

Univerza v Ljubljani
Filozofska fakulteta

Kaja Dobrovoljc

Leksikalne prvine govornega jezika v uporabniških spletnih vsebinah:

Primer večbesednih diskurzivnih označevalcev

Doktorska disertacija

Mentorica: izr. prof. dr. Darinka Verdonik

Somentor: red. prof. dr. Marko Stabej

Študijski program: Humanistika in družboslovje

Področje: Slovenistika

Ljubljana, 2018

ZAHVALA

Ob nastanku tega dela se iskreno zahvaljujem mentorici dr. Darinki Verdonik za pomoč pri zasnovi raziskave, svobodo pri raziskovalnem delu ter vse dragocene komentarje in nasvete, somentorju dr. Marku Stabeju za spodbudo pri ohranjanju kritične drže do navidez samoumevnega, dr. Iztoku Kosmu za administrativno podporo, dr. Simonu Kreku za vse priložnosti in znanje, dr. Tomažu Erjavcu za pomoč pri računalniški infrastrukturi, Katedri za računalniško jezikoslovje Univerze v Uppsali za diskusije, zgled in motivacijo, akcijama evropskega programa COST TextLink in Parseme za možnost dodatnih izobraževanj in izmenjave mnenj ter Javni agenciji za raziskovalno dejavnost Republike Slovenije in Škrabčevi ustanovi za finančno podporo. Še posebej pa bi se rada zahvalila staršem za spodbudo in pomoč skozi celotno izobraževalno pot, Nejcu za neprecenljivo prilagodljivost, potrpežljivost in podporo ter Arnetu za nežen opomnik, da življenje teče dalje.

KAZALO VSEBINE

1	UVOD	9
2	TEORETIČNA IZHODIŠČA	13
2.1	DISKURZNI OZNAČEVALCI	13
2.1.1	<i>Razdrobljenost področja</i>	<i>13</i>
2.1.2	<i>Raziskave v tujem jezikoslovju</i>	<i>15</i>
2.1.2.1	D. Schiffrin in interakcijsko-sociolingvistični pristop	15
2.1.2.2	G. Redeker in model vzporednih komponent	17
2.1.2.3	B. Fraser in slovnično-pragmatični pristop	18
2.1.2.4	K. Fischer in konstrukcijska slovnica	20
2.1.2.5	D. Blakemore in teorija relevantnosti	21
2.1.2.6	Teorije koherentnih povezav	24
2.1.3	<i>Raziskave v slovenskem prostoru</i>	<i>28</i>
2.1.3.1	Konektorji	28
2.1.3.2	Besedilnopovezovalni členki	30
2.1.3.3	Pragmatična analiza diskurza	33
2.1.4	<i>Vpliv aplikativnih pristopov k analizi diskurza</i>	<i>36</i>
2.2	VEČBESEDNI DISKURZNI OZNAČEVALCI	42
2.2.1	<i>Raziskave večbesednih enot v analizi diskurza</i>	<i>42</i>
2.2.1.1	Večbesedne enote v pragmatični analizi diskurza	42
2.2.1.2	Večbesedne enote v besediloslovju	44
2.2.1.3	Raziskave formulaičnega jezika	47
2.2.2	<i>Prepoznavne lastnosti večbesednih diskurznihi označevalci</i>	<i>50</i>
2.2.2.1	Nepropozicijskost	50
2.2.2.2	Diskurzna funkcija	55
2.2.2.3	Večfunkcijskost	66
2.2.2.4	Skladenjska vloga	69
2.2.2.5	Večbesednost	70
2.2.2.6	Ustaljenost	72
2.2.3	<i>Izhodišča raziskave večbesednih diskurznihi označevalci</i>	<i>75</i>
2.2.3.1	Sinteza pregleda literature	75
2.2.3.2	Opredelevanje večbesednih diskurznihi označevalci	79
2.2.3.3	Merila za prepoznavo večbesednih diskurznihi označevalci	81
2.3	DISKURZNI OZNAČEVALCI V UPORABNIŠKIH SPLETNIH VSEBINAH	101

2.3.1	<i>Uporabniške spletne vsebine</i>	101
2.3.2	<i>Značilnosti uporabniških spletnih vsebin</i>	104
2.3.3	<i>Raziskave rabe diskurznihi označevalcev v uporabniških spletnih vsebinah</i>	108
3	METODOLOGIJA	112
3.1	RAZISKOVALNA VPRAŠANJA	112
3.2	GRADIVO.....	113
3.3	STRUKTURA RAZISKAVE.....	116
4	VEČBESEDNI DISKURZNI OZNAČEVALCI GOVORJENE SLOVENŠČINE	118
4.1	DOPOLNITEV SCHEME ZA OZNAČEVANJE VEČBESEDNIHI DISKURZNIHI OZNAČEVALCEV	118
4.1.1	<i>Priprava gradiva</i>	118
4.1.1.1	Vzorčenje korpusa Gos.....	118
4.1.1.2	Uvoz v orodje WebAnno	120
4.1.1.3	Zamejitev nabora prepoznanihi označevalcev	121
4.1.2	<i>Težavnejša mesta kategorizacije</i>	122
4.1.2.1	Funkcijsko problematični izrazi	123
4.1.2.2	Strukturno problematični izrazi	130
4.1.3	<i>Robustna izpeljava funkcij</i>	136
4.1.3.1	Problematičnost funkcijskega razdvoumljanja v kontekstu.....	136
4.1.3.2	Minimalistična funkcijska taksonomija	138
4.2	METODE PREPOZNAVANJA VEČBESEDNIHI DISKURZNIHI OZNAČEVALCEV V BESEDILNIHI KORPUSIHI.....	148
4.2.1	<i>Pregled metod za prepoznavanje večbesednih diskurznihi označevalcev</i>	148
4.2.1.1	Kratek pregled metod za prepoznavo diskurznihi označevalcev v besedilnihi korpusihi	148
4.2.1.2	Kratek pregled metod za luščenje večbesednih enot iz besedilnihi korpusov	150
4.2.1.3	Izbrana metoda pridobivanja seznama večbesednih diskurznihi označevalcev	154
4.2.2	<i>Priprava seznamov potencialnih večbesednih diskurznihi označevalcev</i>	156
4.2.2.1	Luščenje in običajno štetje besednihi nizov.....	156
4.2.2.2	Prilagojeno štetje besednihi nizov	158
4.2.2.3	Mere kolokabilnosti.....	162
4.2.2.4	Prekrivnost seznamov najvišje uvrščenih kandidatov	169
4.2.3	<i>Primerjava metod prepoznavanja večbesednih diskurznihi označevalcev</i>	171
4.2.3.1	Primerjava statističnihi metod	171
4.2.3.2	Primerjava statističnihi metod in ročne prepoznave	178
4.3	SEZNAM VEČBESEDNIHI DISKURZNIHI OZNAČEVALCEV GOVORJENE SLOVENŠČINE	183
4.4	TIPičNOST ZA GOVORJENI JEZIK	189
4.4.1	<i>Izračun pogostosti diskurzne rabe</i>	189

4.4.1.1	Kratek pregled metod za razdvoumljanje diskurznihi označevalcev	190
4.4.1.2	Metoda sklepanja po vzorcu	191
4.4.2	<i>Pogostost rabe v govornem in pisanem slovenščini</i>	194
4.4.2.1	Pogostost rabe v korpusu Gos	194
4.4.2.2	Pogostost rabe v korpusu Kres	197
4.4.3	<i>Tipično govornem večbesedni diskurzni označevalci</i>	200
4.4.3.1	Pregled metod za izračun ključnosti	200
4.4.3.2	Seznam tipično govornem večbesednihi diskurznihi označevalcev	202
5	VEČBESEDNI DISKURZNI OZNAČEVALCI GOVORJENEGA JEZIKA V UPORABNIŠKIH SPLETNIH VSEBINAH	206
5.1	POGOSTOST RABE V UPORABNIŠKIH SPLETNIH VSEBINAH	207
5.2	RABA GLEDE NA TIP UPORABNIŠKIH SPLETNIH VSEBIN	212
5.3	ZNAČILNOSTI RABE V UPORABNIŠKIH SPLETNIH VSEBINAH	218
5.3.1	<i>Raba označevalca a ne</i>	218
5.3.1.1	<i>Raba označevalca a ne</i> v korpusu Gos	218
5.3.1.2	<i>Raba označevalca a ne</i> v korpusu Janes	219
5.3.2	<i>Raba označevalca a veš</i>	221
5.3.2.1	<i>Raba označevalca a veš</i> v korpusu Gos	221
5.3.2.2	<i>Raba označevalca a veš</i> v korpusu Janes	223
5.3.3	<i>Raba označevalca ne vem</i>	226
5.3.3.1	<i>Raba označevalca ne vem</i> v korpusu Gos	226
5.3.3.2	<i>Raba označevalca ne vem</i> v korpusu Janes	227
6	RAZPRAVA	230
7	ZAKLJUČEK IN NADALJNJE RAZISKAVE	235
	POVZETEK	238
	SUMMARY	241
	BIBLIOGRAFIJA	244
	PRILOGE	276

KAZALO TABEL

Tabela 1: Prilagojena funkcijska taksonomija diskurznihi označevalcev po shemi MDMA.	83
Tabela 2: Primerjava tehnoloških dejavnikov računalniško posredovane komunikacije po tipologiji Herring (2007) za tri tipe uporabniških spletnih vsebin.	107
Tabela 3: Velikost in sestava vzorčenega korpusa govorne slovenščine.	120
Tabela 4: Število večbesednih nizov z minimalno relativno pogostostjo 5/milj. in vsaj dvema različnima govorcema v korpusu Gos.	157
Tabela 5: Primerjava štetja pojavnic in različnic za n-grame v korpusu Gos pred in po implementaciji algoritma za izdelavo prilagojenega frekvenčnega seznama.	161
Tabela 6: Načini povprečenja statističnih podatkov prve in druge besede za nize dolžine n besed.	163
Tabela 7: Seznam uporabljenih mer povezovalnosti in njihova generalizacija za večbesedne enote poljubne dolžine (dolžine n besed).	166
Tabela 8: Primerjava 20 najvišje uvrščenih besednih nizov glede na izračun besedne povezovalnosti z merami MI, MI ³ , χ^2 , LL, Dice in t-test.	168
Tabela 9: Delež prekrivnih kandidatov na seznamih 2.000 najvišje uvrščenih večbesednih enot za vsakega izmed različnih statističnih izračunov.	170
Tabela 10: Seznam 79 večbesednih diskurznihi označevalcev, ki so bili prepoznani na podlagi analize kolokacij in se v vzorčenem korpusu govorne slovenščine ne pojavljajo.	180
Tabela 11: Seznam večbesednih diskurznihi označevalcev, ki so bili prepoznani na podlagi ročne analize vzorčenega korpusa in se ne pojavljajo na seznamu 2.000 najvišje uvrščenih enot na nobenem izmed osmih načinov statističnega razvrščanja besednih nizov.	181
Tabela 12: Seznam prepoznanihi večbesednih diskurznihi označevalcev govorne slovenščine.	188
Tabela 13: 30 najbolj tipičnih diskurznihi označevalcev v korpusu Gos glede na korpus Kres na podlagi izračuna ključnosti s pomočjo logaritma verjetnosti (LL).	205
Tabela 14: Pogostost rabe treh najbolj tipično govornih diskurznihi označevalcev v korpusu Janes.	209

KAZALO SLIK

Slika 1: Umestitev uporabniških spletnih vsebin znotraj raziskav računalniško posredovane komunikacije.	103
Slika 2: Prikaz označene izjave v orodju WebAnno.....	121
Slika 3: Primerjava števila različnic za besedne nize z minimalno relativno pogostostjo 5 pojavitev na milijon pojavnici v korpusih Gos in Kres.	158
Slika 4: Število diskurznihi označevalcev med 2.000 najvišje uvrščenimi kandidati izbrane statistične mere.	172
Slika 5: Primerjava natančnosti prepoznavanja večbesednih diskurznihi označevalcev v korpusu Gos za različne statistične mere.	173
Slika 6: Primerjava števila diskurznihi označevalcev na seznamu vseh 1.000 najvišje rangiranihi večbesednih enot za različne skupine označevalcev in različne statistične mere razvrščanja	176
Slika 7: Delež diskurznihi označevalcev, pridobljenih z metodo ročnega pregleda reprezentativnega vzorca korpusa in/ali analize statistično najbolj izstopajočih kolokacij.....	179
Slika 8: Večbesedni diskurzni označevalci glede na funkcijsko skupino.	184
Slika 9: Večbesedni diskurzni označevalci glede na dolžino.	184
Slika 10: Pogostost rabe večbesednih diskurznihi označevalcev v korpusu Gos glede na funkcijsko skupino.	195
Slika 11: Stopnja dvoumnosti večbesednih diskurznihi označevalcev govorjene slovenščine s prikazom števila označevalcev glede na delež diskurznihi rab med vsemi pojavitvami niza v korpusu Gos.	196
Slika 12: Stopnja dvoumnosti večbesednih diskurznihi označevalcev govorjene slovenščine po posameznih funkcijskih skupinah	197
Slika 13: Primerjava pogostosti rabe večbesednih diskurznihi označevalcev v korpusih Gos in Kres.	198
Slika 14: Primerjava pogostosti rabe večbesednih diskurznihi označevalcev v korpusih Gos in Kres za posamezne funkcijske skupine	199

KAZALO PRILOG

Priloga 1: Seznam večbesednih diskurznih označevalcev v vzorčenem korpusu Gos	276
Priloga 2: Seznam 100 najpogostejših besednih nizov v korpusu Gos glede na običajno metodo štetja	280
Priloga 3: Seznam 100 najpogostejših besednih nizov v korpusu Gos glede na prilagojeno metodo štetja.....	283
Priloga 4: Seznam statistično prepoznanih večbesednih diskurznih označevalcev v korpusu Gos.....	287
Priloga 5: Seznam večbesednih diskurznih označevalcev govornjene slovenščine	295
Priloga 6: Vzorčeni primeri rabe označevalca <i>a ne</i> v korpusu Gos.....	301
Priloga 7: Vzorčeni primeri rabe označevalca <i>a veš</i> v korpusu Gos	305
Priloga 8: Vzorčeni primeri rabe označevalca <i>ne vem</i> v korpusu Gos	308
Priloga 10: Vzorčeni primeri rabe označevalca <i>a ne</i> v korpusu Janes	310
Priloga 11: Vzorčeni primeri rabe označevalca <i>a veš</i> korpusu Janes	320
Priloga 12: Vzorčeni primeri rabe označevalca <i>ne vem</i> v korpusu Janes	327

1 Uvod

Pričujoča doktorska disertacija sega na številna različna jezikoslovna področja, od frazeologije do kontrastivne pragmatike, od teoretskih preizpraševanj modeliranja jezika do metodoloških vprašanj korpusnega jezikoslovja, vendarle pa jo lahko glede na temeljne raziskovalne cilje umestimo predvsem na krovno področje **analize diskurza**. Kot eno izmed temeljnih značilnosti tega raziskovalnega področja, ki ga sicer sestavlja izredno širok in raznolik nabor jezikoslovnih disciplin, lahko prepoznamo predvsem usmerjenost v raziskovanje jezikovne rabe nad ravno stavka. Če (poenostavljeno) fonologija raziskuje načine združevanja zvokov v pomenonosne morfeme, morfologija združevanje morfemov v besede, skladnja pa združevanje besed v besedne zveze, stavke in povedi, se analiza diskurza ukvarja zlasti z vprašanjem, kako jezikovni uporabniki stavke, povedi, izjave ali kakorkoli drugače opredeljene sporočilno zaključene enote povezujejo z njihovim kontekstom. Pri tem se različni pristopi razlikujejo predvsem v tem, ali se osredotočajo zlasti na povezovanje s sobesedilom (diskurz kot jezikovni produkt) ali tudi drugimi okoliščinami sporazumevalne situacije, kot so prostorske in časovne okoliščine, udeleženci, odnosi med njimi, njihovi sporazumevalni cilji in podobno (diskurz kot sporazumevalni proces), pri čemer ta razlika pogosto sovпада tudi s tem, ali se raziskovalci osredotočajo predvsem na analizo pisne (diskurz kot besedilo) ali govorne (diskurz kot govorni dogodek) jezikovne rabe.

Ne glede na konceptualna razhajanja v samem razumevanju diskurza pa lahko kot skupno značilnost diskurznoanalitičnih raziskav prepoznamo to, da načine njegovega strukturiranja oziroma upravljanja pogosto preučujejo skozi analizo rabe posebnih jezikovnih sredstev, s katerimi govorci povezavo med vsebino in tako ali drugače opredeljenim kontekstom vzpostavljajo tudi na izrazni ravni. Za tovrstna jezikovna sredstva, katerih pomena oziroma funkcije v jeziku tradicionalne slovnične teorije zaradi svoje osredotočenosti na stavčno in nižje skladijske ravni niso uspele zadovoljivo pojasniti, se sicer uporabljajo različna poimenovanja, med njimi tudi izraz **diskurzni označevalci** (angl. *discourse markers*). Čeprav od prve omembe tega izraza minevajo že skoraj tri desetletja, je še danes eno ključnih vprašanj, povezanih z raziskovanjem diskurzni označevalcev, prav njihova jasna opredelitev in razmejitev, saj skladno s težko opredeljivim konceptom diskurza različne teorije v ospredje postavljajo različne in pogosto le deloma prekrivne lastnosti izrazov, ki naj bi ta diskurz označevali ali nanj kazali. Za

razliko od kopice raziskav diskurzni označevalcev v prvih dveh desetletjih, ki so se osredotočale predvsem na pojasnjevanje njihove vloge v komunikaciji glede na različne teoretske modele, primerjavo njihove rabe v različnih sporazumevalnih okoliščinah ali jezikih ter pojasnjevanje njihovega diahronega razvoja, v zadnjem desetletju raziskave teh jezikovnih sredstev doživljajo nov preporod, saj teoretska, metodološka in aplikativna spoznanja drugih jezikoslovnih disciplin spodbujajo razvoj nekaterih novih raziskovalnih podpodročij, ki smo jim skušali slediti tudi v zasnovi naše raziskave in jih na kratko predstavimo v nadaljevanju.

Najizrazitejši je nedvomno premik od ohlapnih teoretskih opredelitev te jezikoslovne kategorije k vse večji **objektivizaciji in sistematizaciji meril** za prepoznavo izrazov, ki naj bi se vanjo umeščali. S porastom metod korpusnega jezikoslovja se je namreč izkazalo, da prvotnih opredelitev diskurzni označevalcev, ki so pogosto temeljile na abstraktnih naborih njihovih funkcij, idiosinkratičnih modelih diskurza in analizi peščice prototipičnih izrazov, ni mogoče dosledno prenašati na obsežnejše vzorce avtentične jezikovne rabe, kar onemogoča tako napredek na aplikativnem področju, kot so denimo modeliranje jezika za potrebe jezikovnih tehnologij ali leksikološke raziskave za potrebe jezikovnega poučevanja, kot na teoretskem področju, saj ugotovitev dosedanjih raziskav zaradi ohlapnosti izhodišč ni bilo mogoče preverjati, medsebojno primerjati ali empirično nadgrajevati. Ta teoretsko-metodološki premik od teoretsko k empirično utemeljenim jezikoslovnim opisom, ki sicer zaznamuje jezikoslovno znanost nasploh, je še toliko bolj pomemben prav na področju diskurznofunkcijskega besedišča, za katerega je značilna izrazita večfunkcijskost, tj. dejstvo, da lahko posamezni izraz v jeziku opravlja več različnih funkcij, hkrati pa lahko posamezno funkcijo opravlja več različnih in raznolikih izrazov. V primerjavi z jezikoslovnim opisom predmetnopomenskega besedišča, kot so glagoli, samostalniki in pridevniki, kakršnokoli zamejevanje in členjenje te in sorodnih nepropozicijskih kategorij zato predstavlja še toliko bolj zahteven kategorizacijski izziv.

Čeprav se diskurzni označevalci torej vzpostavljajo kot izrazito funkcijska kategorija, pa se v njihovih opredelitvah posredno ali neposredno običajno omenjajo tudi sekundarna izrazna merila. S porastom raziskav formulaičnega jezika, ki dokazujejo, da je jezik prepreden z večbesednimi vzorci, ki vsaj na neki točki jezikovne rabe delujejo kot nerazstavljiva celota in se kot taki tudi shranjujejo v mentalni leksikon govorcev, se kot drugo pomembno novo področje raziskav diskurzni označevalcev kaže tudi premik fokusa z enobesednih na **večbesedne označevalce**. To

je še zlasti pomembno v raziskavah govorjene rabe, ki je zaradi specifičnih okoliščin tvorjenja dokazano bistveno bolj formulaična kot pisni jezik, najpogostejše besedne zveze v govoru pa večinoma opravljajo prav vlogo označevalcev diskurzni in pragmatičnih vidikov sporazumevanja. Tudi s strukturnega vidika tako prihaja do vse večje relativizacije konceptov, kot so ustaljenost, gramatikaliziranost in propozicijska izpraznjenost, s kakršnimi so prvotne raziskave diskurzni označevalcev utemeljevale svojo osredotočenost na peščico najbolj prototipičnih (običajno enobesednih) izrazov, saj je zaradi težke opredeljivosti teh lastnosti težko postavljati ločnico z drugimi tipi besednih zvez, ki v jeziku opravljajo povsem enake vloge, še posebej, če so rabljene nenavadno pogosto. Tako kot so se na področju frazeologije raziskave pomensko najbolj idiomatičnih in ekspresivnih večbesednih enot vedno bolj odpirale proti drugim tipom stalnih besednih zvez, se tudi na področju analize diskurza kaže vse večja potreba po celostnejših raziskavah večbesedne diskurznofunkcijske leksike.

Poleg teoretskih, metodoloških in strukturnih premikov pa se v zadnjih nekaj letih v analizi diskurza odpira tudi povsem novo področje zvrstnostnih raziskav, raziskave t. i. **uporabniških spletnih vsebin** na družbenih omrežjih, forumih, blogih in drugih participativnih spletnih platformah, ki so se vzpostavile s spletom druge generacije. Tako kot za računalniško posredovano komunikacijo nasploh, tudi za uporabniško spletno komunikacijo namreč velja, da zaradi posebnih okoliščin, v katerih običajno nastaja, selektivno vsebuje tako prvine pisnega kot govorjenega jezika, ki se odražajo na različnih jezikovnih ravneh. Pri tem so se dosedanje domače in tuje raziskave prehajanja tipično govorjenih leksikalnih prvin v spletno komunikacijo osredotočale predvsem na vprašanja nestandardnega zapisa, besedišča in skladnje, presenetljivo malo pa je raziskav spletne rabe (tipično govorjenih) diskurzni označevalcev, čeprav sorodne širše raziskave spletnega diskurza izpostavljajo, da govorci v spletni komunikaciji razvijajo zelo specifične strategije upravljanja diskurza, od organizacije tem do menjave vlog in uravnavanja medosebnih odnosov. Vprašanje, v kolikšni meri si pri tem pomagajo z rabo eksplicitnih jezikovnih sredstev, pa doslej ne v slovenskem ne v tujem jezikoslovju še ni bilo globlje raziskano.

Kljub temu da osrednji cilj pričujočega dela – raziskava rabe večbesednih diskurzni označevalcev govorjene slovenščine v uporabniških spletnih vsebinah – narekuje navidezno zelo kompleksen in partikularen preplet raziskovalnih področij, bomo v naši raziskavi rabe

večbesednih diskurznihi označevalci govorne slovenščine v uporabniških spletnih vsebinah skušali sistematično raziskati prav vse tri zgoraj izpostavljene vidike: ponuditi teoretsko in empirično utemeljeno *opredelitev* kategorije diskurznihi označevalci in njihovih prepoznavnih lastnosti, določiti čim bolj izčrpen nabor *večbesednih* diskurznihi označevalci govorne slovenščine in raziskati njihovo rabo v *uporabniških spletnih vsebinah*. Pri tem izhajamo iz treh **delovnih hipotez**:

- (i) da se v govoru tipično uporabljajo večbesedni diskurzni označevalci, povezani s specifičnimi dejavniki tvorjenja govornega diskurza, kot sta prisotnost naslovnika in tvorjenje v realnem času;
- (ii) da se ti zaradi svoje specifične vloge v govorni rabi v uporabniških spletnih vsebinah ne rabijo;
- (iii) oziroma se rabijo predvsem v tistih tipih uporabniških spletnih vsebin, ki izkazujejo podobne sporočanske okoliščine kot govorni diskurz.

Čeprav so vsi trije raziskovalni cilji, še zlasti pa vprašanje večbesednosti in rabe v uporabniških spletnih vsebinah, relevantni in aktualni tudi znotraj širšega jezikoslovnega prostora, naša raziskava predstavlja še toliko pomembnejši prispevek k razvoju slovenistike, kjer se kljub številnim kakovostnim znanstvenim razpravam s področja širše analize diskurza prav opis diskurzne in druge nepropozicijske leksike kaže kot eden najbolj perečih izzivov prihodnosti, zlasti za potrebe slovarskih in drugih aplikativno usmerjenih opisov slovenskega jezika. Naša ambicija torej ni preizpraševanje dosedanjih sorodnih slovenističnih raziskav, ki vsaka na svoj način pomembno prispeva k razumevanju tega kompleksnega področja, temveč njihova nadgradnja z izbiro drugačnega metodološkega pristopa, osredotočenostjo na drugačen tip izrazov in analizo drugačne jezikovne zvrsti, kakršna lahko poleg novih teoretskih spoznanj ponudi tudi in predvsem nove jezikovne vire za nadaljnje sorodne raziskave.

2 Teoretična izhodišča

2.1 Diskurzni označevalci

2.1.1 Razdrobljenost področja

Vse odkar je D. Schiffrin v svoji monografiji z naslovom *Discourse markers* (Schiffrin, 1987) prva empirično načela vprašanje vloge funkcijske leksike v govorjenem diskurzu, zanj ponudila nov teoretski model, ki je presegal tradicionalno besediloslovno usmerjenost v diskurz kot produkt besedilne skladnje in označevalce tako razumljenega diskurza ponazorila z oblikoslovno, skladijsko, funkcijsko in žanrsko raznolikim naborom angleških izrazov *oh, well, and, but, or, because, so, now, then, y'know* in *I mean*, je v devetdesetih letih na širšem področju analize diskurza prišlo do skokovitega porasta znanstvene produkcije, povezane s področjem diskurznihi označevalcev, ki ga je B. Fraser v eni svojih razprav slikovito poimenoval kot »rastoča industrija jezikoslovja« (B. Fraser, 1999, str. 932). Toda kljub dejstvu, da je vsem raziskovalcem v izhodišču skupno preučevanje jezikovnih sredstev, ki se pomensko udejanjajo na ravni diskurza, torej onkraj meja stavkov ali drugače opredeljenih strukturnoskladijskih enot, ti k razumevanju samega koncepta diskurza in z njim povezanihi jezikovnih sredstev pristopajo z izredno **raznolikihi teoretskihi in metodološkihi izhodišči**, ki so med seboj pogosto le delno prekrivna. Njihova pogostost se tako zdi premosorazmerna s težavnostjo njihovega jezikoslovnega opisa (Taboada, 2006, str. 568).

Kot v uvodu v eno temeljnih monografij, posvečenihi razumevanju raznolikosti pristopov k preučevanju teh jezikovnih sredstev, ugotavlja K. Fischer (2006c, str. 1), se danes zdi skoraj nemogoče prebiti se skozi številne raziskave, posvečene diskurznim označevalcem, njihov pregled pa ne omogoča sintetičnega izluščenija jedrne teorije pomenov in funkcij teh jezikovnih izrazov, saj so obstoječe raziskave zaradi raznolikosti obravnavanihi jezikov, izrazov, funkcijskih zamejitev, raziskovalnih vprašanj in metodoloških izbir med seboj le težka primerljive. Zato še danes težko govorimo o ustaljenem pojmovanju diskurznihi označevalcev, saj med raziskovalci »ni konsenza, ali gre za skladijsko ali pragmatično kategorijo, katere vrste izrazov sploh spadajo v to kategorijo, kakšno je razmerje med diskurznimi označevalci in drugimi sorodnimi kategorijami, kot so povezovalci, medmeti, modalni členki ali na govorca usmerjeni prislovi, ter tudi ne glede termina 'diskurzni označevalec' v primerjavi z alternativnimi pojmovanji, kot so

'diskurzni povezovalci', 'pragmatični povezovalci' ali 'pragmatični členek'« (Lewis, 2011, str. 419–420). Ni torej presenetljivo, da so bili odtlej vprašanju težavne opredelitve diskurznih označevalcev posvečeni številni pregledni članki (Schourup, 1999; Dér, 2010; Fischer, 2014), zborniki (Jucker in Ziv, 1998; Fischer, 2006a; Degand idr., 2013a), posebna številka revije (Degand in Simon-Vandenbergen, 2011), poljudnoznanstveni pregledi (Schiffrin, 2001; Blakemore, 2008; Fischer, 2013; Maschler in Schiffrin, 2015) in znanstvena srečanja,¹ če naštejemo le nekaj najodmevnejših.

Če ob strani pustimo vprašanje terminološke zmede na področju diskurznofunkcijskega besedišča, ki jo bomo načenjali ob podrobnejših predstavitev posameznih raziskav v nadaljevanju, lahko razdrobljenost področja raziskav diskurznih označevalcev pripišemo predvsem trem dejavnikom. Prvi je nedvomno dejstvo, da so jezikovna sredstva, ki se jim običajno pripisuje status tako ali drugače opredeljenih diskurznih označevalcev, izrazito **večfunkcijska**, saj lahko ob svoji izpraznjenosti propozicijskega oz. referenčnega pomena v različnih sobesedilnih ali sporazumevalnih kontekstih opravljajo različne vloge. Pri tem se raziskovalci strinjajo predvsem glede tega, da teh vlog ni mogoče pojasniti s tradicionalnimi, skladenjsko usmerjenimi slovničnimi teorijami, manj pa glede tega, kako dejansko pristopiti k njihovemu opisu in razmejevanju glede na druge jezikoslovne kategorije. Diskurzni označevalci se tako pogosto opredeljujejo z negacijo tradicionalnih jezikoslovnih pojmov, kot so propozicijska vsebina, referenčni pomen, skladenjska vpetost in podobno, zaradi česar se ta kategorija pogosto obravnava kot jezikoslovni »koš za odpadke« (Lewis, 2006, str. 44). Poleg večfunkcijskosti in teoretske heterogenosti pa tretja posebnost izvira iz dejstva, da je področje raziskav diskurznih označevalcev za razliko od starejših besediloslovnih teorij, ki so se povečini osredotočale zgolj na pisni diskurz, že od vsega začetka v svoje pojmovanje diskurza in z njim povezanih teorij kot enakovredno ali nemara celo superiorno obliko sporazumevanja vključevalo tudi **govorno interakcijo**. Ker v jezikoslovju doslej še ni bila ponujena teorija diskurza, ki bi celostno pojasnjevala kompleksnost govorne in pisne komunikacije hkrati, se tako tudi

¹ Denimo simpozij *Pragmatic Markers, Discourse Markers and Modal Particles: What do we know and where do we go from here?* leta 2014, simpozij *Identification and annotation of discourse relations in spoken language* (DiSpOL) leta 2015 in konferenca *Discourse Relational Devices* (LPTS) leta 2016.

raziskave diskurznih označevalcev – z izjemo nekaterih uveljavljenih teorij, kot so teorija relevantnosti,² teorija argumentacije ali teorije koherentnih povezav – pogosto utemeljujejo z razmeroma idiosinkratičnimi teoretskimi modeli diskurza, ki skušajo pojasniti načine prepletanja jezikovne rabe in njenega sporazumevalnega konteksta. Kot ugotavlja L. Schourup (1999, str. 228), so prav diskurzni označevalci tako postali nekakšen preizkusni poligon za preverjanje hipotez o razmejevanju pragmatike in semantike ter teorij, ki se ukvarjajo s pojasnjevanjem strukture diskurza in interpretacije izjav.³

Ob vsesplošnem konsenzu o teoretski razdrobljenosti področja diskurznih označevalcev, ki v primerjavi z nekaterimi drugimi jezikoslovnimi disciplinami ne omogoča razmejevanja na jasno prepoznavne teoretske okvire, pregled raziskav v nadaljevanju zato namesto s sintetičnim pregledom najprej začnemo s predstavitvijo najvplivnejših tujih in domačih raziskav ter njihovih teoretskih izhodišč, šele v drugem razdelku pričujočega preglednega poglavja pa posamezne vidike dosedanjih opredelitev (večbesednih) diskurznih označevalcev sintetično primerjamo ob predstavitvi njihovih posameznih prepoznavnih lastnosti. Prav širok nabor teh lastnosti in številni različni načini njihovega kombiniranja namreč nazorneje pojasnjujejo izkazano razdrobljenost področja, ki jo skušamo v svoji končni opredelitvi v sklepnem delu drugega razdelka tudi preseči.

2.1.2 Raziskave v tujem jezikoslovju

2.1.2.1 D. Schiffrin in interakcijsko-sociolingvistični pristop

Čeprav sta izraz diskurzni označevalec (angl. *discourse marker*) v svoji razpravi o posebni vlogi angleškega izraza *well* ('no') kot kazalca na predhodno znano vsebino v okviru analiz psihoterapevtskega diskurza prva omenila že W. Labov in D. Fanshel (1977), za **temeljno delo** na področju raziskav diskurznih označevalcev govorjenega jezika velja monografija D. Schiffrin (1987), v kateri avtorica analizira diskurzne označevalce v korpusu angleških sociolingvističnih

² V slovenističnem prostoru se uporablja tudi izraz teorija relevance (Vogel, 2004; Schlamberger Brezar, 2009), redkeje relevantnostna teorija (Verdonik, 2011).

³ Skozi celotno disertacijo bomo kot prevod angleškega izraza »utterance« uporabljali izraz *izjava*, bolj ali manj sinonimno pa se je v slovenskem prostoru vzpostavil tudi prevod *izrek*. Za pregled različnih opredelitev obeh pojmov, tudi v primerjavi s konceptoma povedi in govornega dejanja, glej npr. (Jakop, 2006, str. 118–120).

intervjujev z ameriškimi judi. V svoji raziskavi diskurzne označevalce v želji po operativnosti namerno ohlapno opredeli kot »zaporedno odvisne elemente, ki razmejujejo enote govora« (Schiffrin, 1987, str. 31), pri čemer so enote govora v različnih kontekstih pojmovane kot povedi, propozicije, govorna dejanja ali intonacijske enote, razmejevanje kot rezultat organizacije družbenih aktivnosti, zaporedna odvisnost pa kot lastnost, ki signalizira njihovo anaforično ali kataforično vpetost v sobesedilni oz. diskurzni kontekst (Schiffrin, 1987, str. 31–40).

Izrazi, ki jih nato podrobneje analizira (*oh, well, and, but, or, because, so, now, then, y'know in I mean*), se še danes pogosto navajajo kot prototipični predstavniki te kategorije, po drugi strani pa je ravno vprašanje njihove homogenosti v jedru teoretskih razhajanj in razdrobljenosti njihovega nadaljnjega raziskovanja. Poleg prvega obsežnega opisa in vzpostavitve termina je ta raziskava na nadaljnji razvoj področja pomembno vplivala tudi zato, ker jo avtorica podpre z **novi teorijo strukture diskurza**, ki področje analize diskurza v nasprotju s predhodnimi raziskavami razširi na spontano govorjeno interakcijo, saj temelji na predpostavki, da diskurz ni zgolj enota jezika, temveč tudi proces socialne interakcije, diskurzna koherenca pa način, kako govorec in naslovnik skupaj integrirata oblike, pomene in dejanja, da osmislita izrečeno (Schiffrin, 1987, str. 49).

Na podlagi tega razvije tezo, da se koherenca diskurza ustvarja skozi razmerje med soležnimi (sosednjimi) enotami diskurza (Schiffrin, 1987, str. 24), a tako diskurza kot njegove koherence ne moremo opazovati kot homogene celote, temveč predlaga delitev diskurza na **pet ločenih ravnin govora** (angl. *planes of talk*),⁴ ki ima vsaka svojo vrsto koherence (Schiffrin, 1987, str. 24–30): predstavno strukturo (angl. *ideational structure*), ki predstavlja organizacijo idej oz. propozicijske vsebine diskurza; strukturo dejanj (*action structure*), ki predstavlja organizacijo govornih dejanj; strukturo izmenjave (*exchange structure*), ki predstavlja strukturo menjavanja vlog med govorcem in naslovnikom; okvir sodelovanja (*participation framework*), ki predstavlja ravnino uravnavanja odnosov med govorcem in naslovnikom ter govorcem in vsebino; ter ravnino informativnosti (*information state*), ki predstavlja védenje govorca in naslovnika ter védenje, ki ga pripisujeta drug drugemu. Na podlagi tovrstnega pojmovanja diskurza nato razvije tezo, da diskurzni označevalci delujejo kot označevalci tega, kako se diskurzna enota, v kateri se

⁴ Prevodi imen posameznih ravnin so prevzeti po Verdonik (2007).

pojaviijo, povezuje z eno ali več ravninami diskurza. Diskurzni označevalci torej delujejo kot **kontekstualne koordinate**, znotraj katerih je izjava tvorjena in znotraj katerih naj bi bil interpretirana, ker pa običajno hkrati kažejo na več ravnin hkrati, s svojo integrativnostjo prispevajo k celostni koherenci diskurza (Schiffrin, 1987, str. 315, 326–330). Pri njihovem pomenskem opisu avtorica zagovarja tezo, da lahko kljub temu posameznemu označevalcu vedno pripišemo prevladujočo oz. primarno ravnino, npr. ravnina informativnosti za označevalca *oh* in *y'know*, okvir sodelovanja za označevalca *well* ali *I mean* ali predstavna ravnina za označevalce *and*, *but*, *or*, *so*, *because* (Schiffrin, 1987, str. 315–316).

2.1.2.2 G. Redeker in model vzporednih komponent

Čeprav D. Schiffrin s svojo analizo izbrane množice izrazov skuša ponazoriti, da predlagani model diskurza z umeščanjem v posamezne ravnine omogoča sistematičen pomenski opis posameznih označevalcev in njihovih konkretnih rab, poznejši raziskovalci izpostavljajo, da teh meril ne upošteva dosledno in ni prenosljiv na širši nabor izrazov. Med njimi denimo G. Redeker (1991) v svoji recenziji monografije D. Schiffrin izpostavi potrebo po jasnejši in doslednejši opredelitvi temeljnih teoretskih izhodišč, kakršna bi omogočala obravnavo širšega nabora kot zgolj peščice arbitrarnih izrazov ter ki bi hkrati pojasnjevala tudi implicitna koherenčna razmerja v diskurzu in koherenčna razmerja, ki jih vzpostavljajo drugi tipi izrazov in ne zgolj tisti, ki jih D. Schiffrin imenuje diskurzni označevalci. Ker meni, da so nekatere izmed petih ravnin, ki jih omenja D. Schiffrin, med seboj neločljivo povezane, težko razmejljive ali h koherenci diskurza ne prispevajo na enak način, na podlagi svojih predhodnih razprav (Redeker, 1990) predlaga **nov, robustnejši model koherence v diskurzu**, ki temelji na predpostavki, da naslovnik v diskurzu vsako izjavo vedno interpretira s treh vidikov: njenega prispevka k vsebini diskurza, njenega prispevka k namenu diskurza in njene umestitve glede na diskurzno celoto. Glede na to G. Redeker predlaga delitev na tri vrste strukture diskurza oz. tri temeljne tipe funkcijskih povezav med enotami diskurza (Redeker, 1991, str. 1667–1668, 2006, str. 354): **predstavno** (vzpostavljanje *semantičnih povezav* med zunajbesedilnimi dogodki, npr. časovno sosledje, vzrok, posledica, tj. povezovanje propozicij), **retorično** (vzpostavljanje *pragmatičnih povezav* med ilokucijskimi namerami, npr. sklep, pojasnilo, utemeljitev, tj. povezovanje ilokucij) in **sekvenčno** (vzpostavljanje *prehodov* med temami ali vmesnimi digresijami, komentarji, popravki itd., tj. povezovanje diskurznihih segmentov).

Sama diskurzne označevalce (ki jih v izogib izključno pragmatičnojezikoslovnim interpretacijam, kakršne so zaznamovale predhodne raziskave, raje imenuje **diskurzni operatorji**, angl. *discourse operators*) opredeli kot izraze, katerih primarna funkcija je naslovnika opozoriti na posebno vrsto **povezave med diskurzno enoto, ki jo uvaja, in njegovim neposrednim diskurznim kontekstom** (Redeker, 1991, str. 1968, 2006, str. 341), pri čemer kot minimalno diskurzno enoto na podlagi koncepta 'enote ideje' (Chafe, 1980) razume prozodično in strukturno zaključeno idejno enoto, ki najpogosteje sovpada s stavkom, diskurzni kontekst pa kot nadpomenko znotraj- in zunajbesedilnega konteksta. S tovrstno opredelitvijo iz kategorije diskurznih označevalcev izključi izraze, ki ne delujejo na ravni celotne izjave oz. enote ideje, temveč se nanašajo samo na enega izmed njenih sestavnih delov (npr. izrazi, ki označujejo spremembo ubeseditve ali samopopravek sredi izjave), po drugi strani pa vanjo vključuje vse anaforične rabe prislovnih določil, kadar te vzpostavljajo povezavo s predhodnim besedilom (npr. *medtem*, *takrat* itd.). Svoj model v poznejših razpravah poimenuje **model vzporednih komponent** (Redeker, 2000, 2006) in ga poveže s konceptom treh metafunkcij sistemsko-funkcijske slovnice (Halliday, 1985) – vsak diskurzni označevalec vedno hkrati deluje na vseh treh ravneh, predstavni (ideacijska oz. predstavna metafunkcija), retorični (medosebna metafunkcija) in sekvenčni (besedilna metafunkcija), a ena vedno prevladuje.

