslu, po olju, v panju, pri norcih, s psom, v filmu, v sanjah, v rogu, v transu, z neba, z živino, za med*. Tudi pri tem načinu iskanja opazimo pomanjkljivosti, npr:

- a) PF se zabeležijo pri več iskalnih pogojih. Pod zvezo *Glag+kot+Sam* se na 1. mestu *kot* najpogostejši recimo pojavi *rasti kot goba*, čeprav gre za PF *rasti kot gobe po dežju* in se zato ponovno pojavi v vzorcu *Glag+kot+Sam+predlSam*. A po drugi strani zaradi tega ne spregledamo dejstva, da se ta PF pojavlja tudi v skrajšani varianti *rasti kot goba / gobe*, v mF na primer 74-krat, kar ni zanemarljivo.
- b) 1000 najpogostejših zvez, najdenih z različnimi skladijskimi pogoji (=vzorci), pomeni, da bo pri različnih skladijskih pogojih upoštevano različno število pojavitev enakih zvez. Tako se npr. pri vzorcu *Glag+kot+Sam* spisek 1000 najpogostejših zvez v GF konča pri 22 pojavitvah zveze, kar pomeni, da zanesljivo nismo zajeli dovolj, sploh pa ne vseh frazemov oz. njihovih variant. Na srečo so po vseh drugih skladijskih pogojih med 1000 najpogostejšimi zvezami zajete pojavitve do 1 ali 2 enakih zvez.
- c) Upošteva se le pogostost levo od *kot* (tj. A-del PF), kar pomeni, da je treba variante iskati na več zaporednih mestih in jih seštevati. Seveda pa je ravno zato mogoče določiti mejo variantnosti, tudi v smislu naštevnosti ali opisnosti.

Nadaljnja obdelava gradiva gre v obrnjeni smeri: od najdenega C-dela h glagolski oz. pridevniški sestavini v A-delu – tako je ne le prihranjeno omenjeno seštevanje, temveč je z

12 V resnici smo najprej iskali po polnopomenskih glagolih, torej [tag="Gg.\*"] (prim. v 4.2), da ne bi mest z navečjo pogostostjo zapolnjevali s pomožnikom pri tvorbi preteklika ali prihodnjika.

ugotavljanjem pogostnosti tudi prikazano število in zaporedje sestavin v A-delu. Na podlagi tega je mogoče dalje ugotavljati, ali gre za dvo- ali tridelni PF, pri tridelnih pa nabor sestavin v A-delu in stopnjo njihove stalnosti.

Preverjanje PF z navedenimi skladenjskimi pogoji bi lahko zapletlo dejstvo, da zlasti pri glagolskih PF zveza med glagolsko in neglagolsko sestavino nikakor ni nujno neposredno stična, kar navedeni skladenjski pogoji predvidevajo.<sup>13</sup> Na srečo se je pokazalo, da večinoma ne velja, kar smo predvidevali v Kržišnik (2014), namreč, »da je praviloma najprej (in običajno to tudi zadošča) treba kot frazno jedro vzeti *kot*+C-del, ki sta neposredno povezana (*kot X*), in z iskanjem »Okolica« pregledati kolokatorje. Tako velja za glagolske in pridevniške kolokatorje, ki so najpogostejši, pa tudi za prislovne, saj so vsi besednoredno (lahko in najpogosteje so) ločeni od *kot*+C-dela (*krasti kot sraka – kradel je / bo kot sraka* itd.)«. Pokazalo se je prav nasprotno, in sicer, da je (vsaj pri glagolskih PF, ki so bili doslej natančneje preverjeni) mogoče z levostičnostjo dobiti večino ali vsaj zadostno količino pojavitev za presojo stalnosti oz. variantnosti. Za primer predstavljamo sliko 22 glagolskih kolokatorjev, ki so rezultat frekvenčnega urejanja 527 konkordanc,<sup>14</sup> dobljenih v mF z iskanjem po pogoju [tag="G.\*"] [word="kot"] [lemma="riba"] [word="v"] [lemma="voda"]:

