

Slovenščina 2.0, 2 (2018)

FORMULAIČNOST V SLOVENSKEM JEZIKU

Kaja DOBROVOLJC

Center za jezikovne vire in tehnologije Univerze v Ljubljani, 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 govorne slovenščine, jih kvantitativno in kvalitativno analiziramo ter primerjamo. Velik delež identificiranih formulaičnih nizov v obeh korpusih dokazuje določeno stopnjo formulaičnosti, še zlasti v govorni 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 govorne slovenščine, ki potrjujejo vpliv specifičnih sporazumevalnih okoliščin in ciljev na razvoj in obseg formulaičnega jezika.

Ključne besede: formulaičnost, besedni nizi, večbesedne enote, pisni jezik, govorni 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 morfolologija 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 novičarskih 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 govornjene 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 govorniki 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 pojavitve v korpusu Kres in 20 pojavitve 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

⁵ Pri 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 slovničnem 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 specifike 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 slovničnimi 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 (Holožan 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, skladijsko 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

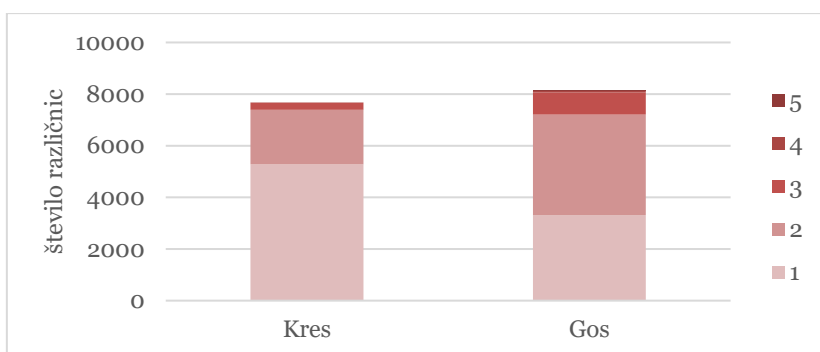
3.1 Delež formulaičnih besednih nizov

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).

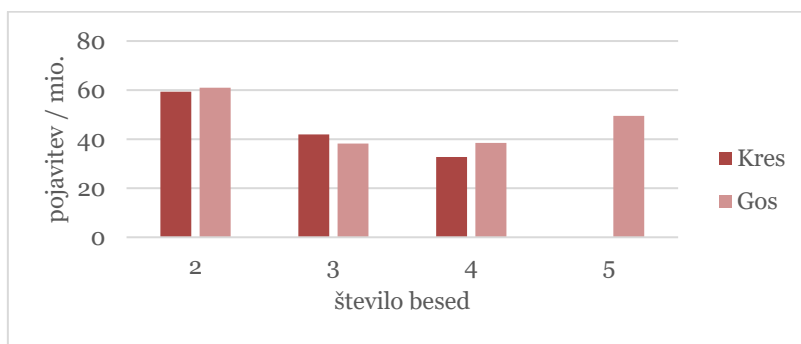


Slika 2: Število različnic na prilagojenem frekvenčnem seznamu korpusov Kres in Gos glede na število besed.

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.

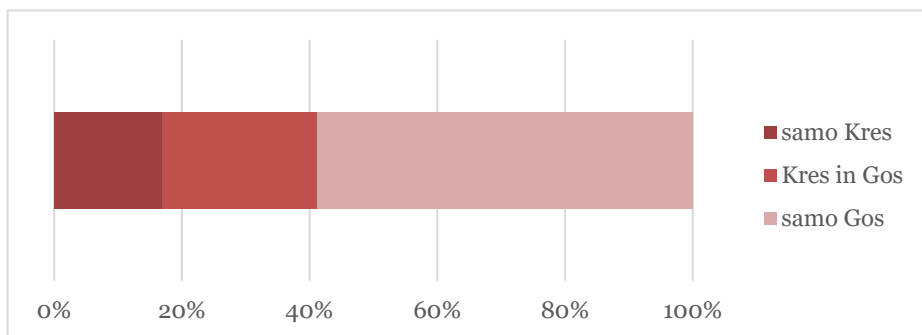
Kot kaže primerjava povprečne pogostosti nizov v vsakem izmed korpusov (Slika 2), se formulaični besedni nizi v povprečju pojavljajo s precej visoko in podobno pogostostjo (povprečno 57 pojavitev na milijon pojavnic v korpusu Kres oz. 56 pojavitev v korpusu Gos), tudi če primerjamo povprečno pogostost 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.