2.1.2.3 B. Fraser in slovnično-pragmatični pristop

Precej bolj se od pristopa D. Schiffrin k razumevanju strukture diskurza in vloge diskurznih označevalcev oddalji B. Fraser (B. Fraser, 1990, 1999, 2006, 2009), ki diskurzne označevalce opazuje in opredeljuje predvsem z vidika njihove vloge pri vzpostavljanju povezav med neposredno povezanimi enotami besedila. Diskurzne označevalce znotraj pristopa, ki ga sam imenuje slovnično-pragmatični pristop, obravnava kot podvrsto širše kategorije pragmatičnih označevalcev (B. Fraser, 1996), tj. izrazno in funkcijsko raznolikih jezikovnih izrazov, ki jim je skupno predvsem to, da se običajno pojavljajo na začetku stavkov in niso del propozicijske besedilne vsebine (govorčevega poročanja o stanju o svetu), temveč v besedilu delujejo kot **jezikovni kazalci govorčevih potencialnih komunikacijskih namer** (B. Fraser, 1996, str. 167). Pragmatične označevalce v prvotnih razpravah razdeli v štiri kategorije: označevalce ilokucijske sile (npr. *Obžalujem, da ...*), označevalce komentiranja vsebine (npr. *na žalost, iskreno povedano ...*), označevalce paralelnih sporočil (npr. *gospod, zdaj, ok*) in diskurzne označevalce, ki

signalizirajo odnos med pomenom diskurznega segmenta, ki ga uvajajo, in predhodnega diskurznega segmenta (B. Fraser, 1999). Lastnost, ki po njegovem diskurzne označevalce loči od drugih pragmatičnih označevalcev, je dejstvo, da ne prispevajo k predstavitvenemu pomenu besedila (angl. *representational meaning*), temveč imajo predvsem **proceduralni pomen**, saj signalizirajo semantično razmerje osnovnega sporočila (angl. *basic message*) do predhodnega besedila in s tem kot dvomestni operatorji določajo, kako naj bo besedilna enota, v kateri se pojavljajo, interpretirana glede na predhodno besedilno enoto (B. Fraser, 1999).

V poznejši razpravi (B. Fraser, 2009) predlagano delitev pragmatičnih označevalcev korenito spremeni tako, da iz nje izključi kategorijo označevalcev paralelnih sporočil in doda novo kategorijo 'označevalcev strukture diskurza', ki signalizirajo govorčevo upravljanje diskurza (*če povzamem; dodajam, da*), spreminjanje teme (*če se vrnemo nazaj*) in pridobivanje pozornosti (*glejte*). Kategorijo diskurzni označevalcev tako v primerjavi z D. Schiffrin in G. Redeker pojmuje precej ožje, saj mednje prišteva samo tiste izraze, ki nastopajo v vlogi povezovalcev **dveh zaporednih besedilnih segmentov (stavčno izraženih ilokucijskih dejanj)**, pri čemer mora biti povezava, ki jo vzpostavljajo, pomenska, ne pa slovnična ali diskurzna (B. Fraser, 2009, str. 7).⁵ Brez posebnih dodatnih teoretičnih utemeljitev označevalce funkcijsko nato razdeli na označevalce kontrasta, npr. *but*, elaboracije, npr. *and*, in inferiranja, npr. *so* (za več primerov glej (B. Fraser, 1999, 2009)), zgolj v eni izmed vmesnih razprav (B. Fraser, 2006) pa omenja še četrto kategorijo označevalcev časovnih razmerij (npr. *then, when*), zato se zdi, da njegova teorija (še) ni dosledno izpeljana.

Fraser v primerjavi s predhodnimi razpravami iz kategorije diskurzni označevalcev tako izloča tipično govorjene izraze z zunajbesedilno vlogo, kot so *oh, y'know* in *I mean*, neleksikalne izraze,

⁵ Za ponazoritev slovničnih povezav navaja izraze, kot so *that, in order that, so as, for ...*, za ponazoritev diskurzni pa izraze, kot so *first, later, incidentally, oh that reminds me ...* Pri tem tudi ni povsem jasno, kako B. Fraser razume razliko med pomenskimi in slovničnimi povezovalci. Na podlagi ponazoritev slovničnih povezovalcev, npr. *in order that, so as, for* (B. Fraser, 2009, str. 7), bi lahko sklepali, da s tem izrazom označuje rabo veznikov kot semantičnih oz. predstavni medpropozicijskih povezovalcev, toda med ponazoritvami v drugih razpravah kot primere diskurzni označevalcev navaja tudi medstavčne veznike, npr. *Harry will not go, unless he is paid an appearance fee* ipd. (B. Fraser, 1999), in tudi sam poudari, da v nasprotju s predhodnimi objavami (B. Fraser, 1990), v katerih je kot diskurzne označevalce obravnaval samo izraze, ki uvajajo samostojne povedi, ta lastnost ne bi smela biti izključujoči pogoj.

kot so zapolnjeni premori (mašila), ter povezovalce daljših besedilnih sekvenc ali tematskega sosledja, ki jih v primerjavi z G. Redeker in D. Schiffrin umešča v širšo kategorijo pragmatičnih označevalcev. Kljub dejstvu, da B. Fraser svojih predlogov taksonomij ne utemeljuje s podrobnejšo teorijo diskurza ali analizami avtentičnega jezikovnega gradiva, v svoji opredelitvi diskurzni označevalcev pa nedosledno združuje posamezne koncepte različnih teoretskih smeri, npr. pragmatičnega jezikoslovja (delovanje na ravni izrekanja), stavčne skladnje (pomen začetnega položaja v izjavi), besediloslovja (razmerja med soležnimi deli besedila) in teorije relevantnosti (proceduralni pomen), pa so njegove razprave še danes med najpogostejše citiranimi, kot eden izmed vplivnih možnih pristopov k opisu diskurzni označevalcev pa se bolj ali manj kritično omenjajo tudi v pregledni literaturi področja (npr. Blakemore, 2008; Schiffrin, 2001).

2.1.2.4 K. Fischer in konstrukcijska slovnica

S povsem drugačnega stališča k opredelitvi diskurzni označevalcev pristopa K. Fischer (2000, 2006b), saj njen pristop k zamejitvi in pomenskemu opisu diskurzni označevalcev temelji na hkratnem združevanju pomenskih in izraznih lastnosti, kakršno izvira iz teorije konstrukcijske slovnice (Kay in Fillmore, 1999; Croft, 2001; Goldberg, 2006). Ta k proučevanju jezika pristopa z vidika jezikovnih konstrukcij, shematskih parov oblike in pomena, ki v jeziku delujejo kot primitivni elementi skladišne reprezentacije, znotraj katerih lahko pojasnujemo posamezne slovnične pojave (Croft, 2000, str. 84). Fischer na podlagi teh izhodišč zagovarja stališče, da tudi leksemi, ki običajno nastopajo v vlogi diskurzni označevalcev, sami po sebi nimajo določenih funkcij, ki naj bi jih opravili v interakciji, kakor tudi nimajo določene besedne vrste, ki ji pripadajo. Ta **pomen se ustvari šele v določeni konstrukciji**, ko se pomenske komponente leksema povežejo s pomenskimi komponentami konstrukcije, pri čemer pa z vidika spoznanj konverzacijske analize **kot enakovredno komponento te analize vključuje tudi širši sporazumevalni kontekst**, ki združuje nabor vseh opravil, ki jih govorci opravljajo v komunikaciji (Fischer, 2006b, 2010).⁶

⁶ Avtorica za poimenovanje tega situacijskega okvira sinonimno uporablja izraza *communicative background frame* (Fischer, 2006b) oz. *communicative situation frame* (Fischer, 2010).

Kot strukturne parametre opisa diskurznihi označevalcev izpostavlja predvsem obliko, položaj izraza v izjavi in položaj v vlogi ter intonacijski potek, in se zato v svojih raziskavah osredotoča zlasti na **kratke, nepregibne, skladenjsko, semantično in pogosto tudi prozodično neintegrirane besede, ki razkrivajo mentalne procese govorca, povezane s samim procesom sporazumevanja ali s sporazumevalno situacijo** (Fischer, 2006b, str. 431, 443), torej izraze, ki se tipično uporabljajo predvsem v govorni interakciji (pastavčni členki ali medmeti, zapolnjeni premori (mašila), označevalci segmentacije). Na podlagi tega sama namesto termina diskurzni označevalec, ki naj bi se pretežno vzpostavljala kot funkcijsko opredeljena kategorija, v katero (lahko) spada izrazno zelo raznolik in težko natančno opredeljiv nabor izrazov, raje uporablja izraz **diskurzni členek** (angl. *discourse particle*), ki jasneje sugerira osredotočenost na njihovo formalno-funcijsko opredeljeno podmnožico (Fischer, 2006c, str. 4–5).

Kot bomo videli v nadaljevanju, tovrstna zamejitev diskurznihi označevalcev ni prisotna le med drugimi raziskovalci, ki se pri kategorizaciji in pomenskem opisu pragmatične leksike naslanjajo na konstrukcijsko slovnico (Fried in Östman, 2005; Diwald, 2006, 2013; Nemo, 2006; Fischer in Alm, 2013), temveč je pogosta tudi pri drugih pragmatičnojezikoslovnih raziskavah govorne interakcije, ki pa ne sledijo nujno tudi enakemu terminološkemu ločevanju diskurznihi členkov od označevalcev. Kljub siceršnji osredotočenosti na raziskave diskurznihi členkov tudi avtorica v novejših razpravah (Fischer, 2014) poudarja pomen ohranjanja **nadkategorije diskurznihi označevalcev**, saj (i) je ta notranje preveč heterogena, da bi jo bilo možno na podlagi izraznih in pomenskih meril zanesljivo ločevati na jasno razmejljive podmnožice, (ii) številni diskurzni označevalci delujejo na več različnih ravneh diskurza, zato osredotočenost na eno samo onemogoča njihovo celostno razumevanje, tudi z vidika njihovega nastanka (dihronnega razvoja), obenem pa (iii) dosleden opis njihove večfuncijskosti ni nemogoč, če upoštevamo različne dimenzije njihove rabe (Fischer, 2014, str. 271–272).

2.1.2.5 D. Blakemore in teorija relevantnosti

Povsem samosvoj pristop je v polje raziskav diskurznihi označevalcev vpeljala D. Blakemore (Blakemore, 1987, 2002), ki diskurzne označevalce analizira v okviru širše **teorije relevantnosti** (Sperber in Wilson, 1986). Jedrna predpostavka te teorije je, da vsako komunikacijsko dejanje spremlja jamstvo, da je v danem kontekstu **optimalno relevantno tako s stališča govorca kot**

naslovnika: v procesu sporazumevanja tvorec skuša svoje sporočilo oblikovati tako, da bo najbolj ustrezno oz. relevantno v danih okoliščinah, prav tako pa tudi naslovnik predpostavlja, da ima izjava zanj največjo možno ustreznost, jo torej že vnaprej jemlje kot optimalno ustrezno in jo zato sam postavi v optimalen kontekst, ki zagotavlja njeno najustreznejšo interpretacijo.

Na podlagi te predpostavke nato D. Blakemore (1987) razvije **ново teorijo pomena**, ki namesto delitve na semantični in pragmatični pomen uvede razliko med dvema tipoma pomena glede na to, kako jezikovna sredstva prispevajo k mentalnim procesom, povezanim z interpretacijo izjav. Medtem ko se **konceptualni pomen** nanaša na koncepte oz. enote konceptualne predstave, se **proceduralni pomen** nanaša na usmerjanje procesa pragmatičnega razumevanja. Če se torej teorija konceptualnega pomena ukvarja s tem, kako se elementi jezikovne strukture preslikavajo na koncepte, tj. na gradnike propozicijskih reprezentacij, ki so podvrženi različnim mentalnim procesom (angl. *computations*), se po drugi strani teorija proceduralnega pomena ukvarja z načinom, kako se elementi jezikovne strukture preslikavajo neposredno na te mentalne procese (Blakemore, 1987, str. 144). Znotraj teorije relevantnosti razlika med proceduralnim in konceptualnim pomenom velja za temeljno ločnico v semantiki, zato imajo vsi izrazi v jeziku bodisi konceptualni pomen bodisi proceduralni pomen.⁷ Čeprav na začetku ločevanje na konceptualno in proceduralno sovпада z ločevanjem na vplivanje in nevpilivanje na resničnostno vrednost, kakršna se običajno uporablja pri razmejevanju med semantiko in pragmatiko oz. razmejevanju med semantičnim in pragmatičnim pomenom (Gazdar 1979), poznejše razprave (D. Wilson in Sperber, 1993; Blakemore, 2002) izpostavijo, da ne gre za popolnoma simetrična razmerja, saj po eni strani nekateri izrazi, ki vplivajo na resničnostno vrednost, ne zaznamujejo

⁷ Čeprav se je izraz *proceduralni pomen* vzpostavil s teorijo relevantnosti, se na področju raziskav diskurznih označevalcev pogosto uporablja tudi z drugačnim pojmovanjem. Tako si denimo B. Fraser izraz sicer izposodi iz teorije relevantnosti, a nasprotuje temeljni predpostavki, da ima lahko vsak izraz v jeziku bodisi konceptualni bodisi proceduralni pomen (B. Fraser, 2009, str. 15–16), saj imajo po njegovem povezovalci, kot so *in contrast*, *as a result*, *after all*, *as consequence* poleg konceptualnega vendarle tudi proceduralni pomen, ki je lahko povsem prekriven s pomenom nekaterih označevalcev, ki jih D. Blakemore umešča med prave diskurzne označevalce, ki naj bi imeli samo proceduralni pomen (npr. *so*). Podobno tudi D. Blakemore (2008, str. 223) opozarja, da pojem proceduralnosti zaradi drugačnih teoretičnih predpostavk ni prenosljiv v teorije, ki se ukvarjajo s koherenco besedila, kakršna je njegova in ki nima jasno razdelanih ločnic med propozicijsko vsebino in vsem ostalim.

pravih konceptov (npr. zaimki), po drugi strani pa nekateri izrazi, ki ne vplivajo na resničnostno vrednost, vendarle zaznamujejo koncepte in ne procesov oz. procedur.

Med slednjimi kot prototipično kategorijo D. Blakemore (Blakemore, 2002) izpostavi prav diskurzne označevalce. Pokaže, da so med izrazi, ki so bili dotlej obravnavani kot predstavniki razreda diskurznih označevalcev (za katere vse velja, da ne vplivajo na resničnostno vrednost) tako taki, ki jim lahko pripišemo proceduralni pomen (*but, however, nevertheless, well*) kot taki, ki v resnici zaznamujejo koncepte (*in contrast, in other words, as a result*). Na podlagi tega trdi, da kategorija diskurznih označevalcev ni kognitivno utemeljena jeziko(slo)vna kategorija, saj vključuje **pomensko nehomogen nabor izrazov**. Pri tem ločnico med enim in drugim pomenom ponazarja z različnimi testi (Blakemore, 2002), ki naj bi ponazarjali, da izraze s proceduralnim pomenom uzaveščamo drugače kot tiste s konceptualnim, npr. ker jih le stežka parafraziramo, jim težko poiščemo sopomenke, v jeziku nimajo pomensko sorodnih homonimov v drugih besednih vrstah ali jih ni mogoče skladenjsko razširjati. Kot predstavnike diskurznih označevalcev nato sama obravnava **samo tiste s proceduralnim pomenom** (in jih preimenuje v **diskurzne povezovalce**, angl. *discourse connectives*).⁸

V primerjavi z drugimi pragmatičnimi pristopi torej teorija relevantnosti k opredeljevanju razreda diskurznih označevalcev pristopa predvsem s semantičnega vidika, ne pa z vidika njihovega prispevka k besedilni, interakcijski ali drugače pojmovani koherenci, saj meni, da je s kognitivnega vidika prepoznavna neke smiselne povezave med dvema deloma besedila *rezultat* razumevanja, ne pa njihov predpogoj. Ker tovrstne koherenčne povezave po mnenju predstavnikov teorije relevantnosti ne obstajajo oz. niso del mentalne predstave besedila, je znotraj te teorije edina kognitivno utemeljena ločnica, na katero se lahko opremo pri kategorizaciji jezikovnih sredstev, torej ločnica med proceduralnim in konceptualnim pomenom.

⁸ Za kritiko vseh treh temeljnih predpostavk relevantnostne teorije, tj. razlike med konceptualnim in proceduralnim pomenom, diskurznih povezovalcev kot nosilcev proceduralnega pomena, in radikalno monosemičnega pristopa, glej razpravo S. P. Bordería (2008).

2.1.2.6 Teorije koherentnih povezav

S teoretskega vidika se teorija relevantnosti na področju analize diskurza vzpostavlja predvsem kot konkurenčna teorija različnim besediloslovnim pristopom, ki se običajno združujejo pod skupnim poimenovanjem teorije koherence (angl. *coherence-based theories*) oz. teorije koherentnih povezav (angl. *theories of coherence relations*), saj v središče svojega zanimanja postavljajo vprašanje, v čem se smiselno povezana (koherentna) zbirka besedilnih gradnikov, tipično stavkov, razlikuje od naključnega niza gradnikov, ne glede na stopnjo njihove površinske povezanosti (besedilne kohezije), kakršno sta izpostavljala M. A. K. Halliday in R. Hasan (1976), ki ni niti nujen niti zadosten pogoj za besedilno koherenco (Sanders in Pander Maat, 2006). Teorije koherence tako izhajajo iz skupne predpostavke, da **med deli koherentnega besedila obstaja bolj ali manj končna množica konceptualnih povezav različnih tipov**, med katerimi tvorci in prejemniki besedila izbirajo v procesu tvorjenja oz. razumevanja besedila. Te povezave imenujejo **koherentne povezave** (angl. *coherence relations*), **diskurzne povezave** (angl. *discourse relations*) ali **retorične povezave** (angl. *rhetorical relations*) in jih torej vidijo kot ključne za razumevanja načina vzpostavljanja koherence v diskurzu.

Med njimi je med najvplivnejšimi sodobnimi teorijami t. i. **teorija retorične strukture**, angl. *rhetorical structure theory* (**RST**) (Mann in Thompson, 1988), ki je bila na podlagi obsežne analize pisnih besedil prvotno razvita za potrebe samodejnega računalniškega tvorjenja besedil, a se je odtlej uveljavila tudi na širšem področju teoretskega in uporabnega jezikoslovja (Taboada in Mann, 2006a, 2006b). Na podlagi pojmovanja koherence kot lastnosti besedila, v katerem ni logično redundantnih enot ali vrzeli, ta teorija predpostavlja, da lahko vsakemu delu besedila (minimalna besedilna enota je običajno stavek) pripišemo določeno funkcijo glede na ostale dele besedila oz. med njimi ustvarimo povezavo, pri tem pa lahko (ni pa nujno) opredelimo tudi njuno hierarhično razmerje glede na pomembnost za celotno razumevanje besedila (t. i. nukleus in satelit). Ta postopek se nato rekurzivno ponavlja na daljših besedilnih enotah, tako da je na koncu celotno besedilo predstavljeno kot hierarhično strukturirano drevo, kjer je vsak del besedila povezan v strukturo in opredeljen z vidika njegove funkcije v tej strukturi. V primerjavi z drugimi teorijami koherence je poleg hierarhizacije pomembnosti besedilnih enot posebnost te teorije zlasti to, da besedilo analizira z vidika avtorjevega namena (*intence*) oz. zavestnega vplivanja na naslovnikovo razumevanja besedila, kar se odraža tudi pri sami opredelitvi tipov

povezav oz. funkcij. Teh avtorja teorije sicer ne pojmujejo kot dokončne množice, zato so bile za prvotno taksonomijo 24 tipov povezav (Mann in Thompson, 1988) predlagane tudi številne nadgradnje.

Če se teorija retorične strukture vzpostavlja predvsem kot analitično orodje za opis strukture koherentnega besedila, pa raziskovalci znotraj t. i. **kognitivnega pristopa h koherentnim relacijam**, angl. *cognitive approach to coherence relations (CCR)*, (Sanders idr., 1992, 1993) z različnimi psiholingvističnimi in korpusnimi analizami dokazujejo, da je ideja koherentnih povezav tudi kognitivno utemeljena, saj vplivajo na kognitivno reprezentacijo diskurza kot jezikovne tvorbe oziroma na samo razumevanje (vsebine) diskurza (Sanders idr., 1992, str. 3). Koherentne povezave so po njihovem mnenju kognitivni mehanizmi, ki se jih avtorji poslužujejo pri tvorjenju besedil, naslovniki pa pri njihovi interpretaciji, na hitrost in način procesiranja besedila oz. oblikovanja njegove mentalne predstave pa vplivajo tako povezave kot njihovi površinski kazalci, ki naslovnika usmerjajo k izbiri ustrezne povezave (Degand idr., 1999; Sanders in Noordman, 2000; Degand in Sanders, 2002). Obenem tudi zagovarjajo stališče, da s kognitivnega vidika ni mogoče predpostavljati, da v zavesti govorcev obstaja določena taksonomija povezav, temveč za njihov opis predlagajo štiri temeljne dimenzije (kognitivne primitivume): polarnost (pozitivna ali negativna), temeljna operacija (dodajanje ali vzrok), izvor koherence (objektivni ali subjektivni) in vrstni red segmentov (osnovni ali neosnovni).⁹ Njihove možne kombinacije nato opisujejo različne podrobnejše tipe povezav, ki bolj ali manj sovpadajo s taksonomijami drugih teorij (Knott in Sanders, 1998; Bunt in Prasad, 2016; Sanders idr., 2016).

Med številnimi drugimi sorodnimi teorijami koherentnih relacij in strukture diskurza (Hobbs, 1985; Kehler, 2002; Asher in Lascarides, 2003) izpostavimo še raziskave v okviru izgradnje

⁹ Polarnost: povezava je pozitivna, če sta propoziciji povezani brez negacije ene izmed njih, s čimer se protivna (*but*) in dopustna razmerja (*although*) ločujejo od drugih razmerij, npr. dodajanja (*and*) ali vzročnosti (*because*). Temeljna operacija: povezava je vzročna, če ena propozicija implicira drugo, kar je denimo značilno za vzročno-posledična in pogojna razmerja; povezava je dodajalna, če med njima ni mogoče razbrati tovrstne povezave, kar je značilno za vezalnost in časovna razmerja. Izvor koherence: povezava je objektivna, če med njima obstaja dejanska povezava v svetu, in subjektivna, če jo vzpostavlja šele tvorec, s čimer se predstavne povezave ločujejo od argumentacijskih (npr. posledica proti sklep). Vrstni red: vrstni red vzročno povezanih propozicij je pozitiven, če v besedilu posledica sledi vzroku, s čimer se uvajanje posledice (*as a result*) razlikuje od uvajanja vzroka (*because*). Več (Sanders idr., 1992, str. 5–17; Knott in Sanders, 1998, str. 7).

diskurzno razčlenjenega korpusa angleških publicističnih besedil **Penn Discourse Treebank (PDTB)** (Prasad idr., 2008, 2014), ki so pomembno vplivale na razvoj računalniškojezikoslovnih pristopov k analizi diskurza. PDTB se vzpostavlja kot teoretsko nevtralen pristop (in ne jezikoslovna teorija), saj se ukvarja zgolj s prepoznavanjem diskurzni povezav med deli besedila, ne pa vprašanjem njihovega nadaljnega združevanja v besedilno celoto. Druga posebnost tega pristopa je, da se pri opredelitvi, prepoznavi in opisu povezav naslanja na koncept (abstraktnih) diskurzni povezovalcev kot diskurzni predikatov, ki jemljejo **natanko dva argumenta v obliki t. i. abstraktnih objektov**, kot so dogodki, stanja in propozicije (Asher, 1993). V besedilih tako najprej prepoznavajo tovrstne **povezovalce** (jih označijo, če so izraženi eksplicitno, ali rekonstruirajo, če so izraženi implicitno) in njihove argumente (dele besedila, ki opisujejo abstraktne objekte), nato pa povezovalcem pripišejo tudi pomen (vrsto povezave, ki jo vzpostavljajo), povezave, povezovalce in argumente pa opredelijo tudi z vidika njihovega izvora (angl. *attribution*; npr. ali jih uvaja avtor besedila ali nekdo drug, ali gre za domneve ali dejstva, ali so zanikani ali ne).

Znotraj teorij koherence imajo tako diskurzni označevalci¹⁰ drugotno vlogo, saj so opazovani predvsem kot površinski kazalci predpostavljenih globljih koherentnih povezav med deli strukturiranega besedila, kot kohezivni elementi, ki sami po sebi še ne zagotavljajo smiselne povezanosti besedila, temveč zgolj signalizirajo konceptualne povezave na globljih ravneh in na ta način vplivajo na interpretacijo besedila in lajšajo njegovo razumevanje. **Diskurzni označevalci so torej le eden izmed možnih načinov signaliziranja tako ali drugače opredeljenih koherentnih povezav**, ki so lahko izražene tudi z drugimi jezikovnimi sredstvi – T. Das in M. Taboada (2013) denimo omenjata še koreferenčne, leksikalne, semantične, morfološke, skladijske, grafične, številčne in žanrske signale – ali pa na površini besedila niso izražene (so implicitne). Kljub temu njihova vloga v razvoju teorij vzpostavljanja besedilne

¹⁰ Izraz diskurzni označevalec (*discourse marker*) se za eksplicitne označevalce koherentnih povezav uporablja predvsem znotraj teorije retorične strukture (RST), medtem ko drugi raziskovalci uporabljajo sorodne izraze, kot so *cue phrases*, *discourse connectives* ali *lexical markers of coherence relations*, a terminološke izbire na področju teorij koherence običajno ne implicirajo bistvenih teoretskih predpostavk in razlik v naboru opazovanih izrazov, kvečjemu z njimi izpostavljajo izrazno motivirano ločnico, npr. med vezniki (*connectives*) in drugimi funkcijsko analognimi jezikovnimi sredstvi (*cue phrases*).

koherence ni nezanemarljiva, saj se prvotni konceptualno utemeljeni nabori relacij danes pogosto dopolnjujejo ali preizprašujejo z leksikalno utemeljenimi pristopi k njihovi zamejitvi, ki predpostavljajo, da lahko o relevantnosti in utemeljenosti posameznih tipov povezav sklepamo prav na podlagi tega, ali so se zanje v jeziku razvila posebna jezikovna sredstva (npr. Knott in Dale, 1994; Louwerse in Mitchell, 2003; Martin, 1992; prim. s Taboada, 2006).¹¹ Ker se v večini teorij koherence kot tipična minimalna enota diskurza običajno navaja stavek oz. stavčno izražena propozicija, za razliko od (nekaterih) pragmatičnih pristopov te teorije kot diskurzne označevalce obravnavajo izrazno zelo raznolik nabor izrazov, od podrednih in prirednih veznikov do prislovov, predložnih zvez in drugih vrst izrazov, kot bomo videli v nadaljevanju.

Kot tradicionalne besediloslovne teorije nasploh so se tudi teorije koherenčnih povezav vzpostavile predvsem z analizo strukture pisnih besedil, vendarle pa je bilo v zadnjih dveh desetletjih tudi nekaj poskusov podobnega **razčlenjevanja dialoških ali monoloških govorjenih besedil**, ki jih, kot ugotavlja M. Taboada (2006, str. 575), lahko razdelimo v dva pristopa: če dialog opazujemo z vidika analitika, vidimo en diskurzni dogodek oz. eno besedilo, če pa dialog opazujemo z vidika udeležencev, ga lahko pojmujejo kot izmenjavo več ločenih besedil. Sama v svoji analizi španskih in angleških dialogov (Taboada, 2004) izbere drugi pristop in tako ravnino menjave vlog loči od retorične strukture oz. strukture informacij. V alternativni prvi pristop pa bi lahko umestili raziskovalce, ki izpostavljajo potrebo po širitvi nabora koherenčnih povezav in njihovih površinskih označevalcev. M. M. Louwersee in H. H. Mitchell (2003) tako med analizirane diskurzne označevalce vključujeta tako označevalce, ki delujejo lokalno kot povezovalci besedila znotraj vloge enega govornika, kot izraze, ki delujejo globalno, denimo vzbujajo naslovnikovo pozornost, kažejo na prevzemanje vloge, izražajo strinjanje ipd. Na podlagi korpusne analize in substitucijskih testov nato za opis označevalcev oz. povezav, ki delujejo na ravni menjave vlog, po vzoru kognitivnih primitivumov teorije CCR predlagata štiri parametre opisa glede na: smer navezovanja, polarnost odnosa do predhodne vsebine, stopnjo (ne)strinjanja in moč odziva. Podobno pri aplikaciji sheme PDTB na korpus spontano govorjene italijanščine S. Tonelli s sodelavci (2010) ugotavlja, da tipični povezovalci dveh abstraktnih

¹¹ G. Redeker in H. Gruber (2014) na podlagi različnih raziskav povzemata, da je v povprečju približno 50 % razmerij eksplicitno signaliziranih s površinskim označevalcem, dejanski delež pa je precej odvisen od žanra.

objektov zelo pogosto nastopajo tudi v drugih vlogah, kot kazalci organizacije, usmerjanja in interaktivne strukture pogovora ali pa v vlogi uvajalcev apozicije. Čeprav teh rab ne vključujejo v analizo (jih ne prepoznavajo kot diskurzne povezovalce, kot jih pojmuje PDTB), pa pri opredeljevanju pravih besedilnih povezovalcev predlagajo podrobnejšo delitev njihovih pragmatičnih rab, ki so bile v izhodiščni taksonomiji povezav predvidene zgolj pri nekaterih tipih besedilnih povezav, prav tako pa niso razlikovale med epistemičnim povezovanjem in povezovanjem govornih dejanj. Obenem opozarjajo tudi na nekaj strukturnih posebnosti, kot je dejstvo, da se implicitne povezave v govoru zaradi fragmentarnosti pogosto vzpostavljajo tudi med besedilnimi enotami, ki si ne sledijo zaporedoma, ali pa se drugi izmed dveh obveznih argumentov zaradi prekinitve ne realizira.

Tudi znotraj besediloslovnih teorij koherence se tako s premikom s pisnega na govorni jezik **rahljajo dotlejšnja pojmovanja koherence v diskurzu**, saj poleg povezovanja soležnih besedilnih enot (ki jih v svoji ozki definiciji denimo ohranja B. Fraser) vključujejo tudi specifične interaktivnega govornega sporazumevanja, s čimer se približujejo širšemu razumevanju diskurza oz. koherence v diskurzu, kakršne sta iz sociolingvističnih in pragmatičnih izhodišč denimo predlagali D. Schiffrin in G. Redeker. Po drugi strani v pragmatični opoziciji še naprej ostajajo raziskovalci znotraj teorije relevantnosti, ki sicer diskurzne označevalce prav tako opazujejo z vidika usmerjanja naslovnikove interpretacije, a zavračajo temeljno idejo diskurza kot strukture in pojmovanja diskurzni označevalcev kot površinskih označevalcev te strukture.

2.1.3 Raziskave v slovenskem prostoru

2.1.3.1 Konektorji

Najširši nabor diskurznofunkcijskih izrazov je doslej v slovenskem prostoru obravnaval V. **Gorjanc** (1998), ki s krovnim poimenovanjem **konektorji** po vzoru van Dijkovega funkcijskega pristopa (van Dijk, 1980) analizira izraze, ki v besedilu povezujejo besedilne enote različnega obsega, vzpostavljajo medstavčna in medpovedna razmerja ali razmerja med delom besedila in njegovim nadaljevanjem (Gorjanc, 1998, str. 367). Konektorji se torej vzpostavljajo predvsem glede na svojo funkcijo povezovalcev propozicij (tj. stavkov z referenčnim pomenom) in višjih besedilnih enot, s čimer pripomorejo k vzpostavljanju besedilnega pomena, s slovničnega vidika pa združujejo različne oblikoslovno-skladenjske kategorije. Glede na prispevek k smiselni

povezanosti besedila jih deli na neopomenjevalne (povezujejo stavke le na ravni površinske besedilne organizacije oz. kohezije, ne vnašajo novih pomenskih enot, npr. relativni in znotraj njih oziralni) in opomenjevalne, ki vnašajo tudi pomenska oz. logična razmerja med povezane dele besedila in na ta način prispevajo k besedilni koherenci. Opomenjevalne konektorje glede na oblikoslovno-skladenjske lastnosti nato razdeli na vezniške, prislovne in členkovne, razpravlja o (težavnih) merilih za njihovo ločevanje in v jedru raziskave analizira posebnosti rabe konektorjev v korpusu znanstvenih člankov s področja matematike. Med najpogostejšimi neopomenjevalnimi konektorji podrobneje analizira *ki* in *da*, med opomenjevalnimi pa pogojne (npr. *kadar*; *če*; *natanko tedaj*, *ko*), vzročne (npr. *ker*, *saj*, *namreč*, *kajti*), sklepalne (npr. *zato*, *tako*, *tedaj*, *na tak način*), pojasnjevalne (npr. *to je*, *in sicer*, *se pravi*, *na primer*) in protivne (npr. *pa*, *vendar*, *toda*) konektorje oz. konektorske zveze.

S konektorji se je pri nas v okviru raziskav slovenskega in hrvaškega jezikoslovnega akademskega diskurza ukvarjala tudi **T. Balažić Bulc** (2009), ki sledi Gorjančevemu funkcijskemu pristopu, a se pri rabi termina konektor omeji samo na **opomenjevalne konektorje**, ki se po Hallidayu in Hasanu (1976) od drugih konektorjev ločijo po tem, da prek svojega pomena in funkcije, ki jo imajo v besedilu, ustvarjajo vezi med deli besedila. Konektorje torej opredeli kot tiste izraze, ki v besedilu izražajo povezave med izjavami, tj. kontekstualiziranimi pomensko-skladenjsko zaokroženimi enotami jezikovne produkcije, ali pa vzpostavljajo organizacijsko strukturo besedila (Balažić Bulc, 2009, str. 36–37). Na podlagi tega konektorje razdeli na **konektorje povezovalce**, ki povezujejo dve sosednji izjavi, in **konektorje organizatorje**, ki delujejo med deli besedila oz. organizirajo dele besedila v smiselno celoto. Glede na opredeljenost izjave kot kontekstualizirane pomensko-skladenjsko zaokrožene enote (Schiffrin, 1994) iz skupine konektorjev izloča tiste, ki vzpostavljajo povezave znotraj izjave, tj. med besednimi zvezami ali posameznimi stavki povedi, npr. pogojnega *če* ali vzročnega *ker*. Ponudi tudi novo tipologijo besedilnih povezav, ki jo oblikuje v procesu analize gradiva (korpusa slovenskih in hrvaških znanstvenih člankov in seminarskih nalog), na podlagi tega pa konektorje razdeli na povezovalce in/ali organizatorje za izražanje nasprotovanja (Balažić Bulc, 2009, str. 51, 62), npr. *pa*, *vendar*, *vendar pa*, *kljub temu*, *a*, *po drugi strani pa*, *sicer pa*, pojasnjevanja (Balažić Bulc, 2009, str. 54), npr. *tj.*, *in sicer*, *tako*, *torej*, *pri čemer*, *to je*, *zaradi česar*), povezovanja (Balažić Bulc, 2009, str. 57, 64), npr. *pa*, *in*, *prav tako*, *poleg tega*, *hkrati pa*,

obenem, ob tem., ravno tako, s tem v zvezi), razlikovanja (Balažic Bulc, 2009, str. 58), npr. *ali pa, oz., ali, oziroma*, razvrščanja (Balažic Bulc, 2009, str. 58, 66), npr. *nato, nadalje; najprej, na koncu, nadalje, nato pa, nenazadnje, ob koncu*, sklepanja (Balažic Bulc, 2009, str. 59, 67), npr. *zato, torej, zaradi tega, potemtakem, in torej, skratka, in tako, tako; zatorej*, utemeljevanja (Balažic Bulc, 2009, str. 60), npr. *saj, namreč, ker, kajti* in ilustriranja (Balažic Bulc, 2009, str. 69), npr. *npr., na primer, kot npr., kot so, recimo, tako na primer*. V sami analizi se nato osredotoči na deset najpogostejših konektorjev v povezovalni ali organizacijski vlogi (*pa, npr., torej, saj, tj., tako, namreč, in sicer, zato, vendar*) in izkazano rabo primerja z njihovimi (nesistematičnimi) opisi v SSKJ in SS. V nadaljnjih raziskavah se podrobneje osredotoča predvsem na slovensko-hrvaško kontrastivno primerjavo njihove vloge v strukturiranju znanstvenega diskurza (Balažic Bulc, 2010; Balažic Bulc in Gorjanc, 2015), pa tudi na vprašanje njihovega pomenskega opisa v enojezičnih razlagalnih slovarjih (Balažic Bulc, 2009).

2.1.3.2 Besedilnopovezovalni členki

Med oblikoskladenjsko raznolikimi konektorskimi sredstvi posebej plodovito področje predstavljajo razprave v povezavi z besedilnopovezovalno vlogo (nekaterih) členkov, tj. (s slovničnega vidika) razmeroma nove besedne vrste, ki je izrazno podobna prislovom, a se po njej ne moremo vprašati, saj v povedi nima vloge samostojnega stavčnega člena, temveč v njej nastopa zgolj kot pomenski modifikator (Toporišič, 1974; Toporišič, 2004, str. 445–449).

Prvi kohezijsko vlogo slovenskega členka podrobneje izpostavlja **A. Skubic** (1999), ki dotedanje, skladenjskopomenske opise členkov po vzoru obravnave (pristavčnih) prislovov (adjunktov) v angleških slovnicaх dopolni z analizo glede na njihovo funkcijo v strukturi in pomenski podstavi stavka, in sicer členke razdeli na dve jedrni skupini: modalne členke, ki izražajo govorčev odnos do oblike ali vsebine povedi, in povezovalne **členke, ki vzpostavljajo medpovedno kohezijo** (Skubic, 1999, str. 212). Kot izhodišče svoje taksonomije besedilnopovezovalne funkcije členkov uporabi sistemsko-funkcijsko slovnico (Halliday, 1985), ki medstavčna razmerja znotraj zloženih povedi deli na krovne funkcije dodelave ali opisa, kjer drugi stavek temeljiteje dodela prvega, tj. ga ponovi z drugimi besedami, ponazori ali opiše podrobneje, funkcije podaljšanja, kjer je drugi stavek dodatek, nasprotje ali alternativa prejšnjega, in funkcijo dopolnitve, kjer novi stavek prejšnjega dopolni z okoliščinskimi informacijami (vzroka, časa, kraja, načina) (Skubic, 1999,

str. 214). Vsaka izmed teh krovnih funkcij je nato členjena še na podrobnejše podtipe, avtor pa vsakega izmed njih na podlagi seznamov členkov iz drugih slovenističnih priročnikov in analize publicističnih besedil ponazori s slovenskimi ustreznici v obliki štirih različnih strukturnih realizacij: veznikov (npr. *zato*), členkov (npr. *torej*, *posledično*, *potem*), t. i. formalno integriranih struktur (npr. *posledica je*) in t. i. semantično integriranih struktur (npr. *zaradi tega*, *ker je tako*). V prvem dodatku razprave izpostavi še t. i. **nadaljevalne členke**, kot so *no*, *ma*, *na*, *ja*, *ne*, *a veš*, *ne?*, *aha* ipd., ki jih izhodiščna slovnica ne omenja, a bi jim zaradi signaliziranja odziva na predhodno povedano poleg fatične prav tako lahko pripisali tudi besedilno vlogo. V drugem dodatku pa poleg povezovalnih na kratko predstavi še različne tipe modalnih (naklonskih) členkov in zavrne sistemsko-funkcijsko ločevanje na modalne adjunkte v ožjem smislu, ki naj bi modificirali zgolj povedek, in opombne adjunkte, ki naj bi modificirali celoten stavek, saj je za razliko od angleščine stava modalnih členkov v slovenščini poljubna, domet njihove modifikacije pa odvisen tudi od poudarka v govoru.

Tudi **M. Smolej** (2004a) na podlagi skladenjskih, besedilnih in pragmatičnih meril ponudi alternativno taksonomijo in združevanje nekaterih členkovnih skupin, ki jih je glede na prevladujoči pomenski vidik predlagala Slovenska slovnica (Toporišič, 2004). Členke glede na temeljno funkcijo razdeli v tri skupine: (i) členke s funkcijo modifikacije, tj. pomenske dopolnitve vsebine oz. referencialnega pomena, ki jih nadalje razvrsti na modalne členke (*morda*, *verjetno*, *gotovo*), členke čustvovanja (*žal*, *na srečo*), nikalne členke in tvorce stalnih sporočanj oblik povedi (*saj*, *če*, *da*); (ii) členke s prekrivno funkcijo modifikacije in besedilnega povezovanja, kot so poudarjalni členki (*blizu*, *celo*, *izključno*, *kar prav*, *ravno*, *tako rekoč*), in (iii) členke s funkcijo besedilnega povezovanja, ki jih imenuje **besedilni povezovalci**. Ti se od drugih členkov razlikujejo predvsem po tem, da ne vstopajo v pomensko razmerje s posameznimi stavčnimi členi oz. deli povedi, ampak izražajo besedilnoskladenjska razmerja oz. logične odnose med deli besedila ali povedi (Smolej, 2004a, str. 49). To skupino nadalje razdeli na sekundarne nosilce vsebine, ki prenašajo vsebino predhodnih delov besedila (so anaforični), npr. *kakor koli že*, *kljub vsemu*, *konec koncev*, *namreč*, *potem*, *pravzaprav*, *sicer* itd., napovedovalce modifikacije oblike, ki vzpostavljajo most med dvema pomensko prekrivnima besedilnima enotama, npr. *bolje rečeno*, *ali natančneje*, *z drugimi besedami* ipd., in delilne signale, ki skrbijo za organizacijo besedila, npr. *prvič*, *dalje*, *na koncu*, *nato*, *za začetek*, *zdaj*, *no*,

nič. V komplementarni razpravi (Smolej, 2004b) členke nadalje deli tudi z vidika stopnje njihove obveznosti znotraj stavka oz. povedi in/ali besedila. Glede na to, ali izpustitev členka vpliva zgolj na prekinitev besedilnega toka oz. smiselnosti besedila ali pa tudi na samo strukturo stavka oz. povedi, v kateri se pojavijo, členke deli na enostopenjsko in dvostopenjsko obvezne členke, pri čemer se med prve poleg nekaterih drugih skupin členkov umeščajo tudi besedilni povezovalci, npr. *skratka* (Smolej, 2004b, str. 149).