lemma	Frekvence	Items: 21   Total frequency: 526
p   N počutiti kot riba v voda	354	
p   N znajti kot riba v voda	81	
p   N biti kot riba v voda	37	
p   N plavati kot riba v voda	17	
p   N gubati kot riba v voda	8	
p   N živeti kot riba v voda	7	
p   N videti kot riba v voda	4	
p   N delovati kot riba v voda	3	
p   N prepoznati kot riba v voda	2	
p   N počutiti kot riba v voda	2	
p   N zavezati kot riba v voda	1	
p   N uživati kot riba v voda	1	
p   N splavati kot riba v voda	1	
p   N spadati kot riba v voda	1	
p   N smukati kot riba v voda	1	
p   N pučtiti kot riba v voda	1	
p   N prosperirati kot riba v voda	1	
p   N očutiti kot riba v voda	1	
p   N morati kot riba v voda	1	
p   N izkazati kot riba v voda	1	
p   N izgledati kot riba v voda	1	

Še najpogosteje levostičnost pri PF zaplete ločilo (vejica) pred veznikom *kot*, za katero v slovenščini velja nekoliko »mehanično« ortografsko pravilo, ki zahteva, da vejice pred *kot* ni, če osebna glagolska oblika ni eksplicitno izražena, torej vejice ni v primeru, kot je naveden zgoraj ali npr. pri *bati se česa kot hudič križa*. Zlasti (nikakor pa ne samo) v korpusu družbenih omrežij Janes se take vejice kljub temu pogosto pojavljajo – in seveda pri iskanju predstavljajo eno vmesno mesto.

**3.1.** Za bolj plastično predstavo o razliki med dvo- in tridelnimi PF ter dalje med tridelnimi v A-delu bolj variantnimi in nevariantnimi (ali zelo omejeno variantnimi) objavljamo »slike« frekvenc za vsako od teh kategorij mF, ki je združen korpus: pod (a) dvodelni

13 V tej fazi smo kolokativnost preizkusili tudi z iskanjem kolokacij v korpusu, in sicer 2 mesti levo in desno od t. i. fraznega jedra, tj. nevariantnega dela frazema (termin frazno jedro je pri analizi frazeologije v okviru korpusnega pristopa pri nas uporabljen npr. v Gantar 2007: 37 idr.). Izkazalo se je, da je zlasti pri kvantitativno obsežnem naboru kolokatorjev odčitavanje relevantnih oteženo, ker kolokatorjev ni mogoče urejati po besednih vrstah.

14 Vseh z iskanjem [word="kot"] [lemma="riba"] [word="v"] [word="vodi"] najdenih v mF je 671.

frazem *kot za stavo*, (b) v A-delu variantni frazem v okviru danega semantičnega polja *deževati, liti, uliti se, usuti se, scati, polivati, prati ... kot iz škafa* in pod (c) *kot sraka* – ki je C-del dveh tridelnih PF *krasti kot sraka in drati se, vreščati, vpiti kot sraka* – oba A-dela sta razmeroma stabilna. Pozornosti vredno je tudi razmerje med številom konkordanc in številom enot dane skladske zveze, ki je zapisano nad vsako sliko.

### Glag+*kot za stavo*

mF 2624 : 493

lemma	Frekvence	Items: 493    Total frequency: 2,625
p   N biti kot za stava	411	
p   N zadevati kot za stava	248	
p   N padati kot za stava	211	
p   N zmagovati kot za stava	64	
p   N deževati kot za stava	54	
p   N pokati kot za stava	44	
p   N pihati kot za stava	41	
p   N snežiti kot za stava	40	
p   N zabijati kot za stava	35	
p   N liti kot za stava	31	
p   N tresti kot za stava	28	
p   N prehitovati kot za stava	27	
p   N rasti kot za stava	25	
p   N menjati kot za stava	25	
p   N leteti kot za stava	25	
p   N teči kot za stava	22	
p   N pobirati kot za stava	22	
p   N izgubljati kot za stava	21	
p   N zvoniti kot za stava	20	
p   N vrstiti kot za stava	20	