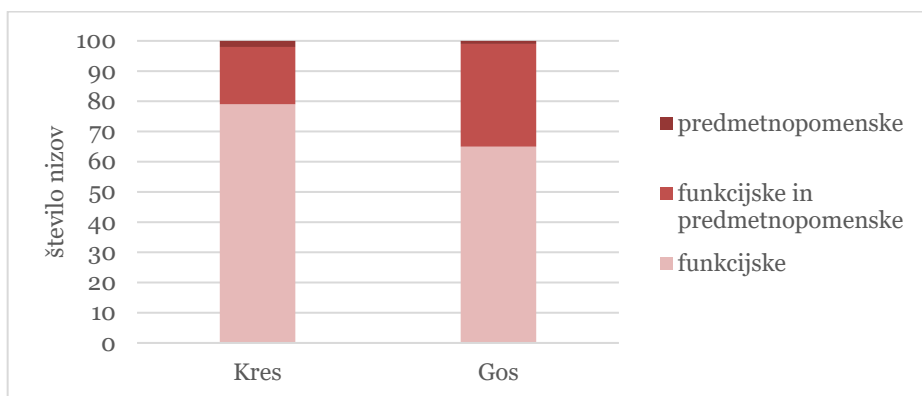
Slika 3: Prekrivnost formulaičnih nizov v korpusih Kres in Gos.

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 najpogosteje 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).

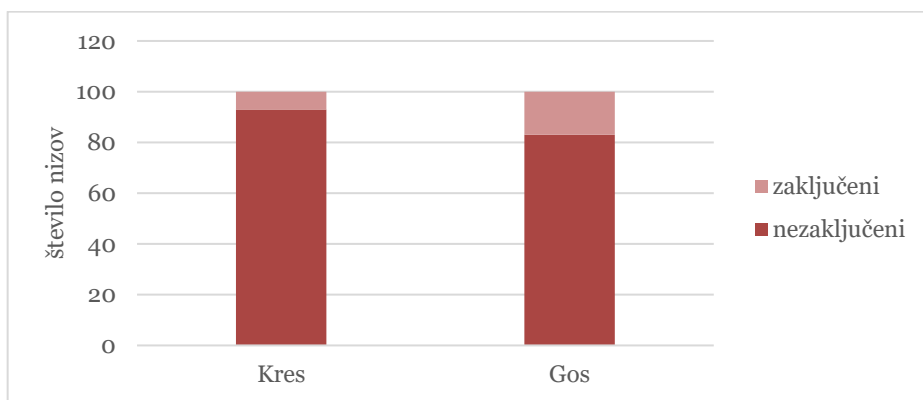


Slika 4: Vrsta besed v formulaičnih nizih pisne in govorjene slovenščine.

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. diskurzivnih 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 govorne slovenščine.

Velik delež strukturno nezaključenih formulaičnih nizov sovpada 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 nerazstavljeni 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 pomenskovozljivostnih 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>

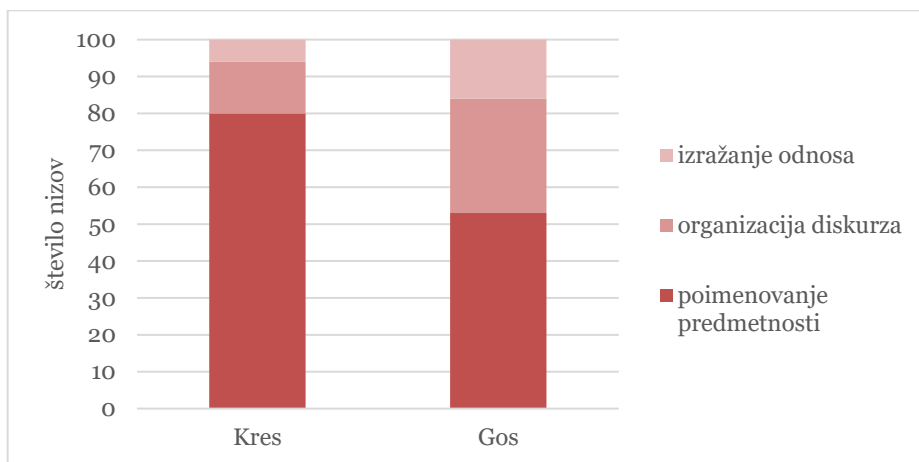
Tabela 3: Tri najpogostejše skladenjske strukture formulaičnih nizov govornjene slovenščine (G = glagol, V = veznik, Z = zaimek).