Podobno temeljno funkcijsko delitev na povezovalne (besedilne) in naklonske (medosebne) členke prevzame tudi **Slovar členkov** (Žele, 2014b), ki členke glede na njihovo skladijsko vlogo opredeli kot besedno vrsto brez predmetnega ali slovničnega pomena, ki v stavku deluje kot skladijski modifikator celotnega povedja ali posameznih stavčnih členov, in sicer tako, da kot ubesedeni govorni dogodek vzpostavlja smiselno razmerje med besedilom in sobesedilom in drugimi sporočanjскими okoliščinami (Žele, 2014b, str. 9–10). **Povezovalni členki** vzpostavljajo besedilno koherenco in kohezijo ter se glede na pomen delijo še na podskupine dodajalnih (npr. *celo, kaj šele, še več*), izbirnih (npr. *drugače, sicer pa*), izvzemalnih (npr. *edino, le, sicer*), navezovalnih oz. nadaljevalnih (npr. *kakorkoli že, potemtakem, vsekakor*), nadomestnih (npr. *namesto tega, nasprotno*), nasprotovalnih (npr. *zato pa*), pojasnjevalnih (npr. *to se pravi, torej*), ponazarjalnih (npr. *namreč*), popravnih (npr. *ali bolje, namreč*), poudarnih (npr. *pravzaprav, predvsem*), povzemalnih (npr. *skratka, torej*) in zastranitvenih (npr. *mimogrede*) členkov. Po drugi strani se **naklonski členki** osredotočajo na udeležence, na okoliščine, na glagolski proces ali na količino in so razdeljeni na čustvenostne (npr. *končno, saj, začuda*), pozivne (npr. *dejansko, kajne, naj*), vrednotenjske (npr. *baje, morda, nemara*) in nikalne (npr. *nikar*) členke. Za razliko od predhodnih razprav, usmerjenih v notranjo funkcijsko členitev členkov, slovar pri določanju njihovega dejanskega leksikalnega nabora med pripadnike te kategorije poleg t. i. prvotnih oz. pravih členkov uvršča zelo širok nabor jezikovnih sredstev, ki v besedilih (lahko) nastopajo v podobnih funkcijah kot prvotni členki, npr. vezniki, prislovi, povedkovniki, samostalnik, pridevnik in medmet, kar kaže na zapletenost opredeljevanja te besednovrstne kategorije in bolj ali manj dosledno prepletanje raznolikih meril opredelitve, kot so izvor oblike, jedrni pomen, skladijska vloga, besedilna vloga, funkcija v sobesedilu, diahroni razvoj, domet modifikacije itd. (Smolej, 2015a; Vidovič-Muha, 2015).

2.1.3.3 Pragmatična analiza diskurza

Poleg zgoraj naštetih slovenističnih razprav, ki se z diskurzno leksiko ukvarjajo predvsem z vidika njene umestitve v jezikovni sistem, korelacije z drugimi ravnmi jezikovnega opisa in funkcijskim diapazonom posameznih besednih vrst, so v slovenskem prostoru razmeroma pogosti tudi strukturalistično razbremenjeni funkcijski opisi leksike, kakršni so se vzpostavili na širšem področju pragmatičnega jezikoslovja. Za razliko od V. Gorjanca, ki se v skladu z izbranim tradicionalnim besedilnoskladenjskim pristopom ukvarja s konektorji kot funkcijsko raznolikimi linearnimi povezovalci stavčnih propozicij, povedi in večjih besedilnih enot, **M. Schlamberger Brezar** (2007; 2009) denimo v središče svojega zanimanja zgolj postavlja njihovo podmnožico – pragmatične konektorje oz. **pragmatične povezovalce**. Te s sintezo dognanj različnih teorij, kot sta ženevska šola pragmatike (Roulet idr., 1985) in teorija argumentacije (Anscombre in Ducrot, 1983), opredeli kot tiste izraze, ki povezujejo izreke oz. govorna dejanja tako, da izražajo govorčev odnos in osebno interpretacijo (delujejo na ravni izrekanja), s čimer iz svoje opredelitve pragmatičnih povezovalcev izloči konektorje, ki (ne glede na vrsto pomena) delujejo na nižjih ravneh besedila (vezniki), ter konektorje, ki med izreki vzpostavljajo semantične povezave (operatorji).¹² Pri svoji opredelitvi se ne opira na oblikoskladenjske lastnosti pragmatičnih povezovalcev, saj ti pripadajo različnim oblikoslovnim razredom, kot so priredni in podredni vezniki, prislovne zveze, členki ipd., nobeden od teh razredov pa ni sestavljen izključno iz povezovalcev (Schlamberger Brezar, 2009, str. 45–46), zato so izrazi, ki lahko nastopajo v vlogi pragmatičnih povezovalcev, pogosto večpomenski in lahko nastopajo tudi kot operatorji ali vezniki, npr. *in*, *ali*, *ker* (Schlamberger Brezar, 2009, str. 91). Na podlagi delitve povezovalce nato razdeli na zaznamovalce ilokucijske funkcije (npr. *rad bi vas nekaj vprašal*), zaznamovalce interakcijske funkcije (npr. *namreč*, *kajti*, *saj*; *čeprav*, *ampak*) in zaznamovalce strukture konverzacije (*no*, *in*), na gradivu govorjene posnetkov televizijskih oddaj pa nato podrobneje analizira izbrane povezovalce v francoščini z vidika njihove vloge v argumentaciji, pojasnjuje njihovo ločevanje glede na druge rabe ter jih sopostavlja s slovenskimi ustrezniki.

¹² Za ponazoritev kompleksnega prepletanja skladenjskih in semantičnih meril glej denimo analizo rabe veznika/povezovalca *et* 'in' (Schlamberger Brezar, 2009, str. 99–115) ali *parce que* 'ker' (Schlamberger Brezar, 2009, str. 117–129).

Tudi raziskave **A. Pisanski Peterlin** (2005, 2007, 2011) se umeščajo na področje kontrastivnega jezikoslovja, natančneje angleško-slovenske kontrastivne retorike, ki jo A. Pisanski Peterlin opazuje v okviru **teorije metabesedilnosti** (Hyland, 2005). Po tej teoriji so metabesedilni elementi vsi tisti deli besedila, ki jim je skupno to, da ne prispevajo nič k propozicijski vsebini besedila, sicer pa so funkcijsko razpršeni na raznolike vloge organizacije besedila ali izražanja odnosa med piscem in bralcem, tj. prehodi (*in, ampak*), označevalci okvira (*nazadnje, če zaključim*), endorfični označevalci (*omenjen zgoraj*), dokazovalci (*X trdi*), tolmači (*namreč, npr.*), omejevalci (*morda, možno*), ojačevalci (*jasno je, da*), označevalci odnosa do vsebine (*žal, presenetljivo*), samoomembe (*jaz, moj, naš*), označevalci odnosa do bralca (*lahko vidite, da*). Med temi krovnimi metabesedilnimi kategorijami se je A. Pisanski Peterlin v svojih analizah akademskega pisnega diskurza večinoma osredotočala na dve kategoriji, tj. endorfične označevalce oz. kažipote (npr. *pokazali bomo da, 2. poglavje*) in označevalce odnosa do vsebine, v novejših raziskavah pa tudi na označevalce epistemske naklonskosti oz. omejevalce (Pisanski Peterlin, 2010, 2015c), nagovore (Pisanski Peterlin, 2014) in prehode oz. povezovalce (Pisanski Peterlin, 2015b).

Z analizo diskurza v spontanem govoru se v slovenskem prostoru podrobno ukvarja zlasti **D. Verdonik** (Verdonik, 2006, 2007, 2008; Verdonik idr., 2007), ki v slovenistično terminologijo po prvi omembi (Kranjc, 1999) tudi vpelje izraz diskurzni označevalec. Te funkcijsko opredeli kot tiste izraze, ki ne prispevajo veliko k propozicijski vsebini sporočila, temveč imajo v diskurzu **predvsem pragmatično vlogo** (Verdonik, 2007, str. 47; Verdonik idr., 2007, str. 154–155). V primerjavi s predstavljenimi tujimi raziskavami diskurzni označevalcev iz opredelitve tako izloči izraze, ki delujejo na ravni besedila (npr. vzpostavljajo odnose med propozicijami, govornimi dejanji ali besedilnimi sekvencami) in se osredotoči zgolj na omejen nabor najpogostejših in propozicijsko najbolj izpraznjenih diskurzni členkov, kot so izrazi *ja, mhm, aha, no, dobro, v redu, eee, mmm, glejte, veste, mislim, zdaj, seveda, ne?, a ne?* ipd. (Verdonik idr., 2007, str. 175). V poznejši razpravi (Verdonik, 2008) nato v kontekstu označevanja vrste diskurzni označevalcev v korpusih za potrebe jezikovnotehnoloških aplikacij predlaga nekoliko širšo opredelitev, tj. izrazi, ki ne vplivajo bistveno na pomen sporočila, temveč opravljajo predvsem komunikacijske, pragmatične oz. proceduralne funkcije (Verdonik, 2008, str. 25), ter njihovo delitev na štiri temeljne tipe diskurzni označevalcev: **ideacijske** (ti primarno delujejo na

predstavni oz. ideacijski ravni diskurza; imajo predvsem besedilnopovezovalno vlogo, npr. *in, pa, torej, in sicer*), **interakcijske** (ti primarno delujejo na interakcijski oz. sociološki ravni; imajo vlogo vzpostavljanja in razvijanja odnosa med sogovorniki; npr. *mhm, ja, no, dobro, glejte*), označevalce **procesov tvorjenja** (ti primarno delujejo na kognitivni ravni; označujejo iskanje ustrezne besede, odločanje glede nadaljevanja, ohranjanje vloge, npr. *eee, mmm, bom rekel, ne vem*) in **interpretacijske** označevalce (tudi ti delujejo na kognitivni ravni in usmerjajo naslovnika pri interpretaciji, mednje pa umesti predvsem označevalce odprte propozicije, kot so *in tako naprej, pa tako, ali pa kaj* itd.). To je tudi skupina, ki jo pomensko in izrazno nato analizira še podrobneje (Verdonik in Kačič, 2012; Verdonik, 2015b), obenem pa diskurzne označevalce posredno omenja tudi v okviru razprav o izzivih slovarskih opisov tipično govorjene leksike (Verdonik, 2015a; Verdonik in Maučec, 2016), tudi z vidika njihovega pomenskega opisa (Verdonik, 2015c).

Druga raziskava, ki se je v slovenskem prostoru ukvarjala z diskurznofunkcijsko leksiko spontanega govora, pa je monografija **M. Smolej** (2012), ki v okviru širše raziskave značilnosti besedilnih vrst v spontanem govoru izpostavi tudi nekatere metabesedilne elemente (Smolej, 2012, str. 87–125), ki v besedilu opravljajo vlogo vzdrževanja stika s sogovorcem (medosebna funkcija) in/ali vlogo organiziranja besedila v koherentno celoto (besedilna funkcija). Izbrane metabesedilne elemente razdeli v dve jedrni skupini, in sicer z izrazom **govorni signali** poimenuje tiste izraze, katerih primarna vloga je vzdrževanje komunikacijskega razmerja med govorcema, hkrati pa besedilo tudi členijo ali označujejo konec posameznih besedilnih enot. To skupino še podrobneje deli na konativne signale, ki služijo za uravnavanje naslovnikove pozornosti in preverjanje razumevanja (npr. *(a) ne, (a) veš, ej, čak-i, predstavljaj si, moram rečt, razumeš?, ti povem, ti*), in na psevdofatične signale, s katerimi govorec svoje izjave poudarja (npr. *šit, hudiča, seveda, a daj daj daj, mislim, neverjetno, okej*). V drugo skupino obravnavanih metabesedilnih elementov pa umesti t. i. **kazalce stopenj besedilne zgradbe**, ki se v govoru udejanjajo predvsem na ravni organizacije besedila, in sicer kot kazalci temeljnega začetka (npr. *ja, skratka, torej, dobro*), kazalci konca (npr. *tko da, to je to*), kazalci nadaljevanja (npr. *in, in pol, no v glavnem, v bistvu, pol pa*), kazalci novega začetka (npr. *no, zdaj*), kazalci preobrata (npr. *a veš kaj ...*) in kazalci vpete informacije (*aja*), pri čemer pri večini skupin poleg posameznih leksemov ali zvez navaja tudi njihove funkcijske stavčne ustrezničke (npr. *no, bom povedala, kaj*

se je zgodil). Pri njihovi nadaljnji analizi z vidika vloge v določanju vrste govornih besedil ugotavlja, da je raba govornih signalov odvisna predvsem od posameznega govorca, manj pa od vrste besedila, medtem ko kazalci stopenj besedilne zgradbe korelirajo tudi s posameznimi besedilnimi tipi, kot so poročila ali anekdote.

2.1.4 Vpliv aplikativnih pristopov k analizi diskurza

Pregled najvplivnejših teoretskih pristopov k analizi diskurznofunkcijskega besedišča potrjuje relevantnost in razmah tega raziskovalnega področja tako v tujini kot pri nas, a hkrati razkriva, da k preučevanju navidezno prekrivnega predmeta pristopajo na zelo raznolike načine, saj glede na izbrani teoretski okvir in raziskovalne cilje v središče svojega zanimanja postavljajo zelo različne nabore jezikovnih sredstev oz. njihovih prepoznavnih izraznih in pomenskih lastnosti. Te so lahko bolj ali manj prekrivne (npr. povezovalci predstavne vsebine pri besediloslovnih in pragmatičnih teorijah koherence) ali pa se med seboj celo izključujejo (npr. medstavčni vezniki B. Fraserja in govornjeni členki K. Fischer). Dejanska ocena njihove prekrivnosti ali komplementarnosti pa je v praksi okrnjena, saj se večina teoretskih razprav osredotoča predvsem na obrazložitev izbranega modela diskurza in/ali utemeljitev opazovanih izrazov kot koherentne funkcijske kategorije, manj pa na vprašanje njihovega izčrpnega leksikalnega nabora, objektivnega ločevanja od drugih kategorij in sistematičnega pomenskega opisa. Tovrstna teoretska preizpraševanja posledično vodijo v vse večjo razdrobljenost področja in **onemogočajo napredek tako na teoretskem kot aplikativnem področju**, saj predlaganih teoretskih modelov, ki pogosto temeljijo na abstraktnih definicijah, izmišljenih primerih rabe ali analizi peščice prototipičnih izrazov, ni mogoče prenesti na obsežnejše gradivo, ki bi omogočilo njihovo empirično validacijo, nadgradnjo ali nadaljnjo aplikativno izrabo.

Tudi področje analize diskurza se v zadnjih dveh desetletjih namreč vse bolj **povezuje z drugimi interdisciplinarnimi področji uporabnega jezikoslovja**, kot so leksikografija, jezikovna didaktika in klinično jezikoslovje, saj porast jezikovnopragmatičnih in psiholingvističnih raziskav na področju usvajanja in poučevanja jezika dokazuje, da diskurzni označevalci predstavljajo pomemben del razvoja jezikovne zmožnosti v prvem oz. drugem/tujem jeziku (Chaudron in Richards, 1986; Flowerdew in Tauroza, 1995; Simmons-Mackie in Damico, 1996; Sankoff idr., 1997; Degand in Sanders, 2002; Hellermann in Vergun, 2007). Obenem so raziskave diskurznih

leksikalnih prvin dragocene tudi z vidika razvoja raznih aplikacij za strojno procesiranje naravnih jezikov (jezikovnih tehnologij), kot so sistemi za razpoznavo govora (Heeman in Allen, 1999), diskurzno razčlenjevanje (Schilder, 2002), samodejno generiranje besedil (Stede in Umbach, 1998; Callaway in Lester, 2002),¹³ samodejno povzemanje besedil (Marcu, 2000; Maskey in Hirschberg, 2005; Louis idr., 2010), samodejno poenostavljanje besedil (Candido Jr. idr., 2009) ali strojno prevajanje (Marcu idr., 2000; Meyer in Popescu-Belis, 2012)), ki morajo za svoje optimalno delovanje preseči razumevanje jezika kot zgolj vreče besed, besednih nizov ali stavkov (Webber idr., 2012). Tovrstne aplikacije svoj model znanja pogosto gradijo na večjih količinah ročno označenih avtentičnih besedil (učnih korpusih) na različnih jezikoslovnih ravneh, ki poleg nižjih ravni jezikoslovnega opisa (lematizacija, oblikoskladenjsko označevanje, skladenjsko razčlenjevanje, pomensko razčlenjevanje) običajno vključujejo tudi diskurzno razčlenjeno raven, z identificiranimi besedilnimi segmenti, njihovimi razmerji in tudi morebitnimi eksplicitnimi označevalci.

Pomen prenosa teorije v prakso so na področju analize diskurza tako prepoznali zlasti raziskovalci znotraj teorij koherentnih povezav, saj so bili v zadnjih petnajstih letih izdelani številni **diskurzno označeni oz. razčlenjeni korpusi** za različne jezike, znotraj teorije retoričnih povezav denimo angleška korpusa RST Discourse Treebank (Carlson idr., 2003) in RST Signalling Corpus (Taboada in Das, 2013), nemški Potsdam Commentary Corpus (Stede in Neumann, 2014), španski RST Spanish Treebank (da Cunha idr., 2011), baskovski Euskal RST Treebank (Iruskieta idr., 2013), portugalski CST (Cardoso idr., 2011) in korpus nizozemskih besedil (Redeker idr., 2012); znotraj kognitivnega pristopa CCR nizozemski korpus DiscAn (Scholman idr., 2016); znotraj teorije SDRT angleški korpus Discor (Reese idr., 2007), korpus pisne francoščine Annodis (Afantenos, Asher, Benamara, Bras idr., 2012) in francoski korpus spletnih pogovorov STAC (Afantenos, Asher, Benamara, Cadilhac idr., 2012); angleški korpus Discourse GraphBank (Wolf in Gibson, 2005), ki temelji na Hobbsovem seznamu povezav (Hobbs, 1985); angleško-nemški korpus GECCo (Lapshinova-Koltunski in Kunz, 2014b), v katerem so označeni kohezijski elementi (Halliday in Hasan, 1976); ter tudi številni korpusi, ki so

¹³ Več referenc tudi v (Taboada in Mann, 2006a, str. 568).

nastali po vzoru angleškega korpusa Penn Discourse Treebank (Prasad idr., 2008), npr. diskurzno razčlenjeni korpusi hindijščine (Oza idr., 2009), arabščine (Al-Saif in Markert, 2010), kitajščine (Zhou in Xue, 2012), turščine (Zeyrek idr., 2010); češčine (Poláková idr., 2013), danščine (Buch-Kromann in Korzen, 2010), nemščine (Versley in Gastel, 2012), francoščine (Danlos idr., 2012) in govorjene italijanščine (Tonelli idr., 2010).

S porastom številnih diskurzno označenih korpusov za različne jezike pa se kljub uspešnemu prenosu (nekaterih) teorij v prakso vedno bolj kaže tudi potreba po njihovem **poenotenju** oz. identifikaciji skupnih točk, ki bi omogočila primerjavo ali hkratno izrabo jezikovnih virov, nastalih na podlagi različnih teorij, jezikov in diskurzni tipov, ter s tem pripomogla k celostnejšemu razumevanju tvorjenja in interpretacije diskurza. S tem namenom je bila oblikovana tudi akcija evropske raziskovalne mreže COST Textlink,¹⁴ ki si med drugimi cilji, povezanimi z empiričnimi pristopi k analizi diskurza, prizadeva tudi za vzpostavitev skupnega konceptualnega okvira za opisovanje funkcij diskurzni označevalcev v različnih evropskih in neevropskih jezikih ter usklajenih meril za njihovo prepoznavo v (pisnih in govorjenih) besedilih, zanje pa namenoma uporablja teoretsko nevtralen izraz *discourse structuring devices*. Na podlagi podobnih izhodišč znotraj konzorcija ISO istočasno nastajata tudi predloga za poenoten semantični opis temeljnih koherenčnih povezav v diskurzu in drugih vprašanj, povezanih z njegovim razčlenjevanjem (Prasad in Bunt, 2015), ter standard za semantično razčlenjevanje dialoških izmenjav (Bunt idr., 2012).

Tudi na področju raziskav govorjenega diskurza se tako v luči aplikativnih prenosov in kontrastivnih raziskav pojavlja vse večja potreba po **teoretsko razbremenjenih pristopih k funkcijskem opisu diskurzni označevalcev v govoru**. Eden takih predlogov nastaja v okviru projekta MDMA univerze UC Louvain (Bolly idr., 2014, 2015), ki si prizadeva za vzpostavitev sistematične empirične metode za prepoznavo in označevanje diskurzni označevalcev v različnih besedilnih tipih, jezikih in prenosniških zvrsteh, pri čemer izhajajo iz stališča, da je to mogoče razviti šele na podlagi empirične analize najširšega nabora *potencialnih* diskurzni označevalcev. Diskurzne označevalce tako (namenoma) ohlapno opredelijo kot skladenjsko

¹⁴ <http://textlink.ii.metu.edu.tr/>

izpustljive proceduralne izraze, ki naslovníku sporočajo, kako besedilno enoto, v kateri se pojavijo, integrirati v mentalno reprezentacijo tekočega diskurza (Bolly idr., 2014; Crible, 2016), nato pa jih na podlagi unije najvplivnejših teoretskih pristopov funkcijsko zamejijo na štiri temeljne skupine z jasno opredeljenimi funkcijami: ideacijske (povezave med zunajbesedilnimi dogodki, npr. vzrok ali kontrast), retorične (povezave med govornimi dejanji, npr. sklep, in metabesedilne funkcije, npr. poudarjanje), sekvenčne (strukturiranje besedilnih segmentov, npr. menjava teme, prevzemanje vloge) in medosebne (upravljanje odnosa med govorcem in naslovníkom, npr. preverjanje pozornosti). Posebnost predlaganega pristopa je, da ne ostaja na ravni ohlapnih funkcijskih definicij, temveč je bil zanj razvit podroben označevalni protokol (Crible, 2014), ki na podlagi označevanja dvojezičnega korpusa in analize neujemanj med dvema neodvisnima strokovnjakoma (Crible in Zufferey, 2015) eksplicitno opredeljuje meje z drugimi kategorijami nepropozicijske leksike (pragmatičnimi označevalci) in načine ločevanja med posameznimi funkcijami. Tudi nasploh se na področju korpusne analize diskurza v zadnjem času pojavlja vse več raziskav, ki opozarjajo na pomembnost **evalvacije teoretskih opisov in kategorizacij z označevalnimi eksperimenti na avtentičnem gradivu**, v katerih se primerjajo odločitve več kompetentnih oz. izurjenih strokovnjakov (Verdonik, Žgank idr., 2007; Spooren in Degand, 2010; Zufferey idr., 2012; Scholman idr., 2016).

V **slovenskem prostoru** kljub številnim, raznolikim in tudi nekaterim gradivno dopolnjenim teoretskim razpravam o opredelitvi, zamejitvi in pomenskem opisu diskurzne leksike doslej še ni bilo poskusov njihovega sistematičnega prenosa na obsežnejše, uravnoteženo in širše dostopno korpusno gradivo, vsaj ne v smislu diskurzno razčlenjenega korpusa ali celostnega označevanja izrazov, ki naj bi pripadali določeni diskurznofunkcijski skupini, saj se tudi gradivno utemeljene raziskave pogosto osredotočajo zgolj na posamezne izraze ali specifične žanre. Da so nekateri **slovenistični teoretski modeli težko prenosljivi na konkretno in obsežnejše jezikovno gradivo** navsezadnje potrjuje tudi raznolikost pristopov k besednovrstni kategorizaciji in pomenskemu opisu diskurzno oz. pragmatično relevantne funkcijske leksike v obstoječih slovenskih slovarskih priročnikih, kot so prva in druga izdaja Slovarja slovenskega knjižnega jezika, Slovenski pravopis ali Slovar slovenskih členkov. Kot enotno izpostavljajo pregledne razprave v okviru načrtovanja bodočega enojezičnega razlagalnega slovarja, opis funkcijskega besedišča predstavlja enega največjih slovenističnih izzivov prihodnosti, saj noben od naštetih

priročnikov tega vprašanja doslej ne rešuje na zadovoljiv oz. dosleden način. Med slovarskimi priročniki tako prihaja do precejšnjih razlik, denimo v naboru in pomenskem opisu veznikov (Pisanski Peterlin, 2015a) in členkov (Balažic Bulc, 2015) ter načinu njihovega medsebojnega razmejevanja (Smolej, 2015a), pa tudi razmejevanja z drugimi besednimi vrstami, zlasti ob upoštevanju izrazne možnosti slovenščine, da lahko s pomensko konverzijo v funkcijsko povsem enakih vlogah (medmetnih, členkovnih, vezniških) nastopa oblikoskladenjsko zelo širok nabor izrazov, hkrati pa lahko posamezna oblika nastopa v več različnih funkcijah.

Med nedavnimi aplikativnimi pristopi k reševanju te problematike v slovenščini velja omeniti še **jezikovnotehnološko usmerjen slovnični opis** slovenskega jezika, kakršen je bil vzpostavljen za potrebe pridobivanja jezikovnih podatkov iz besedilnih korpusov znotraj projekta JOS¹⁵ (Erjavec in Krek, 2008; Erjavec idr., 2010). Pri besednovrstni kategorizaciji diskurzne leksike namreč poleg pomena in skladenjske vloge, iz kakršnih izhajajo prevladujoči slovarski opisi, pri razvrščanju leksike upošteva tudi stopnjo njene dvoumnosti oz. možnost prehajanja oblike med kategorijami. Tako je kategorija členka denimo zreducirana zgolj na oblike, ki ne glede na kontekst nikoli ne omogočajo zastavljanja vprašalnice (npr. *češ*), oblike, ki v rabi izkazujejo tudi prislovne interpretacije (npr. *končno*), pa so opredeljene kot prislovi. Podoben princip je upoštevan tudi pri razmejevanju medmetov od drugih oblik, ki lahko nastopajo tudi v medmetni skladenjski vlogi (npr. samostalnik *hvala*), razmejevanju členkov od veznikov, veznikov od prislovov in podobno. Konkretna kategorizacijske odločitve glede posameznih oblik so beležene v zalednem leksikonu besednih oblik Sloleks (Dobrovoljc, Erjavec idr., 2015; Dobrovoljc, Krek idr., 2015), za oblike, ki imajo v leksikonu pripisanih več možnih kategorij (npr. členek in veznik *sicer*, prislov in veznik *zato*, členek in prislov *prav*), pa so podrobnejša navodila za razvrščanje oz. razdvoumljanje v kontekstu podana v navodilih za označevanje učnega korpusa (Holožan idr., 2008). Sistem JOS tako v izhodišču ponuja nekoliko sistematičnejši pristop k jezikoslovnemu razvrščanju posameznih oblik, a je zaradi opiranja na diahrono pogojeno stopnjo oblikovne dvoumnosti s pomenskega vidika manj predvidljiv (prim. npr. prislov *mogoče* in členek *morda*), hkrati pa tudi ne povsem dosleden, saj denimo nepojasnjeno ostaja razlikovanje med prislovom in

¹⁵ <http://nl.ijs.si/jos/>

členkom *zlasti* (v primerjavi z »nedvoumnim« členkom *predvsem*), med prislovom in členkom *vsekakor* (v primerjavi s členkom *seveda*), razvrstitev veznika *vendarle* (v primerjavi s prislovom *navsezadnje*), razvrstitev členka *sicer* (v primerjavi s prislovom *torej*, kadar se ne pojavljata kot veznika) itd. Tudi v gradivno usmerjenih pristopih so torej ne glede na raven slovničnega opisa ali tip raziskave potrebne nadaljnje izboljšave, pri čemer se ob izkazani težavnosti razmejevanja diskurznofunkcijskih kategorij kot ena izmed možnosti kaže tudi združevanje posameznih besednih vrst, ki bi razlikovanje identičnih oblik na podlagi pomenskih ali distribucijskih lastnosti vpeljalo šele na raven podkategorij (glej npr. dve možni izpeljavi nabora kategorij JOS v (Krek, 2010)).

2.2 Večbesedni diskurzni označevalci

2.2.1 Raziskave večbesednih enot v analizi diskurza

Znotraj ožjega področja raziskav diskurzni označevalcev se je vprašanju večbesednih enot doslej **redko namenjala posebna pozornost**, saj se večbesednost običajno omenja le kot ena izmed izraznih možnosti te oblikoslovno in skladnejsko raznolike funkcijske kategorije. Kljub temu da so prvotne pragmatičnojezikoslovne opredelitve diskurzni označevalcev pogosto poudarjale njihovo kratkost, enozložnost in nenaglašenost, se danes z izjemo raziskav diskurzni členkov govornega jezika dolžina izrazov le redko omenja kot njihova izhodiščna prepoznavna lastnost, saj s funkcijskega vidika tovrstna zamejitev ni upravičljiva – številne raziskave namreč kažejo, da funkcije, ki se običajno pripisujejo prototipičnim enobesednim označevalcem, pogosto opravljajo tudi druge vrste struktur (Heine, 2013, str. 1211), njihov nabor pa je odvisen tudi od izraznih možnosti posameznih jezikov (Fischer, 2014, str. 173), zato je z izraznega vidika diskurzne označevalce ustrezneje obravnavati kot bolj ali manj odprto kategorijo (Schourup, 1999, str. 236), diskurzno funkcijo pa kot diahroni funkcijski potencial izrazno zelo raznolikega nabora enot (Hansen, 2006, str. 27). Vendarle pa v praksi tudi deklarativno funkcijski pristopi kratkost diskurzni označevalcev še naprej pogosto omenjajo kot *tipično* lastnost in se tudi v svojih analizah **običajno osredotočajo le na najpogostejše in najbolj ustaljene jezikovne enote** (Fischer, 2006c, str. 5), kar se odraža tudi v izbranih ponazoritvah večbesednih enot v določeni diskurzni ali pragmatični funkciji. Razmeroma redke raziskave na širšem področju analize diskurza v slovenističnem in tujem jezikoslovju, ki se ukvarjajo prav z vprašanjem večbesednosti, pa lahko glede na širši teoretski okvir oz. raziskovalne cilje razdelimo v tri ločene skupine.

2.2.1.1 Večbesedne enote v pragmatični analizi diskurza

Na področju pragmatičnih raziskav govornega jezika se z večbesednimi enotami v funkciji diskurzni označevalcev največ ukvarjajo **diahrone raziskave**, ki analizirajo procese postopne izgube propozicijskega pomena in razvoja novih, pragmatičnodiskurzni funkcij. Znotraj širše skupine besednozvezni diskurzni označevalcev, med katere L. J. Brinton in E. C. Traugott (Brinton in Traugott, 2005) denimo umeščata širok nabor izrazov, od večbesednih veznikov do vlnjastostnih obrazcev, je bilo doslej največ pragmatičnih raziskav posvečenih funkcijsko

opredeljenemu podtipu **stavčnih parentez**,¹⁶ za katere se v tuji literaturi konceptualno bolj ali manj prekrivno običajno uporabljata izraza *comment clause* (Brinton, 2008) ali *parenthetical* (Blakemore, 2006; Dehé in Kavalova, 2007; Kaltenböck, 2007). L. J. Brinton (2008) jih v kontekstu opazovanja njihovega diahronega razvoja opredeli kot stavke z izraženim glagolom, ki se od drugih stavčnih tipov razlikujejo po tem, da so skladenjsko razmeroma neodvisni od besedilne enote, v kateri se pojavijo, njihov položaj v izjavi je razmeroma prost, so prozodično ločeni in semantično izpušljivi, v diskurzu pa opravljajo podobne besedilne in medosebne funkcije kot drugi diskurzni oz. pragmatični označevalci, npr. *you see, so you see, if you will, as it were, what's more, I gather, I mean* itd.

Poleg raziskav parentez se na področju pragmatičnih raziskav večbesednih izrazov govornega jezika v zadnjem času izpostavlja še diskurzno-pragmatična vloga **konverzacijskih obrazcev** (Aijmer, 1996), kot so kletvice, vzkliki, pozdravi in vljudnostne fraze, **označevalcev odprte propozicije** (Overstreet, 2000, 2014; Verdonik, 2015b), npr. *and stuff like that*, in **vprašalnih pristavkov** (Moore in Podesva 2009), npr. *isn't it?*, sicer pa tako kot med pragmatičnojezikoslovnimi raziskavami diskurzni označevalci nasploh prevladujejo raziskave **posamičnih večbesednih enot**, torej njihovega podrobnega funkcijskega opisa, distribucije, razmejevanja s propozicijsko rabo in opazovanjem procesa gramatikalizacije, npr. izrazov, kot so *I think* (Aijmer, 1997; Simon-Vandenberg, 2000; Kaltenböck, 2009), *I believe* (Dehé in Wichmann, 2010), *you know* (B Erman, 2001; Macaulay, 2002), *I mean* (Tree in Schrock, 2002), *in fact* (Schwenter in Traugott, 2000), *of course* (Aijmer, 2013) ipd. Tej skupini raziskav je torej skupno, da skušajo pojasniti vlogo in/ali nastanek najpogostejših parentetičnih nepropozicijskih izrazov v govoru, njihova strukturna zgradba oz. večbesednost pa pri tem ni osrednjega pomena (pri nas podobno velja za obravnavo večbesednih izrazov v raziskavah D. Verdonik in M. Smolej).

V skupino pragmatičnojezikoslovnih analiz večbesednih enot s področja analize diskurza lahko v slovenskem prostoru delno uvrstimo tudi razpravo N. Jakop (2006), ki pod krovnim izrazom **pragmatični frazemi** raziskuje leksikalizirane nestavčnočlenske besedne zveze z ustaljeno in

¹⁶ Pri nas o parentezi kot komunikacijskem pojavu denimo B. Vičar (2009; 2011).

uslovarjeno pragmatično funkcijo (so celovita sporočila oz. uslovarjena govorna dejanja). Za razliko od drugih pomensko prenesenih frazemov pragmatični frazemi (v ožjem smislu) nimajo denotativnega (predmetnega) pomena, temveč lahko njihov (pragmatični) pomen opišemo zgolj z upoštevanjem okoliščin njihove rabe, torej razmerja govorca do naslovnika, zunajbesedilnih okoliščin ali besedila (Jakop, 2006, str. 144–151). Pri tem N. Jakop med pragmatične frazeme z leksikaliziranim razmerjem med govorcem in besedilom umešča funkcijsko raznolike tipe zvez v nekaterih medmetnih in členkovnih vlogah, vključno z nekaterimi zvezami, s katerimi govorec strukturira besedilo oziroma vzpostavlja razmerje do sobesedila, npr. *hočem reči, v resnici, kakor smo videli, se pravi, in tako naprej, in sicer*.

2.2.1.2 Večbesedne enote v besediloslovju

V drugo skupino raziskav večbesednih enot na področju analize diskurza lahko uvrstimo besediloslovne raziskave pisnega ali govorjenega diskurza, ki diskurzne označevalce opazujejo z vidika njihove vloge pri vzpostavljanju logičnih povezav med diskurznimi segmenti oz. njihovega členjenja. Tudi te so najprej večbesednost obravnavale le kot **eno izmed možnih izraznih variant ustaljenih jezikovnih sredstev z besedilnopovezovalno funkcijo**, kot so podredni in priredni večbesedni vezniki (*as soon as, as though, so that*), povezovalne predložne zveze (*on the other hand, in other words, in fact, after all, for example*), ali (redkeje) medmetne zveze (*sort of, you know, that's right*), kakršni se denimo pojavljajo na seznamih funkcijsko in strukturno bolj ali manj sorodno zamejenih označevalcev koherence pisnega oz. govorjenega jezika (Swan, 1992; Knott, 1996; Marcu, 2000; Louwerse in Mitchell, 2003; Prasad idr., 2007). Podobno velja tudi za slovenistične raziskave, ki pri teoretskem utemeljevanju, opisovanju ali razvrščanju jezikovnih sredstev z besedilnopovezovalno vlogo običajno vključujejo tudi ustaljene

večbesedne enote,¹⁷ pri čemer je bila v zadnjem času izpostavljena tudi potreba po dopolnitvi dosedanjih seznamov večbesednih diskurznofunkcijskih enot s korpusnimi raziskavami.¹⁸

Z odmikom od tradicionalnih, skladenjsko-funkcijskih pristopov k izraziteje funkcijsko motiviranemu preučevanju jezikovnih sredstev, ki signalizirajo koherenčne povezave med deli besedila, pa se nabor preučevanih izrazov na področju analize diskurza vedno bolj širi tudi na **druge vrste večbesednih enot, ne glede na njihovo skladenjsko vlogo, skladenjsko strukturo ali ustaljenost**. Med razčlenjevanjem prve različice korpusa Penn Discourse Treebank se je tako kmalu izkazalo, da vzpostavljanja koherenčnih povezav ni mogoče pojasniti zgolj s prepoznavo izrazov z gramatikalizirano besedilnopovezovalno funkcijo, ki pripadajo določenim skladenjskim kategorijam, na eni strani ali vstavljanjem manjkajočih povezovalcev v primeru implicitnih povezav na drugi, saj je tip povezave med dvema stavkoma na površini pogosto eksplicitno signaliziran, toda z drugimi jezikovnimi sredstvi (Prasad idr., 2010). Namesto prvotne osredotočenosti na diskurzne povezovalce (povezovalne veznike, prislove in predložne zveze) tako v širšo kategorijo t. i. označevalcev diskurznih povezav (angl. *discourse relational markers*), tj. leksikalno odprte množice sredstev, ki lahko ali pa ne pripadajo jasno opredeljivim skladenjskim kategorijam (Prasad idr., 2010, str. 1024), poleg eksplicitnih in implicitnih

¹⁷ Slovenska slovnica tako omenja različne vrste **večbesednih veznikov** (Toporišič, 2004, str. 426), kot so *zato ker, češ da, kljub temu da, takoj ko*, **dvodelne veznike** (Toporišič, 2004, str. 429), kot so *ne samo- ampak tudi; tako-kakor; ali-ali*, **besedilnozgradbene večbesedne členke** (Toporišič, 2004, str. 447–448), kot so *poleg tega, na primer*, implicitno pa tudi nekatere **predložne zveze** (Toporišič, 2004, str. 424), kot sta *v kolikor in v primeru, da*. V. Gorjanc (1998) govori o **frazoloških konektorjih** različnih oblikoskladenjskih tipov (*in sicer, to je; kljub temu da; na ta način*), Smolej (2004a) in Balažic Bulc (2009) v svojih ponazoritvah med zgledi ob enobesednih navajata tudi **večbesedne členkovne besedilne povezovalce** (npr. *kakor koli že; bolje rečeno; za začetek*) oz. **konektorje različnih pomenskih tipov** (npr. *vendar pa, kljub temu, in obratno, po drugi strani, pri tem, prav zaradi tega, v skladu s tem, ob koncu*), kot besedilnopovezovalne enote pa so opredeljene tudi nekatere **večbesedne iztočnice ali podiztočnice** v slovarjih SSKJ, SSKJ2, SP in SSČ.

¹⁸ A. Pisanski Peterlin pri primerjavi obravnave veznikov v SSKJ1, SP in SSKJ2 denimo ugotavlja, da »se zelo pomembno vprašanje odpira v zvezi z obravnavno vezniških zvez oz. večbesednih veznikov: SP in SSKJ/SSKJ2 se glede tega, katere vezniške zveze je smiselno obravnavati kot ločena gesla, ne ujemata v celoti: korpusna analiza, narejena na osnovi skladenjsko razčlenjenega korpusa, bi utegnila pomembno prispevati k bolj sistematičnemu odločanju na tem področju.« (Pisanski Peterlin, 2015a, str. 521). T. Balažic Bulc prav tako pri primerjavi različnih pristopov k leksikografski obravnavi členkov izpostavlja, da »poseben premislek zahtevajo večdelni členki oz. členkovni frazemi, predvsem zato, ker so v nekaterih klasifikacijah poleg večdelnih členkov, kot so npr. *tako ali tako, kakorkoli že, vsekakor pa* ipd., kot členki navedeni tudi jezikovni elementi, ki sodijo k besedilnim kategorijam metabesedila oz. metadiskurza, kot je npr. *po mojem mnenju, z drugimi besedami, se pravi, po drugi strani*.« (Balažic Bulc, 2015, str. 538).

povezovalcev PDTB uvršča še t. i. **alternativne leksikalizacije**. Med njimi prevladujejo večbesedne enote, saj ta skupina vključuje tako skladenjsko bolj ali manj zamejene leksikalizirane izraze (npr. *in response, quite the contrary; what's more, never mind that, to begin with*) kot skladenjsko in leksikalno povsem odprto množico besednih nizov (*A major reason is, That compares with, The increase was due mainly to, That is why*). Podobno tudi pri nas A. Skubic (1999) znotrajpovedna in medpovedna razmerja sistemsko-funkcijske slovnice poleg navajanja vezniških in členkovnih besed oz. zvez ponazarja tudi z dvema drugima skupinama funkcijsko analognih (večinoma večbesednih) izrazov. Formalno integrirane strukture so glavni stavki in odvisniki različnih (navideznih) tipov, ki so integrirani le formalno in niso del propozicije (npr. *se pravi, da; to pomeni, da; če ponazorim, kot sem rekel, ob tem lahko omenim še, da*), semantično integrirane strukture pa so besedilnopovezovalne strukture, ki so s propozicijo povezane tako izrazno kot pomensko, saj v njej sodelujejo kot prislovna določila ali prislovni odvisniki (npr. *kljub vsemu, na začetku, zaradi tega, ker je tako, v tistem trenutku, na ta način*).¹⁹

Z izjemo pristopa PDTB in na njem temelječih raziskav podobnih struktur v drugih jezikih – v češčini jih denimo M. Rysová in K. Rysová (2014, 2015) obravnavata pod pojmom sekundarni povezovalci – se sicer v korpusnih analizah označevalci diskurzni povezav proste zveze le redko eksplicitno označujejo ali vključujejo na končne sezname diskurzni označevalcev, saj se besedilnopovezovalna funkcija tovrstnih zvez pojasnjuje predvsem s **kohezivno vlogo njihovih posameznih sestavin**, npr. pomenom posameznih leksemov, njihovim pomenskim razmerjem do leksemov v sobesedilu ali navezovalnostjo (Taboada in Das, 2013; Redeker in Gruber, 2014, str. 5–6). Čeprav se pomenska nerazstavljivost, nepropozicijski pomen oz. izguba propozicijskega pomena običajno tudi v pragmatičnojezikoslovnih opredelitvah diskurzni označevalcev navaja kot lastnost, po kateri se diskurzni označevalci ločijo od drugih večbesednih jezikovnih sredstev s

¹⁹ S podskupino formalno integriranih struktur se v obravnavi (sicer naklonskih, ne besedilnopovezovalnih) stavčnih modifikatorjev propozicije oz. členkovnih stavkov ukvarja tudi Smolej (Smolej, 2015b), ki izpostavlja, da členkovni stavki pogosto nastopajo v vlogi odvisnikov, pomensko pa na propozicijsko vsebino nadrednega stavka ne vplivajo, temveč jo samo modificirajo (npr. *če kdaj, potem prav leta 1988 ni bilo razloga za beg iz Slovenije*). To neskladje med strukturno in semantično ravno imenuje psevdohipotaksa.

podobno diskurzno ali pragmatično funkcijo,²⁰ pa smiselnost in objektivno določljivost tega merila v zadnjem desetletju pod vprašaj postavljajo spoznanja korpusnojezikoslovnih raziskav formulaične narave jezika.