### Glag+*kot iz škafa*

mF 1520 : 33

lemma	Frekvence	Items: 33    Total frequency: 1,521
p   N liti kot iz škaf	904	
p   N biti kot iz škaf	206	
p   N uliti kot iz škaf	189	
p   N deževati kot iz škaf	84	
p   N viliti kot iz škaf	31	
p   N scati kot iz škaf	25	
p   N padati kot iz škaf	19	
p   N ulivati kot iz škaf	15	
p   N miniti kot iz škaf	7	
p   N zlivati kot iz škaf	6	
p   N priti kot iz škaf	4	
p   N vlivati kot iz škaf	3	
p   N teči kot iz škaf	3	
p   N zalivati kot iz škaf	2	
p   N usuti kot iz škaf	2	
p   N uscati kot iz škaf	2	
p   N nalivati kot iz škaf	2	
p   N dežavati kot iz škaf	2	
p   N zliti kot iz škaf	1	
p   N zgrmeti kot iz škaf	1	



- c) Glag+kot sraka: *krasti kot sraka*  
*dreti se kot sraka*

mF 360 : 20

lemma	Frekvence	Items: 20    Total frequency: 360
krasti kot sraka	251	
biti kot sraka	38	
dreti kot sraka	27	
vreščati kot sraka	9	
kričati kot sraka	8	
vpiti kot sraka	5	
lagati kot sraka	5	
pokrasti kot sraka	3	
peti kot sraka	2	
goljufati kot sraka	2	
čvekati kot sraka	1	
režati kot sraka	1	
rezati kot sraka	1	
razdreti kot sraka	1	
presedati kot sraka	1	
prepisovati kot sraka	1	
okrasti kot sraka	1	
lastniniti kot sraka	1	
jokati kot sraka	1	
hreščati kot sraka	1	

#### 4. PFZ V NOVEJŠIH KORPUSIH

Na gradivu PFZ bomo preverili rezultate v tč. 3 opisanega postopka. Najprej bomo skupine, določene glede na razpon pogostnosti rabe v Kržišnik (2014) primerjali s stanjem v vseh treh novejših korpusih, nato bomo skušali ugotoviti, koliko in katere PFZ dobimo v 1000 najpogostejših zvezah izbranih skladijskih pogojev in vzpostavili interpretativno primerjavo z naborom PFZ, ki smo ga uporabili v Kržišnik (2014).

**4.1.** Petstopenjsko delitev PFZ glede na količino pojavitev v Gigafidi 1.0 in s tem povezano interpretacijo vsake skupine (Kržišnik 2014) smo dopolnili s podatki iz novejših korpusov GF, J in sIW:<sup>15</sup>

- 1) 0 pojavitev (14,9 %) – frazemi niso več v rabi v pisnih tekstih – časovna markiranost: *debel kot polh* (GF 0, J 1, sIW 0), *hud kot sršen* (GF 0, J 0, sIW 0), *siten kot breja mačka* (GF 0, J 0, sIW 0), *potiti se kot miš v kopi* (GF 0, J 0, sIW 0), *peti kot škrjanček* (GF 1, J 3, sIW 0);
- 2) 1–9 pojavitev (30 %) – frazemi se umikajo iz splošne rabe – časovna markiranost ali neknjižnost, zlasti dialektalna označenost PFZ: *smrdeti kot medmašni kozel* (GF 0, J 0, sIW 1), *spolzek kot jegulja* (GF 6, J 0, sIW 3), *pijan kot muha* (GF 1, J 0, sIW 0), *ležati kot krava* (GF 3, J 4, sIW 2), *jesti kot pujs* (GF 3, J 1, sIW 4), *žreti kot prasec* (GF 1, J 1, sIW 0);
- 3) 10–19 pojavitev (17,5 %) – raba je omejena – regionalno, starostno (generacijsko) ali zaradi premočne ekspresije: *len kot fuks* (GF 5, J 2, sIW 10), *premlatiti (koga) kot mačka* (GF 1+pretepsti 6, J 0+nalomiti 1, sIW 0+sinon. *prefukati* 1, *prebutati* 1, *pretepsti* 1), *kot ovce brez pastir-*