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 govornjeni 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 specifične 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 govornikovih 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 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 najpogosteje ponavljajočih se besednih nizov v referenčnih korpusih pisne in govorne 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 govorno 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. *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 govorni slovenščini izstopajo predvsem deiktični nizi za poročanje oz. pripovedovanje (npr. *to je bilo, to je pa, in to je*). Prav tako v govorni 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 zasnovu 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 govorjenega jezika v uporabniških spletnih vsebinah: primer večbesednih diskurzivnih 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 slovnice: 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 sporočanjejski frazemi. *Jezik in slovstvo*, 46(5): 191–200.
- Tremblay, A., B. Derwing, G. Libbern in C. Westbury. (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	<i>da bi se</i>	26	<i>ki so ga</i>	51	<i>da se ne</i>	76	<i>v tem času</i>
2	<i>ki ga je</i>	27	<i>ki je bila</i>	52	<i>na drugi strani</i>	77	<i>je v tem</i>
3	<i>ki se je</i>	28	<i>se mu je</i>	53	<i>kot da bi</i>	78	<i>mislím da je</i>
4	<i>da se je</i>	29	<i>da je v</i>	54	<i>da gre za</i>	79	<i>kar pomeni da</i>
5	<i>ki jih je</i>	30	<i>je da se</i>	55	<i>uradni list rs št.</i>	80	<i>o tem da</i>
6	<i>ki jo je</i>	31	<i>da je bila</i>	56	<i>da je bilo</i>	81	<i>ne glede na to</i>
7	<i>pa se je</i>	32	<i>v zvezi s</i>	57	<i>je bila v</i>	82	<i>glede na to da</i>
8	<i>ko se je</i>	33	<i>ne bi bilo</i>	58	<i>da bi lahko</i>	83	<i>ki mu je</i>
9	<i>se je v</i>	34	<i>v tem primeru</i>	59	<i>pa je tudi</i>	84	<i>so se v</i>
10	<i>ki so se</i>	35	<i>se je na</i>	60	<i>v tem da</i>	85	<i>da bi bil</i>
11	<i>ki je bil</i>	36	<i>člena tega zakona</i>	61	<i>medtem ko je</i>	86	<i>iz prejšnjega odstavka</i>
12	<i>je da je</i>	37	<i>za to da</i>	62	<i>pa je da</i>	87	<i>v kateri je</i>
13	<i>v skladu z</i>	38	<i>ki so jo</i>	63	<i>se ji je</i>	88	<i>naj bi bil</i>
14	<i>ki so jih</i>	39	<i>naj bi se</i>	64	<i>pa so se</i>	89	<i>zato ker je</i>
15	<i>da se bo</i>	40	<i>pa je bil</i>	65	<i>to pomeni da</i>	90	<i>ko je bil</i>
16	<i>ki naj bi</i>	41	<i>da ne bi</i>	66	<i>pa naj bi</i>	91	<i>se je da</i>
17	<i>ki je v</i>	42	<i>ne glede na</i>	67	<i>da bi ga</i>	92	<i>to je bilo</i>
18	<i>da je bil</i>	43	<i>je bil v</i>	68	<i>pa je bila</i>	93	<i>v katerem je</i>
19	<i>ne da bi</i>	44	<i>da so se</i>	69	<i>se mi zdi</i>	94	<i>na to da</i>
20	<i>da je to</i>	45	<i>tako da je</i>	70	<i>to da je</i>	95	<i>naj bi bila</i>
21	<i>se mi je</i>	46	<i>tako da se</i>	71	<i>pa je bilo</i>	96	<i>na ta način</i>
22	<i>v skladu s</i>	47	<i>je bilo v</i>	72	<i>da bi bilo</i>	97	<i>da ne bo</i>
23	<i>in s tem</i>	48	<i>ki so v</i>	73	<i>s tem da</i>	98	<i>in da se</i>
24	<i>v zvezi z</i>	49	<i>ki so bili</i>	74	<i>je še vedno</i>	99	<i>v primerjavi z</i>
25	<i>d. o. o.</i>	50	<i>in da je</i>	75	<i>če bi se</i>	100	<i>pa je v</i>

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/>