2.2.1.3 Raziskave formulaičnega jezika

V tradicionalnem jezikoslovju je dolgo časa prevladovalo prepričanje, da govorci v svoj mentalni leksikon shranjujemo posamične besede, ki jih take tudi prikličemo iz spomina in nato po kognitivno ločenem sistemu slovničnih pravil združujemo 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 (Firth, 1957; Bolinger, 1976; Coulmas, 1979; Pawley in Syder, 1983), nove metodološke možnosti, kakršne so se s tehnološkim napredkom in pojavom digitalizacije vzpostavile v zadnjih treh desetletjih, pa so povzročile pravi razmah empiričnih, zlasti korpusnojezikoslovnih in psiholingvističnih raziskav različnih vidikov besedne povezovalnosti, ki dokazujejo, da govorci v svoj spomin shranjujemo precej večje in kompleksnejše enote besedišča kot zgolj posamične besede. Za to sicer izredno heterogeno področje raziskav, ki se na kakršenkoli način ukvarjajo s pojasnjevanjem vzorcev besednega povezovanja v jeziku, se je v zadnjem času vzpostavilo krovno poimenovanje **raziskave t. i. formulaičnega jezika** (*formulaic language*), ki v svoji najsplošnejši opredelitvi označuje »prekinjene ali neprekinjene« (Wray, 2005) »kombinacije besed, ki vsaj v nekem vidiku niso popolnoma predvidljive, denimo zaradi svojega pomena, ki je bolj ali manj nenapovedljiv na podlagi pomena posamičnih besed, zaradi svoje funkcije, ki jo lahko doseže samo zveza kot celota, ali pa zaradi posebnosti v zgradbi, kot sta netipična morfologija ali besedni red« (Wray, 2013, str. 317). Raziskavam in teorijam formulaičnega jezika je tako skupno predvsem to, da skušajo v jeziku prepoznavati leksikalne vzorce, ki na neki točki jezikovne rabe delujejo kot nerazstavljiva enota in so kot taki tudi shranjeni v spominu govorcev, ne glede na njihovo

²⁰ G. Redeker (1991, str. 1168–1169) denimo med diskurzne označevalce prišteva samo skladenjsko neintegrirane besede in besedne zveze, kot so vezniki, prislovi, stavčne parenteze in medmeti, ne pa stavčnih kazalcev strukture diskurza, npr. *let me tell you a story, as I said before, since this is so*, ki sicer opravljajo enako (sekvenčno) funkcijo.

skladenjsko, pomensko ali drugače opredeljeno celovitost.²¹ Preko te neobremenjenosti s strukturnimi, pomenskimi in drugimi okviri besedne povezovalnosti se formulaični jezik vzpostavlja tudi kot nadpomenka frazeologije, discipline, ki v središče svojega zanimanja postavlja predvsem skladenjsko zaključene leksikalizirane večbesedne enote ter njihovo pomensko (ne)prozornost oz. (ne)razstavljivost.

Čeprav torej področje formulaičnega jezika zajema široko področje raziskav, od frazeologije do leksikogramatičnih slovničnih teorij, so razvoj področja najbolj zaznamovala predvsem spoznanja korpusnojezikoslovnih raziskav, ki se osredotočajo na analizo **statistično izstopajočih besednih nizov** (*n-grams*, *lexical bundles*, *chunks*, *clusters*), ne glede na njihovo pomensko, strukturno ali intuitivno celovitost, saj izhajajo iz prepričanja, da je prav pogostost rabe tista, ki kaže na njihovo posebno vlogo v procesu komunikacije. Na nezanemarljivost formulaičnih besednih nizov med prvimi podrobneje opozori J. Sinclair (1991), ki v svoji teoriji o prepletanju načela idiomatskosti (angl. *idiom principle*) in proste izbire (angl. *open-choice principle*) predpostavlja, da ima govorec v svojem mentalnem leksikonu na voljo številne bolj ali manj vnaprej pripravljene besedilne vzorce, pri oblikovanju besedil pa nato nenehno preklaplja med sprotnim tvorjenjem po skladenjskih pravilih (načelo proste izbire) in že pripravljenimi besednimi kombinacijami (načelo idiomatskosti), četudi se slednje zdijo razstavljive (Sinclair, 1991, str. 110). To hipotezo posredno ali neposredno potrjujejo številne korpusnojezikoslovne raziskave, ki z različnimi metodološkimi pristopi ugotavljajo, da naj bi delež vnaprej pripravljenih kombinacij predstavljal med tretjino in polovico človeškega diskurza (za pregled glej (Conklin in Schmitt, 2012, str. 2–3). Ni presenetljivo, da se pri tem večina raziskav hkrati osredotoča tako na pisni kot (spontano) govorjeni diskurz, saj je spontani govor zaradi specifičnih

²¹ Da je formulaična narava jezika izrazito multidisciplinarno raziskovalno področje, dokazuje tudi množica izrazov, ki se v strokovni literaturi uporabljajo za poimenovanje različnih jezikovnih vzorcev, bodisi sinonimno z izrazom formulaični jezik bodisi za opisovanje posameznih podskupin tega krovnega izraza, pri čemer raziskovalci dosledno opozarjajo na izmuzljivost in pogosto arbitrarnost njihovega zamejevanja oz. razvrščanja, npr. *collocation*, *idiom*, *(fixed) phrase*, *fixed expression*, *multiword (lexical) unit*, *multiword expression*, *multiword combination*, *colligation*, *chunk*, *cluster*, *formula*, *formulaic sequence*, *prefabricated routine*, *prefab*, *ready-made utterance*, *lexical phrase*, *binomial*, *morpheme equivalent unit*, *polywords*, *institutionalized utterance*, *sentence frame*, *lexical frame*, *lexicalized sentence stems*, *(language, lexical, syntactic, prefabricated) pattern*, *speech formula*, *polyword listemes*, *conversational routine*, *recurrent string*, *phraseology*, *phrase-frame*, *phrasal construction*, *conccgram*, *meaning shift unit* itd.

okoliščin tvorjenja pod časovnih pritiskom izrazito bolj formulaičen kot pisni jezik (Aijmer, 1996; Kuiper, 1996; Biber idr., 1999; Britt Erman in Warren, 2000; Biber, 2009), hkrati pa so ti nizi v govoru tudi bistveno manj variabilni, tj. ne dopuščajo paradigmatskih leksikalnih variacij posameznih sestavin (Biber, 2009; Gray in Biber, 2013).

Z vidika preučevanja večbesednih enot na področju analize diskurza je še zlasti pomembno to, da pri nadaljnjih funkcijskih opisih izluščenih nizov raziskovalci izpostavljajo prav njihovo pogosto **vlogo organizatorjev diskurza**. D. Biber s sodelavci (Biber idr., 2004) na podlagi funkcijske delitve najpogostejših besednih nizov (leksikalnih skupov, angl. *lexical bundles*) na skupe za izražanje odnosa (*stance expressions*, npr. *I don't know if, I want you to*), skupe za organiziranje besedila (*discourse organizers*, npr. *going to talk about, if you look at, on the other hand* itd.) in skupe za poimenovanje predmetnosti (*referential expressions*, npr. *those of you who, one of the things, or something like that* itd.) s primerjavo v dveh govorjenih in pisnih vrstah diskurza ugotavlja, da se raba posameznih skupin v posameznih diskurzni tipih zelo razlikuje. Raba skupov za organiziranje diskurza je izrazito pogosta v univerzitetnih predavanjih in nekoliko manj v neformalnih pogovorih ter bistveno redkejša v univerzitetnih učbenikih in znanstveni literaturi. D. Biber (2006) nato s podrobnejšo analizo pokaže, da do precejšnjih razlik prihaja tudi v posameznih podžanrih govorjenega in pisnega jezika. Čeprav v nobenem od diskurzni tipov organizatorji besedila niso najpogostejši v primerjavi z drugima dvema skupinama (v govoru prevladujejo nizi za izražanje odnosa, v pisnem jeziku pa tisti za izražanje predmetnosti), H. Nesi in H. Basturkmen (2006) v svoji raziskavi **kohezivne vloge skupov** v govorjenem akademskem diskurzu opozarjata, da logične odnose med deli diskurza ali njihovimi prehodi signalizirajo tudi številni skupi, ki so jih D. Biber in sodelavci (Biber idr., 2004) na podlagi polnopomenskih leksikalnih sestavin sicer umestili med skupe za poimenovanje predmetnosti ali izražanje odnosa, in opozarjata, da je delež formulaičnih nizov s prevladujočo diskurzno vlogo v resnici še precej večji oziroma večinski.

Raziskave formulaičnih besednih nizov s svojim teoretsko neobremenjenim, popolnim korpusnim pristopom, ki se pri načrtovanju in analizi korpusnih podatkov ne naslanja na ugotovitve ali hipoteze predhodnih jezikoslovnih teorij (Tognini-Bonelli, 2001, str. 84–87), vključno z avtomatsko ali ročno pripisanimi podatki o lemi, besedni vrsti in drugih slovničnih interpretacijah (Biber, 2009, str. 276–278), tako dokazujejo, da je organizacija posredovanih sporočil ena

temeljnih funkcij v procesu sporazumevanja, pri njenem uresničevanju pa se govorci pogosto zatekamo k rabi formulaičnih besedilnih vzorcev, ki brišejo mejo med besediščem in slovnico jezika. Ni torej presenetljivo, da **se termin diskurzni označevalec v izrazito funkcijskih pristopih vse bolj odpira tudi proti vključevanju formulaičnih nizov**, kjer sta pomenska razstavljivost in strukturna enovitost drugotnega pomena. Za razliko od pojmovanja alternativnih leksikalizacij v PDTB (Prasad idr., 2010), ki vključujejo tudi proste besedne zveze, pa je po drugi strani v raziskavah formulaičnega jezika ključna njihova **pogostost oz. institucionaliziranost** (Sag idr., 2002). D. Siepmann (2005) v eni redkih raziskav, posvečenih prav večbesednim diskurznim označevalcem, na podlagi Sinclairjeve teorije te večbesedne enote (imenuje jih **diskurzni označevalci drugega nivoja**) opredeli kot »ustaljene izraze ali kolokacije, sestavljene iz dveh ali več tiskanih besed, ki delujejo kot ena sama enota. Njihova funkcija je lajšanje procesa interpretiranja koherentnih povezav med elementi, nizi ali besedilnimi segmenti in/ali vidiki komunikacijske situacije« (Siepmann, 2005, str. 52). Primerja rabo večbesednih diskurznih označevalcev v pisnih besedilih rojenih in tujih govorcev angleškega, francoskega in nemškega jezika, in sicer se osredotoči predvsem na formulaične označevalce ponazoritve (*for the sake of illustration ...*), reformulacije (*to put it crudely...*) in sklepanja (*so it turns out that ...*), v tej in poznejši razpravi (Siepmann, 2006) pa opozori tudi na pomembnost njihovega vključevanja v enojezične in večjezične razlagalne slovarje.

2.2.2 Prepoznavne lastnosti večbesednih diskurznih označevalcev

2.2.2.1 Nepropozicijskost

Kot ena temeljnih prepoznavnih lastnosti različnih opredelitev diskurznih označevalcev se zlasti v pragmatičnojezikoslovnih teorijah pogosto pojavlja pojmovanje diskurznih označevalcev kot izrazov, ki ne prispevajo veliko k vsebini diskurza, temveč v diskurzu opravljajo neko drugo funkcijo. Za to lastnost različni avtorji uporabljajo različne ubeseditve, ki se običajno vzpostavljajo kot **negacija prispevka k tako imenovani propozicijski vsebini diskurza**, npr. da diskurzni označevalci ne vplivajo na propozicijski pomen sporočil, delujejo izven propozicije, ne vplivajo na resničnostno vrednost izjave, imajo nepropozicijsko funkcijo, ne prispevajo k propozicijski vsebini enot, v katerih se pojavljajo, so semantično neobvezni, njihov izpust ne vpliva na resničnostno vrednost sporočila in podobno. Čeprav večina avtorjev v svojih

opredelitvah diskurzni označevalcev idejo nepropozicijskosti jemlje kot splošno sprejeto jezikovno danost, brez posebnih pojasnil ali opiranja na teoretski kontekst, v okviru katerega te pojme razumejo, nadaljnji funkcijski opisi in ponazoritve nabora in funkcij diskurzni označevalcev implicitno razkrivajo neenotno razumevanje tega, kaj je propozicija in na njej temelječe pojmovanje (propozicijske) vsebine diskurza.

Med najožjimi pristopi k pojmovanju nepropozicijskega v besedilu je skladijsko-semantično pojmovanje **propozicije kot pomenske podstave površinskih skladijskih struktur besedila** (stavkov ali povedi), kakršno se je vzpostavilo tudi v slovenskem jezikoslovju (Kunst-Gnamuš, 1981; Toporišič, 2004). Propozicija je tako opredeljena kot skupek pomena povedja (npr. dejanje, stanje ipd.), delovalnikov (npr. vršilec dejanja, prejemnik dejanja, prizadeti) ter pomensko obveznih ali neobveznih okoliščin (npr. kraj, čas, namen, orodje ipd.), kar se na površini običajno prevede v kombinacijo glagola (povedka) in njegovih vezljivostnih argumentov oz. dopolnil (stavčnih členov). Propozicija je torej vmesna stopnja med govorčevim dojetjem sestave dogodka, dejanja, stanja ali dogajanja v zunanjem svetu in njegovo končno ubeseditvijo v besedilu, ki se ji lahko glede na komunikacijski namen poleg ubeseditve enot propozicije dodajo še drugi izrazi, ki posamezne enote ali celotno propozicijo modificirajo.²² Na podoben način, a z manj predpostavkami o obstoju globinskih pomenskih struktur in njihovih interakcij s površinskim besedilom, propozicijo opredeljuje tudi sistemska-funkcijska slovnica (Halliday in Matthiessen, 2014, str. 138), kot rabo stavka za izmenjavo informacij, nečesa, kar lahko potrdimo ali zanikamo, v kar lahko podvomimo, čemur lahko nasprotujemo, kar lahko sprejmemo, obžalujemo itd.²³

²² Pri tem je slovenski izraz pomenska podstava *povedi* v tem smislu nekoliko zavajajoč, saj ima posamezna poved (če jo razumemo kot zaključeno sporočilo oz. izjavo) lahko več pomenskih podstav (propozicij), če opisuje več ločenih dogodkov ali stanj. Tudi J. Toporišič (2004, str. 491) navaja primere povedi iz več pomenskih podstav: npr. podstave [1. biti vroče], [2. oditi, jezero, vsi], [3. kopati se, vsi] se lahko realizirajo v večstavčni povedi *Bilo je vroče, zato so vsi odšli k jezeru in se kopali*. Obenem tudi ni nujno, da so na površini vse pomenske podstave dejansko realizirane kot glagolski stavki, saj je povedje lahko izraženo tudi samostalniško ali pridevniško (npr. *Zaradi vročine so se vsi odšli kopat v jezeru*).

²³ S tem tudi Halliday implicira razumevanje propozicije kot nekega potencialnega dejstva v svetu, kakršnega je mogoče preverjati s filozofskim testom resničnosti, tj. določanjem pogojev, v katerih propozicija drži. Ta test onkraj teorij formalne semantike danes sicer ni več merodajen, saj so ga številni avtorji zaradi neupoštevanja kontekstualnih vidikov, ki vplivajo na pomen jezikovnih sredstev, za določanje obsega in pomena povezovalcev zavrgli kot

Kot lahko razberemo iz pregleda A. Pisanski Peterlin (2007, 2011, str. 15–16), je prav Hallidayeva slovnica v zadnjih nekaj desetletjih močno vplivala tudi na ločevanje propozicijske in drugih ravni besedila na širšem področju analize diskurza, kot so denimo raziskave metabesedilnosti. Toda bolj kot prek zgoraj omenjene dejanske opredelitve propozicije kot sporočilne pomenske vrednosti nekaterih stavčnih tipov (sistemska-funkcijska slovnica se namreč ukvarja predvsem s pojasnjevanjem pojavov in funkcij na ravni stavka), ki ni neposredno prenosljiva na nadstavčno raven, je bila na področju analize diskurza vplivna Hallidayeva uvodna opredelitev treh temeljnih (meta)funkcij jezika: predstavne, ki se nanaša na zunajjezikovno človeško izkustvo, medosebne, ki se nanaša na odnose med udeleženci, in besedilne, ki se nanaša na tematsko in kohezijsko povezanost besedila (Halliday in Matthiessen, 2014, str. 30–31). V splošnem se torej v teorijah, ki temeljijo na Hallidayevem funkcijskem modelu, **propozicijska vsebina povezuje s predstavno funkcijo, nepropozicijska pa z medosebno in besedilno funkcijo**, saj izrazi s prevladujočo medosebno in besedilno funkcijo niso del propozicije, temveč jo le modificirajo. Z vidika razvrščanja besedilnopovezovalne leksike pa je treba pri tem izpostaviti, da Halliday predstavno funkcijo z vidika opisovanja funkcij stavčnih gradnikov dodatno razdeli še na dve podkategoriji: **izkustveno in logično** (Halliday in Matthiessen, 2014, str. 361–362), in sicer logična označuje rabo jezika za vzpostavljanje strukturnih razmerij med posameznimi enotami stavka, ki so lahko v primeru stavčnega kompleksa oz. zložene povedi (*clause complex*) tudi medstavčni vezniki. Če torej povezovalcem samostojnih stavkov, stavčnih kompleksov oz. povedi (*conjunctive adjuncts*, povezovalni adjunkti) pripiše besedilno metafunkcijo, povezovalcem stavkov v stavčnem kompleksu oz. zloženi povedi (*conjunctions*, vezniki) zaradi njihove strukturne vloge pripiše predstavno-logično funkcijo, priznava pa, da s **semantičnega vidika obe skupini opravljata enako vlogo**.²⁴ Podobno skladenjsko utemeljeno

neuporabnega, zlasti pod vplivom teorije relevantnosti, ki raziskuje razmerje med konceptualnim in neresničnostnopolgojnim (za pregled glej (Blakemore, 2002)).

²⁴ »The conjunctive Adjuncts (often called 'discourse Adjuncts'), as noted above, cover roughly the same semantic space as the conjunctions; but whereas conjunctions set up a grammatical (systemic-structural) relationship with another clause, which may be either preceding or following, the relationship established by conjunctive Adjuncts, while semantically cohesive, is not a structural one (hence they can relate only to what has gone before).« (Halliday in Matthiessen, 2014, str. 110). Sopostavljanje besedilnih in logičnih povezovalcev s sorodnim pomenom je nazorno prikazano in povzeto v (Halliday in Matthiessen, 2014, str. 670–673).

argumentacijo lahko v slovenistični slovnični tradiciji denimo zasledimo pri ločevanju vezniških besed od besedilnopovezovalnih členkov.

Še dodatno pojem propozicijske vsebine zamejijo novejša teorije metabesedilnosti (Hyland 2005), ki iz metabesedilnosti proti propozicijskosti potisnejo vse, kar ima referenco v zunanjem svetu, s čimer med propozicijsko vsebino uvrstijo tudi tiste povezovalce s (Hallidayevo) besedilno funkcijo, ki med povedmi vzpostavljajo zunanja, objektivna oz. semantična razmerja, kot metabesedilne elemente pa prepoznavajo zgolj tiste, ki delujejo znotraj besedila in to besedilo strukturirajo.²⁵ Znotraj teorij metabesedilnosti se torej **nepropozicijskost povezuje predvsem s subjektivnim oz. pragmatičnim upravljanjem diskurza** (glej tudi poglavje 2.2.2.2.3 *Izvor povezave* v nadaljevanju), ne pa z drugimi vrstami vzpostavljanja medpovednih ali medstavčnih povezav.

Kot smo že omenili, umeščanje diskurzni označevalcev glede na dihotomijo propozicijsko – nepropozicijsko prevladuje predvsem v pragmatičnojezikoslovnih pristopih. Teorije koherence v diskurzu po drugi strani povezovalce umeščajo na raven propozicijske vsebine, a jih opredeljujejo kot nosilce **posebne vrste propozicij** oz. pomenov. Mann in Thompson (1986) tako poleg propozicij, ki jih izražajo posamezni stavki (imenujeta jih leksikalne propozicije), vpeljeta še pojem relacijskih propozicij (retoričnih povezav), ki nastajajo iz njihovih kombinacij in so pogosto implicitne, npr. vzrok, pojasnilo, pogoj ipd. Medtem ko leksikalne propozicije sestavljajo semantični odnosi med povedkom in njegovimi argumenti, so **relacijske propozicije funkcijski odnosi med stavki, povedmi in daljšimi besedilnimi elementi**, zaradi vzpostavljene povezave pa je pomen povezane celote več kot le seštevek pomena posameznih enot (Sanders idr., 1992, str. 2). Podobno pri nas tudi A. Žele (Žele, 2014a) omenja dve glavni sestavini besedilnega sporočila: vsebino denotatov (pomen sporočila) in vsebino denotatnih razmerij (smisel in logičnost sporočila), slednje pa lahko opazujemo tako na ravni besedila kot na ravni propozicij. Glede na to, ali gre za razmerja med več ločenimi povedmi na eni strani ali za razmerja med stavki v zloženi povedi na drugi, jih imenuje besedilna oz. povedna medpropozicijska razmerja.

²⁵ A. Pisanski Peterlin (2011, str. 17) denimo navaja metabesedilne rabe *povezovalcev in, toda, namreč, npr., nazadnje* ipd.

Relacijske propozicije se tako vzpostavljajo kot krovno poimenovanje tako za medstavčna kot medpovedna razmerja, s čimer se s funkcijskega vidika briše Hallidayeva meja med logično (medstavčni povezovalci) in besedilno (medpovedni povezovalci) funkcijo, hkrati pa tudi metabesedilna meja med zunanjimi/semantičnimi in notranjimi/pragmatičnimi povezovalci, saj vsa ta sredstva signalizirajo relacijske propozicije v diskurzu, le drugačnega tipa.

Če torej povzamemo, kljub pogostim omembam nepropozicijskosti kot temeljne prepoznavne lastnosti diskurzni označevalcev različne teorije propozicijsko vsebino razumejo različno in glede na izbrana izrazna in pomenska merila vanjo vključujejo ali iz nje izključujejo različna jezikovna sredstva,²⁶ vključno z različnim naborom diskurzni označevalcev glede na to, kakšne strukturne enote povezujejo in kakšna je narava teh povezav. Sami bomo v nadaljevanju izraz *propozicija* tako uporabljali v skladenjsko-semantičnem pomenu, torej kot stavčno izražene informacije o glagolskem dogodku, njegovih udeležencih in okoliščinah, ne le zato, ker je to minimalni teoretski konsenz, kakršnega v tem sintetičnem pregledu iščemo, ampak tudi zaradi težavne dosledne implementacije meril, kakršne pri svojih opredelitvah propozicijske ravni omenjajo druge teorije, zlasti pri načrtovanem delu z obsežnejšim govorjenim gradivom (npr. pojmovanje stavka pri Hallidayu; razdvoumljanje med zunajbesedilno in metabesedilno rabo pri Hylandu), kar podrobneje ponazorimo v nadaljevanju. Za posredovanja izkustvenih informacij o svetu pa namesto izraza propozicijska vsebina raje uporabljamo izraz *predstavna ali ideacijska vsebina* diskurza, ki vključuje tako propozicije kot njihova objektivna pomenska medpropozicijska razmerja. V naslednjem poglavju raziščemo, kam različni pristopi v širokem spektru leksike izven stavčno izražene propozicije umeščajo diskurzne označevalce oz. izraze s prevladujočo diskurzno funkcijo, torej kaj v širokem naboru nepropozicijske leksike je dejansko v funkciji diskurza.

²⁶ Če kot primer vzamemo štiri različne ubeseditve vzročno-posledične povezave med Markovo boleznijo in njegovim odhodom domov – [1] *Marko je zbolel, zato je šel domov*, [2] *Ker je Marko zbolel, je šel domov*, [3] *Marko je zbolel, torej je šel domov*, [4] *Marko je zbolel, šel je torej domov* – bi sistemsko-funkcijska slovnica besedilno (nepropozicijsko) vlogo pripisala vsem primerom razen podrednemu vezniku *ker* v [3], ki opravlja logično predstavno (propozicijsko) funkcijo, teorija metabesedilnosti bi iz metabesedilne (nepropozicijske) vsebine izločila še veznik *zato* v primeru [1], saj ta za razliko od subjektivne sklepalne funkcije veznika *torej* v primerih [3] in [4] vzpostavlja semantično povezavo med dvema propozicijama, teorije koherence pa med posameznimi primeri ne ločujejo, saj vsi konektorji opravljajo vlogo vzpostavljalcev medpropozicijskih povezav (različnih podtipov).

2.2.2.2 Diskurzna funkcija

2.2.2.2.1 Povezovalnost

V različnih opredelitvah diskurzni označevalcev se kot njihova osnovna pomenska lastnost, ki jo v nadaljevanju imenujemo diskurzna funkcija, običajno izpostavlja dejstvo, da ti izrazi sami po sebi nimajo referenčnega, propozicijskega ali konceptualnega pomena, temveč v diskurzu opravljajo vlogo deiktičnih sredstev, ki jezikovni kontekst, v katerem se pojavijo, povezujejo z drugimi deli diskurza (Schourup, 1999, str. 231). Kot konsezualno pomensko lastnost diskurzni označevalcev lahko torej prepoznamo idejo besedilno vpete **povezovalnosti** (Schourup, 1999, str. 231), **deiktičnosti** (Fischer, 2006c, str. 11–12) oziroma **indeksikalnosti** (Schiffrin, 1987, str. 322–325), ki diskurzne označevalce sicer razmeroma jasno ločuje od nekaterih drugih nepropozicijskih oz. pragmatičnih izrazov, kot so čustvenonaklonski izrazi ali vzkliki (Blakemore, 2008, str. 221–222),²⁷ a hkrati odpira zelo široko polje interpretacij, kaj sploh so enote diskurza, ki naj bi jih ti izrazi povezovali, označevali ali nanje kazali (vprašanje dometa), ter tudi, kako pomembno je pri tem razmerje med besedilom in njegovim tvorcem (vprašanje izvora). Kot bomo videli v nadaljevanju, sta prav ti dve vprašanji ključni za razumevanje razdrobljenosti področja raziskav diskurzni označevalcev, saj poleg temeljnih konceptualnih razlik med različnimi teorijami pojasnjujejo tudi pogosta razhajanja med raziskovalci s sorodnimi izhodišči ali raziskovalnimi cilji.

2.2.2.2.2 Domet

Vprašanje dometa diskurzni označevalcev je pravzaprav dvojno: **kaj so temeljni gradniki diskurza, ki naj bi jih označevalci povezovali (diskurzne enote)**, po drugi strani pa tudi, **ali so te enote res zgolj enote besedila**. Različne pristope k opredelitvam funkcije diskurzni označevalcev glede na njihov domet lahko v splošnem umestimo na različne točke med dvema skrajnima poloma: razumevanjem diskurza kot niza besedilno izraženih propozicij na eni strani in razumevanjem diskurza kot procesa interakcije. Pri tem poudarimo, da to pojmovanje ni nujno

²⁷ Še širši koncept, ki poleg diskurzne povezovalnosti vključuje tudi funkcije drugih nepropozicijskih izrazov, je denimo refleksivnost (Aijmer idr., 2006, str. 105–106).

pogojeno s tipom preučevanega diskurza (npr. pisni proti govorjeni jezik), temveč je odvisno predvsem od izbranih teoretskih predpostavk ali raziskovalnega cilja. Vprašanja funkcijskega dometa diskurzni označevalcev, ki so se prvotno izpostavljala zlasti pod vplivom raziskav govorjenega jezika, namreč niso nujno vezana zgolj na govorjeni jezik, so pa zanj bistveno značilnejša.

2.2.2.2.2.1 Diskurzni označevalci kot povezovalci besedilnih enot

V prvi skupini raziskav, kamor se tipično umeščajo teorije koherence in druge besediloslovne raziskave strukture (prevladujoče pisnega) diskurza, prevladuje predvsem vprašanje, kaj so pravzaprav **temeljne enote besedila**, ki prispevajo k mentalni izgradnji besedilnega sveta in med katerimi se implicitno ali eksplicitno vzpostavljajo koherenčne povezave. Za razliko od starejših besediloslovnih teorij, ki kot najmanjšo enoto besedilne skladnje jemljejo (eno- ali večstavčno) poved (Halliday in Hasan, 1976; Halliday in Matthiessen, 2014), je teorijam koherence skupno to, da **kot tipično minimalno diskuržno enoto običajno omenjajo stavek**, pri čemer pa se konkretne zamejitve razlikujejo od izbranih skladenjskih in/ali semantičnih kriterijev. W. C. Mann in S. Thompson (1988) v postavitvi teorije retorične strukture denimo kot minimalne enote diskurza opredelita stavke, z izjemo stavčno izraženih osebkih, predmetnih in omejevalnih oziralnih odvisnikov, ki jih obravnavata kot del nadrednih struktur in ne kot samostojne enote, po drugi strani pa kot samostojne enote poleg glavnih stavkov obravnavata druge vrste podrednih stavkov, denimo prislovne odvisnike. Izhajata namreč iz predpostavke, da je skladenjsko podredje stavka zgolj posledica gramatikalizacije določene retorične povezave in h koherenci diskurza prispeva na povsem enak način kot prirednost (Matthiessen in Thompson, 1988). Določanje dejanskih semantičnih hierarhičnih razmerij zato opisujeta šele na ravni določanja nadrejenega in podrejenega elementa retorične povezave.

Semantični pristopi k segmentaciji diskurza po drugi strani stavek omenjajo kot *tipično* enoto diskurza, a je ta lahko izražena tudi z nestavčnimi strukturami. Asher (1993), na čigar teoriji denimo temelji pristop PDTB, tako diskurz raje opisuje kot niz **abstraktnih objektov**, ki označujejo dogodke, stanja, dejstva, možnosti, propozicije ipd. (Asher, 1993, str. 57), ki so lahko na površinski ravni izraženi z različnimi skladenjskimi strukturami, od nominalizacij, povedkovih prilastkov in apozicij do celotnih odstavkov. Tudi M. Taboada in W. C. Mann (2006b, str. 429–

430) ugotavljata, da je lahko skladenjsko pogojena stavčna segmentacija za posamezne raziskovalne cilje problematična, saj je lahko bodisi preveč ali premalo podrobna, zaradi vezanosti na skladenjske lastnosti posameznega jezika onemogoča večjezično primerjavo in tudi ni neposredno prenosljiva v raziskave govornega jezika, kjer se kot enote običajno uporabljajo prozodična in ne nujno skladenjska merila.

Kot v enem izmed novejših prispevkov k empirično utemeljenemu razumevanju segmentacije govornega diskurza ugotavljata L. Degand in A. C. Simon (2009), kljub relevantnosti raziskovalnega cilja doslej v literaturi namreč še ni konsenza, kaj točno naj bi bila **besedilna enota govornega diskurza** in kako jo prepoznati. V pregledu različnih pristopov (Degand in Simon, 2009, str. 2–6), med katerimi jih nekaj pri nas predstavlja tudi M. Schlamberger Brezar (2009, str. 57–77), avtorici povzemata, da se pri segmentaciji govornega diskurza odvisno od teoretskega modela poudarja bodisi razčlenjevanje glede na semantične entitete (posredovane ideje) bodisi glede na kontekstualno pogojene površinske segmente (enote besedila ali vloge). Pri njihovem določanju so pomembni tako skladenjski kot prozodični namigi: skladenjske strukture se pogosto povezujejo s konceptualno strukturo diskurza kot **nosilci semantične vsebine** (ideje, reference, dogodki), prozodične strukture pa se povezujejo s postopnimi **koraki v razvoju diskurza**. Pri določanju temeljne minimalne enote diskurza se teorije torej razlikujejo v tem, ali dajejo prednost skladnji (diskurzu kot produktu), kot npr. ženevska šola (Filliettaz in Roulet, 2002) ali teorije koherentnih povezav, ali intonaciji (diskurzu kot procesu), kot npr. konverzacijska analiza (Selting, 2000; Schegloff, 2007), v omenjeni razpravi pa avtorici predlagata nov pristop, ki združuje oba vidika. Na podlagi rezultatov neodvisnega dvonivojskega razčlenjevanja korpusa govorne francoščine temeljno diskurzno enoto namreč opredelita kot rezultat hkratnega sovpadanja skladenjske in prozodične meje, ki se lahko zato realizira v več različnih tipih temeljnih diskurznih enot: enote, kjer ena skladenjska enota (stavek) sovпада z eno intonacijsko enoto; enote, pri katerih en stavek razpade v več intonacijskih enot; in enote, kjer se več stavkov pojavlja znotraj ene intonacijske enote. Tudi v raziskavah govornega jezika različni prepleti skladenjskih, semantičnih in prozodičnih meril vodijo v raznolika in bolj ali manj objektivno določljiva pojmovanja diskurznih enot ter z njimi povezanih pojmov, kot so **izrek, izjava, enota govora, idejna enota, enota sporočila, govorno dejanje, ilokucijsko dejanje**

ipd., ki pa se ne glede na izbrani teoretski model pogosto zelo pomanjkljivo pojasnjeno navajajo v opredelitvah vloge diskurzni označevalcev.

Kot povezovalci različno opredeljenih minimalnih besedilnih enot ali njihovih kompleksnih kombinacij diskurzni označevalci znotraj besedilnoskladenjskih teorij koherence tako tipično vključujejo **podredne** (npr. *although*) in **priredne** (npr. *but*) **veznike**, povezovalne **prislove** oz. **členke** (npr. *nevertheless, nonetheless*) ali predložne zveze (npr. *for this reason*), na svoji spodnji meji pa včasih tudi **predloge** (npr. *in order to, despite*).

2.2.2.2.2 Diskurzni označevalci kot povezovalci besedila in sporazumevalnega konteksta

Bolj kot vprašanje razčlenjevanja govornega diskurza na posamične *besedilne* enote pa so raziskave govornega diskurza vplivale na širjenje prvotnega pojmovanja strukture diskurza kot niza soležnih besedilnih enot oz. koherence kot rezultata njihove smiselne povezanosti. G. Redeker (2000, str. 1–3) tako opozarja, da omejevanje raziskav na razmeroma ekstremne žanre, kot so predstavitvena pisna besedila na eni strani in izrazito interaktivna dialoška besedila na drugi, povzroča pretirano osredotočanje na specifične posameznega diskurzne tipa, ki onemogoča oblikovanje enotne teorije koherence v diskurzu. Kot smo izpostavili že v predstavitvi njenih teoretskih izhodišč (poglavje 2.1.2.2 *G. Redeker in model vzporednih komponent*), pri tem izpostavlja predvsem potrebo po širšem razumevanju strukture diskurza, ki poleg vzpostavljanja predstavni (ideacijska struktura) in retoričnih (retorična struktura) povezav med soležnimi besedilnimi enotami, kakršne izpostavljajo tudi druge besediloslovne teorije, vključuje **tudi vzpostavljanje povezav med besedilno enoto in njenim širše razumljenim diskurznim kontekstom**, kot so prehodi med temami, digresije, parafraze, prekinitve, odzivi, menjave vlog ipd. (sekvenčna struktura), ki so zlasti pogoste v govornih besedilih.²⁸ V tovrstnem pojmovanju diskurza je torej jezik njegova nujna sestavina, a njegov neločljivi del predstavlja tudi zunajbesedilni kontekst (Redeker, 2006, str. 357).²⁹

²⁸ Za seznam koherentnih povezav znotraj vsake izmed treh struktur glej Redeker (2000, str. 15), za njihovo sovpadanje in seznam označevalcev tudi Redeker (2006, str. 345, 354).

²⁹ Podobno tudi Hansen (1998a, 2006) v svoji opredelitvi diskurzni označevalcev kot nepropozicijskih in metadiskurzni izrazov, ki naslovniku dajejo navodila, kako naj izjavo gostiteljico, v kateri se pojavljajo, integrira v

Podobno tudi U. Lenk (1998) izpostavlja ločnico med **lokalnim in globalnim dometom** označevalcev, saj diskurzni označevalci v govorjenem jeziku poleg vzpostavljanja lokalne koherence (povezovanja med dvema soležnima izjavama) vzpostavljajo tudi globalno koherenco, tj. označujejo povezavo z drugimi segmenti diskurza, kot so predhodne teme, tema pred digresijo, teme, ki bodo šele sledile, pa tudi s situacijskim ali splošnim védenjem, ki predhodno v diskurzu ni bilo omenjeno (Lenk, 1998, str. 247). Sama med diskurzne označevalce na podlagi te ločnice umesti samo izraze z globalnim dometom, podobno pa tudi R. Weltereit (2006) izpostavlja, da naj bi bila ključna prepoznavna lastnost diskurznih označevalcev, na podlagi katerih jih sam tudi ločuje od veznikov, dejstvo, da njihovega dometa ni mogoče opisati s slovničnimi kategorijami, kot sta stavek ali poved, temveč z vidika tvorčevih diskurznih strategij oz. namer (Waltereit, 2006, str. 64–66). Čeprav teorija relevantnosti v svojih opredelitvah diskurznih označevalcev ne razlikuje med različnimi dometi, pa je globalni domet pogosto izpostavljen kot argument zavračanja predpostavk, na katerih temeljijo teorije koherence, saj naj bi prav raba diskurznih označevalcev v navezovanju na zunajbesedilni kontekst dokazovala, da jih ne moremo razumeti kot vzpostavljavce besedilne strukture, temveč kot usmerjevalce interpretacije, kar npr. V. Rouchota (1996) ponazori s primeri, kjer je povezovalac rabljen v prvi izrečeni izjavi dialoga kot reakcija na zunajjezikovni dogodek (npr. raba sklepalnega označevalca *torej* v izjavi »*Torej skušaš ostajati fit*«, s katero se govorec odzove na naslovnikov povratek z rekreativnega teka).

S širitvijo ciljnega dometa diskurznih označevalcev s sobesedilnega na širše pojmovani kontekst se glede na težavno opredelitev samega koncepta sporazumevalnega konteksta (za pregled glej (Saint-Georges, 2013), pri nas (Verdonik, 2013)) seveda odpirajo nova vprašanja, povezana z zamejitvijo in funkcijskim opisom diskurznih označevalcev. Med njimi se v zadnjem času izpostavlja predvsem vprašanje razmejčitve diskurznih označevalcev od t. i. **modalnih členkov** (angl. *modal particles*).³⁰ Izraz se je v literaturi s področja diskurznih označevalcev uveljavil

svoj mentalni model diskurza, da bo ta izjava optimalno koherentna (Hansen, 2006, str. 25–26), izpostavlja, da mentalno predstavo diskurza poleg védenja, ki se je ustvarilo z besedilom neločljivo soustvarjata tudi situacijski in kognitivni kontekst, pri čemer pa za razliko od Redekerjeve loči med transakcijsko koherenco in interakcijsko koherenco (Hansen, 1998a, str. 180), s čimer iz razreda diskurznih označevalcev, katerih primarna naloga je vzpostavljanje transakcijske koherence, denimo izloči označevalce menjave vlog.

³⁰ Termin ni povsem prekriven s slovenističnim pojmovanjem členkov, modalnih členkov ali naklonskih členkov (prim. npr. različne opredelitve v (Skubic, 1999; Smolej, 2004a; Žele, 2014b)), temveč zajema le eno izmed njihovih

zlasti preko raziskav nemškega jezika, kjer se ta kategorija izrazov (nem. *Modalpartikel*, tudi *Abtönungspartikel*) običajno omenja kot samostojna besedna vrsta, ki jo sestavljajo skladenjsko integrirane, običajno nepoudarjene nepregibne besede, ki se pojavljajo za pregibno obliko povedka in modificirajo pomen celotnega stavka tako, da govorno dejanje (tj. ilokucijsko silo propozicije)³¹ **povezujejo s skupnim védenjem oz. prepričanjem, ki ga govorec pripisuje naslovniku** (Diewald, 2006; Waltereit in Detges, 2007; Bross, 2012).³² Ker so modalni členki tako diahrono (v obeh kategorijah lahko pogosto nastopajo isti izrazi) kot funkcijsko (obema kategorijama je skupno povezovanje izjave z njenim kontekstom) podobni tako ali drugače opredeljenim diskurznim označevalcem, v literaturi ni enotnega konsenza, ali gre za dve ločeni jezikoslovni kategoriji ali pa so modalni členki le ena izmed podkategorij diskurznih označevalcev (Degand idr., 2013b).

Med zagovorniki ločevanja ene in druge skupine so zlasti raziskovalci diskurznih členkov znotraj konstrukcijske slovnice (Aijmer, 2013; Diewald, 2013; Fischer in Alm, 2013), ki menijo, da se različne distribucijske lastnosti (skladenjski položaji) istega izraza povezujejo tudi s specifičnimi, med seboj povsem ločenimi pomeni oz. funkcijami.³³ Pri tem raziskave v drugih jezikih kažejo, da izrazi z ustaljeno funkcijo označevanja že znane informacije niso omejeni zgolj na nemščino ali njej sorodne jezike, temveč se pojavljajo tudi v številnih drugih tipološko različnih jezikih (glej zbornik raziskav (Degand idr., 2013a)), a v drugačnih skladenjskih položajih, slovnice posameznih jezikov pa jih pogosto umeščajo med prislove (prim. npr. *Kako si torej to naredil?* proti *How did you do it, then?*). Po drugi strani avtorji, kot so Schoonjans (2013), Cuenca

podskupin oz. specifične rabe posameznih oblik. Za izčrpno primerjavo nemških modalnih členkov in njihovih slovenskih ustreznikov glej (Petrič, 1999).

³¹ Običajno se posamezni modalni členki povezujejo s specifičnimi vrstami govornih dejanj, npr. vprašanji, ukazi. trditvami ipd., za pregled posameznih tipov govornih dejanj in členkov, ki se povezujejo z njimi, glej (Bross, 2012, str. 188–189).

³² Bross (2012) na podlagi različnih klasifikacij omenja dva razreda: jedrne modalne členke (*aber, auch, bloß, denn, doch, eben, eigentlich, etwa, halt, ja, mal, nur, schon, vielleicht and wohl*), ki se lahko pojavljajo zgolj v sredinski poziciji, a imajo homonime v drugih besednih vrstah, ter obrobne modalne členke (*fein, ganz, gerade, gleich, einfach, erst and ruhig*), ki sicer nimajo homonimov, a se lahko pojavljajo tudi na začetku stavka.

³³ V nemščini naj bi tako povezave med besedilnimi enotami vzpostavljali skladenjsko integrirani vezniki ali povezovalni prislovi, povezave z nepropozicijskimi vidiki sporazumevanja (menjava vlog, upravljanje govora in medosebnih odnosov, struktura tem, okvir sodelovanja ipd.) skladenjsko neintegrirani diskurzni členki, povezave z besedilno neizraženim skupnim védenjem pa modalni členki (Diewald, 2013).