<sup>15</sup> Podatki v Gigafidi 1.0 so po njenem obsegu – 1.409.753.568 pojavnic – najbolj neposredno primerljivi s podatki v GF, in sicer v2.0, treba pa je upoštevati, da gre pri prvi za nededuplicirano verzijo, medtem ko je druga deduplicirana.

ja (GF 2, J 1, sIW 3), *trmast kot bik* (GF 4, J 11, sIW 4)<sup>16</sup>, *iti / lepiti se (na koga / kaj) kot muhe na drek* (GF 4, J 3, sIW2);

4) 20 – 60 pojavitev (22 %) – frazemi so v rabi v standardni slovenščini:<sup>17</sup>

*iti / lepiti se (na koga / kaj) kot muhe na med* (GF 21, J 7, sIW 11), *imeti spomin kot zlata ribica* (prim. rezultate v 2.2), *zapreti se kot školjka* (GF 31, J 2, sIW 14)<sup>18</sup>, *rdeč kot kuhan rak* (GF 82, J 16, sIW 41), *reven kot cerkvena miš* (GF 209, J 43, sIW 80), *je koga kot rusov* (GF 39, J 2, sIW 23)<sup>19</sup>;

5) nad 60 pojavitev (14,5 %) – tisti del PFZ, ki so zaradi velike pogostosti v slovenščini zadnjih 20 let zanimivi tudi sociolingvistično:

*počutiti se kot riba v vodi* (GF 233, J 18, sIW 98), *krasti kot sraka* (GF 79, J 102, sIW 62); *lagati, kot pes teče* (GF 47, J 61, sIW 82); *delati (s čim / kom) kot svinja z mehom* (GF 77, J 61, sIW 60);<sup>20</sup> *garati kot črna živina* (GF 102, J 16, sIW 46), *(obnašati se) kot slon v trgovini s porcelanom* (GF 47, J 9, sIW 25), *piti kot žolna* (GF 57, J 22, sIW 30), *kot pes*.<sup>21</sup>

Ugotavljamo, da so podatki v 1. in 5., deloma tudi 4. skupini količinsko primerljivi, medtem ko je za 2. in 3. skupino treba predvidevati premik za eno pogostnostno skupino nižje glede na pogostnost (izjeme med njimi so označena s krepkim tiskom).

**4.2.** Kaj torej za izbiranje (npr. za slovar) PFZ pomeni 1000 najpogostejših zvez po izbranih skladijskih pogojih: koliko PFZ je med njimi in katere so to? Zaradi omejenega prostora, predstavljamo nabor le v dveh skladijskih vzorcih / pogojih: prvega, Glag+kot+Samobč, v katerem je največ PF (predvidoma pa jih tudi največ manjka, prim. v 3 b), in šestega, Glag+kot+Samobč+Samobč, ki jih ima najmanj. Rezultati veljajo za stanje v GF 2.0 dedup, ki je po velikosti (in tudi zvrstno) najbolj primerljiva z Gigafido 1.0:

1. Glag+kot+Samobč [tag="Gg.\*"] [word="kot"] [tag="So.\*"]<sup>22</sup>

*počutiti se / znajti se kot riba [v vodi], boriti se kot lev, držati se kot klop, kot pes, delati / ravnati kot svinja [z mehom], krasti kot sraka, dvigniti / vstati kot feniks, obna-*

16 Med frekvencami je v GF in sIW na prvem mestu *močen kot bik* (GF 45, sIW 11).

17 Termin standardna slovenščina je rabljen namenoma in presega obseg knjižnega jezika predvsem v smeri pogovornosti in govorjenosti.