(Cuenca in Marín, 2012; 2013) ali Squartini (2013) opozarjajo, da meje med obema kategorijama zaradi prekrivnih funkcij in kontekstualno pogojenih interpretacij niso zlahka določljive, zato jih je smiselneje opazovati kot točke na istem funkcijskem kontinuumu, nekateri avtorji pa pri svojih opredelitvah dometa diskurznih označevalcev niti ne razlikujejo med povezovanjem z že znanim skupnim védenjem in z védenjem, ki se je dinamično oblikovalo šele v procesu sporazumevanja (Roulet, 2006, str. 120; Zeevat, 2006).³⁴ Razmerje med označevalci skupnega védenja in razredom diskurznih označevalcev je tako odvisno predvsem od raziskovalčeve osredotočenosti na posamezna funkcijska in/ali izrazna merila (Degand idr., 2013b, str. 15).

2.2.2.2.3 Diskurzni označevalci kot označevalci procesa govorjene interakcije

Če torej prva skupina raziskav domet diskurznih označevalcev zameji na povezovanje med besedilno enoto in njenim ožjim ali širšim sobesedilnim kontekstom, druga skupina na povezovanje med besedilno enoto in njenim sobesedilnim in zunajbesedilnim kontekstom, pa se tretja skupina v svojem pojmovanju dometa diskurznih označevalcev razbremeni predpostavke, da se diskurzni označevalci povezujejo (zgolj) s površinsko organizacijo besedila, temveč delujejo predvsem kot »kazalci globlje kognitivne, ekspresivne, besedilne in družbene organizacije diskurza« (Maschler in Schiffrin, 2015, str. 204). Tem raziskavam, ki se večinoma povezujejo s **sociolingvističnimi ali konverzacijskoanalitičnimi raziskavami govorjene interakcije**, je skupno torej predvsem to, da vlogo diskurznih označevalcev opisujejo z vidika njihovega **signaliziranja posameznih vidikov sporazumevalnega procesa**, kot so odnos med naslovnikom in govorcem, odnos med govorcem in besedilom, namere, proces izmenjave, proces tvorjenja ipd. Bolj kot domet v besedilu (začetna točka) je torej pomembno to, katerega od teh vidikov ali več njih označevalec razkriva (končna točka), te končne točke pa so običajno opredeljene kot različne ravni (Schiffrin, 1987), stopnje (Frank-Job, 2006), domene (Sweetser,

³⁴ Zeevat (2006) tako označevanje skupnega védenja (common ground markers, presupposition triggers) pojmuje celo kot krovno prepoznavno lastnost širšega razreda diskurznih označevalcev, saj na podlagi dejstva, da se skupno védenje dinamično nadgrajuje tudi v procesu sporazumevanja, to lastnost pripisuje tako označevalcem že znane informacije, npr. *indeed*, kot diskurznim povezovalcem, kot je *but*, ki se nanašajo na védenje, ki je bilo ustvarjeno s predhodno vsebino. Tudi Roulet (2006) kot informacijo diskurznega spomina, na katero naj bi kazali proceduralni besedilni povezovalci, opredeljuje kot informacijo, ki izhaja bodisi iz predhodnega besedila, neposrednega kognitivnega okolja ali védenja o svetu (*I will not go to the pictures; after all, I am too tired*).

1990) ali dimenzije diskurza (Fischer, 2006c, str. 9), komponente komunikativne situacije (Diewald, 2006, str. 408, 2013, str. 5) ali komponente jezikovne rabe (Louwerse in Mitchell, 2003). Tudi Fischer (2006c, str. 11–12) tako ugotavlja, da je pri opredeljevanju dometa diskurzni označevalcev v tem kontekstu bolj kot o povezovalnosti smiselno govoriti o deiktčnosti, katere podskupina je tudi (besedilna) povezovalnost, bolj kot o enotah diskurza pa o enotah diskurzne situacije oz. o pogojih ali okoliščinah diskurza. Za razliko od materializirane, besedilne ravni pa tovrstno pojmovanje diskurza obenem odpira pomembno metodološko vprašanje njegovega objektivnega določanja in funkcijskega opisa označevalcev, kadar ti delujejo na več ravneh hkrati (Fischer, 2014, str. 281).

Kljub skupni izhodiščni razbremenjenosti pojmovanja diskurza kot besedilne strukture se ti pristopi pri samih opredelitvah diskurzni označevalcev za razliko od prvih dveh skupin notranje precej razlikujejo, predvsem glede na izbrana izrazna merila, kot je položaj označevalca v izjavi ali njegova skladenjska integriranost, ki posredno vpliva tudi na zamejitev njihovih funkcij, npr. ločevanje med predstavno in drugimi ravnmi (prim. npr. diskurzne označevalce pri D. Schiffrin, ki vključujejo celoten spekter možnih domotov,³⁵ in interakcijske členke pri K. Fischer, ki se omejuje samo na pragmatične funkcije, ki so jih razvili gramatikalizirani skladenjsko neintegrirani izrazi). Po drugi strani pa je vsem tovrstnim teorijam skupna zlasti podmnožica **tipično govorjenih, ustaljenih, kratkih in skladenjsko neintegriranih členkov, medmetov, mašil in fraz**, ki se tipično pojavljajo zgolj v spontanem govoru³⁶ in ki jih drugi dve skupini iz svojih opredelitev diskurzni označevalcev večinoma izločata.³⁷ To je tudi tista množica, ki ji

³⁵ Diskurzne označevalce kot nadpomenko povezovalcev pojmujejo tudi nekateri drugi raziskovalci, npr. Pons Bordería (2006), ki v skladu s celostnim večteoretskim pristopom diskurzne označevalce opredeljuje kot nadpomenko povezovalcev, fatičnih označevalcev interakcije in modalnih členkov, ali Roulet (2006) in Rossari (2006), ki označevalce diskurzni povezav glede na predpostavke ženevske šole delita na jezikovne (skladenjske in semantične povezave), besedilne in situacijske, odvisno od opazovane ravni organizacije diskurza.

³⁶ Aijmer (2002, str. 2) po Stenström (1994) za angleščino denimo navaja naslednji seznam diskurzni členkov: *actually, ah, all right, anyway, God, goodness, gosh, I mean, I see, I think, mhm, no, now, oh, OK, please, quite, really, right, sort of, sure, vprašalni pristavki, that's rights, yes/yeah, you know, you see, well*, ki jim je torej skupna visoka stopnja ustaljenosti in vezanost na interakcijski diskurz, vključujejo pa raznolik nabor izrazov – od vljudnostnih formul do vzklikov in mašil.

³⁷ Redeker (1991, str. 1665–1666) podobno kot Fraser (1990, str. 390) v kritiki dela D. Schiffrin poudarja, da čeprav diskurzne označevalce opredeli kot izraze, ki razmejujejo enote govora, med analiziranimi označevalci pogosto navaja primere, kjer je njihov domet besede ali besedna zveza znotraj same izjave, zato jih težko opredelimo kot

status diskurznega označevalca pogosto pripisujejo druge, zlasti skladenjsko usmerjene jezikoslovne discipline, pri poimenovanju tipično govornih, skladenjsko neintegriranih jezikovnih sredstev, ki jih ni mogoče umeščati v tradicionalnejše, formalno opredeljene skladenjske ali pomenske kategorije.

S pojmovanjem diskurzni označevalcev kot označevalcev interaktivnega sporazumevalnega procesa, pri katerih je njihova povezava s sobesedilom drugotnega pomena, pa se odpira vprašanje meje z drugo pogosto obravnavano, a prav tako težko opredeljivo kategorijo nepropozicijske funkcijske leksike, t. i. **pragmatičnimi označevalci**. V Enciklopediji uporabnega jezikoslovja (Romero-Trillo, 2013) so ti denimo opredeljeni kot tisti izrazi v govornih interakciji, ki v splošnem ne prispevajo k semantičnemu pomenu sporočila, temveč so razvili komunikativni status, ki v diskurzu povezuje govorca, naslovnika in kontekst sporočila, npr. *you know, I mean, you see, well, yeah*. Ni torej presenetljivo, da med obema izrazoma pogosto prihaja do prekrivanj in raznolikih poskusov razmejevanj. V raziskavah, ki pri opredelitvah diskurzni označevalcev ločujejo med funkcijami, vezanimi na besedilo, in drugimi pragmatičnimi funkcijami, se danes pragmatični označevalec **večinoma vzpostavlja kot nadpomenka izraza diskurzni označevalec**, a s širšim pojmovanjem pragmatičnih označevalcev kot enot, ki poleg zgoraj naštetih izrazov (pragmatičnih členkov) vključujejo še druge vrste nepropozicijskih izrazov, ki niso nujno vezani zgolj na govorni diskurz (B. Fraser, 1996; Hansen, 2006; Aijmer in Simon-Vandenberg, 2011; Degand idr., 2013b) in tudi ne nujno na razumevanje pragmatike kot področja, ki se začne, kjer se skladnja in semantika končata.

2.2.2.2.3 Izvor povezave

Čeprav se različne teorije razlikujejo v pojmovanju dometa označevalcev in z njim povezanih taksonomijah posamičnih funkcij ali tipov povezav, pa med njimi večinoma obstaja konsenz, da lahko v splošnem ločujemo med dvema temeljnima načinoma vzpostavljanja povezav med deli

kazalce strukture (govorjenega) diskurza, npr. *oh, y'know* ipd. Sama zato iz svoje definicije izključuje rabe, ki imajo domet znotraj izjave in ne zunaj nje. Schifffrinova (2001, str. 66–67) na te očitke s ponazoritvami rabe izraza *y'know* odgovarja, da ta označuje tranzicije iz ene faze pogovora v drugega, npr. kadar je rabljen na koncu argumentacije, povezuje sklep s predhodnim dokazom; kadar začenja novo temo, prejšnjo zgodbo povezuje z novo; kadar uvaja novo referenco, s tem opozori na znanost informacije.

besedila: **semantičnim in pragmatičnim**. To razliko je v svoji razpravi s ponazoritvijo na povezovalcih *and, but, or, so, if* prvi podrobneje izpostavil van Dijk (1979), ki semantične povezovalce opredeli kot povezovalce izkazanih dejstev oz. propozicij (npr. *I was sick, so I stayed in bed.*), pragmatične povezovalce pa kot povezovalce govornih dejanj (npr. *John is sick. So, let's start.*). E. Sweetser (1990, str. 77) nato to delitev še dopolni, saj poleg semantičnih oz. vsebinskih povezovalcev, ki vzpostavljajo razmerja med dogodki v resničnem svetu (npr. *John came back because he loved her*, kjer povezovalec uvaja vzrok za njegovo vrnitev), pragmatične razdeli na epistemske povezovalce, ki vzpostavljajo subjektivne argumentacijske povezave (npr. *John loved her, because he came back*, kjer povezovalec uvaja svoje pojasnilo za subjektivni sklep v prvi trditvi), in povezovalce govornih dejanj (v ožjem smislu), ki vzpostavljajo povezave med dvema ločenima izjavama (npr. *What are you doing tonight, because there's a good movie on*, kjer povezovalec uvaja pojasnilo za govorno dejanje v prvi izjavi).³⁸ Še podrobnejšo delitev epistemskih rab pa sta glede na stopnjo govorčevega subjektivnega in zavestnega vzpostavljanja medbesedilnih povezav predlagala H. Pander Maat in L. Degand (2002).

Ločnico med vzpostavljanjem semantičnih in pragmatičnih povezav, za katero se znotraj teorij koherence pogosto uporablja tudi poimenovanje **izvor koherence** (Sanders idr., 1992), upoštevajo tudi druge teorije, bodisi na ravni razvrščanja bodisi na ravni opisa diskurznih označevalcev in/ali diskurznih povezav, zanj pa uporabljajo različna poimenovanja, poleg dihotomije semantično – pragmatično (van Dijk, 1979; Prasad idr., 2008; Schiffrin, 1987) tudi objektivno – subjektivno (Sanders idr., 1992), predstavno – retorično (Redeker, 1990), predmetno – predstavitveno³⁹ (Mann in Thompson, 1988) ali zunanje – notranje (Halliday in Hasan, 1976; Martin, 1992), vendar pri dejanskem vzpostavljanja ločnice glede na to, ali povezave vrednotijo z vidika tvorčevih namer ali stanja v svetu, med njimi prihaja do večjih ali manjših razlik (Taboada

³⁸ Primerov namenoma ne prevajam, saj tako kot v nekaterih drugih jezikih slovenščina med vsebinskim in epistemičnim vzrokom lahko ločuje že na ravni izbire veznika (*John je bolan, torej/zato začnimo.*, John jo je moral imeti rad, *saj/ker* se je vrnil.), toda doslednost takega razlikovanja v rabi doslej še ni bila empirično preverjena.

³⁹ RST (Mann in Thompson 1988: 256-257) retorične povezave deli na predmetne (angl. *subject matter*), npr. elaboration, circumstance, solutionhood, cause, restatement, in predstavitvene (angl. *presentational*), npr. motivation, background, justification, concession, glede na želeni efekt na bralca: predmetne so tiste, pri katerih želimo, da naslovnik prepozna relacijo, predstavitvene pa tiste, kjer želimo pri naslovniku prebuditi nek vzgib (*inclination*), npr. da informacijo v jedrnem argumentu sprejme, vanjo verjame ali jo pozitivno vrednoti.

in Mann, 2006b, str. 436; Sanders idr., 2016), zlasti pri opredeljevanju medbesedilnih povezav, ki za razliko od zgoraj prikazanih razlikovanj med posledico in sklepom ali med vzrokom in pojasnilom niso okoliščinske, temveč so ne glede na kontekst vedno usmerjene v besedilo (npr. sprememba ubeseditve, povzemanje) ali neločljivo povezane z zunajbesedilnim, kulturološko pogojenim védenjem (npr. dopustnost). Tako tudi že omenjeni predlog poenotenja nabora diskurzni povezav znotraj konzorcija ISO (Prasad in Bunt, 2015; Bunt in Prasad, 2016) ne predvideva razvrščanja na pragmatično in semantično na ravni nabora možnih povezav, temveč zgolj na ravni njihovega (poljubnega) dodatnega opisa.

Čeprav van Dijk (1979, str. 449) v svoji razpravi sicer omenja, da naj bi pragmatične povezovalce zaznamovale različne skladijske in prozodične omejitve, kot je pogosto pojavljanje na začetku izjav ali posebna intonacija, so nadaljnje empirične raziskave pokazale, da lahko **eno in drugo rabo največkrat prepoznamo šele v kontekstu**, ob upoštevanju pomena sobesedila in njegove kulturološko pogojene interpretacije. Že Halliday in Hassan (1976) namreč opozarjata, da je v dejanskih besedilih pogosto težko ločevati med zunanjim pomenom, ki je inherenten pojavom, o katerih z jezikom govorimo, in notranjim pomenom, ki je inherenten procesu sporazumevanja (Halliday in Hasan, 1976, str. 241), npr. pri ločevanju med naštevanjem dogodkov v zunanjem svetu in dodajanjem v sporazumevalnem procesu z veznikom *and*, in da je ustrežnejše kot o dihotomiji govoriti o kontinuumu (Halliday in Hasan, 1976, str. 267).

Kljub temu se prav **na tej ločnici pogosto utemeljujejo nekateri poskusi zamejitev diskurzni označevalcev od drugih besedilnih povezovalcev**. Tako kot v slovenističnem prostoru Schlamberger Brezar (2009) v skladu z izhodišči pragmatične analize diskurza med analizirane povezovalce prišteva zgolj besedilnopovezovalna sredstva, ki povezujejo izreke oz. govorna dejanja tako, da izražajo govorčev odnos in osebno interpretacijo, tudi D. Lewis v svoji opredelitvi diskurzni označevalcev poudarja, da mednje prišteva zgolj tiste besedilne povezovalce, ki so usmerjeni v govorca (Lewis, 2006, str. 55): če veznik *but* v povedi »*It's not green, but red*« označuje dejansko neskladje obeh barv (ni diskurzni označevalec), isti veznik v povedi »*I like red, but my sister likes blue*« razkriva govorčevo mnenje o neskladju dveh idej, v povedi »*J's going away next month, but you probably knew that*« pa govorčevo zavedanje neskladja med ravnokar izrečenim in stanjem v svetu, torej je v obeh primerih veznik rabljen kot diskurzni označevalec. Podobno tudi C. Bazzanella (2006, str. 456–457) med diskurzne

označevalce umešča zgolj tiste nepropozicijske izraze, ki nosijo pragmatični pomen, tj. v diskurzu opravljajo predvsem kognitivne, interakcijske in metabesedilne funkcije.

2.2.2.3 Večfunkcijskost

Kot smo doslej že večkrat izpostavili, je ena najbolj tipičnih lastnosti diskurzni označevalcev dejstvo, da lahko ista fonološka oz. ortografska oblika v različnih skladenjskih, besedilnih ali situacijskih kontekstih zaradi svoje diahrono pogojene pomenske izpraznjenosti pogosto opravlja različne funkcije, prav ta lastnost pa je pogosto vzrok za razhajanja med posameznimi opredelitvami in interpretacijami teh izrazov, saj različni raziskovalci glede na različna izbrana merila ali teoretske modele diskurza mejo med diskurznimi in nediskurznimi rabami iste oblike vzpostavljajo na različnih mestih, obenem pa različno pristopajo tudi k samemu opisu posamičnih diskurzni funkcij. Tudi če je neka oblika splošno prepoznana kot diskurzni označevalec (npr. *but*, *because* ali *now*), se različni raziskovalci torej razhajajo v tem, v katerih rabah se oblika dejansko pojavlja kot diskurzni označevalec, ali med posameznimi diskurznimi rabami istega označevalca prihaja do pomenskih razlik, pa tudi, ali jim je v konkretni rabi sploh mogoče pripisati zgolj eno, prevladujočo funkcijo. Četudi raznolikost pristopov dokazuje, da široko pojmovani kategoriji diskurzni označevalcev ne moremo pripisati ene same krovne funkcije, ki bi zaobjela celoten spekter njihovih rab, pa prav to dejstvo dokazuje, da je morda njihova **temeljna skupna lastnost**, ki hkrati narekuje njihovo celostno obravnavo, prav večfunkcijskost (Fischer, 2014).

Raznolikost pristopov k obravnavi večfunkcijske narave diskurzni označevalcev se v literaturi običajno pojasnjuje z dihotomijo med **minimalističnimi oz. monosemičnimi pristopi** na eni strani in **maksimalističnimi oz. homonimnimi pristopi** na drugi (Hansen, 1998b; Fischer, 2006c). Minimalističnim pristopom je skupna predpostavka, da lahko posamični obliki diskurznega označevalca določimo nespremenljive pomenske komponente (jedrni pomen), ki so prisotne v vseh njegovih rabah, posamezne pomenske odtenke pa pripisujejo interakciji te oblike s kontekstom vsakokratne rabe. Po drugi strani si maksimalistični pristopi prizadevajo za identifikacijo teh subtilnih kontekstualno pogojenih pomenskih razlik, a med različnimi identificiranimi pomeni ne vzpostavljajo nikakršnih povezav. Kot ugotavlja M.-B. M. Hansen (1998b, str. 240, 2006, str. 24), je minimalistični pristop s teoretskega vidika privlačnejši, saj

breme podrobnejše interpretacije prepušča pragmatiki, a so tovrstni minimalistični opisi jedrnega pomena običajno preveč abstraktni in presplošni za kakršenkoli prenos v prakso ali razlikovanje z drugimi diskurznimi označevalci, obenem pa tudi ne upoštevajo morebitne sistematičnosti njihovega diahronega razvoja (prim. denimo koncept proceduralnega pomena znotraj teorije relevantnosti). Maksimalistični pristopi, kakršnega denimo zagovarja konverzacijska analiza, pa po drugi strani s pojasnjevanjem pomenskih lastnosti označevalca s pomenskimi komponentami konteksta, v katerem se pojavlja, brišejo mejo med pomenom oblike in pomenom konteksta ter prav tako ne upoštevajo intuitivne in diahrone pomenske povezanosti teh izrazov.

Večina sodobnih pristopov k pomenskemu opisu diskurzni označevalcev se tako raje poslužuje vmesnega, **polisemičnega pristopa**, ki priznava, da ima lahko posamezna oblika več različnih prepoznavnih pomenov, ki niso odvisni zgolj od vsakokratnega konteksta njihove rabe, a med temi pomeni obstaja določena povezava. Vendarle pa se polisemični pristopi notranje precej razlikujejo glede na to, kako pristopajo k pojasnjevanju povezanosti posameznih pomenov oz. funkcij in kje vzpostavljajo ločnico med diskuržno funkcijo in drugimi rabami posamezne oblike: od razlikovanj na podlagi splošnih diahronih procesov v jeziku (npr. metaforični ali metonomični prenosi, leksikalizacija implikatur) do razlikovanj na podlagi kombinacije več meril, kot so skladijski položaj, pomenske lastnosti sobesedila, domet, prozodija, zunajbesedilni pragmatični parametri ali povezovanje s posameznimi teoretsko predpostavljenimi ravnmi oz. komponentami diskurza (za pregled nekaterih pristopov in avtorjev na področju pragmatičnega jezikoslovja denimo glej (Fischer, 2006c, str. 13–20). Pri tem izpostavimo, da večfunkcijskosti ne smemo razumeti zgolj v pragmatičnem maksimalističnem smislu iskanja korelacij med rabo leksema in kompleksnim zunajbesedilnim kontekstom, temveč se večfunkcijskost v enaki meri kot ena izmed pogostih lastnosti diskurzni označevalcev izpostavlja tudi znotraj tradicionalnih besediloslovnih teorij, saj posamezni besedilni povezovalci lahko vzpostavljajo različne tipe koherentnih povezav, bodisi z vidika njihovega besedilnega dometa (npr. *so* kot vzpostavljaalec povezave med dvema propozicijama, uvajalec pojasnila oz. popravka ali kazalec tematskega prehoda med dvema deloma besedila), vira koherence (npr. *so* kot uvajalec objektivnega vzročno-posledičnega odnosa ali subjektivnega sklepa) ali konkretnega semantičnega tipa povezave (npr. *since* kot označevalec časovnega in pojasnjevalnega razmerja).

Problematičnost opisa (več)funkcijske leksike se implicitno razkriva tudi v slovenskem prostoru, kjer so bile v slovničnih in slovarskih priročnikih ter posamičnih razpravah ponujene **različne besednovrstne kategorizacije in pomenske členitve** izrazov, ki se lahko v diskurzu pojavljajo v različnih skladenjskih položajih in diskurzni vlogah. Če za primer vzamemo izraza *torej* in *vendar*, ki bi jo vsaj v eni izmed svojih rab med diskurzne označevalce uvrstile vse doslej predstavljene domače in tuje teorije, lahko npr. vidimo, da slovarja SSKJ in SSKJ2 besedo *vendar* v vseh skladenjskih položajih obravnavata kot večpomenski prislov (SSKJ1) oz. členek (SSKJ2), tudi v vezniški vlogi (glej npr. *morda se bo vendar kaj našlo zate proti težave niso bile velike, vendar obstajajo*), po drugi strani pa besedo *torej* glede na skladenjski položaj delita na dve homonimni iztočnici: kot veznik je obravnavana, kadar stoji na začetku besedilne enote in uvaja vzročno-sklepalno razmerje (npr. *pozni smo, torej gremo čim prej dalje*) ali pojasnilo (npr. *predstava bo prvega, torej v nedeljo*), kot členek pa kadar se pojavlja v drugih skladenjskih položajih, četudi s pomenskega vidika prav tako izraža vzročno-sklepalno razmerje (npr. *zmagali so v svoji skupini in bodo torej igrali v finalu*). Čeprav *torej* tudi Slovenski pravopis glede na skladenjski položaj in ne glede na prekriven pomen opredeli kot dve iztočnici, vezniško in členkovno, pa *vendar* po drugi strani za razliko od enoiztočniške večpomenske obravnave v SSKJ deli na tri homonimne iztočnice, prislovno (*oglasil se, morda se bo vendar kaj našlo zate*), vezniško (*primera sta podobna, vendar ne enaka*) in členkovno (*odpri vendar*). Spet tretjo rešitev pa ponuja Slovar slovenskih členkov, ki oba izraza v vseh možnih rabah obravnava kot večpomenska členka, pri čemer je *vendar* v vlogi uvajalca nasprotja s prej povedanim (*težave niso velike, vendar obstajajo*) opredeljen kot povezovalni členek, *torej* v vlogi označevalca vzročno-sklepalnega razmerja (*Nismo jim pomagali, torej smo sami krivi*) pa kot vrednotenjski naklonski, in ne povezovalni členek.

Na potrebo po sistematičnejši kategorizaciji in pomenskem opisu diskurznofunkcijskega besedišča opozarjajo tudi nekatere novejši slovenistične leksikološke razprave, tako na področju opisa veznikov in členkov (Balažic Bulc, 2009; Balažic Bulc, 2015; Pisanski Peterlin, 2015a; Smolej, 2015a) kot diskurzni, pragmatični ali metabesedilni izrazov, ki se tipično pojavljajo predvsem v govorjenem jeziku, kot je denimo kontekstualno občutljivi označevalec *zdej* (Verdonik, 2015c).

2.2.2.4 Skladenjska vloga

Čeprav se diskurzni označevalci prevladujoče vzpostavljajo kot funkcijska kategorija in opisujejo predvsem z vidika njihove vloge v diskurzu, se ne glede na izbrani teoretski okvir funkcijske oz. pomenske opredelitve običajno dopolnjujejo tudi z izraznimi, običajno skladenjskimi lastnostmi, bodisi na ravni same opredelitve bodisi na ravni opisa prototipičnih predstavnikov kategorije. Kot smo videli že v pregledu literature, se je zlasti v pragmatičnojezikoslovnih opredelitvah označevalcev med tipičnimi skladenjskimi lastnostmi pogosto navajal **položaj** na začetku stavka oz. diskurzne enote (Schiffrin, 1987, str. 328; B. Fraser, 1999, str. 938; Schourup, 1999, str. 233), toda ta lastnost danes običajno ni več izpostavljena, zlasti pod vplivom kontrastivnih raziskav, ki kažejo, da se v enakih skladenjskih položajih v različnih jezikih razvijajo različni pomeni, hkrati pa se isti pomeni v različnih jezikih realizirajo v različnih skladenjskih položajih (Fischer, 2013).

Prav tako se v pragmatičnojezikoslovnih opredelitvah označevalcev skladno z idejo nepropozicijskosti oz. njihove semantične neobveznosti med tipičnimi skladenjskimi lastnostmi pogosto omenja njihova **skladenjska izpustljivost** (Schiffrin, 1987, str. 328), šibka povezanost s soležno stavčno enoto (Sankoff idr., 1997; Schourup, 1999) oz. dejstvo, da njihova odstranitev ne vpliva na slovničnost enote, v kateri se pojavijo (Hansen, 1998b, str. 237), a tudi ta lastnost je v prakso razmeroma težko prenosljiva, saj so s semantično-skladenjskega vidika (ne pa nujno tudi z vidika pragmatične sprejemljivosti) izpustljivi tudi nepropozicijski elementi na nižjih (npr. epistemski pridevniki tipa *tako imenovana*) ali višjih skladenjskih nivojih (npr. formulaični stavčni nastavki tipa *to pomeni, da*). O skladenjski izpustljivosti težko govorimo tudi v primeru izrazov, ki signalizirajo pojasnila, pristavke, popravke ipd. nižjih stavčnih enot, zato je pri določanju skladenjskih lastnosti diskurznih označevalcev bolj kot o skladenjskem položaju ali skladenjski izpustljivosti ustrezno govoriti o njihovi strukturni vpetosti oz. integriranosti.

Kot izpostavlja K. Fischer (2006c, str. 9–12, 2014), je prav strukturna vpetost, ki jo sama imenuje **prozodična, skladenjska in semantična (ne)odvisnost**, ena izmed ključnih izraznih lastnosti, s katero lahko pojasnimo raznolikost pristopov k obravnavi diskurznih označevalcev in njihovih izhodiščnih funkcijskih opredelitev. Na enem koncu kontinuuma najdemo raziskave **strukturno integriranih enot**, kot so priredni ali podredni vezniki, povezovalni prislovi ipd., ki se pri opredelitvah diskurznih označevalcev običajno osredotočajo predvsem na njihovo

besedilnopovezovalno funkcijo oz. vlogo pri vzpostavljanju besedilne koherence, tako v pisnem kot govorjenem jeziku (npr. Fraser, Redeker, Roulet, Rossari, Hansen). Tovrstne raziskave skladijsko integriranih enot tako v svojih opredelitvah pogosto poudarjajo povezanost z besedilno enoto, v kateri se pojavljajo, npr. »diskurzni segment, ki jih gosti« (B. Fraser, 2006, str. 190), »skladijski gostitelj« (Lewis, 2006, str. 55), izjava ali enota gostiteljica (Hansen, 1998b, str. 236) in podobno. Na drugo stran kontinuuma pa se umeščajo zlasti raziskave spontane konverzacije, ki v središče svojega zanimanja že v izhodišču postavljajo skladijsko, pomensko in prozodično povsem **neintegrirane, parentetične izraze** ter funkcije, ki so jih ti izrazi razvili v procesu izgube svojega propozicijskega pomena, kot so upravljanje dialoških izmenjav, organizacija tem, uravnavanje medosebnih odnosov ipd. (npr. Fischer, Diewald, Frank-Job, Aijmer, pri nas Verdonik).

Pri tem K. Fischer (2006c, str. 10–11) poudarja, da številne raziskave izbirajo zgolj posamezne vidike te štiristopenjske dihotomije (tj. ločevanja (i) med integriranimi in neintegriranimi enotami, (ii) med povezovanjem in upravljanjem diskurza, (iii) med osredotočenostjo na besedilo ali na dialog, (iv) med povezovanjem enota ali okoliščin diskurza) ali pa v svoji opredelitvi sploh ne razlikujejo med različnimi skladijskimi vlogami, zlasti ob upoštevanju dejstva, da se ista fonološka oz. ortografska oblika pogosto pojavlja v različnih oblikah skladijske integriranosti. Nenazadnje lahko strukturno-funkcijski diapazon pojasnimo prav s tem, da so diskurzni označevalci v temeljni razpravi področja (Schiffrin, 1987) v svoji ohlapni opredelitvi izrazov, ki razmejujejo enote govora, vključevali tako integrirane stavčne povezovalce, ki delujejo na predstavnici ravni (*and, but, because*), kot nekatere neintegrirane, tipično pogovorne izraze, ki delujejo na drugih ravneh (*oh, you know*),⁴⁰ kar je odpiralo zelo široko polje nadaljnjih raziskav in reinterpretacij.

2.2.2.5 Večbesednost

Če večbesedne diskurzne označevalce opredelimo kot označevalce, ki se od drugih označevalcev razlikujejo po tem, da so sestavljeni iz dveh ali več besed, si v tem poglavju na kratko ogledamo

⁴⁰ Pri tem Schiffrinova izraz *you know* kot označevalec analizira tudi v propozicijsko obveznih integriranih rabah, npr. *you know these tea rooms?* (Schiffrin, 1987, str. 268–294).

problematičnost opredeljevanja besede, zlasti v govoru, in ga hkrati tudi povežemo z razumevanjem tega koncepta znotraj naše raziskave. Slovenska slovnica (Toporišič, 2004, str. 103) besedo opredeljuje kot **najmanjšo pomensko samostojno enoto** stavčne povedi, ki jo v pisnem jeziku sestavlja nerazdružen niz črk, v govoru pa nepretrgan niz glasov. Delov besede ni mogoče ločiti, zamenjati zaporedja njenih zlogovnih in pomenskih sestavin, lahko pa jim zamenjujemo mesto v stavku. V obrobno skupino besed, ki jih ne moremo opredeliti zgolj grafemsko, kot niz črk med dvema presledkoma, J. Toporišič (2004, str. 104) uvrsti tudi nekatere stalne zveze, ki se v rabi pišejo bodisi skupaj bodisi narazen, npr. *avto cesta*, *sivo zelen*, *Slovenija avto*, *bog ve*, *C vitamin*, po drugi strani pa kot zveze več samostojnih besed navaja nekatere črkovne nize, ki se v rabi pišejo skupaj, npr. *ponavadi*, *nalepem*, *semtertja*, *čimbolj*, *odkod*, *nase*, *nočem* ipd., saj kljub pomenski samostojnosti omogočajo nekatere vrste skladenjskih pretvorb in variacij (Toporišič, 2004, str. 104–105). Prav tako med besede uvrsti nize s t. i. prostimi morfemi (npr. *smejati se*, *misliti si*, *polomiti ga*).

Za razliko od tovrstnih pravopisnih, besedotvornih, skladenjskih ali pomenskih meril, ki brišejo mejo med besedo kot zapisovalno enoto in besedo kot leksikalno enoto (leksemom), tj. temeljno poimenovalno enoto, enoto slovarja (Vidovič-Muha, 2000, str. 17; Gantar, 2007, str. 60–61), se bomo sami pri pojmovanju besede in z njo povezane večbesednosti v skladu z izbranim popolnim korpusnim pristopom opirali zlasti na njene objektivno prepoznavne izrazne lastnosti, kot je nepretrgan črkovni niz (**grafemska beseda**), ki sicer prav tako odpira vprašanja mejnosti, kot so obravnava črkovnih nizov, ki se v rabi pojavljajo tako v zapisu skupaj kot narazen, npr. *kdor*(*jkoli*), ali nizov, ki se pišejo s stičnimi ali nestičnimi ločili, npr. *slovensko-ameriški*, *AC/DC*, *O'Brian*, *d.o.o.* (za več primerov glej (Krek, 2010, str. 28–45)).

Še bistveno kompleksnejše pa je opredeljevanje besede v govorjenem jeziku, v katerem ta funkcionira izključno kot prozodični in ne grafemski pojav. Kot povzema H. Fraser (2013), je še danes pogosto razširjeno prepričanje, da v govoru besede prepoznavamo na podlagi prepoznavanja posameznih fonemov, ki jih nato povežemo v pomensko smiselne celote. To prepričanje se torej opira na predpostavke, da so besede v govoru ločene z jasnimi premori, so vedno izgovorjene na enak način in da jih je mogoče razstaviti v nize posameznih zvokov (fonemov), ki so prav tako samostojni in razmeroma nespremenljivi. Toda že v petdesetih letih prejšnjega stoletja se je pri poskusu prenosa tovrstnih teorij v računalniške sisteme za obdelavo

govora izkazalo, da so te predpostavke napačne. Govor namreč ni niz medsebojno razmejljivih fonemov, temveč nepretrgan zvočni niz, v katerem **ni sistematičnih premorov med besedami** in njenimi posameznimi zvoki. Tudi grafemski prepis govora v obliki zapisanih zvokov, besed, izjav, vlog in tudi nebesednih sporazumevalnih vidikov v obliki transkripcij moramo tako razumeti predvsem kot **rezultat konsenzualnih konvencij** (Dittmar, 2013) in ne jezikovne danosti. Tudi sami se bomo tako znotraj pričujoče raziskave v želji po objektivizaciji in dolgoročni uporabnosti rezultatov pri določanju grafemskih besed naslonili na dogovorna načela transkribiranja uporabljenega gradiva (poglavje 3.2 *Gradivo*).

2.2.2.6 Ustaljenost

Pojav ustaljenosti se v jezikoslovju povezuje z več različnimi teorijami in koncepti historičnega jezikoslovja, zlasti procesi leksikalizacije, gramatikalizacije in pragmatikalizacije. V svoji najbolj splošni opredelitvi **leksikalizacija** označuje diahroni proces, katerem se niz besed in/ali morfemov sčasoma ustali in razvije lastni specializirani pomen in/ali funkcijo (Moon, 1998, str. 36), s čimer beseda oz. besedna zveza postane enota besedišča (slovarja, leksikona) nekega jezika, ki naj bi jo kot tako govorci tudi shranjevali v svoj spomin. Čeprav nekateri avtorji leksikaliziranost razumejo kot krovno lastnost vseh ustaljenih jezikovnih sredstev (Brinton in Traugott, 2005), se ta prevladujoče uporablja predvsem za pojasnjevanje nastanka predmetnopomenskega besedišča, tj. enot propozicijskega, referenčnega, predmetnega oz. denotativnega pomena. Z vidika pojasnjevanja nastanka večbesednih izrazov se torej lastnost leksikaliziranosti običajno pripisuje predvsem stalnim besednim zvezam s samostojnim glagolskim, samostalniškim, pridevniškim in prislovnim pomenom. Poleg ustaljenega (tudi nerazstavljivega, osamosvojenega) pomena se kot kazalec leksikaliziranosti pogosto navajata tudi ustaljenost zgradbe in skladske vloge.

Na drugi strani se razvoj funkcijskega besedišča – kamor se poleg besed ali besednih zvez, ki se funkcijsko uresničujejo na ravni stavka (npr. predlogi ali vezniki), umešča tudi pragmatično besedišče, ki se funkcijsko uresničuje šele v širšem sobesedilnem ali situacijskem kontekstu (npr. medmeti ali členki) – običajno pojasnjuje skozi proces **gramatikalizacije**. Ta označuje »spremembo, s katero leksikalne enote in konstrukcije vstopajo v določene jezikovne kontekste z namenom opravljanja **slovničnih funkcij** in po začetni gramatikalizaciji razvijajo nove slovnične

funkcije» (Hopper in Traugott, 2003, str. 18). V teorijah, ki pragmatiko vključujejo kot enakovredno komponento slovničnega modeliranja, se tako tudi nastanek diskurznih označevalcev običajno pojasnjuje kot rezultat procesa gramatikalizacije. Poleg koalescence (oženja dometa) in fonološke redukcije, ki naj v primerjavi z nastankom drugih funkcijskih izrazov za diskurzne označevalce ne bi bili značilni, L. J. Brinton (2010, str. 62) navaja deset tipičnih mehanizmov procesa gramatikalizacije: dekategoriizacijo (izguba oblikoslovnih ali , str. skladenjskih lastnosti izvirne besedne vrste), prehod iz odprte v zaprto besedno vrsto, zamrznitev oblike, desementizacijo (izpraznitev pomena), prehod iz referenčnega (propozicijskega) v nereferenčni (pragmatični ali proceduralni) pomen, konvencionalizacijo možnih inferenc, subjektifikacijo (vezanost na govorca), divergenco (razvoj več vzporednih pomenov oz. funkcij), plastenje (razvoj več izrazov za isto funkcijo) in obstojnost (sledi propozicijskega pomena).

Ker pa pragmatično oz. diskurzno besedišče vendarle izkazuje nekoliko drugačne pomenske in diahrone lastnosti v primerjavi s funkcijskimi izrazi, ki delujejo na stavčni ravni, se je za pojasnjevanje njihovega nastanka znotraj teorij, ki jasneje poudarjajo ločnico med slovnico na eni in pragmatiko na drugi strani, vzporedno vzpostavil tudi izraz **pragmatikalizacija**, ki označuje »proces, s katerim besedna zveza ali beseda v danem kontekstu propozicijski pomen spremeni v prevladujoče metasporazumevalni, diskurznointerakcijski pomen« (Frank-Job, 2006, str. 397), torej kot proces »gramatikalizacije diskurznih funkcij« (Diewald, 2011). Kot povzema B. Heine (2013, str. 1218), naj bi se pragmatikalizirani izrazi, kot so diskurzni označevalci, od gramatikaliziranih izrazov, kot so znotrajstavčni slovnični izrazi, razlikovali v tem, da so skladenjsko izolirani (niso integrirani), se ne povezujejo z drugimi stavčnimi enotami, svoj kontekstualni domet širijo (in ne ožijo), so pomensko in skladenjsko izpušljivi, ne vplivajo na resničnostno vrednost in jih ni mogoče uvrstiti v nobeno izmed tradicionalnih slovničnih kategorij (ne omogočajo paradigatskega opisa).

Kot v izčrpnem preglednem članku različnih pristopov k razumevanju gramatikalizacije in/ali pragmatikalizacije diskurznih označevalcev ugotavljata L. Degand in J. Evers-Vermeul (2015), so ti odvisni predvsem od same konceptualizacije slovnice (razmejevanja slovnice od pragmatike) in od nabora mehanizmov, s katero naj bi se gramatikalizacija povezovala, kar smo večkrat sicer izpostavili tudi v pregledu (raznolikih) pristopov k razumevanju drugih

prepoznavnih lastnosti te jezikoslovne kategorije. V skladu s sodobnimi funkcijskoslovničnimi in leksikogramatični pristopi k opisu jezika avtorici menita, da ni mogoče vzpostaviti jasne meje med slovnico na eni strani in pragmatiko na drugi (saj lahko pragmatične funkcije pripišemo tudi tradicionalnim slovničnim kategorijam, pragmatične kategorije pa izkazujejo tudi slovnične specifikke), zato tudi sami zagovarjata stališče, da **ni potrebe po umetnem razmejevanju procesov nastanka diskurznih od drugih funkcijskih enot** in vse povezujeta s krovnim procesom gramatikalizacije.

Vendarle pa je v kontekstu *večbesednih* diskurznih označevalcev potrebno problematizirati tudi sam pojem gramatikaliziranosti kot njihove izhodiščne prepoznavne lastnosti. Tako kot je na področju večbesednih leksikal(izira)nih enot prišlo do **relativizacije konceptov ustaljenosti** (pomena, zgradbe, skladenjske vloge), **uslovarjenosti** ter pomenske **nerazstavljivosti** (nenapovedljivosti glede na pomen sestavin),⁴¹ ki izpostavljajo problematičnost objektivnega ločevanja med večbesednimi leksikalnimi enotami in drugimi ustaljenimi izrazi, med frazeologijo v ožjem in širšem smislu, se tudi na področju diskurzno funkcijske leksike pod vprašaj postavlja osredotočenost na najbolj gramatikalizirane, tj. pomensko izpraznjene in izrazno zamrznjene izraze. Kot smo izpostavili že v pregledu dosedanjih raziskav večbesednih enot na področju analize diskurza (poglavje 2.2.1 *Raziskave večbesednih enot v analizi diskurza*), namreč vedno več raziskav dokazuje, da v jeziku diskurzne funkcije opravlja precej širši nabor (strukturno zaključenih ali nezaključenih) izrazov kot zgolj najbolj gramatikalizirani diskurzni označevalci, za katere bi prav tako lahko trdili, da so v določeni meri ustaljeni oz. **institucionalizirani**, saj se v rabi pojavljajo izstopajoče pogosto.