18 Nedovršniška varianta *zapirati se* je veliko redkejša in ni upoštevana.

19 Kot omenjeno že v Kržišnik (2014) je ta frazem problematičen, ker se zelo uveljavlja nova motivacija, ki se kaže z variantnim pisanjem male oz. velike začetnice, kar ta PF počasi izriva iz skupine PFŽ (prim. *rus* 'ščurku podobna žuželka' in *Rus* 'pripadnik naroda'). V Kržišnik (2014: 252) je navedeno razmerje *rus* : *Rus* 5:36, v novejših korpusih so razmerja naslednja: GF 6 : 33, J 1 : 1, sIW 1 : 13.

20 Upoštevana je le varianta z *delati*, zelo pogosta pa je tudi varianta *ravnati*, enako velja za naslednji frazem, kjer je upoštevana le varianta *garati*, čeprav je *delati* le malo manj pogost; pri *kot slon v ...* je upoštevana samo varianta *obnašati se*, pogosta je tudi varianta s sinonimom *vesti se*, prim. navedene variante v 4.2.

21 Intenzifikator *kot pes* bi zahteval posebno raziskavo – najprej zato, da bi ga kot intenzifikator zares potrdili.

22 Zapis zaporedja frazemov ustreza njihovi pogostnosti (pri glagolskih frazemih to velja za najpogostejšo glagolsko varianto). V oglatih oklepajih je naveden tisti del frazema, ki ga najdemo tudi s katerim drugim skladijskim pogojem (prim. 3 a). Zunajfrazemska določila zaradi omejenega prostora opuščamo.

*šati se kot slon [v trgovini s porcelanom], skrivati kot kača [noge], cepati kot muhe, boriti se kot levinja, piti kot žolna, hoditi kot mačka [okrog vrele kaše], molčati kot riba, zapreti se kot školjka, garati kot živina, lepiti se kot muhe [na med], plavati kot riba, šopiriti se kot pav, zebe kot psa k., peti kot slavček.*

6. `Glag+kot+Samobč+Samobč [tag="Gg.*"][word="kot"][tag="So.*"][tag="So.*"]`  
*skrivati kot kača noge, vstati kot pišč feniks, držati kot pes ježa.*

Navedimo še PFZ, ki smo jih dobili, a v Kržišnik (2014) zaradi ne vključenosti v slovar ali pre nizke pogostnosti v korpusu niso bile upoštevane: *(za)rjuti / zavpiti kot ranjena zver, izgledati kot oskubljena kura / kokoš, držati kot pes ježa, igrati / poigravati se kot mačka z mišjo, vleče se kot zgodba o steklem polžu.*

Tako smo z 9 skladijskimi vzorci / pogoji (ki niso vsi skladijski vzorci glagolskih PF) dobili 50 PFZ. V Kržišnik (2014) je glagolskih PFZ 40 % ok. 250, in če ob tem upoštevamo še, da je med vsemi PFZ 15 % takih, ki se v sodobnih besedilih ne rabijo več, torej sodijo v 1. skupino (gl. 4.1), lahko rečemo, da smo na splošno med 1000 najpogostejšimi glagolskimi zvezami dobili nekako 50–60 % glagolskih PFZ, rabljenih v sodobnih besedilih. To bi pomenilo, da 1000 najpogostejših zvez v vsakem skladijskem pogoju ni zadostna količina, da bi z njo pokrili nabor PF za sodobni slovar slovenščine – to vsekakor velja za 1. skladijski pogoj (vzorec `Glag+kot+Sam`). Drugo vprašanje pa je, ali lahko iz PFZ delamo posplošitev na vse PF, čisto mogoče namreč je, da količina rabe PFZ upada.