Tudi gramatikaliziranosti ter z njo povezanih konceptov pomenske izpraznjenosti in izrazne ustaljenosti tako pri vrednotenju prepoznanih lastnosti *večbesednih* diskurznih označevalcev ne moremo obravnavati kot absolutno določljive in splošno uveljavljene prepoznavne lastnosti, zato se tudi pri njenem prenosu na dejansko gradivo kaže potreba po **natančnejši določitvi** oz. razmejitvi glede na druge institucionalizirane besedne zveze s prepoznavno oz. ustaljeno diskurzno funkcijo. Potreben je torej premik od raziskovalčeve subjektivne oz. intuitivne presoje

⁴¹Za pregled in številne primere v slovenščini glej (Gantar, 2007).

stopnje gramatikaliziranosti in pomenske izpraznjenosti k **objektivnemu prepoznavanju ustaljenega** v rabi, kjer se kot uporaben parameter poleg izraznih meril⁴² kaže predvsem podatek o pogostosti.

2.2.3 Izhodišča raziskave večbesednih diskurzni označevalcev

2.2.3.1 *Sinteza pregleda literature*

Pregled tujih in domačih raziskav diskurzni označevalcev in sorodnih diskurznofunkcijskih jezikoslovnih kategorij potrjuje izrazito razdrobljenost tega raziskovalnega področja glede na (ne)izbrani teoretski model diskurza, raziskovalne cilje, preučevano gradivo in različne načine sopostavljanja pomenskih in izraznih prepoznavnih lastnosti. Pri tem je podrobnejši vpogled v različne opredelitve njihove krovne funkcije potrdil, da se pri zamejitvi funkcije diskurzni označevalcev od funkcij drugih propozicijsko neobveznih jezikovnih sredstev različni raziskovalci strinjajo, da je temeljna prepoznavna lastnost diskurzni označevalcev njihova povezovalnost, a pri tem ponujajo različne opredelitve glede na razumevanje pojma diskurz in z njim povezanega dometa označevalcev (od vzpostavljalcev povezav med sporočili na predstavniki ravni do označevalcev nejezikovnih vidikov sporazumevalnega procesa), glede na potrebo po razlikovanju med semantičnim in pragmatičnim upravljanjem diskurza ter tudi glede na način samega razlikovanja med semantiko in pragmatiko. Številne različne kombinacije teh treh vidikov, bolj ali manj teoretsko idiosinkratične, tako ponujajo širok nabor možnih opredelitev funkcije diskurznega označevanja in z njo povezane zamejitve jezikovnih sredstev v vlogi diskurzni označevalcev.

Čeprav bi glede na temeljni pristop k preučevanju diskurza posamezne raziskave lahko združili v splošnejše teoretske skupine, kot so besedilnoskladenjske raziskave strukture diskurza, ki se osredotočajo predvsem na eksplicitne vzpostavljalce pomenskih povezav med deli besedila (npr. M. A. K. Halliday in R. Hasan, teorije koherentnih povezav, B. Fraser; pri nas V. Gorjanc in T.

⁴² Pri izdelavi leksikona diskurzni povezovalcev francoščine (Roze idr., 2012) je bil denimo za ločevanje večbesednih diskurzni označevalcev od drugih formulacijskih nizov s prepoznavno diskurznimi vlogo implementiran t. i. test razstavljivosti, po katerem se med diskurzne veznike umešajo samo tiste zveze, ki jih ni mogoče modificirati s pridevnikom, jim spremeniti števila ali določnega člena, npr. *en tout cas* ('v vsakem primeru') ali *résultat* ('posledično').

Balažic Bulc); pragmatičnobesediloslovne raziskave koherence sporazumevalnega procesa, ki se osredotočajo predvsem na vzpostavljalce povezav med besedilom in njegovim znotraj- in zunajbesedilnim kontekstom (npr. G. Redeker, D. Schiffrin, M. M. Louwerse in H. H. Mitchell; pri nas M. Smolej 2012); raziskave konverzacijske analize, ki se osredotočajo predvsem na skladijsko neintegrirane označevalce mentalnih procesov jezikovnega sporazumevanja (npr. K. Fischer, K. Aijmer; pri nas D. Verdonik); raziskave subjektivnega upravljanja besedila, ki se osredotočajo na označevalce govorčevega subjektivnega vzpostavljanja medbesedilnih povezav (npr. D. Lewis, C. Bazzanella; pri nas M. Schlamberger Brezar, A. Pisanski Peterlin); ali raziskave teorije relevantnosti, ki se ukvarjajo z nosilci proceduralnega pomena (npr. D. Blakemore, V. Rouchota, L. Schourup), tudi med raziskovalci s sorodnimi raziskovalnimi cilji pogosto prihaja do razhajanj v konkretnih opredelitvah in zamejitvah diskurzni označevalcev. Te so najpogostejše pogojene z vzpostavljanjem sekundarnih izraznih meril, kot so zamejitve besedilne enote (povezovanje stavkov, povedi, izjav, govornih dejanj, besedilnih sekvenc), zamejitve znotrajbesedilnega dometa (modifikacija besedilnih enot ali členjenje besedila), zamejitve glede na stopnjo skladijsko-semantične vpetosti, pri premiku na večbesedne diskurzne označevalce pa tudi glede na stopnjo njihove strukturne celovitosti in ustaljenosti. Pri tem se posamezne odločitve včasih utemeljujejo s predpostavkami specifičnega teoretskega modela diskurza (tj. njegovega členjenja ali razslojevanja), najpogostejše pa prevladuje izrazito semaziološki pristop – analiza rabe vnaprej določene peščice prototipičnih predstavnikov opazovane jezikovne kategorije in naknadno posploševanje na kategorijo kot celoto.

Če to razdrobljenost plastično ponazorimo še na prototipičnem naboru izrazov *oh, well, and, but, or, because, so, now, then, y'know* in *I mean*, ki ga je v svojem temeljnem delu ponudila D. Schiffrin (1987), bi tako besediloslovne raziskave med diskurzne označevalce umestile izraze *and, but, or, because, so, then*, kadar nastopajo v vlogi povezovalcev dveh soležnih besedilnih enot, ne glede na subjektivnost/objektivnost vzpostavljanje povezave (a z različno zamejenimi koncepti besedilne enote v razponu od abstraktnih objektov do besedilnih sekvenc), raziskave subjektivnega upravljanja diskurza bi se omejile zgolj na njihove subjektivne rabe na ravni izjav, konverzacijskoanalitične raziskave diskurzni členkov bi obravnavale vse izraze, a zgolj v skladijsko neintegriranih, parentetičnih, tipično govorjenih rabah, relevantnostna teorija bi se ukvarjala samo z *and, but, or, because, so* kot nosilci proceduralnega pomena ne glede na

semantično ali pragmatično rabo, pragmatičnobesediloslovne raziskave koherence govornega diskurza pa bi prav tako obravnavale celoten nabor *oh, well, and, but, or, because, so, now, then, y'know* in *I mean* ne glede na izvor koherence ali skladenjsko vpetost, a bi se medsebojno razlikovale glede tega, ali upoštevajo zgolj rabe, ki modificirajo celotno propozicijo, ali tudi tiste, ki modificirajo zgolj njen del (npr. razlikovanje med rabo *you know* ali *I mean* na začetku ali na sredi izjave).

Ob tovrstni razpršenosti partikularnih raziskav, ki sicer vsaka na svoj način dragoceno prispeva k razumevanju sporazumevalne vloge teh izredno pogosto rabljenih večfunkcijskih izrazov,⁴³ tako ni presenetljivo, da so v zadnjem desetletju vse pogostejši tudi poskusi njihove konsolidacije. Zlasti aplikativno usmerjene raziskave si tako prizadevajo za oblikovanje operativnega modela, ki bi omogočal metodološko čim objektivnejši prenos dosedanjih spoznanj na dejansko gradivo in bil hkrati dovolj teoretsko, žanrsko in jezikovno neobremenjen oz. univerzalen, da bi v povratni zanki omogočal tudi empirično utemeljen nadaljnji razvoj analize diskurza in njenih posameznih poddisciplin, tako na teoretskem kot aplikativnem področju. Ta potreba je enako izrazita tudi v slovenskem prostoru, saj so se tudi dosedanje slovenistične raziskave diskurznofunkcijske leksike večinoma osredotočala na partikularna vprašanja, kot so besednovrstno opredeljevanje in razmejevanje, pomenski opisi posameznih kategorij, posebnosti posameznih diskurznihih tipov, semantično-pragmatično razslojevanje diskurza in podobno, zato težko govorimo o obstoju razmeroma stabilne teoretske ali aplikativne osnove za nadaljnje empirične raziskave diskurznofunkcijske leksike, kot je naša.⁴⁴

⁴³ Prispevke posameznih področij nazorno povzema Pons Bordeira, ki prav tako zagovarja celostni, večteoretski in večdimenzionalni pristop k analizi teh večfunkcijskih izrazov: »*Text Linguistics highlighted the supersentential values of connectives and how they help to build coherence relations. Conversation Analysis displayed the interactional aspects of their behaviour and Politeness Theory did the same for the image of the self/others. Relevance Theory showed the cognitive role played by connectives. French studies within and around Argumentation Theory discovered that connectives do not merely display a relationship between propositions but also between arguments and conclusions, explicit or otherwise. From a different perspective, neo-Gricean developments insisted on what is encoded and what is not, what is said and what is implicated. Finally, the diachronic carried out in grammaticalization studies established universal clines of evolution linked to cognitive processes.*« (Borderia, 2006, str. 79)

⁴⁴ Kot na primeru slovarske obravnave členkov ugotavlja Balažic Bulc, dosedanja praksa klasificiranja na temelju jezikovnih občutkov, drugih klasifikacij in iskanja potrditev v praksi ni obsodila sadov, zato bi jo bilo v bodoče smiselno izvesti na podlagi rezultatov korpusne analize in ne obratno (Balažic Bulc, 2015, str. 538).

Dokler torej na širšem področju analize diskurza ni prevladujočega konsenza glede tega, kaj je diskurz in kako se ustvarja njegova mentalna reprezentacija, kje se konča semantika in kje začne pragmatika, ali je med njima sploh smiselno razlikovati, v kolikšni meri je koherenca diskurza odvisna od besedila samega, v kolikšni meri pa od njegovega sporazumevalnega konteksta, v kakšni soodvisnosti sta skladnja in pomen, ali sta upravljanje govorenega in pisnega diskurza primerljiva procesa, ali se analiza diskurza začne na ravni propozicije, izjave ali kje drugje, bo tako tudi razumevanje diskurznihi označevalcev znotraj pričujoče raziskave temeljilo na načelu iskanja najmanjšega skupnega večkratnika, ki posamezne vloge teh večfunkcijskih izrazov vpeljuje šele na raven njihove podrobnejše analize v kontekstu, ne pa vnaprejšnje opredelitve oz. zamejitve. S tem naša raziskava ne sovпада zgolj s trenutnimi smernicami na področju raziskav diskurznihi označevalcev⁴⁵ in aplikativne analize diskurza, temveč se obenem približa tudi pojmovanju diskurznihi označevalcev znotraj prvotnih teorij koherence v (govornem) diskurzu, kjer se je ta izraz tudi vzpostavil.

Teoretsko manj obremenjen, vključujoč pristop k opredelitvi razreda diskurznihi označevalcev je še toliko pomembnejši na podpodročju raziskav večbesednihi diskurznihi označevalcev, ki je tako v tujem kot slovenskem jezikoslovju bistveno slabše raziskano. Poleg funkcijskih razhajanj, kakršno zaznamuje področje nasploh, se pri večbesednihi enotah dodatno odpira predvsem vprašanje obveznosti njihove strukturne celovitosti in ustaljenosti, zlasti pod vplivom raziskav formulaičnega jezika, ki kažejo, da podobne funkcije kot jedrna skupina gramatikaliziranihi diskurznofunkcijskih stalnihi besednihi zvez opravljajo tudi daljše, skladenjsko bolj ali manj zaključene sintagme, ki so v rabi nezanemarljivo pogoste. Čeprav se bomo sami v pričujoči raziskavi vendarle osredotočali predvsem na skladenjsko zaključene večbesedne enote, pa se

⁴⁵ Potrebo po širokem pojmovanju kategorije diskurznihi označevalcev namesto zamejevanja na posamezne podskupine lepo povzema K. Fischer (2014, str. 287–289): »*To sum up, I want to suggest that as soon as we open up our view to the whole spectrum of candidates for discourse markers, distinctions and defining criteria for possible subclasses of discourse markers appear as natural products of clustering, and this enhances our understanding of discourse markers instead of hindering it. [...] As much as we have to agree with Schououp's description of the diversity of definitions, the conclusion drawn here from the current state of the art is not that the field is too young to be described in precise terms, but that we need a theory that accounts precievesly fort he lack of a precise definition of the field. In particular, we need a theory of discourse functions before a defiition of discourse marker scan succeed. Such a theory is also vital for cross-linguistic comparison (...). Furthermore, also for methodological reasons it is essential not to separate the range of phenomena considered initially since the most striking characteristics of discourse markers do not become apparent if too much of the focus is on subclasses only.*«

bomo pri vzpostavljanju ločnice med razstavljivimi in nerazstavljivimi besednimi nizi v duhu metodološkega objektivizma bolj kot na težko opredeljiva merila pomenske razstavljivosti, prenesenosti ali izpraznjenosti opirali predvsem na merilo pogostnosti v rabi. S tem bomo v pričujočo pilotno raziskavo diskurznofunkcijskih večbesednih enot v slovenskem prostoru zajeli širši nabor izrazov, ki bo omogočil analizo onkraj peščice prototipičnih izrazov, obenem pa bi lahko koristil tudi drugim nadaljnjim slovničnim ali funkcijskim raziskavam posameznih vidikov njihovega opisa.

2.2.3.2 *Opredelitev večbesednih diskurznih označevalcev*

Čeprav je bil v slovenskem prostoru izraz diskurzni označevalec doslej rabljen predvsem v pomenu označevalca nekaterih kognitivnih in medosebnih vidikov govorjene interakcije (Verdonik, 2007; Verdonik idr., 2007) ter pozneje nekoliko razširjen tudi na besedilne povezovalce (Verdonik, 2008),⁴⁶ bomo sami po vzoru tujih pragmatičnobesediloslovnih raziskav izraz diskurzni označevalec uporabljali v širšem pomenu, ki pojem diskurza ohranja v stiku z besedilom oz. jezikovna sredstva vrednoti (tudi) z vidika njihove vloge v organizaciji besedilno posredovanih sporočil, tako v pisnem kot govorjenem diskurzu. Večbesedne diskurzne označevalce na podlagi teh izhodišč in v pregledu izpostavljenih sklepov opredelimo kot **propozicijsko neobvezne ustaljene kombinacije dveh ali več zaporednih grafemskih besed s funkcijo diskurznega povezovanja, strukturiranja ali upravljanja**, ki poleg povezovanja besedilnih enot z njihovim sobesedilnim kontekstom in z njim povezanega védenja vključuje tudi povezovanje besedilnih enot z njenim širšim sporazumevalnim kontekstom. Pri tem pri sami opredelitvi diskurznih označevalcev ne ločujemo med predstavno in drugimi ravnmi diskurza, med objektivnim in subjektivnim besedilnim povezovanjem, med skladenjsko integriranimi in

⁴⁶ Iz opredelitve ideacijskih označevalcev v (Verdonik, 2008) ni razvidno, ali so bili v razred ideacijskih označevalcev naštetí vezniki in prislovi uvrščeni v vseh ali samo določenem tipu (pragmatičnih) rab. Glede na nabor lahko sicer sklepamo, da gre predvsem za povezovalce izjav in ne propozicij, zato poimenovanje ideacijski morda ni najustreznejše, saj se ta v tuji literaturi povezuje z medpropozicijskimi povezavami na predstavní oz. semantični ravni. Če po drugi strani ta skupina zajema pragmatične povezovalce, bi med njimi pričakovali tudi nekatere druge argumentacijske oz. pragmatične veznike (npr. *ker*, *čeprav* ipd.), kakršne v predhodnih slovenističnih raziskavah izpostavljata M. Schlamberger Brezar in T. Balažic Bulc.

neintegriranimi izrazi, temveč nekatere izmed teh vidikov vključujemo na nižje ravni njihovega opisa.

S tovrstno opredelitvijo se zelo približamo teorijam koherence v govoru, znotraj katerih se je izraz diskurzni označevalec tudi vzpostavil, a se v primerjavi s interakcijsko-sociolingvističnim pristopom (Schiffrin, 1987) osredotočimo zgolj na tiste, ki se povezujejo s tvorčevim predhodnim ali nadaljnjim besedilom (izločimo odzive, kot so odgovori na vprašanja, čustveni odzivi, vzkliki, oporni signali, vljudnostni obrazci, nagovori ipd., ki jih obravnavamo kot samostojna sporočila),⁴⁷ v primerjavi z modelom paralelnih komponent (Redeker, 1990, 2000, 2006) pa ne vpeljujemo pogoja, da mora biti ta besedilna enota celotna izjava, saj nekaterih metabesedilnih sredstev ni mogoče opisati z vidika njihovega stavčnega dometa (npr. uvajalci apozicij, samopopravkov ali mašila), a hkrati z zagotavljanjem tekočnosti govora na podoben način prispevajo k smiselni povezanosti posredovanega besedila. Taka funkcijska zamejitev diskurznihih označevalcev se torej vzpostavlja tudi kot nadpomenka dosedanjih slovenističnih terminov, kot so opomenjevalni konektorji (Gorjanc, 1998; Balažic Bulc, 2009), členki s kohezivno vlogo (Skubic, 1999), pragmatični povezovalci (Schlamberger Brezar, 2009), členkovni besedilni povezovalci (Smolej, 2004a), govorni signali (Smolej, 2012), kazalci stopenj besedilne zgradbe (Smolej, 2012), diskurzni označevalci v ožjem smislu (Verdonik, 2007, 2008), ki vsak s svojega vidika sicer skupaj pokrivajo že dobršen del v nadaljevanju preučevanih izrazov.

Ker smo v pregledu literature že večkrat opozorili, da so krovne opredelitve diskurznihih označevalcev, kot je zgornja, glede na konceptualno izmuzljivost jezikoslovnih konceptov, kot so diskurz, kontekst ali ustaljenost, preveč abstraktne za metodološko dosledno razmejitev razreda

⁴⁷ S tem sicer iz naše opredelitve kljub deklarativnemu iskanju najmanjšega skupnega večkratnika prevladujočih teorij izločimo izraze, ki bi jih med diskurzne členke umeščali raziskovalci gramatikaliziranih označevalcev procesa interakcije (glej poglavje 2.2.2.2.3 *Diskurzni označevalci kot označevalci procesa govorjene interakcije*), npr. vljudnostni obrazci, odzivi, vzkliki ipd. Sami namreč menimo, da lahko v primerjavi z drugimi diskurznimi označevalci, ki se vsaj v nekem vidiku vpenjajo v govorčevo (in ne sogovorčevo) besedilo, tem izrazom pripišemo status samostojnih sporočil. Z drugimi besedami, izhajamo iz stališča, da brez jasno razdelanega modela diskurza ni smiselno vsega, kar je tipično govorjeno, privzeto umeščati na raven diskurza, temveč lahko to razumemo tudi kot njegovo (formulaično) vsebino. Ponujena opredelitev tako ponuja možnost nadaljnje razširitve, a v tem primeru zahteva premišljeno alternativno ubeseditev krovne funkcije diskurznihih označevalcev ter tudi razdelano razmerje do drugih nepropozicijskih oz. pragmatičnih izrazov, kot so eliptični označevalci odnosa do vsebine (npr. *na srečo*, *seveda*, *mogoče*), ki so prav tako pogosto rabljeni kot parentetični formulaični odzivi na sogovorčevo vsebino.

diskurzni označevalci od drugih propozicijsko neobveznih izrazov, svoje razumevanje diskurzne funkcije v nadaljevanju ponazorimo še s konkretnjšo funkcijsko taksonomijo, ki nas bo usmerjala pri prepoznavi in opisu večbesednih diskurzni označevalci v pričujoči raziskavi, prav tako pa podrobneje predstavimo izbrana strukturna merila, na podlagi katerih bomo večbesedne diskurzne označevalce ločevali od drugih funkcijsko analognih formulaičnih nizov. Pri tem obenem izpostavljamo, da je predlagana opredelitev diskurzni označevalci zgolj prvi korak k celostnejši obravnavi diskurznofunkcijske leksike in je kot taka odprta za nadaljnje dopolnitve, zlasti na podlagi metodološko ponovljivih korpusnih, psiholingvističnih in drugih empirično utemeljenih znanstvenih dognanj.

2.2.3.3 Merila za prepoznavo večbesednih diskurzni označevalci

2.2.3.3.1 Funkcijski opis

Kot je bilo v pregledu literature že večkrat izpostavljeno, na širšem področju raziskav diskurznofunkcijske leksike obstajajo številne in medsebojno zelo raznolike tipologije funkcij, ki naj bi jih ti izrazi opravljali v diskurzu, saj so te odvisne tako od njihove opredelitve kot siceršnjih raziskovalnih ciljev, izbranega gradiva in metod. V teoretskih razpravah se tako funkcije diskurzni označevalci običajno navajajo zgolj v obliki ilustrativnih seznamov odprtega tipa ali v obliki abstraktnjših opisov splošnih funkcijskih domen, kot so besedilno povezovanje, govorčevo besedilno upravljanje, strukturiranje, usmerjanje procesa interpretacije, povezovanje s sogovornikom in podobno, v aplikativno usmerjenih raziskavah, ki poudarjajo pomen prenosa teoretskih izhodišč na obsežnejše avtentično gradivo, pa se običajno navajajo končni spiski, podprti s podrobnejšimi označevalnimi navodili, ki poleg splošnejših opisov funkcij vključujejo tudi ponazoritve na avtentičnem gradivu in podrobnejše razlage težavnejših mest jezikoslovne kategorizacije.

Med slednjimi je glede na cilje in gradivo naše raziskave (teoretsko neobremenjena prepoznavna diskurzni označevalci v korpusu govorne slovenščine) najbolj relevantna tipologija, ki je v okviru projekta MDMA nastala na univerzi UC Louvain (Crible, 2014, 2016) in smo jo na kratko že predstavili v poglavju 2.1.4 *Vpliv aplikativnih pristopov k analizi diskurza*. Ta tipologija je bila namreč že v izhodišču razvita za jezikovno neodvisno označevanje funkcij diskurzni označevalci tako v pisnem kot govornem diskurzu, na podlagi sinteze nekaterih predhodnih

tipologij splošnih funkcijskih domen (Halliday in Hasan, 1976; Redeker, 1990; Sweetser, 1990; Gonzalez, 2005) in posameznih funkcij (Gonzalez, 2005; Prasad idr., 2008; Cuenca, 2013) ter analiz večkratnega neodvisnega korpusnega označevanja, samo kategorijo diskurznih označevalcev pa avtorji sheme prav tako opredelijo razmeroma teoretsko neobremenjeno in široko, kot skladenjsko neobvezne izraze, ki usmerjajo proces interpretacije tako, da označujejo, kako se besedilna enota, v kateri se pojavijo, umešča v mentalno reprezentacijo nastajajočega diskurza, ne glede na njihov znotraj- ali zunajbesedilni domet (Crible, 2016, razdelek 3.2.).

Funkcijska taksonomija MDMA (Crible, 2014) združuje 30 kontekstualno razlikovalnih funkcij diskurznih označevalcev. Podobno kot model vzporednih komponent (Redeker, 1990, 2000) tipologija na prvi ravni ločuje med **predstavnimi** (ideacijskimi) **diskurznimi označevalci**, ki vzpostavljajo razmerja med dejanskimi, zunajbesedilnimi dogodki (npr. vzrok, posledica, časovno razvrščanje, kontrast), **retoričnimi diskurznimi označevalci**, ki razkrivajo govorčevo metabesedilno upravljanje diskurza, tj. vzpostavljajo pragmatična razmerja med izjavami oz. diskurznimi dejanji (npr. utemeljevanje, sklepanje, pojasnjevanje), **sekvenčnimi diskurznimi označevalci**, ki organizirajo strukturiranje daljših besedilnih segmentov, tako na mikro- kot makroravni (npr. vračanje k temi, prevzemanje vloge, zapolnjevanje vrzeli), k tem trem pa podobno kot M. González (2005) v svoji analizi vloge označevalcev v pragmatični strukturi diskurza dodaja še četrto skupino, **medosebne diskurzne označevalce**, ki služijo za uravnavanje medosebnih razmerij med govorcem in naslovnikom (npr. preverjanje pozornosti, izražanje strinjanja ali nestrinjanja).

Kot prikazuje povzetek funkcij v Tabeli 1, smo izvorno tipologijo prilagodili tako, da funkcijo omiljevanja iz retorične skupine umestimo v skupino medosebnih označevalcev ter funkcijo uvajanja komentarja (digresije) tako kot G. Redeker (2000) iz retorične umestimo med sekvenčne označevalce, sicer pa sledimo izhodiščni razvrstitvi domen in funkcij.

2.2 Večbesedni diskurzni označevalci

predstavni označevalci	retorični označevalci	sekvenčni označevalci	medosebni označevalci
vzrok	utemeljitev	začetek	preverjanje
posledica	sklep	zaključek	ublažitev
protivnost	nasprotje	nadaljevanje	strinjanje
dopustnost	specifikacija	menjava teme	nestrinjanje
pogoj	relevantnost	komentar	elipsa
alternativa	reformulacija	navajanje	omilitev
čas	poudarek	dodajanje	
izjema	omilitev	naštevanje	
	komentar	zapolnjevanje vrzeli	

Tabela 1: Prilagojena funkcijska taksonomija diskurzni označevalcev po shemi MDMA.

Za lažje razumevanje posameznih funkcij in načina njihovega razlikovanja vsako izmed funkcij v nadaljevanju na kratko predstavimo in ponazorimo. Razen če ni drugače navedeno, opredelitve in primere povzemamo in prevajamo na podlagi izhodiščne označevalne sheme MDMA (Crible, 2014, str. 19–25, 30–35), ki se sicer tudi sama pogosto sklicuje na označevalno shemo korpusa PDTB (Prasad idr., 2007). Za lažjo predstavo jih mestoma dopolnjujemo tudi z nedavno predlaganim standardno (ISO) klasifikacijo semantičnih tipov v diskurzu (Prasad in Bunt, 2015; Bunt in Prasad, 2016) ter sopostavljamo z bolj ali manj podrobnimi diskurznofunkcijskimi tipologijami slovenskega prostora, ki so bile razvite na podlagi korpusnih raziskav pisne (Skubic, 1999; Balažic Bulc, 2009) ali govorjene slovenščine (Verdonik, 2007, 2008; Smolej, 2012).

2.2.3.3.1.1 Predstavni označevalci

Diskurzni označevalci **vzroka** (angl. *cause*) označujejo povezavo med dvema argumentoma,⁴⁸ med katerima obstaja logična vzročna povezava, pri čemer sta argumenta dva zunajbesedilna

⁴⁸ Za razliko od pristopa PDTB, ki izraz argument uporablja za označevanje stavčno izraženih abstraktnih objektov, kot so dogodki, stanja in propozicije (Prasad idr., 2007, str. 1), označevalna shema MDMA argument v kontekstu opisa dometa povezovalnih označevalcev razume širše, saj lahko ti besedilno enoto, v kateri se pojavijo, povezujejo (i) s soležno besedilno, tipično stavčno, enoto, (ii) z besedilno sekvenco več takih enot, kot je tematska ali sporočilna enota, ali (iii) s kontekstualno domnevo, ki besedilno ni bila izražena ali iz predhodnega besedila ni jasno razvidna (Crible, 2014, str. 15).

realna ali hipotetična dogodka. Ta funkcija izključuje označevalce pogoja. Parafraza: »To se je zgodilo, ker ...«. Primer: »*se mučijo ker je za nekatere to prva tovrstna izkušnja*«. // Tako Halliday kot ISO funkcijo vzroka obravnavata kot del združene vzročno-posledične kategorije (Skubic, 1999, str. 226). Tipični označevalci v slovenščini: *ker, kajti, saj, namreč*, kadar vzpostavljajo predstavno vzročno-posledično povezavo med dvema situacijama.

Diskurzni označevalci **posledice** (angl. *consequence*) označujejo povezavo med dvema argumentoma, kjer en argument opisuje logično posledico situacije, opisane v drugem argumentu, vključno z označevalci rezultata in namena. Parafraza: »Kot posledica tega se je zgodilo to.«. Primera: »*Najprej sem želel delati v marketingu ali oglaševanju, ampak sem ugotovil, da to ni zame. Zato sem opravil triletni tečaj marketinga*»; »*V podjetju Northeast pravijo, da bodo vlogo ponovno vložili in upajo, da jo bo FERC obravnaval prednostno, da bodo lahko nakup opravili že do naslednjega poletja*«. // Ta funkcija glede na tipologijo ISO združuje funkcije namena, (posledice) vzroka in (posledice) pozitivnega pogoja ter je tako sorodna Hallidayevi funkciji posledice (vzroka, razloga, namena ali pozitivnega pogoja) znotraj vzročno-pogojne kategorije (Skubic, 1999, str. 226–227). Tipični označevalci v slovenščini: *zato, torej, potem, da*, kadar vzpostavljajo predstavno vzročno-posledično povezavo med dvema situacijama.

Diskurzni označevalci **pogoja** (angl. *condition*) označujejo povezavo med dvema argumentoma, kjer en argument predstavlja (dejanski, pretekli, neuresničljiv) pogoj, drugi pa njegovo posledico, če je ta pogoj uresničen. Parafraza: »Le pod tem pogojem ...«. Primer: »*Poleg tega so pri podjetju Black & Decker rekli, da bodo odprodali še dva druga Emhartova obrata, če jim bo ponujena prava cena*«. // Ta funkcija ustreza funkciji (pozitivnega) pogoja v tipologiji ISO (povezavi med neuresničeno situacijo v prvem argumentu, ki bi v primeru uresničitve povzročila situacijo v drugem argumentu), A. Skubic pa je v prenosu Hallidayeve tipologije ne navaja, saj ta običajno ni del medpovedne skladnje, lahko pa bi jo obravnavali kot funkcijo pozitivnega pogoja znotraj vzročno-pogojne kategorije, kjer je trenutno omenjena zgolj posledica pogoja (Skubic, 1999, str. 227). Tipični označevalci v slovenščini: *če*, kadar vzpostavlja predstavno pogojno-posledično povezavo med dvema situacijama.

Diskurzni označevalci **protivnosti** (angl. *contrast*) označujejo povezavo med dvema argumentoma, ki imata skupno povedje ali lastnost, med njima pa so poudarjene razlike ali

nasprotja v vrednostih, ki jih tej lastnosti pripisujeta. Parafraza: »X je to, medtem ko je Y to.«. Primera: »*to lahko storiš na konkreten način, lahko pa to storiš tudi nekoliko bolj implicitno*«, »...imamo scenarij. Ampak na voljo imaš tudi dovolj časa, da sliko povečaš in si ogledaš, kaj se dogaja.« // Ta funkcija ustreza funkciji protivnosti v tipologiji ISO (poudarjanje ene ali več razlik med argumentoma v razmerju do njunih povedij ali posameznih lastnosti, ki jih omenjata) ter funkciji protivnega dodajanja v Hallidayevi tipologiji (Skubic, 1999, str. 221–222). Tipično označevalci v slovenščini: *pa*, *ampak*, *toda*, *vendar*, kadar vzpostavljajo predstavno nasprotje med dvema soležnima propozicijama.

Diskurzni označevalci **dopustnosti** (angl. *concession*) za razliko od označevalcev protivnosti, ki se nanašajo na eksplicitno izražene razlikovalne lastnosti, vpeljujejo zanikanje jasno razločljivega pričakovanja, ki ga je ustvaril eden izmed argumentov (bodisi realna situacija bodisi domneva). Parafraza: »Čeprav ..., se je zgodilo nekaj drugega.« Primer: »*v kraju, ki se imenuje Sutton, ki je v bistvu del Londona, ampak spada pod Surry*«. ⁴⁹ // Ta funkcija ustreza funkciji dopustnosti v tipologiji ISO (en argument zanika ali razveljavi pričakovano vzročno povezavo z drugim argumentom) in funkciji dopustnosti v Hallidayevi razdelitvi (Skubic, 1999, str. 228). Tipični označevalci v slovenščini: *čeprav*, *ampak*, *vendar pa*, kadar vzpostavljajo negacijo pričakovane vzročne povezave med dvema soležnima propozicijama.

Diskurzni označevalci **alternative** (angl. *alternative*) označujejo povezavo med dvema alternativnima situacijama, ki se lahko medsebojno izključujeta ali pa ne, pri čemer govorec ene ne vrednoti kot bolj ustrezne, temveč zgolj poroča o možnih dejstvih. Parafrazi »po eni strani ... po drugi ...«, »namesto tega«. Primera: »*ni dovoljeno distribuirati vedno znova Namesto tega bo ABC moral prodati pravice za enkratno plačilo*«. »Seveda veliko ljudi svojo poezijo objavlja

⁴⁹ Čeprav pri drugih tipih povezav shema MDMA upošteva položaj konektorja glede na enoti, ki ju povezuje, denimo v vzročno-posledičnih razmerjih razlikuje med označevalci vzroka (*ker*) in posledice (*zato*) ali v pogojno-posledičnih med označevalci pogoja (*če*) in njegove potencialne posledice (*potem*), pri dopustnosti ne ločujejo med vpeljevanjem enote pričakovanja (*čeprav*) in označevanjem enote njenega zanikanja (*ampak*). V izvorni shemi PDTB (Prasad idr., 2007, str. 34) to razliko označujejo s podtipoma *concession:expectation* (*although*) in *concession:contra-expectation* (*but*), pri čemer pa shema MDMA funkcijo *contra-expectation* po eni strani omenja v opredelitvi pragmatičnega nasprotja (*opposition*), po drugi strani pa jo s prevedenim zgledom z veznikom *but* oz. *ampak* ponazarjajo pri opisu funkcije dopustnosti. V izogib pretiranim modifikacijam izhodiščne sheme tudi sami znotraj iste kategorije dopustnosti ohranjamo obe rabi (»*čeprav* x, y« in »x, *ampak* y«), a smo iz prevoda opredelitve pragmatičnega nasprotja odstranili omembo protipričakovanja.

na spletu, bodisi na MySpaceu ali Facebooku, ali pa če so člani skupin in organizacij, ki dejansko objavljajo na spletu«. // Ta funkcija po tipologiji ISO ustreza funkcijam disjunkcije (povezava med dvema izključujočima ali neizključujočima alternativama), (posledice) negativnega pogoja (povezava med neuresničeno situacijo v enem argumentu, ki bi v primeru neuresničitve povzročila situacijo v drugem pogoju), podobno pa tudi v Hallidayevi tipologiji funkcijam spremembe-nadomestitve in spremembe-zamenjave, pa tudi posledice negativnega pogoja (Skubic, 1999, str. 222–223, 227). Tipični označevalci v slovenščini: *ali (pa)*, *ali-ali*, *sicer*, *oziroma*, *drugače*, *namesto tega*, kadar uvajajo predstavno ločno povezavo med dvema situacijama.

Diskurzni označevalci **časa** (angl. *temporal*) označujejo časovna razmerja med situacijami, opisanimi v posameznem argumentu, ki se bodisi prekrivajo ali si sledijo. Parafraza: »Po/pred/med tem, potem ...«. Primer: »in potem ko se je Theo takrat rodil, ali je vaš mož imel ...«. // Ta funkcija je v tipologiji ISO razdeljena na funkcijo sinhronosti (med povezanima argumentoma obstaja določena stopnja prekrivanja) in funkcijo asinhronosti (en argument se zgodi pred drugim), v Hallidayevi tipologiji pa funkciji istočasne, predhodne ali poznejše časovne dopolnitve (Skubic, 1999, str. 223–224). Tipični označevalci v slovenščini: *ko*, *potem*, *nato*, *hkrati*, kadar uvajajo predstavno časovno povezavo med dvema dogodkoma.

Diskurzni označevalci **izjeme** (angl. *exception*) označujejo povezavo med dvema argumentov, kjer en argument opisuje izjemo glede na posplošitev v drugem argumentu. Parafraza: »z izjemo ...«, »razen«. Primer: »samo naglas mogoče, ampak drugače govorimo skoraj isti jezik«. // Tudi v tipologiji ISO je tej funkciji ekvivalentna funkcija izjeme (povezava med argumentom, ki navaja okoliščine, v katerem opisana situacija drži, in argumentom, ki izpostavlja primer, v katerem ta ne drži), v Hallidayevi tipologiji pa bi bila tej najbližja funkcija spremembe-izvzemanja (Skubic, 1999, str. 222). Tipični označevalci v slovenščini: *samo da*, *razen tega*, *drugače pa*, kadar vzpostavljajo predstavno povezavo med ugotovitvijo in njeno izjemo.

2.2.3.3.1.2 Retorični označevalci

Diskurzni označevalci **utemeljevanja** (angl. *motivation*) uvajajo pragmatični vzrok (epistemski vzrok, vzrok med govornimi dejanji), ki se nanaša na subjektivno vsebino trditve ali govornega dejanja. Parafraza: »To pravim, ker ...«. Primer: »in vi ste bili prav na univerzi v Birminghamu

ker če prav razumem je v Birminghamu več univez?«. // Tako po tipologiji ISO kot Hallidayevi tipologiji bi tej funkciji ustrezale enake funkcije kot pri funkciji vzroka, saj ne razlikujejo med semantično in pragmatično rabo. Po taksonomiji Balažic Bulc to funkcijo opravljajo konektorji povezovalci utemeljevanja, ki povezujejo dve izjavi, od katerih druga argumentira trditev iz prve izjave (Balažic Bulc, 2009, str. 60–61). Tipični označevalci v slovenščini: *saj, namreč, ker, kajti*, kadar vzpostavljajo retorično povezavo med izjavo in njenim predhodnim besedilnim kontekstom.

Diskurzni označevalci **sklepa** (angl. *conclusion*) uvajajo pragmatično posledico (epistemska posledica, posledica med govornimi dejanji), ki vključuje sklepalno povzemanje, a izključuje preprosto parafraziranje (glej označevalce spremembe ubeseditve). Sklep običajno sledi kompleksnejšemu predhodnemu kontekstu, uvaja oceno ali posplošitev, povezavo med obema segmentoma pa vzpostavlja predvsem govorec. Parafraza: »Zdaj lahko rečemo, da ...«. Primer: »in pripoveduje o različnih vrstah, no, kotičkov v naravi, tako da je res krasna«. // Po tipologiji ISO torej tudi ta funkcija spada v splošno vzročno-posledično kategorijo, v Hallidayevi tipologiji pa v funkcijo povzetka in podkategorijo sklepalne posledice (Skubic, 1999, str. 226). Po taksonomiji Balažic Bulc funkciji sklepanja ustrezajo tako konektorji povezovalci sklepanja, ki v drugi izjavi uvajajo zaključen argument prve izjave, kot konektorji organizatorji sklepanja, ki uvajajo sklepno misel besedilne sekvence in hkrati odpirajo nadaljnji potek besedila (Balažic Bulc, 2009, str. 59–60, 67–68). Tipični označevalci v slovenščini: *zato, torej, potemtakem, tako, skratka*, kadar vzpostavljajo retorično povezavo med izjavo in njenim predhodnim besedilnim kontekstom.

Diskurzni označevalci **nasprotja** (angl. *opposition*) uvajajo pragmatično protivnost ali dopustnost (epistemska protivnost, protivnost govornih dejanj). Parafraza: »Čeprav sem to rekel, v resnici ...«. Primer: »...je precej majhen. V bistvu je prav lep, to je lep kraj za živeti«. // Tako tipologija ISO kot Hallidayeva tipologija ne razlikujeta med semantičnim in pragmatičnim nasprotjem, zato bi to kategorijo lahko umestili v splošnejši kategoriji nasprotja in dopustnosti. Balažic Bulc (2009, str. 51–52, 62–63) po drugi strani ločuje med konektorji povezovalci, ki izražajo nasprotje med dvema izjavama, od katerih je druga protiargument prvi, obe pa se nanašata na istega nanosnika (ti torej ustrezajo predstavnim funkcijam protivnosti in dopustnosti), in konektorji organizatorji, ki uvajajo nasprotje med besedilnima sekvencama znotraj istega

tematskega polja in torej ustrezajo funkciji retoričnega nasprotja. Tipični označevalci v slovenščini: *pa, toda, vendar, vendar pa, nasprotno pa*, kadar vzpostavljajo retorično povezavo med izjavo in njenim predhodnim besedilnim kontekstom.

Diskurzni označevalci **relevantnosti** (ang. *relevance*) uvajajo pragmatični pogoj, ko argumenta nista medsebojno vzročno povezana oz. eden ne vpliva na drugega, temveč pogoj v enem argumentu pojasnjuje situacijsko relevantnost govornega dejanja v drugem argumentu. Parafraza: »To lahko rečem samo v kontekstu ...« Primer: »Če si žejen, je v hladilniku pivo.« // Tipični označevalec v slovenščini: *če*, kadar povezuje dve izjavi, med katerima ni vzročnega razmerja.

Diskurzni označevalci **spremembe ubeseditve** (ang. *reformulation*) označujejo enakovrednost dveh preprostih solednjih besedilnih enot z drugačno ubeseditvijo, kar vključuje tako parafraziranje kot dejansko spremembo ubeseditve, pri kateri govorec drugo ubeseditve vrednoti kot bolj ustrezno (pravilno, relevantno). Parafraza: »Z drugimi besedami«, »Hočem reči ...«. Primeri: »*pa imaš več dela?* Mislim, ali je vedno več povpraševanja po prevajanju?«, »...vsak semester, torej trikrat letno«, »kako lahko sam jezik sproži drugačne, no, en kup senzoričnih izkušenj.« // Znotraj tipologije ISO tej funkciji ustrezata funkciji ponovne ubeseditve (*restatement*; oba argumenta opisujeta isto situacijo, a z različnih perspektiv) in zamenjave (*substitution*; argumenta sta alternativni možnosti, pri čemer je drugi argument vrednoten kot izbrana ali preferenčna izbira). V Hallidayevi tipologiji ji ustrezata funkciji preoblikovanja in pojasnila-popravka, Smolej pa jih v svoji tipologiji besedilnopovezovalnih členkov imenuje napovedovalci modifikacije oblike oz. parafraz (Smolej, 2004a, str. 50–51). V kontekstu raziskav samopopravljanj v spontanem govoru pa jih D. Verdonik imenuje tudi popravljalna sredstva (Verdonik, 2007, str. 133). Tipični označevalci v slovenščini: *se pravi, torej, namreč, oziroma, z drugimi besedami, mislim*, kadar vzpostavljajo povezavo med dvema pomensko podobnima besedilnima enotama.

Diskurzni označevalci **opisa** (ang. *specification*) označujejo povezavo med dvema argumentoma, kjer en argument vsebino drugega opiše še podrobneje ali pa jo ponazori s primeri. Parafraza: »Kar je/so/počne ...«, »na primer«, »še posebej«. Primera: »Učim na šoli, ki se imenuje Osnovna šola Devonshire, se pravi gre za osnovno šolo, v kateri so otroci stari od ...«, »In tako je na vseh ravneh. Tako da, na primer, mislim, čeprav je morda nekoliko kontroveržno, mislim, da je v

resnici precej pomembno za dijake.« // V tipologiji ISO bi tej opredelitvi ustrezale funkcije podobnosti (poudarjanje ene ali več podobnosti med dvema argumentoma), načina (en argument opisuje nastanek drugega), ponazoritve (en argument opiše okoliščine dane situacije, drugi pa njihove posamezne uresničitve) in elaboracije (oba argumenta opisujeta isto situacijo, a z več ali manj podrobnostmi), v Hallidayevi taksonomiji pa funkciji ponazoritve in pojasnila-izpostavitev, delno tudi načina (Skubic, 1999, str. 215, 219–220, 225). T. Balažic Bulc tovrstne izraze loči na konektorje organizatorje za izražanje ilustriranja, ki v besedilo eksplicitno vpeljujejo zgled (Balažic Bulc, 2009, str. 69–70), in konektorje povezovalce za izražanje pojasnjevanja, ki povezujejo dve izjavi, od katerih druga natančneje pojasnjuje oz. informira o argumentu, podanem v prvi izjavi (Balažic Bulc, 2009, str. 54–56), pri čemer glede na navedene zgled funkcijo pojasnjevanja razume nekoliko širše. Tipični označevalci v slovenščini: *in sicer, tj., torej, na primer, kot da*, kadar povezujejo dve izjavi, kjer druga podrobneje opiše prvo.

Diskurzni označevalci **poudarka** (angl. *emphasis*) izhodiščne sheme krepijo propozicijsko vrednost sosednje pragmatične funkcije in se največkrat pojavljajo v kombinaciji z drugim, pomensko manj izpraznjenim označevalcem. Poudarek se nanaša na drug sobesedilni izraz, ki ga krepi. Parafraza: ustreza prozodičnemu ali grafičnemu poudarku elementa, ki mu sledi. Primer: »*ampak v bistvu bomo ekspertizo pridobili tudi od zunaj*«. // Če to opredelitev nekoliko preubesedimo, so označevalci poudarka torej izrazi, ki naj bi poudarjali ilokucijsko silo besedilne enote, v kateri se pojavijo, npr. trditev, protiargument, vprašanje, pojasnilo ipd., kakršno seveda lahko soustvarja morebitni sosedni diskurzni označevalec.⁵⁰ Glede na to, da ne gre za funkcijo, ki bi povezovala dve besedilni enoti, tipologija ISO zanjo ne predvideva posebne funkcijske kategorije, prav tako tudi shema PDTB ugotavlja, da izraza, kot sta *in fact* in *indeed*, znotraj

⁵⁰ Dejstvo, da shema kot eno izmed možnih funkcij diskurznega upravljanja (upravičeno) predvideva tudi poudarjanje ilokucijske sile, pa pod vprašanje postavlja vprašanje njihovega razmejevanja od modalnih členkov, torej izrazov, ki ilokucijsko silo propozicije povezujejo s skupnim védenjem oz. s prepričanjem, ki ga govorec pripisuje naslovniku (glej kratek pregled opredelitev kategorije modalnih členkov v poglavju 2.2.2.2.2 *Diskurzni označevalci kot povezovalci besedila in sporazumevalnega konteksta*). Avtorji sheme namreč modalne členke obravnavajo kot samostojno, od diskurznihih označevalcev ločeno podkategorijo pragmatičnih označevalcev, predvsem na podlagi dejstva, da so modalni členki v angleščini in francoščini bistveno redkejši pojav kot v nekaterih drugih jezikih (Crible 2016: razdelek), vendar prav ponazoritev rabe označevalca *actually* kot uvajalca nasprotja z zunajbesedilnim predpostavljanim védenjem (*čeprav bi kdo mislil drugače, bomo v bistvu ...*) kaže na nesmiselnost ločevanja členkov v modalni in diskurzni vlogi, saj se ti razlikujejo zgolj v dometu.

njihove sheme v prihodnje ne bi smela biti obravnavana kot povezovalca (Prasad idr., 2007, str. 8). Po drugi strani funkcija poudarka ustreza Hallidayevi funkciji podkrepitve (Skubic, 1999, str. 220), funkcijo poudarjanja ilokucijske sile in/ali krepitve pomena soležnega veznika pa pri nekaterih členkih (npr. *pa*) izpostavlja tudi SSKJ. Tipični označevalci v slovenščini: *pa, pravzaprav, v bistvu, dejansko*.

2.2.3.3.1.3 Sekvenčni označevalci

Diskurzni označevalci **odpiranja** (angl. *opening boundary*) običajno označujejo začetek nove vloge in s tem prevzem položaja govorca, lahko pa tudi začetek nove sekvence znotraj posamezne teme, če ga ni mogoče uvrstiti v katero izmed drugih sekvenčnih funkcij. Primer: »... *imeli tradicionalni poročni zajtrk?*« // »*Torej, kup stvari. Imeli bodo stvari, kot so ...*«. // Čeprav Hallidayeva izvorna taksonomija povezovalnih adjunktov te funkcije ne omenja, Skubic na podlagi poznejših razprav v svojem članku izpostavlja tudi skupino nadaljevalnih oz. fatičnih adjunktov, ki signalizirajo odziv na predhodno vsebino (Skubic, 1999, str. 229–230), ki delno lahko opravljajo tudi funkcijo odpiranja, delno pa druge vrste fatičnih procesnih funkcij. Prav tako funkcijo odpiranja pogosto omenjajo raziskave govornega diskurza, denimo znotraj skupine t. i. zaznamovalcev strukture konverzacije, ki organizirajo diskurz kot celoto z nakazovanjem odpiranja in zapiranja ter medsebojnega navezovanja manjših enot (Schlamberger Brezar, 2009, str. 84–86) ali t. i. kazalcev stopenj besedilne zgradbe, kot so kazalci temeljnega začetka (Smolej, 2012, str. 113–116). Tipični označevalci v slovenščini: *no, ja, dobro, torej, tako, zdaj*.

Diskurzni označevalci **zapiranja** (angl. *closing boundary*) označujejo namero zaključevanja seznama, tematskega sklopa ali vloge govorca in se tako vedno pojavljajo na koncu ali kot samostojne enote. Parafraza: »*ta tema/vloga je zdaj zaključena*«. Primer: »*in to so opazili tako otroci kot jaz, tako da*«. // Tudi ta funkcija se v tipologiji M. Schlamberger Brezar umešča med zaznamovalce strukture konverzacije, v tipologiji M. Smolej pa med kazalce konca (Smolej, 2012, str. 116–119), zlasti kazalce odprtega konca, ki so rabljeni samostojno oz. brez besednega nadaljevanja. Tipični označevalci v slovenščini: *tako da, to je to, in tako naprej*.

Diskurzni označevalci **nadaljevanja** (angl. *resuming*) označujejo namero povezovanja segmenta, ki sledi, s predhodno temo, povratka k temi po digresiji, oklevanju ali nerelevantnem ekskurzu.

Parafraza: »če se vrnemo nazaj k naši temi«. Primer: *»posamezni tipi reakcij in posamezne senzorične izkušnje. Tako da, v povezavi s poezijo seveda, lahko to storiš tudi v konkretnem smislu«, »se je dejansko uničila, se mi zdi, kar se tiče vzorca. Ampak, ja, življenje se je obrnilo na glavo. // Tudi to funkcijo M. Schlamberger Brezar navaja kot del zaznamovalcev strukture konverzacije, M. Smolej pa kot delilne signale nadaljevanja (Smolej, 2012, str. 120–123). Tipični označevalci v slovenščini: no, v glavnem, skratka.*

Diskurzni označevalci **menjave teme** (angl. *topic-shifting*) označujejo spremembo teme znotraj vloge ali med dvema vlogama. Za razliko od funkcije nadaljevanja je morebitna povezava s predhodnim kontekstom zelo šibka oz. je ni. Nova tema je lahko podtema prejšnje, ki pa mora biti zaključena. Parafraza: *»pojdimο zdaj na ...«*. Primera: *»tako prakse tudi delujejo, so partnerstvo enakopravnih ljudi. In vi želite, vprašali ste me o podpornem osebju. Običajno ...«, »to je zelo surova pokrajina, zelo prostrana, tako da. Ampak moj naglas je zelo specifičen za dve vasi, saj leži na meji med Yorkshirom in Lancashirom. // V Hallidayevi tipologiji tej funkciji delno ustreza funkcija pojasnila-opustitve (Skubic, 1999, str. 219), v tipologiji M. Smolej pa funkcija kazalcev novega začetka, ki nakazujejo pripoved o novih dogodkih, ki so s predhodnimi povezani le na tematski ravni (Smolej, 2012, str. 123–124). Tipični označevalci v slovenščini: no, zdaj, sicer pa, kakorkoli že.*

Diskurzni označevalci **komentarja** (angl. *comment*), ki smo ga za razliko od izhodiščne retorične interpretacije umestili med sekvenčne označevalce, označujejo uvajanje digresije, parenteze ali naknadnega dodatka, ki ni neposredno povezan z vsebino sporočanega, a prispeva k njenemu celostnemu razumevanju. Parafraza: *»mimogrede...«*. Primer: *»ena izmed stvari, ki se po mojem spreminjajo v vseh parkih, in tudi sami se zagotovo premikamo v to smer, je to, da se želimo vrniti nazaj k podjetjem v povojih«*. // Ta funkcija ustreza Hallidayevi funkciji pojasnila-zastranitve, M. Smolej pa jo omenja pri kazalcih vpete informacije, s katerimi govorec sporoča, da bo sledila digresija, ki bo dopolnila že sporočeno informacijo (Smolej, 2012, str. 125). Tipični označevalci v slovenščini: *mimogrede, samo tako, aja.*

Diskurzni označevalci **navajanja** (angl. *quoting*) označujejo začetek poročanega govora, svojega ali govora tretje osebe. Parafraza: *ustreza narekovajem. »you would understand oh the horse*

comes on«. ⁵¹ // To funkcijo slovenistične raziskave omenjajo manj pogosto, vendar D. Zuljan Kumar v analizi briških pogovorov denimo opozarja na pogosto rabo spremnih stavkov premega govora, ki pripomorejo k ustvarjanju vtisa, da upovedeno dogajanje poteka v tistem trenutku (Zuljan Kumar, 2007, str. 45). Tipični označevalci v slovenščini: *kao*.

Diskurzni označevalci **naštevania** (angl. *enumeration*) označujejo zaporedno urejanje diskurznihi dogodkov v obliki seznama, s čimer se deiktčno navezujejo na druge diskurzne dogodke.

Parafraza: »prvič ... drugič ...«. Primer: »*to območje, ki ga uporabljamo za športni park v Surreyju. Kot prvo, je malo ven iz glavnega kampusa, tako da to pomaga. Kot drugo, ...*«. // Ob širokem razumevanju koncepta argumentov bi lahko to funkcijo v tipologiji ISO umestili med sinhrono in asinhrono povezavo, v Hallidayevi tipologiji pa je za funkcijo urejanja znotrajbesedilnih odnosov predvidena posebna notranja (endorfična) kategorija dopolnitev, kjer funkciji naštevania sovpadata predvsem podkategoriji začetnosti, naslednosti in sklepnosti (Skubic, 1999, str. 225). Tudi T. Balažic Bulc pri funkciji razvrščanja ločuje med povezovalci dveh izjav, ki si pomensko sledita v določenem zaporedju (označevalci predstavne časovne funkcije), ter konektorji organizatorji, ki razvrščajo besedilne sekvence (Balažic Bulc, 2009, str. 66) in torej ustrezajo funkciji sekvenčnega naštevania. Tipični označevalci v slovenščini: *najprej, prvič, za začetek, drugič, dalje, nenazadnje*.

Diskurzni označevalci **dodajanja** (angl. *addition*) v drugem argumentu vpeljujejo dodatne, nove informacije, ki so povezane s situacijo v prvem argumentu, in jih ni mogoče tolmačiti z drugo retorično funkcijo, kot je sklepanje ali podrobnejši opis. Parafraza: logični operator »+«. Primera: »*kurikulum je precej usmerjen v igro in angleščine se res hitro naučijo*«, »*Ime mi je Fiona Doloughan in sem učiteljica*«. // Tej funkciji po tipologiji ISO ustrezata tako funkcija vezave (*conjunction*, oba argumenta sta v enakem razmerju do neke druge situacije v diskurzu, pri čemer njuna vezava kaže na to, da glede na to situacijo oba argumenta opravljata isto vlogo ali jo počneta skupaj) kot razširitve (*expansion*, drugi argument uvaja nadaljnji opis ene ali več entitet

⁵¹ Zaradi specifičnih funkcijskih lastnosti medmeta *oh* v angleščini primer ni povsem prenosljiv v slovenski prevod. lahko pa bi ga ponazorili s funkcijo označevalca *kao*, rabljenega v primeru: *a po sem pomislila veš zato k on kadar mu je slabo a veš zmeri tko dela [1 glas] ne in je mogoče mislu da k sem jz [1 glas] ne da mi je slabo in da je pršu mogoče kao ja boga moja a veš*.

prvega argumenta, s čimer se fokus pripovedi premakne proti drugemu argumentu), v Hallidayevi tipologiji pa poleg funkcije dodajanja, kjer ena poved prejšnji doda novo, s prvo povezano informacijo, še funkciji hkratnega notranjega (znotrajbesedilnega) razvrščanja in ozirne dopolnitve, ki označuje, da je treba novo poved razlagati v oziru oz. tematskem okviru prejšnje (Skubic, 1999, str. 220–221, 225, 229). Glede na to, da shema MDMA ne razlikuje med različnimi dometi označevalcev dodajanja, bi lahko v to kategorijo iz tipologije T. Balažic Bulc (2009) uvrstili tako konektorje povezovalce kot konektorje organizatorje za izražanje povezovanja. Medtem ko prvi povezujejo dve izjavi, kjer druga dodaja novo informacijo o nanosniku iz prve (Balažic Bulc, 2009, str. 56–57), drugi najavljajo naslednjo besedilno sekvenco, ki se tematsko navezuje na predhodno (Balažic Bulc, 2009, str. 64–65). Tipični povezovalci: *in, pa, pri tem, ob tem, prav tako, poleg tega, glede tega, hkrati*.

Diskurzni označevalci **zapolnjevanja vrzeli** (angl. *punctuation*) signalizirajo željo pa zadrževanju vloge med načrtovanjem nadaljnjega govora ali zaradi katerega koli drugega razloga, ki ga ne pokriva nobena izmed drugih sekvenčnih funkcij. Parafraza: *ustreza vejicam*. Primer: »... *malo bolj s poti so, ne vem, so kar posebne.*« // V slovenskem prostoru se je z njihovo analizo ukvarjala zlasti D. Verdonik (2007, 2008; Verdonik idr., 2007). Tipični označevalci v slovenščini: *eee*,⁵² *mmm, ne vem, bi rekel*.

2.2.3.3.1.4 Medosebni označevalci

Diskurzni označevalci **spremljanja** (angl. *monitoring*) označujejo preverjanje razumevanja in pozornosti v obliki eksplicitnega nagovora sogovornika, vključno z označevalci skupnega védenja oz. že znane informacije. Ker so izrazi v vlogi označevalcev spremljanja pogosto avtomatizirani do te mere, da jih je moč interpretirati tudi kot zapolnjevalce vrzeli, shema daje prednost njihovi izrazni usmerjenosti v govorca in jih ne glede na poudarjenost ali izpraznjenost fatične funkcije umešča med označevalce spremljanja. Parafraza: »saj veš, kaj mislim«. »in

⁵² Čeprav izhodiščna shema nejezikovna mašila izloča iz diskurznih označevalcev in jih umešča na ločeno raven označevanja netekočnosti v govoru (Crible, 2014, str. 4), jih na podlagi njihove sorodne funkcije z leksikalnimi zapolnjevalci vrzeli (Swerts, 1998; Clark in Tree, 2002; Tottie, 2011) tako kot Verdonik (2007, 2008) umeščamo med diskurzne označevalce.

potem se spremeni, a ne, v bolj praktično angleščino v resnici«. ⁵³ // Ta funkcija je sorodna funkciji konativnih signalov nadaljevanja, ki služijo za uravnavanje naslovnikove pozornosti, za pozivanje k poslušanju, poudarjanje in preverjanje razumevanja (Smolej, 2012, str. 89–111), prav tako jih v svojih analizah kot označevalce odnosa med sogovornikoma omenja Verdonik (2007, str. 111–112). Tipični označevalci v slovenščini: *(a) ne (?)*, *(a) veš, razumeš, glejte*.

Diskurzni označevalci **ublažitve** (angl. *face-saving*) izražajo spoštljivost, vljudnost in preprečujejo ogrožanje govorcevega ali naslovnikovega dostojanstva. Parafraza: »to je občutljiva tema«. Primer: »*prihajam iz okolja, ki, saj veš, verjetno bi rekli, da moja družina spada v srednji razred*«. ⁵⁴ // Te funkcije slovenistične raziskave govorjenega jezika ne omenjajo; M. Smolej (2012, str. 102) kot različico konativnih signalov *veš* in *a veš* sicer navaja tudi zvezo *saj veš*, a ji pripiše vlogo opozarjanja naslovnika na že znano informacijo, ki po tipologiji MDMA ustreza funkciji spremljanja (zgoraj), saj tudi med navedenimi primeri ni rabe te zveze v funkciji izogibanja občutljivi vsebini.

Diskurzni označevalci **strinjanja** (angl. *agreeing*) označujejo razumevanje. Ta funkcija se ne pripisuje izrazom, ki nastopajo v vlogi polnih odgovorov na vprašanje ali jih lahko razumemo kot zaključena sporočila. Parafraza: »razumem, se strinjam«. Primer: »*Regeneracija torej ni namenjena zgolj gradnji bivališč in stavb, ampak tudi izgradnji parkov in ustvarjanju okoljsko prijaznejše prihodnosti? / »absolutno, ja. [...] do leta 2015, tako da, ja, zelo je usmerjena v okolje*«. // Izraze v funkciji označevalcev strinjanja in nestrinjanja (spodaj), ki pogosto opravljajo tudi druge diskurzne funkcije, v slovenskem prostoru omenja zlasti D. Verdonik (2007). Tipični označevalci v slovenščini: *ja, dobro, okej*.

Diskurzni označevalci **nestrinjanja** (angl. *disagreeing*) označujejo izražanje nestrinjanja z vsebino diskurza predhodnega govorca. Ta funkcija se ne pripisuje izrazom, ki nastopajo v vlogi polnih odgovorov na vprašanje ali jih lahko razumemo kot zaključena sporočila. Parafraza: »ne«.

⁵³ Originalni zgled: »and then it'll turn into, you know, more practical English really«.

⁵⁴ Originalni zgled: »I come from a background where, you know, now I guess my family you would say is middle class«.

Primer: »*torej si ti ekvivalent Cantone?*« »*No, nisem tako dober kot je on.*« // Tipični označevalec v slovenščini: *no, ne vem.*

Diskurzni označevalci **elipse** (angl. *ellipsis*) označujejo namerno nedoločnost, nakazujejo vključitev drugih elementov predhodne kategorije brez konkretnega poimenovanja. Običajno so to označevalci odprte propozicije. Parafraza: »in take stvari«. Primer: »*Bilo je veliko trgovanja. Mislim, da veliko začimb in tobaka in take stvari tako da očitno ...*«. // D. Verdonik te izraze umešča med t. i. interpretacijske diskurzne označevalce oz. označevalce odprte propozicije (Verdonik, 2008, 2015b; Verdonik in Kačič, 2012). Tipični označevalci v slovenščini: *in tako naprej, pa tako, pa to, ali pa kaj.*

Diskurzni označevalci **omilitve** (angl. *approximation, hedging*) označujejo namerno izogibanje natančnosti oz. omiljevanje govorčeve zavezanosti k veljavnosti izrečenega, z izjemo omiljevanja zaradi zgoraj opisane ublažitve. Parafraza: »približno«, »ne dobesedno«. Primer: »*Tega ne učim kako bi rekla posebej.*«⁵⁵ // V slovenskem prostoru se je ta funkcija v kontekstu diskurznih označevalcev doslej omenjala zgolj kot del funkcijskega opisa označevalcev odprte propozicije (Verdonik in Kačič, 2012; Verdonik, 2015b), znotraj širše analize izrazov gotovostne oz. epistemske naklonskosti pa jo omenja tudi A. Pisanski Peterlin (2015c). Tipični označevalci v slovenščini: *nekako, na nek način, ne vem.*

2.2.3.3.1.5 Razdvoumljanje večfunkcijskosti

Predstavljena funkcijska taksonomija MDMA (z minimalnimi prilagoditvami) pokriva zelo širok spekter posameznih funkcij, ki jih lahko diskurzni označevalci opravljajo v diskurzu, hkrati pa omogoča teoretsko pogojeno filtriranje na posamezne podskupine: besediloslovne raziskave koherence pisnega diskurza bi tako kot relevantne običajno prepoznale izraze, ki delujejo na predstavi in retorični ravni, raziskave ožje pojmovane koherence govornega diskurza tudi take, ki delujejo na sekvenčni ravni, raziskave širše pojmovane koherence v govoru pa izraze, ki delujejo na vseh štirih ravneh. Po drugi strani bi raziskave metabesedilnosti iz svojega zanimanja

⁵⁵ Originalni zgled: »I don't teach that sort of separately.«

izključile predstavno raven, raziskovalci diskurzni členkov znotraj konverzijske analize pa bi se najpogosteje osredotočili zgolj na sekvenčno in medosebno raven.

Vendarle pa lahko že iz samih primerov razberemo, da je takšno funkcijsko razvrščanje diskurzni označevalcev možno le ob natančni analizi širšega diskurznega konteksta, tako besedilnega kot situacijskega, še zlasti ob upoštevanju dejstva, da v različnih funkcijah pogosto nastopajo enake oblike. Za angleški označevalec *so* tako označevalna shema MDMA (Crible, 2014, str. 36–39) denimo navaja primere, v katerih ta izraz opravlja funkcije izražanja posledice, sklepa, opisa, spremembe ubeseditve, nadaljevanja, prevzemanja ali zaključevanja vloge, za označevalec *and* funkcije dodajanja, opisa, posledice, časovnega razvrščanja, protivnosti, menjave teme in začenjanja nove vloge (prim. Schiffrin, 2006), za označevalec *but* funkcije protivnosti, dopustnosti, nasprotja, nadaljevanja teme, menjave teme ali zapiranja vloge, za označevalec *well* pa funkcije zapolnjevanja vrzeli, spremembe ubeseditve, odpiranja in izražanja nestrinjanja.

Shema MDMA zato na več mestih vključuje specifična navodila za razdvoumljanje med posameznimi funkcijami (npr. opis proti sprememba ubeseditve), uvaja načelo preferenčne izbire v primeru težjega razdvoumljanja (npr. predstavna interpretacija pred retorično ali retorična pred sekvenčno), dopušča možnost dveh funkcijskih oznak v primeru dvoumnosti oz. izrazite večfunkcijskosti (npr. v primeru prekrivanja medosebnih in sekvenčnih funkcij), še zlasti pa se posveča razlikovanju med pomensko sorodnimi predstavni in drugimi (pragmatičnimi) funkcijami (npr. predstavna proti retorična protivnost). Ločnica med predstavno in pragmatično (tj. retorično, sekvenčno ali medosebno) rabo povezovalcev se je namreč tako v aplikacijah te sheme (Crible in Zufferey, 2015) kot drugih shem, kot je denimo PDTB (Zufferey in Degand, 2013), izkazala kot točka, na kateri se interpretacije kompetentnih označevalcev najpogosteje razhajajo, čeprav se prav na tej ločnici pogosto utemeljujejo različne opredelitve diskurzni označevalcev in navsezadnje tudi vzpostavljajo celotna raziskovalna področja, kot so denimo raziskave metabesedilnosti, pragmatična analiza diskurza in podobno.

Ker tudi v nadaljevanju predstavljena metodologija naše raziskave poleg kvalitativnega vidika, kjer bi bilo mogoče razdvoumljati tovrstne funkcije glede na podrobno analizo sobesedila in širšega sporazumevalnega konteksta, vključuje tudi kvantitativni vidik, kjer ni mogoče

pričakovati kontekstualno občutljivih razdvoumljanj,⁵⁶ vprašanje zanesljivega funkcijskega razdvoumljanja na našem konkretnem gradivu podrobneje načnemo v predstavitvi rezultatov empiričnega dela.

2.2.3.3.2 Strukturna zamejitev

Glede na to, da prvotno izhajamo iz teoretsko nevtralne minimalistične besedilnofunkcijske opredelitve, ki morebitne pomenske odtenke označevalca v odvisnosti od skladenjskega položaja ali stopnje integriranosti vključuje na raven njegovega (gradivno utemeljenega) opisa in ne vnaprejšnje zamejitve, v svoji opredelitvi diskurzni označevalcev ne bomo razlikovali med skladenjsko neintegriranimi in integriranimi izrazi. Ker pa slednji odpirajo širok nabor struktur, v nadaljevanju podrobneje izpostavimo **tri mejne kategorije**, povezane z izrazno zamejitvijo večbesednih diskurzni označevalcev, kot bodo pojmovani znotraj te raziskave, od drugih funkcijsko analognih formulaičnih večbesednih enot.

2.2.3.3.2.1 Formalno integrirani označevalci

Formalno integrirane strukture lahko opredelimo kot strukture, ki na površinskoskladenjski ravni nastopajo kot glavni stavki ali odvisniki različnih tipov, a na pomenskосkladenjski ravni niso del propozicije (Skubic, 1999, str. 217). Formalno integrirani izrazi, kot so *to pomeni*, *da*, *lahko bi rekli*, *da*, *če ponazorim*, *če se vrnem* ipd., na površinski ravni torej bodisi odpirajo bodisi zapolnjujejo vezljivostna mesta, s pomenskega vidika pa se zdijo izpustljivi, saj ne vplivajo na skladenjsko in sporočilno celovitost preostalega besedila. Kot smo izpostavili že v pregledu raziskav, korpusnojezikoslovne raziskave besednih nizov dokazujejo, da tovrstne strukture v jezikovni rabi izkazujejo visoko stopnjo formulaičnosti, med njimi pa prevladujejo zlasti take s pragmatično in diskurzno funkcijo. Čeprav priznavamo, da z izključno funkcijskega vidika, kakršnega bi morale v ospredje postavljati nadaljnje besediloslovne raziskave koherentnosti, vzpostavljanje tovrstne ločnice ni potrebno, bomo v pričujoči raziskavi iz razreda večbesednih

⁵⁶ Podobno nezmožnost strojnega razdvoumljanja specifičnih funkcij in njen vpliv na teoretske tokove izpostavljata Webber in Joshi (2012, str. 49): »When computational discourse processing turned to machine learning methods based on reliably identifiable features, it abandoned (at least temporarily) the centrality of pragmatics and speaker intentions to discourse. That is, there were few or no features that directly indicated or could serve as reliable proxies for what role speaker intended his/her utterance to play in the larger discourse.«

diskurzni označevalci ne glede na njihovo pogostost **izključili tiste formulaične nadredne stavke** (v tujem jezikoslovju se zanje včasih uporablja tudi izraz *matrix clause*), ki odpirajo paradigmatško variabilna vezljivostna mesta (npr. po Skubicu *to pomeni, da; posebej kaže izpostaviti, ob tem lahko omenim*, toda ne *se pravi da, a veš*), saj jih obravnavamo kot nosilce samostojnih propozicij.

Po drugi strani med diskurzne označevalce **vkjučujemo vse druge tipe formalno integriranih struktur s prepoznano diskurzno funkcijo**, če izkazujejo zahtevano stopnjo ustaljenosti, ki jo znotraj te raziskave razumemo kot hkratno **sovpadanje funkcijske celovitosti, propozicijske izpustljivosti, skladenjske zaključenosti in pogostosti**.⁵⁷ Poleg formalno integriranih odvisnikov različnih tipov, kakršne izpostavljata A. Skubic (1999) in M. Smolej (Smolej, 2015b), to vključuje tudi druge vrste navidezno integriranih enot, ki niso nujno omejene zgolj na stavčne enote (npr. *kot rečeno, če ne drugega*), temveč se lahko pojavljajo tudi na nižjih skladenjskih ravneh (npr. *ali nekaj takega*), med katerimi so nekatere v slovenščini že povsem gramatikalizirane (npr. *kljub temu da, zato ker*). Med slednje umeščamo tudi t. i. modificirane povezovalce (Prasad idr., 2007, str. 9), tj. vezniške zveze, ki jih sestavljajo kombinacije veznika in pred njim stoječega prislova ali členka (Toporišič, 2004, str. 427). Čeprav slovenski priročniki naštevajo le omejen nabor tovrstnih izrazov (npr. *brž ko, medtem ko, še ko, potem ko, takoj ko, češ da, samo da, le da, s tem da* ipd.), V. Gorjanc (1998, str. 377) denimo izpostavlja, da so tovrstni frazeologizirani konektorji s kataforičnim modifikatorjem pogosto rabljena stilistična sredstva matematičnega diskurza, npr. *'tedaj in le tedaj, kadar', 'natanko tedaj, ko', 'natanko takrat, kadar'*, zato jih bomo ob izkazani ustaljenosti v rabi tudi sami obravnavali kot zveze in ne kot skladenjsko razstavljive kombinacije prislova in uvajalca njegovega prilastkovega odvisnika.

2.2.3.3.2 Semantično integrirani označevalci

Nekoliko več pozornosti je bilo v slovenističnem in širšem besediloslovnem prostoru namenjenih drugi skupini, semantično integriranim strukturam, ki so tako na površinski kot pomenski ravni

⁵⁷ Kot večbesedne diskurzne označevalce torej prepoznavamo tiste nize besed, ki jim lahko pripišemo eno od zgoraj opredeljenih funkcij, niso nosilci propozicije ali njen sestavni del, ne odpirajo slovnično ali pomensko vezljivostnih mest in so v rabi razmeroma pogosti. Pri tem na podlagi pregleda frekvenčnih pragov v poglavju o ustaljenosti kot prag za določanje ustaljenosti v rabi predlagamo 5 pojavitev na milijon pojavnic v referenčnem korpusu.

del stavčno izražene propozicije, denimo kot prislovi ali prislovni odvisniki, po katerih se lahko vprašamo, a jim lahko zaradi navezovanja na okoliščinsko pomensko sestavino, izraženo v predhodnem besedilu pripišemo tudi besedilnopovezovalno vlogo (Skubic, 1999, str. 216–217), npr. *najprej, nazadnje, v tistem trenutku, zaradi tega, na ta način, v tem primeru* ipd. Tudi V. Gorjanc (1998) kot posebno skupino frazeoloških konektorjev izpostavlja tiste, ki z navezovalno sestavino združujejo konektorsko in kohezivno funkcijo, pri čemer mednje poleg izrazov s kataforično sestavino, ki kažejo v besedilu naprej (te smo omenili že med formalno integriranimi strukturami, npr. *zato ker*), navaja tudi konektorje z anaforično sestavino (npr. *zaradi tega, na ta način, v tem primeru*), prav tako nekatere izmed teh izrazov (*poleg tega, prav tako*) med členke uvršča tudi Slovenska slovnica (Toporišič, 2004, str. 449). V tujih raziskavah se vključevanje tovrstnih anaforičnih izrazov razlikuje glede na to, ali v jezikoslovni analizi ločujejo med vzpostavljanjem koherence na ravni medleksemskih povezav (referenčna koherenca) in vzpostavljanjem koherence na ravni medbesedilnih povezav (relacijska koherenca; glej npr. Webber idr. (2003), Stede in Grishina (2016)), pa tudi glede na zahtevano stopnjo izpraznjenosti propozicijskega pomena (gramatikaliziranosti). Sami bomo v pričujoči raziskavi med diskurzne označevalce prištevali **tudi ustaljene izraze s kohezivno sestavino**, saj te pogosto kažejo na kompleksnejši besedilni kontekst, ustaljenost njihove diskurzne funkcije pa navsezadnje dokazuje tudi diahroni razvoj nekaterih slovenskih prislovov in veznikov, ki so nastali s sklapljanjem predloga in anaforičnega kazalnega zaimka (npr. *zato, potem*). Pri tem pa za razliko od Skubičeve ponazoritve semantično integriranih struktur mednje prištevamo zgolj tiste, ki izkazujejo zahtevano stopnjo ustaljenosti, kažejo v besedilu nazaj ali naprej in se navezujejo na besedilno izraženo propozicijo ali širši besedilni kontekst (npr. *zaradi tega, kljub vsemu, na ta način, za začetek*, ne pa npr. *ob naslednji priložnosti*).

2.2.3.3.2.3 Kolokacije diskurznih označevalcev

Tretja mejna strukturna kategorija potencialnih večbesednih diskurznih označevalcev pa so kolokacije diskurznih označevalcev, tj. statistično izstopajoče formalno in semantično neintegrirane sopojavitve dveh ali več diskurznih označevalcev, ki se s skladenjskega vidika zdijo razstavljive, s funkcijskega vidika pa odpirajo vprašanje, ali lahko njihovo funkcijo v diskurzu razumemo kot seštevek funkcij posameznih sestavin ali kot zveze kot celote. Že L. Schourup (1999, str. 233) izpostavlja, da je tovrstno gručenje diskurznih označevalcev pogosto, a še ne

povsem pojasnjeno, zato bi utegnile nadaljnje raziskave te problematike pomembno prispevati k razumevanju semantičnega statusa posameznih označevalcev. Med redkimi raziskavami tega vprašanja B. Fraser (B. Fraser, 2011, 2013) denimo izpostavlja pogosto sopojavljanje označevalcev nasprotja (npr. *'but, instead', 'yet, on the other hand'*), C. Koops in A. Lohman (2015) pa se osredotočata na vprašanje razmerja med zaporedjem posameznih označevalcev in njihovo gramatikalizacijo, pri čemer ugotovita, da vrstni red ni naključen, temveč je odvisen od diahronega razvoja posameznega označevalca.

V slovenskem prostoru pogosto sopojavljanje diskurzni označevalcev v korpusu telefonskih pogovorov izpostavlja predvsem D. Verdonik in sodelavci (2007, str. 167–168), ki prav tako ugotavljajo, da njihovo nizanje ni povsem poljubno (npr. *ja glejte*), nekateri označevalci pa v kombinacijah nastopajo pogostejše kot drugi. Ne glede na pogostost sopojava ta raziskava kolokacije označevalcev torej obravnava kot razstavljive kombinacije, medtem ko M. Smolej (2012) v svoji ponazoritvi govornih signalov navaja celotne formulaične kombinacije, kot so denimo delilni signali nadaljevanja *no v glavnem, no ampak, no in potem, tako da pol* ipd. V raziskavah pisne slovenščine ta problematika doslej eksplicitno še ni bila izpostavljena, seznam obravnavanih večbesednih izrazov v raziskavah A. Skubica (1999) in T. Balažic Bulc (2009) pa razkrivajo, da se kot nerazstavljive kombinacije obravnavajo predvsem zveze z označevalcem *pa*, npr. *obenem pa, vendar pa, sicer pa*.

Ker nobena od naštetih raziskav ne ponuja operativne rešitve za dosledno razločevanje kompozicijske in nekompozicijske interpretacije, se bomo v pričujoči raziskavi opirali predvsem na kriterij **strukturne in funkcijske samostojnosti** – kot večbesedne diskurzne označevalce torej prepoznavamo tiste statistično izstopajoče sopojava dve ali več diskurzni označevalcev, kjer noben izmed elementov ni skladenjsko integriran v predhodno ali nadaljnje besedilo (kot npr. *pa* v pogosti kolokaciji *no pa*), hkrati pa funkcije zveze kot celote ne moremo opisati s kombinacijo diskurzni funkcij, ki bi jih pripisali njenim posamičnim gradnikom (prim. npr. nerazstavljivo zvezo *no ja* na eni in razstavljivo zvezo *no ampak* na drugi strani). Morebitna težavna mesta tovrstnega razmejevanja ustrezno izpostavljam v empiričnem delu naše raziskave.

2.3 Diskurzni označevalci v uporabniških spletnih vsebinah

2.3.1 Uporabniške spletne vsebine

S splošnim tehnološkim napredkom druge polovice prejšnjega stoletja, vzpostavitvijo področja informacijske tehnologije, razvojem računalnikov in drugih elektronskih naprav za medčloveško komunikacijo se je v zadnjih treh desetletjih izrazilo povečal tudi jezikoslovni interes za preučevanje jezikovne rabe v teh novih sporazumevalnih okoliščinah. Kot njeno krovno poimenovanje se je že v začetku devetdesetih vzpostavil izraz **računalniško posredovana komunikacija** (angl. *computer-mediated communication*), ki v svoji najsplošnejši opredelitvi označuje kakršnokoli »komunikacijo, ki nastane ob medčloveški interakciji v obliki sporočil, posredovanih preko omrežno povezanih ali mobilnih računalnikov« (Herring in Androutsopoulos, 2015, str. 127), v katero se sicer uvršča zelo širok in raznolik nabor komunikacijskih tipov, od mobilne komunikacije v obliki sms sporočil do elektronske pošte, spletnih klepetalnic, forumov in družbenih omrežij, če naštejemo le nekaj najpogostejše preučevanih. Termin računalniško posredovana komunikacija je še danes eden najbolj uveljavljenih (navsezadnje iz njega izhaja tudi naslov ene najvplivnejših znanstvenih revij področja, revije *Journal of Computer-Mediated Communication*), kljub dejstvu, da pojem komunikacije obsega širši nabor pojavov kot zgolj komunikacijo z jezikom, koncept računalniške posredovanosti pa glede na laično razumevanje pojma računalnik neupravičeno sugerira izključenost sodobnih mobilnih 'računalnikov', kot so pametni telefoni ali druge vrste prenosnih naprav (Crystal, 2011, str. 2). S podobnimi argumenti so bila zato doslej predlagana še številna druga sorodna poimenovanja, kot so elektronska oz. elektronsko posredovana komunikacija (N. S. Baron, 2008), digitalna oz. digitalno posredovana komunikacija (*The Routledge Handbook of Language and Digital Communication*, 2015) in digitalni diskurz (Thurlow in Mroczek, 2011) ali komunikacija od-tipkovnice-do-zaslona (Jucker in Dürscheid, 2012), pa tudi poimenovanja, ki bolj kot prenosnik izpostavljajo novost komunikacijskih kanalov, npr. jezik novih medijev (Aitchison in Lewis, 2003; Thurlow in Mroczek, 2011).

S skokovitim razvojem in porastom rabe interneta se zato med slednjimi v zadnjem desetletju fokus s široko razumljene računalniško posredovane komunikacije vse bolj usmerja v raziskave jezikovne rabe na spletu, za kar so bila doslej prav tako ponujena različna poimenovanja in z

njimi povezane konceptualne zamejitve, npr. netspeak (Crystal, 2006), internetni jezik (Crystal, 2011) ter internetna oz. internetno posredovana komunikacija (Beißwenger, 2007; Yus, 2011), pri čemer se je v slovenskem prostoru bolj kot internetni jezik uveljavil izraz **spletni jezik** oz. jezik na spletu (Michelizza, 2015).⁵⁸

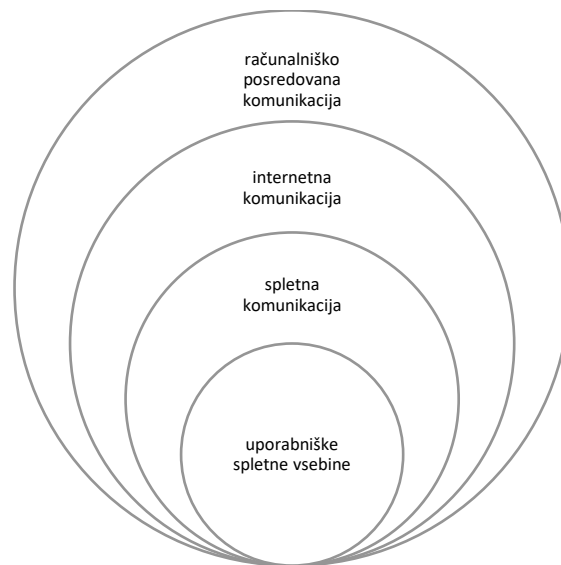
Znotraj širokega in žanrsko raznoliknega polja jezikovne produkcije, posredovane preko internetnih ali spletnih protokolov, pa je v zadnjih nekaj letih prišlo do vzpostavitve še enega pomembnega podpodročja raziskav računalniško oz. spletno posredovane komunikacije, in sicer raziskav t. i. uporabniških spletnih vsebin. Izraz **uporabniške spletne vsebine** (angl. *user-generated content*, tudi *user-created content*)⁵⁹ v svojem najširšem pomenu označuje »kakršnokoli obliko vsebine [...], ki jo ustvarijo uporabniki spletnih sistemov ali storitev, pogosto v okviru platform družbenih medijev«, pri čemer se jezikoslovne in jezikovnotehnološke raziskave osredotočajo predvsem na analizo *besedilnih* uporabniških spletnih vsebin, kot so spletni dnevniki (blogi), družbena omrežja (npr. Twitter in Facebook), forumi, kolaborativna spletišča (wikiji) in druge participativne spletne strani. Nastanek tovrstnih platform je namreč neločljivo povezan z razvojem t. i. spleta druge generacije oz. Spleta 2.0 (O'Reilly, 2007), ki za razliko od začetne vloge spleta kot sistema za povezovanje informacij, v katerem so uporabniki večinoma nastopali v vlogi pasivnih bralcev statičnih vsebin razmeroma maloštevilnih avtorjev, označuje skokovit razmah spletnih platform za povezovanje, izmenjavo informacij in sodelovanje, v katerih uporabniki spletnih vsebin nastopajo tudi kot njihovi aktivni soustvarjalci.

⁵⁸ Čeprav sta pogosto rabljena sinonimno, pridevnika *spletni* in *internetni* označujeta dva različna pojava informacijske tehnologije. Kot povzema članek na slovenski Wikipediji (<https://sl.wikipedia.org/wiki/Internet>), izraz internet (tudi medmrežje) označuje kakršnokoli računalniško omrežje, ki povezuje več omrežij, kot lastno ime (Internet) pa se rabi za poimenovanje največjega tovrstnega mednarodnega omrežja računalnikov, skupaj z informacijami in uslugami za uporabnike, kot so svetovni splet, elektronska pošta ali neposredni klepet (instant messaging). Internet (ki temelji na protokolu TC/IP) torej predstavlja omrežno povezanost strojne opreme in v njej shranjenih podatkov, svetovni splet (ki temelji na protokolu HTTP) pa je le ena izmed storitev, ki jo taka mreža omogoča, zaradi česar se pridevnik *internetni* torej vzpostavlja kot nadpomenka *spletnega*.