## 5. SKLEP

Graditi slovar samo na podatkih iz korpusov verjetno ni dovolj, je pa neizogibno. Kar pa se metode dela s korpusnimi podatki na tukaj opisani način tiče, zagotovo velja, da je z ustrežno (do določene mere avtomatizirano) metodologijo znotraj korpusnega jezikoslovja to mnogo boljše in lažje izvedljivo, kakor z delno korpusno metodo, v kateri korpus predstavlja le besedilno gradivo. Ko bo slovar slovenskih kolokacij dokončan, bomo lahko primerjali, koliko so s to raziskavo dobljeni rezultati relevantni.

## LITERATURA

- Fink-Arsovski, Željka, *Poredbena frazeologija: pogled izvana i iznutra*, FF press, Zagreb, 2002.
- Gantar, Polona, *Stalne besedne zveze v slovenščini*, Založba ZRC, ZRC SAZU, Ljubljana, 2007.
- Kosem, Iztok (ur.), *Kolokacije v slovenščini*, Znanstvena založba Filozofske fakultete Univerze v Ljubljani, Ljubljana, 2021., <https://e-knjige.ff.uni-lj.si/znanstvena-zalozba>
- Kržišnik, Erika, „Zoonimi v slovenskih frazeoloških primerah“, v: *Životinje u frazeološko ruhu*, Ivana Vidović Bolt (ur.), FF-press, Zagreb, 2014., str. 243–262, <https://openbooks.ffzg.unizg.hr/index.php/FFpress/catalog/book/115>

## SPLETNI VIRI

- eSSKJ = *Slovar slovenskega knjižnega jezika*, 2016–, <https://fran.si> (Dostop: nov. 2021.).
- Gigafida 1.0 = Slovenski referenčni korpus, 1990 – 2011, <https://www.clarin.si/noske> (Dostop: nov., dec. 2021.).
- Gigafida 2.0 = Slovenski referenčni korpus Gigafida v2.0, 1990-2018, brez bližnjih duplikatov, <https://www.clarin.si/noske> (Dostop: nov., dec. 2021.).
- Janes = Korpus besedil s slovenskih družbenih omrežij Janes v1.0, <https://www.clarin.si/noske> (Dostop: nov., dec. 2021.).
- Kolokacijski slovar sodobne slovenščine*, <https://viri.cjvt.si/kolokacije/slv> (Dostop: nov., dec. 2021.).
- metaFida = Slovenski združeni korpus v0.1, <https://www.clarin.si/noske> (Dostop: nov., dec. 2021.).
- slWaC = Korpus slovenskega spleta slWaC v2.1, 2014, <https://www.clarin.si/noske> (Dostop: nov., dec. 2021.).
- SSF = Keber, Janez, *Slovar slovenskih frazemov*, 2015, spletna izdaja <https://www.fran.si> .
- SSKJ<sup>2</sup> = *Slovar slovenskega knjižnega jezika*, 2014, spletna izdaja [www.https://fran.si](http://www.https://fran.si) .

## SIMILES IN USE AND IN DICTIONARIES

The aim of the contribution is a continuation of the article *Zoonyms in similes* (Kržišnik 2014), but with a reversed hierarchy: we will pay particular attention to similes in general, but the corpus data on the use of zoonyms in similes will serve as a comparative starting point for further checks of usage in more recent Slovenian text corpora that differ from the point of view of stylistic variety, i.e., Gigafida v2.0 (1990–2018), slWaC v2.1 (2014) and Janes. The eSSKJ (e-Dictionary of Standard Slovenian) is a growing dictionary (the compilation started in 2016) “with a completely new comprehensive description of the lexis of standard Slovenian from the second half of the 20<sup>th</sup> century to the present day” ([www.fran.si](http://www.fran.si)); in its manual of style, it lists 100 hits in corpus sources that differ with regard to stylistic variety as a threshold for the inclusion of a (single-word) lexeme into the dictionary, while for phraseological units, the lowest possible threshold is 5 hits. Our aim was to explore what this means for similes.

**Key words:** text corpus; phraseology; dictionary; similes