⁵⁹ Čeprav se v slovenskem jezikoslovnem prostoru sinonimno z izrazom *uporabniške spletne vsebine* (Fišer idr., 2015; Pollak, 2015b; Škrjanec idr., 2015) pojavlja tudi termin *spletne uporabniške vsebine* (Fišer idr., 2016), bomo sami v nadaljevanju uporabljali prvo različico, saj menimo, da ustrežneje izpostavlja obravnavano podmnožico spletnih vsebin glede na njene tvorce (spletne vsebine uporabnikov), ne pa tip uporabniških vsebin glede na medij (vsebine uporabnikov na spletu).

2.3 Diskurzni označevalci v uporabniških spletnih vsebinah

Kot povzemamo na Sliki 1, je torej jezikovna raba uporabniških spletnih vsebin le ena izmed podzvrsti širšega polja računalniško, internetno oz. spletno posredovane komunikacije, za katero S. C. Herring (2012a) ugotavlja, da jo med drugim zaznamujejo nove oblike vsebin (posodobitve statusov, metaoznačevanje vizualnih vsebin, urejanje wiki-jev), nove okoliščine (masovno občinstvo v več jezikih, geografske prilagoditve družbenih omrežij), novi vzorci rabe (sodelovanje, večopravilnost, večavtorstvo), nove tehnološke možnosti (npr. klepetalnice v spletnih igrah, igre družbenih omrežij, označevanje drugih uporabnikov ipd.) ter nove uporabniške prilagoditve (npr. interaktivna raba znakov @ in #, pojav retvitanja ipd.). V primerjavi s starejšimi oz. drugimi tipi računalniško posredovane komunikacije uporabniške spletne vsebine tako po eni strani izkazujejo nekatere podobnosti (npr. nestandardni zapis ali črkovanje, preklapljanje kodov, razlike med spoloma ipd.), po drugi pa se od njih tudi razlikujejo, bodisi s prenovljeno rabo že znanih prvin računalniško posredovane komunikacije (npr. posodabljanje statusa, navajanje in posredovanje vsebin drugih uporabnikov, izmenjava vlog, medbesedilnost ipd.), bodisi s pojavom povsem novih vidikov jezikovne rabe (npr. dinamično kolaborativno soustvarjanje besedil, multimodalna komunikacija z besedilom, grafiko, avdio- in videovsebinami ipd.).



Slika 1: Umestitev uporabniških spletnih vsebin znotraj področja raziskav računalniško posredovane komunikacije.

Kljub razmeroma jasni zamejitvi koncepta uporabniških spletnih vsebin z vidika obdobja nastanka, družbene vloge in tipov komunikacijskih platform v nadaljevanju raziščemo, kakšne so značilnosti jezikovne rabe uporabniških spletnih vsebin in v kolikšni meri je te sploh mogoče posploševati.

2.3.2 Značilnosti uporabniških spletnih vsebin

Eno osrednjih raziskovalnih vprašanj, ki raziskave računalniško posredovane komunikacije zaznamuje že vse od njihovih začetkov v devetdesetih letih prejšnjega stoletja, je vprašanje, **ali je računalniško posredovana komunikacija bolj podobna pisni ali govorjeni jezikovni rabi?**

Kot povzema D. Crystal (2011), je bila namreč ločnica med obema tipoma jezikovne rabe (medijema ali modalnostma) do pojava računalniško posredovane komunikacije obravnavana kot ena temeljnih dihotomij jezikoslovja z jasno opredeljenimi razlikami: če so za govor značilni omejenost s časom, dinamičnost, minljivost in prisotnost naslovnikov, pisni jezik po drugi strani zaznamujejo omejenost s prostorom, statičnost in obstojnost, naslovnik pa je tvorcu običajno oddaljen ali celo neznan. Vse te okoliščine jezikovne produkcije in recepcije so vplivale na razvoj specifičnih jezikovnih prvin ene ali druge modalnosti (za njihov pregled glej Crystal, 2011, str. 17–19; Sindoni, 2014, Poglavje 1), pri čemer je glede na raznolikost pisnih in govorjenih komunikacijskih tipov ustreznejše kot o absolutni dihotomiji govoriti o večdimenzionalnem kontinuumu med dvema skrajnima točkama (prim. Biber, 1988; Chafe in Tannen, 1987; Roberts in Street, 1998).

Čeprav so prvotne raziskave računalniško posredovane komunikacije to jezikovno zvrst skušale karseda natančno umestiti na ta kontinuum, bodisi v smeri pisne bodisi v smeri govorjene rabe (Murray, 1990; Ferrara idr., 1991; Maynor, 1994; Yates, 1996; N. S. Baron, 1998), je kopica kontrastivnih raziskav zadnjih dveh desetletij (Baym, 1995; Herring, 1996, 2007; Cherny, 1999) dokazala predvsem pretirano poenostavljenost izhodiščnega raziskovalnega vprašanja, saj tudi računalniško posredovane komunikacije zaradi njene heterogenosti **ni mogoče obravnavati kot ene enotne jezikovne zvrsti**, diskurznega tipa, žanra ali kakorkoli drugače poimenovane kategorije jezikovne rabe, temveč je raba jezika v posameznih tipih računalniško posredovane komunikacije (in s tem tudi delež prehajanja tipičnih prvin pisnega in govorjenega jezika)

odvisna od številnih dejavnikov. Ali, kot jedrnato povzema D. Crystal (2011, str. 21), »internetni jezik ni enak ne govoru ne pisanju, temveč selektivno in prilagodljivo izkazuje lastnosti obeh«.

Kot enega izmed možnih pristopov k razumevanju in sistematičnejšemu opisu kompleksnosti računalniško posredovane komunikacije, S. C. Herring (2007) tako predlaga začetno taksonomijo kontekstualnih dejavnikov oz. faset (angl. *facets*), ki vplivajo na jezikovno rabo računalniško posredovane komunikacije. Med **tehnološkimi dejavniki** denimo izpostavlja stopnjo sinhronosti (sinhrono in asinhrono), način posredovanja sporočila (enosmerni ali dvosmerni prikaz tvorjenja sporočila), obstojnost besedila (arhiviranje ali sprotno brisanje), velikost sporočila (omejenost s količino podatkov ali številom znakov), raznolikost komunikacijskih kanalov (prisotnost grafičnih elementov, video- ali avdiovsebin), stopnjo anonimnosti (tvorca in naslovnikov), stopnjo zasebnosti (nabor naslovnikov), možnost filtriranja (tvorcev sporočil), možnost navajanja (posredovanja vsebin drugih uporabnikov) in format sporočila (vrstni red sporočil, grafično oblikovanje). Med situacijskimi oz. **družbenimi dejavniki** na drugi strani pa avtorica izpostavlja dejavnike, kot so struktura sodelovanja (število aktivnih in potencialnih udeležencev komunikacije ter povezane stopnje zasebnosti, anonimnosti in aktivnosti), lastnosti udeležencev (demografske značilnosti, poznavanje medija, izkušnost, status, splošno védenje, ideološka in druga prepričanja), namen (splošni namen medija in cilj konkretnih interakcij), tema (nabor tem medija in konkretnih interakcij), ton (stopnja formalnosti), vrsta aktivnosti (konvencionalni način doseganja ciljev interakcije, žanr), norme (družbene organiziranosti, družbene sprejemljivosti in jezikovne rabe) in jezikovni kod (jezik/dialekt, zvrst in zapisovalni sistem). Vsi naštetí dejavniki (njihov nabor sicer ostaja odprt) torej z večjo ali manjšo težo sovplivajo na rabo jezika v posameznih tipih računalniško posredovane komunikacije ali celo v njenih posameznih uresničitvah.

Tudi uporabniških spletnih vsebin s tega vidika ne moremo obravnavati kot homogene zvrsti. Če skozi prizmo zgoraj predstavljene taksonomije primerjamo tri najpogostejše preučevane tipe uporabniških spletnih vsebin, tvite, bloge in forume, povzetek primerjave (tehnoloških dejavnikov) v Tabeli 2 kaže, da se v vseh treh besedilnih tipih običajno prevladuje asinhrona komunikacija, enosmerni prikaz nastajanja sporočila, arhiviranje posredovanih sporočil, kombiniranje s prvinami nebesedne komunikacije, možnost navajanja besedil drugih uporabnikov in prikazovanje v časovnem zaporedju, po drugi strani pa se tipi med seboj tudi razlikujejo,

denimo z vidika prostorske omejenosti posredovanih sporočil, anonimnosti tvorcev in naslovnikov, stopnje zasebnosti komunikacije, možnosti omejevanja naslovnikov ali oblikovnih specifik posredovanih sporočil. Poleg tega pogosto variirajo tudi značilnosti posameznih tipov, saj so te odvisne tako od tehnoloških možnosti platform, na katerih se pojavljajo, kot od individualnih izbir samih uporabnikov.

Kot primer primerjave na tem mestu med naštetimi dejavniki podrobneje izpostavimo stopnjo sinhronosti, ki naj bi sodeč po nekaterih raziskavah najbolj vplivala na stopnjo prehajanja govorjenih prvin; sinhrona oz. kvazisinhrona⁶⁰ komunikacija, kot je spletni klepet v resničnem času, naj bi bila namreč bolj podobna govorjeni, asinhrona komunikacija, kot sta elektronska pošta ali bloganje, pa pisni jezikovni rabi (Herring, 2001; Crystal, 2006; Condon in Čech, 2010; N. S. Baron, 2010). Tudi z novimi tehnološkimi možnostmi blogi še naprej ostajajo razmeroma statičen komunikacijski tip, namenjen asinhroni enosmerni komunikaciji, pri kateri tvorjenje besedila ni odvisno od časa njegove recepcije, prostor za morebitne odzive naslovnikov ali nadaljnjo diskusijo pa je od prvotnega besedila vsebinsko in oblikovno ločen. Čeprav je tudi Twitter primarno zasnovan za asinhrono komunikacijo, torej tvorjenje kratkih zaključenih besedil (mikroblogov), pa se glede na drugačen namen in format medija, tip aktivnosti ter z njo povezano strukturo sodelovanja na tvite naslovniki odzivajo pogostejše in predvsem v krajšem časovnem obdobju. Če se udeleženci diskusije na tvite soudeležencev odzivajo v realnem času, se lahko torej tudi prvotno asinhrona komunikacija sprevrže v precej sinhrono, dialoško izmenjavo sporočil. Podobno tehnološko možnost sicer omogočajo tudi (nekateri) forumi in razdelki s komentarji na blogih, a jo tam uporabniki zaradi drugačne strukture sodelovanja (npr. prisotnost administratorja), komunikacijskega namena in ustaljenih žanrskih konvencij redko izkoriščajo.

Že hitra primerjava enega izmed vidikov komunikacije (stopnje sinhronosti) v različnih tipih uporabniških spletnih vsebin torej kaže, da lastnosti te komunikacijske zvrsti ni mogoče

⁶⁰ Nekateri avtorji (Garcia in Baker Jacobs, 1999; Jucker in Dürscheid, 2012) opozarjajo, da lahko o pravi sinhronosti govorimo samo v primeru komunikacije, ki je posredovana znak za znakom (kjer torej naslovnik torej v realnem času spremlja samo nastajanje besedila), ne pa v obliki zaključenih izjav ali sporočil, zato kot ustrežnejše terminološko poimenovanje predlagajo izraz *kvazisinhrono*. Sami bomo v tej disertaciji pojem sinhronosti razumeli v enakem pomenu kot S. C. Herring (2007, str. 13), torej kot simultano komunikacijo med dvema ali več hkrati prisotnimi uporabniki, ne glede na obseg posredovanih sporočil.

2.3 Diskurzni označevalci v uporabniških spletnih vsebinah

posploševati, saj je odvisna od različnih tehnoloških in družbenih dejavnikov, zato je pri jezikoslovnih analizah uporabniških spletnih vsebin nujno zavedanje heterogenosti njenih posameznih podtipov, ki je bila dokazana tudi že v prvih primerjavah rabe različnih jezikovnih prvin v posameznih tipih slovenskih uporabniških vsebin (Fišer idr., 2015).

Tip vsebin Dejavnik	Twitter	forumi	blogi
T1. Sinhronost	asinhrono ali sinhrono	običajno asinhrono	asinhrono
T2. Način posredovanja	enosmerno	enosmerno	enosmerno
T3. Obstočnost	arhivirano	odvisno od platforme , običajno arhivirano	odvisno od platforme, običajno arhivirano
T4. Velikost	140 znakov	odvisno od platforme, običajno neomejeno	odvisno od platforme, običajno neomejeno
T5. Komunikacijski kanali	večinoma besedilo, tudi grafične, avdio in video vsebine	odvisno od platforme, večinoma besedilo	besedilo, tudi grafične, avdio in video vsebine
T6. Anonimnost	odvisno od uporabnika, običajno identificiran	odvisno od uporabnika, običajno anonimen	odvisno od uporabnika, običajno identificiran
T7. Zasebnost	javno, zamejeno ali zasebno	običajno javno	običajno javno
T8. Filtriranje	možnost zamejevanja soudeležencev (naslovnikov, tvorcev)	običajno brez možnosti filtriranja soudeležencev	odvisno od platforme, običajno brez možnosti filtriranja soudeležencev
T9. Navajanje	možnost navajanja predhodnih vsebin	običajno možnost navajanja predhodnih vsebin	običajno brez možnosti navajanja predhodnih vsebin
T10. Format	časovno nizanje, možnost gnezdenja	časovno nizanje, možnost gnezdenja odvisna od platforme	časovno nizanje

Tabela 2: Primerjava tehnoloških dejavnikov računalniško posredovane komunikacije po tipologiji Herring (2007) za tri tipe uporabniških spletnih vsebin.⁶¹

⁶¹ Za razčlenbo vseh (tako tehnoloških kot družbenih) dejavnikov rabe na blogih in geselskih člankih Wikipedije na podlagi iste sheme glej tudi (Michelizza, 2015, str. 244–256) .

2.3.3 Raziskave rabe diskurznihih označevalcev v uporabniških spletnih vsebinah

Čeprav se glede na novost področja jezikoslovne analize računalniško posredovane komunikacije pogosto osredotočajo predvsem na najbolj izstopajoče jezikovne pojave, kot so posebnosti v zapisovanju, črkovanju, besedotvorju ali stavčni skladnji (za pregled glej Bieswanger 2013; Crystal 2001, 2006, 2011, 57–77; Herring 2012), tudi v slovenščini (glej zbornik konference *Slovenščina na spletu in v novih medijih 2015*), v zadnjih petnajstih letih **narašča tudi delež diskurznoanalitičnih raziskav** (Yus, 2011; Herring idr., 2013b; Herring in Androutsopoulos, 2015), ki – skladno s široko opredelitvijo samega področja analize diskurza – načenjajo zelo raznolik nabor tem, od načinov strukturiranja diskurza in problematičnosti določanja temeljne enote analize (Condon in Čech, 2010; N. S. Baron, 2010; Jucker in Dürscheid, 2012), upravljanja menjave vlog (Cherny, 1999; Garcia in Baker Jacobs, 1999; Herring, 1999; Condon in Čech, 2010), tematske organizacije (Condon in Čech, 2010; Bou-Franch idr., 2012; Rashid, 2016), analize tipov govornih dejanj in strategij njihove rabe (Herring idr., 2005; Hancock, 2009; Harrison in Allton, 2013) do raziskav vpliva posameznih diskurznihih strategij na jezikovno rabo na nižjih jezikoslovnih ravneh (Sotillo, 2000).

Med redkimi raziskavami, ki tovrstna vprašanja raziskujejo prav skozi prizmo rabe diskurznofunkcijske leksike, se jih tudi danes večina osredotoča na 'starejše' tipe računalniško posredovane komunikacije, denimo primerjavo strategij rabe označevalcev menjave vlog v spontanem pogovoru ali spletnem klepetu (Degand in van Bergen, 2016) ali primerjavo rabe subjektivnih in objektivnih označevalcev vzroka v pisnih, govorjenih in klepetalniških besedilih (Sanders in Spooren, 2015), pri čemer obe našteti študiji izpostavljata precejšnjo podobnost opazovanih diskurznihih strategij v pogovorni in klepetalniški rabi. Kot v uvodu v zbornik pragmatičnojezikoslovnih raziskav računalniško posredovane komunikacije izpostavlja S. C. Herring s sodelavci (2013a, str. 4–5), obravnavano področje »zaznamuje izrazita vrzel v raziskavah fenomenov spleta druge generacije, kot so wikiji, mikroblogi in družbena omrežja.« Ni torej presenetljivo, da v pregledu literature najdemo **zelo malo jezikoslovnih raziskav rabe**

diskurzni označevalci⁶² v uporabniških spletnih vsebinah, četudi sorodne

jezikovnotehnološke raziskave nenehno izpostavljajo pomembno vlogo diskurzni označevalcev pri segmentaciji diskurza (Gimpel idr., 2011), zaznavanju nestrinjanja (Somasundaran in Wiebe, 2009; Abbott idr., 2011; Misra in Walker, 2013; Allen idr., 2014), analizi sentimenta (Mittal idr., 2013; Stepanov in Riccardi, 2015), modeliranju argumentacije (Peldszus, 2014; Eckle-Kohler idr., 2015; Habernal in Gurevych, 2016), slovničnem označevanju (Beißwenger idr., 2016) in drugih **aplikativnih ciljih računalniškega procesiranja uporabniških spletnih vsebin**.

J. E. Fox Tree (Fox Tree, 2015) je v svoji raziskavi rabe diskurzni označevalcev v različnih tipih pisne rabe izvedla dva ločena eksperimenta. V prvem je primerjala rabo različnih diskurzni označevalcev v odvisnosti od komunikacijskih namer tvorca (nestrinjanje-strinjanje, napad-podpora, občutenja-dejstva, nesramnost-prijaznost, nevtrarno-sarkastično) v korpusu spletnih forumskih diskusij in ugotovila, da z izjemo označevalca *well* (ki je sicer rabljen najpogosteje) vseh 17 analiziranih označevalcev kaže tendenco pojavljanja z določenimi tipi zapisov, npr. *oh* z nestrinjanjem, poudarjanjem občutenj, in sarkazmom, *actually* pa z nestrinjanjem, a ne sarkazmom. Drugi eksperiment, v katerem je bilo pet izbranih diskurzni označevalcev (*um*, *uh*, *you know*, *like*, *well*, *oh*) primerjanih še z vidika rabe v neposrednih in spletnih pogovorih med prijatelji, je avtor ponudil več zanimivih ugotovitev, in sicer se označevalci *like*, *you know*, *um* in *uh* bistveno pogosteje pojavljajo v govoru, vendar njihova raba na spletu kljub temu ni nezanemarljiva, kar kaže, da ti označevalci niso zgolj pomensko izpraznjeni tiki, ki nastajajo pod pritiskom govorjenega jezika. Po drugi strani je raba označevalcev *well* in *oh* v obeh tipih pogovorov podobno pogosta, kar kaže, da njuna raba ni pogojena s specifičnimi okoliščinami tvorjenja spontanega govora, temveč imajo poudarjeno besedilnopovezovalno vlogo, ki je pomembna tako v neposrednem kot spletnem klepetu.

⁶² V pregledu literature izpostavljamo raziskave, ki se osredotočajo na rabo izrazov, ki so kot diskurzni označevalci opredeljeni znotraj naše raziskave. V skladu z raznolikimi opredelitvami diskurzni označevalcev, ki smo jo izpostavili že pri splošnem pregledu raziskav te jezikoslovne kategorije, se tudi na področju raziskav računalniško posredovane komunikacije ta termin namreč uporablja za poimenovanje zelo raznolikih skupin nepropozicijske leksike, vključno z nebesednimi pojavi, kot so emotikoni, ki opravljajo tako vlogo označevalcev segmentacije besedila kot uravnavalcev medosebnih razmerij (Knox, 2009 v Zappavigna, 2012, str. 72).

S primerjavo rabe zapolnjenih premorov (*äh* in *ähm*) v neformalni govorjeni nemščini in nemških tvitih tudi **I. Rehbein (2015)** ugotavlja, da so ti v spletni pisni rabi rabljeni bistveno redkeje, a oba analizirana označevalca v obeh korpusih izkazujeta enake funkcije, tj. označevanje oklevanja, popravljanja, ponavljanja ali nedokončanih izjav, vendar v tvitih prevladujeta predvsem funkciji oklevanja in popravljanja. Ti rezultati torej prav tako dokazujejo, da raba mašil ni zgolj posledica kognitivne obremenitve, tipične za spontano govorjeno interakcijo, problem jezikovne performance, kar avtorica pojasnjuje s tezo, da so mašila besede z izven-propozicijskim pomenom, s katerimi tvorec svojemu besedilu namerno dodaja podtekst, kot so izražanje negotovosti, načrtno omiljevanje k zavezanosti izrečenega, humorni učinek ali izražanje negativnega odnosa. Prav tako sta **I. Rehbein in J. Ruppenhofer (2013)** v komplementarni raziskavi kot del diskurznihih označevalcev (poleg medmetov in opornih signalov) primerjala rabo vprašalnih pristavkov, ki so v korpusu tvitov rabljeni devetkrat redkeje kot v korpusu neformalnih pogovorov med najstniki, a le enkrat redkeje kot v korpusu pogovorov med poslovnimi partnerji.

Kljub pogostim raziskavam rabe besedilnih povezovalcev v tradicionalnih raziskavah pisnih besedil je njihovo rabo v uporabniških spletnih vsebinah doslej problematizirala le pilotna raziskava **T. Scheffler in sodelavcev (2016)**, ki so analizirali rabo besedilnih konektorjev v korpusu nemških tvitov. Po splošni ugotovitvi, da se v slabih dveh petinah vseh tvitov v povprečju pojavi vsaj en konektor, se nato osredotočijo zgolj na rabo različnih tipov vzročnih konektorjev, med katerimi v tvitih prevladuje raba konektorja *weil* (Scheffler, 2014), ki se tipično pojavlja predvsem v govoru, vendar se za razliko od govorjene komunikacije, kjer ta povezovalec enako pogosto nastopa v vlogah povezovalca propozicij, epistemskih izjav in govornih dejanj (glej poglavje 2.2.2.2.3 *Izvor povezave*), na twitterju prevladujoče rabi kot semantični povezovalec propozicij, redkeje pa v obeh pragmatičnih vlogah, kar se sicer razlikuje od ugotovitve raziskave T. Sanders in W. Spooren (2015), da v spletnem klepetu ne prihaja do razlik v deležu semantične in pragmatične rabe posameznih označevalcev.

Na koncu izpostavimo še raziskavo **Lin (2013)**, ki se osredotoča prav na večbesedne diskurzne označevalce, in sicer znotraj širše raziskave formulaičnih tribesednih nizov v neposrednih pogovorih in spletnih diskusijah britanskih in tajvanskih najstniških govorcev angleščine, primerjalno z rabo v dveh referenčnih angleških korpusih govorjene in spletne interakcije.

Analiza rabe najpogostejših tribesednih nizov s prevladujočo diskuržno funkcijo je pokazala, da je v referenčnih korpusih njihova raba pogostejša v govorjeni kot v spletni rabi, medtem ko so diskurzne formule v neposredni in spletni interakciji najstnikov rabljene približno enako pogosto. Ne glede na pogostost rabe pa se nizi med obema tipoma komunikacije kvalitativno razlikujejo, saj v obeh korpusih spletnih diskusij prevladujejo povezovalni nizi (npr. *and I am, as well as, and we were*), v obeh korpusih neposrednih pogovorov pa so pogosti tudi nizi z drugimi diskurznimi funkcijami, npr. vzpostavljanje tekočnosti (*I mean I*), ponazarjanje (*sort of thing*), podajanje mnenja (*to be honest*).

V slovenskih raziskavah uporabniških spletnih vsebin ali drugih oblik računalniško posredovane komunikacije diskurzni označevalci doslej niso bili podrobneje preučevani. Vendarle pa **Zwitter Vitez in Fišer (2015b)** v svoji pilotni raziskavi rabe tipično govorjenih besed (referenčnega govornega korpusa Gos glede na referenčni pisni korpus Kres) v primerjavi s tipičnim besediščem uporabniških spletnih vsebin (korpusa Janes glede na korpus Kres) ugotavljata, da je besedišče uporabniških spletnih vsebin v splošnem sicer bližje pisni slovenščini, med podobnostmi z govorjeno slovenščino pa med drugim izpostavljata predvsem pogosto rabo prislovov, medmetov in členkov v vlogi diskurzni označevalcev (npr. *ampak, itak, pač*).

Če povzamemo, so uporabniške spletne vsebine torej le ena izmed oblik spletne besedilne komunikacije, ki se od drugih oblik računalniško oz. spletno posredovane komunikacije razlikuje predvsem po tem, da jo tvorijo uporabniki različnih platform spleta druge generacije. Na jezikovno rabo uporabniških spletnih vsebin vplivajo številni tehnološki in družbeni dejavniki, zato te ni mogoče posploševati, tudi ne z vidika podobnosti s pisno ali govorjeno rabo. Kljub porastu raziskav na področju analize računalniško posredovanega diskurza, je bilo doslej rabi diskurznofunkcijskega besedišča v računalniško posredovani komunikaciji, še zlasti pa v uporabniških spletnih vsebinah, namenjene le malo pozornosti. Čeprav so ugotovitve dosedanjih raziskav le delno prekrivne, prav vse potrjujejo pogosto rabo diskurzni označevalcev v uporabniških spletnih vsebinah in vsaj delno sorodnost s strategijami upravljanja diskurza v govorjeni rabi, ne glede na funkcijsko skupino označevalcev ali komunikacijski tip.

3 Metodologija

3.1 Raziskovalna vprašanja

Da bi potrdili osrednje hipoteze pričujoče disertacije, tj. (i) da se v govoru tipično uporabljajo večbesedni diskurzni označevalci, povezani s specifičnimi dejavniki tvorjenja govornega diskurza, kot sta prisotnost naslovnika in tvorjenje v realnem času, (ii) da se ti zaradi svoje specifične vloge v govorni rabi v uporabniških spletnih vsebinah ne rabijo (iii) oziroma se rabijo predvsem v tistih tipih uporabniških spletnih vsebin s podobnimi sporočalnjskimi okoliščinami, bomo na podlagi zgoraj predstavljenega pregleda širšega področja sorodnih raziskav in izbranih teoretičnih izhodišč z nadaljnjo empirično analizo večbesednih diskurzni označevalcev skušali odgovoriti na naslednja raziskovalna vprašanja znotraj treh krovnih raziskovalnih ciljev:

- A. Opredelitev sheme za označevanje večbesednih diskurzni označevalcev v slovenskem jeziku:
 - 1. Ali merila, izdelana na podlagi predhodnih raziskav, omogočajo nedvoumno prepoznavo in funkcijski opis večbesednih diskurzni označevalcev v korpusu govorne slovenščine?
- B. Določitev nabora tipično govornih večbesednih diskurzni označevalcev govorne slovenščine:
 - 2. Katera metoda je najustreznejša za prepoznavanje večbesednih diskurzni označevalcev v besedilnih korpusih?
 - 3. Kateri večbesedni diskurzni označevalci se rabijo v govorni slovenščini?
 - 4. Kateri večbesedni diskurzni označevalci so tipični za govorno rabo?
- C. Analiza rabe tipično govornih večbesednih diskurzni označevalcev v slovenskih uporabniških spletnih vsebinah:
 - 5. Kako pogosto se najbolj tipični govorni večbesedni diskurzni označevalci rabijo v uporabniških spletnih vsebinah?

6. V kolikšni meri je pogostost rabe najbolj tipičnih govorjenih večbesednih diskurznih označevalcev odvisna od tipa uporabniških spletnih vsebin?
7. Ali najbolj tipični govorjeni večbesedni diskurzni označevalci v uporabniških spletnih vsebinah opravljajo enako vlogo kot v govorjeni rabi?

V skladu z izbranim korpusnim pristopom bomo za ta namen uporabili tri referenčne korpusne sodobne slovenščine: referenčni korpus govorjene slovenščine Gos, referenčni korpus slovenskih uporabniških spletnih vsebin Janes in referenčni korpus pisne slovenščine Kres. V nadaljevanju predstavimo njihovo podrobnejšo zgradbo in izbrani postopek analize.

3.2 Gradivo

Korpus govorjene slovenščine **Gos** je največja prostodostopna zbirka posnetkov in transkripcij spontanega govora v slovenskem jeziku, ki je nastala v okviru projekta Sporazumevanje v slovenskem jeziku (Verdonik idr., 2013; Verdonik in Zwitter Vitez, 2011). Vsebuje transkripcije približno 120 ur posnetkov (1 milijon pojavnic, 287 govornih dogodkov) 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. Slednji se v splošnem delijo na štiri diskurzne tipe: javni informativni in izobraževalni govor (35 % pojavnic), kot so televizijske in radijske oddaje, intervjuji, diskusije, šolske ure, univerzitetna predavanja; javni razvedrilni govor (22 %), kot so pogovorne oddaje, jutranji radijski program, športni prenos; nejavni nezasebni (15 %), kot so delovni sestanki, svetovanja, prodajne in druge storitve; in nejavni zasebni (28 %), kamor se uvrščajo pogovori med družinskimi člani in prijatelji. Korpus Gos vsebuje ročni prepis govora v dveh različicah (Verdonik in Zwitter Vitez, 2011, str. 57–68): pogovorni (ortografski zapis besede po načelu »zapiši, kot slišiš«, npr. *tko*) in standardizirani (standardni zapis po načelu »zapiši, kot pišemo«, npr. *tako*),⁶³ pri čemer bomo v tej raziskavi kot

⁶³ Pri tem sta z vidika opredelitve grafemske besede (pojavnice) obe ravni zapisa običajno med seboj simetrični (ena zapisana beseda oz. korpusna pojavnica v standardiziranem zapisu ustreza eni pojavnici v pogovornem zapisu), v nekaterih primerih pa je bila ena izgovorna beseda standardizirana v dve ali več pravopisnih besed ali pa je bilo več izgovornih besed pretvorjenih v eno samo pravopisno besedo. V prvi skupini so najpogostejše zveze členkov, zaimkov ali veznikov in reduciranih oblik pomožnega glagola biti (*toj* → *to_je*) ali drugih pogostih enozložnih

referenčnega obravnavali standardizirani zapis,⁶⁴ ki nevtralizira narečno, zvrstno ali drugače pogojene izgovorne posebnosti ter omogoča dolgoročno primerljivost s sorodnimi raziskavami v drugih korpusnih virih. Poleg leksikalnih pojavnic (besed) korpus Gos vsebuje tudi pojavnice, ki označujejo zapolnjene premore (mašila), nedokončane ali nerazumljive besede, osnovne prozodične informacije (npr. vprašalna intonacija), daljše premore, prekrivni govor, glasovne (npr. kašljanje) ali neglasovne zvoke (npr. zvonjenje). Korpus je bil ročno segmentiran na vloge, tj. enote govora enega govorca, in izjave, tj. pomensko, skladenjsko in akustično zaključene enote, standardiziranim pojavnici pa so bile strojno pripisani še podatki o osnovni obliki oz. lemi, besedni vrsti in drugih oblikoskladenjskih lastnostih (Grčar idr., 2012). V tej raziskavi smo uporabljali različico 1.4, kakršna je dostopna preko prostodostopnega konkordančnika noSketchEngine⁶⁵ in se od prvotne različice, dostopne preko specializiranega konkordančnika na uradni spletni strani,⁶⁶ ki omogoča tudi hkratno poslušanje izvornih posnetkov, razlikuje v tem, da vsebuje nekaj hevrističnih ročnih popravkov nekaterih najbolj problematičnih kategorij zapisa, lematizacije in/ali oblikoskladenjskih lastnosti (Verdonik, 2014).

Korpus slovenskih uporabniških spletnih vsebin **Janes** nastaja v okviru istoimenskega projekta, posvečenega jezikoslovni analizi nestandardne slovenščine na spletu. Najnovejša različica korpusa, Janes v0.4 (Fišer idr., 2016), ki smo jo uporabili tudi v tej raziskavi, vsebuje nekaj več kot 200 milijonov pojavnic (9 milijonov besedil različnih dolžin) iz šestih različnih zvrsti uporabniških spletnih vsebin: tvitov (107 milijonov pojavnic oz. 51,4 % pojavnic celotnega korpusa), forumov s treh specializiranih forumskih portalov (47 mil. oz. 22,6 %), blogov in komentarjev nanje na dveh blogerskih portalih (34,5 mil. oz. 16,6 %), komentarjev na novice treh novičarskih portalov (14,8 mil. oz. 7,1 %) ter pogovorne strani na slovenski Wikipediji (4,8 mil.

glagolskih oblik (*nem* → *ne_vem*), vezniški oz. drugi sklopi (*lihtak* → *glih_tako*) in zveze, pri katerih tudi zapis v pisnem jeziku kljub pravopisnim priporočilom niha med zapisom skupaj in narazen (*karkoli* → *kar_koli*, *pomoje* → *po_moje*), v drugi, bistveno manjši skupini pa prevladujejo predvsem črkovane kratice tipa *es_i* → *si*.

⁶⁴ V primeru nedoslednosti ali bližnjih parov v standardiziranih transkripcijah (npr. *po pa* – *pol pa* ali *se prav* – *se pravi*) teh ne popravljamo, temveč vsakega izmed variantnih zapisov obravnavamo kot svojo obliko, pri čemer bi v morebitnih nadgradnjah korpusa prepise tovrstnih navideznih variant veljalo poenotiti.

⁶⁵ http://nl.ijs.si/noske/all.cgi/first_form?corpname=gos

⁶⁶ <http://www.korpus-gos.net/>

oz. 2,3 %).⁶⁷ Korpus Janes tako za razliko od drugih oblik uporabniških spletnih vsebin vsebuje javno objavljeno, večinoma asinhrono komunikacijo, za razliko od zasebne, pogostejše sinhrono komunikacije klepetalniških družbenih omrežij, kot so Facebook Messenger, Viber ali WhatsApp. Besedila korpusa Janes dodatno vsebujejo še številne metapodatke, od podatkov o izvoru, avtorju, času nastanka, stopnji tehnične in jezikovne (ne)standardnosti, sentimentu in drugih specifikah izvirnega besedila, do avtomatsko pripisanih podatkov o normaliziranih oblikah izvornih pojavnic (npr. normalizirana oblika *jaz* za izvirne oblike *jest*, *jst*, *jas*, *js* ipd.) ter njihovih lemah in oblikoskladenjskih podatkih (Ljubešić idr., 2014; Ljubešić in Erjavec, 2016). Zaradi varovanja avtorskih pravic, zaščite osebnih podatkov in pogojev uporabe spletnih platform je korpus trenutno dostopen samo projektnim partnerjem, preko konkordančnika (no)SketchEngine.⁶⁸

Uravnoteženi korpus sodobne pisne slovenščine **Kres** (Logar Berginc idr., 2012) vsebuje približno 120 milijonov pojavnic, 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 %).⁶⁹ Tako kot druga dva obravnavana korpusa tudi korpus Kres vsebuje avtomatsko pripisane podatke o lemah, besedni vrsti in oblikoskladenjskih lastnostih posameznih pojavnic (Grčar idr., 2012), v isti različici (v1.0) pa je dostopen tako preko

⁶⁷ Celotni korpus Janes v0.4 sicer vsebuje večje število pojavnic (215 milijonov), saj so mednje štete tudi besedila spletnih novic (poleg komentarjev nanje). Tako kot avtorji korpusa (Fišer idr., 2016) tudi sami znotraj te raziskave podkorpusa novic v analizah ne bomo upoštevali, saj te ne spadajo med uporabniško ustvarjene spletne vsebine.

⁶⁸ http://nl.ijs.si/noske/janes04.cgi/first_form?corpname=janes.04

⁶⁹ Čeprav bi prepise sej državnega zbora, podnapise in druge oblike postproduksijskih besedil, ki se umeščajo v kategorijo *drugo* (5 % korpusa Kres) lahko obravnavali kot prepise izvirno govorjenih besedil, jih v nadaljnji analizi iz korpusa ne izločamo, saj menimo, da jih je zaradi večje ali manjše stopnje prilagoditev ustreznejše obravnavati kot poseben tip pisnih besedil, ne pa kot neposredno transkripcijo govora, kot je to značilno za besedila korpusa Gos.

konkordančnika na uradni spletni strani⁷⁰ kot preko specializiranega orodja noSketchEngine,⁷¹ preko katerega smo do korpusa dostopali tudi za potrebe te raziskave.

3.3 Struktura raziskave

Glede na prepletenost različnih raziskovalnih področij bo v skladu s predstavljenimi sedmimi raziskovalnimi vprašanji naša raziskava potekala v sedmih medsebojno povezanih metodoloških korakih. V **prvem** bomo z ročnim označevanjem večbesednih diskurznihih označevalcev v reprezentativnem vzorcu korpusa Gos v obsegu približno 30.000 pojavnic najprej kvalitativno **ovrednotili izhodiščna merila za prepoznavo in funkcijski opis** diskurznihih označevalcev, kakršna smo oblikovali na podlagi pregleda sorodnih raziskav, ter jih nato dopolnili z natančnejšo opredelitvijo problematičnih primerov, ki doslej v literaturi niso bili podrobneje obravnavani (poglavje 4.1.1.3 *Zamejitev nabora prepoznanihih označevalcev*), in ponovnim premislekom izhodiščne funkcijske taksonomije z vidika njene uporabnosti v kvantitativnih korpusnih raziskavah (poglavje 4.1.3 *Robustna izpeljava funkcij*).

V **drugem koraku** bomo izhodiščni nabor večbesednih diskurznihih označevalcev, prepoznanihih v procesu kvalitativne analize vzorčenega korpusa Gos, na podlagi istih meril **dopolnili še z avtomatskih luščenjem večbesednih diskurznihih označevalcev iz celotnega korpusa Gos** oz. ročnim pregledom seznamov statistično najbolj izstopajočih besednih nizov. Pri tem bomo glede na splošno neraziskanost tega področja uporabili več različnih statističnih metod za razvrščanje izluščenih besednih nizov, od različnih načinov štetja pogostosti do različnih izračunov besedne povezovalnosti. S primerjavo deleža diskurznihih označevalcev na seznamu najvišje uvrščenih kandidatov vsake izmed statističnih metod bomo nato določili tisto, ki je za prepoznavanje večbesednih diskurznihih označevalcev najbolj uporabna (poglavje 4.2.3.1 *Primerjava statističnih metod*), hkrati pa bomo s primerjavo rezultatov statističnih metod na eni strani in ročne analize vzorčenega korpusa na drugi primerjali tudi splošno uporabnost kvantitativnih metod na

⁷⁰ <http://www.korpus-kres.net/>

⁷¹ http://nl.ijs.si/noske/all.cgi/first_form?corpname=kres

področju, na katerem sicer izrazito prevladujejo kvalitativne korpusne metode (poglavje 4.2.3.2 *Primerjava statističnih metod in ročne prepoznave*).

Unijo rezultatov statističnih in ročnih metod prepoznavanja diskurzni označevalcev bomo v **tretjem koraku** nato združili v enoten seznam (poglavje 4.3 *Seznam večbesednih diskurzni označevalcev govorne slovenščine*), ki ga bomo v nadaljevanju raziskave obravnavali kot izhodiščni **nabor večbesednih diskurzni označevalcev govorne slovenščine**.

V **četrtem koraku** bomo nato za ta seznam skušali določiti podmnožico tistih označevalcev, ki se tipično uporabljajo predvsem v govorni rabi. Za ta namen bomo najprej z izbrano metodo razdvoumljanja diskurzni označevalcev izračunali tako pogostost njihove rabe v govornem korpusu Gos kot v korpusu pisne slovenščine Kres, na podlagi obeh podatkov pa bomo nato z izbrano metodo za izračun ključnosti označevalce razvrstili glede na njihovo **tipičnost za govorni korpus Gos** glede na pisni korpus Kres (poglavje 4.4.3.2 *Seznam tipično govornih večbesednih diskurzni označevalcev*).

V sklepnem delu raziskave bomo nato za tri najbolj tipične označevalce podrobneje analizirali še njihovo rabo v uporabniški spletni komunikaciji, in sicer bomo v **petem koraku** najprej ob upoštevanju zapisovalnih posebnosti na spletu izračunali **pogostost rabe** označevalcev v korpusu Janes ter jo primerjali s pogostostjo v korpusu Gos (poglavje 5.1 *Pogostost rabe v uporabniških spletnih vsebinah*). V **šestem koraku** bomo nato podrobneje analizirali še njihovo **rabo po posameznih zvrsteh** korpusa Janes (tviti, forumi, blogi, komentarji blogov,⁷² komentarji novic, wiki) in skušali **pojasniti morebitne razlike** v rabi glede na besedilno zvrst (poglavje 5.2 *Raba glede na tip uporabniških spletnih vsebin*). V zadnjem, **sedmem koraku** bomo s kvalitativno analizo vzorcev rabe vseh treh označevalcev v korpusu Gos in Janes skušali odgovoriti še na vprašanje, ali izbrani trije označevalci v korpusu Janes izkazujejo **enake funkcije** kot v korpusu Gos (poglavje 5.3 *Značilnosti rabe v uporabniških spletnih vsebinah*).

⁷² V raziskavi torej blogovske zapise in komentarje nanje obravnavamo kot dva ločena besedilna tipa.

4 Večbesedni diskurzni označevalci govorne slovenščine

4.1 Dopolnitev sheme za označevanje večbesednih diskurznih označevalcev

V tem poglavju izhodiščna merila za prepoznavo in funkcijski opis večbesednih diskurznih označevalcev govorne slovenščine, ki smo jih izdelali na podlagi sorodnih tujih in domačih raziskav (poglavje 2.2.3 *Izhodišča raziskave večbesednih diskurznih označevalcev*), dopolnimo z ugotovitvami, ki izhajajo iz procesa ročnega označevanja teh jezikovnih sredstev v reprezentativnem vzorcu korpusa govorne slovenščine Gos, in s tem odgovorimo na prvo raziskovalno vprašanje: »Ali merila, izdelana na podlagi sorodnih raziskav, omogočajo nedvoumno prepoznavo in funkcijski opis večbesednih diskurznih označevalcev v korpusu govorne slovenščine?«.

4.1.1 Priprava gradiva

4.1.1.1 Vzorčenje korpusa Gos

V prvem koraku priprave gradiva za ročno označevanje večbesednih diskurznih označevalcev v korpusu govorne slovenščine smo izdelali vzorec referenčnega korpusa govorne slovenščine Gos, pri čemer smo algoritem za njegovo izdelavo zasnovali tako, da je vzorec čim bolj reprezentativen glede na celoto (s čimer zagotovimo njegovo dolgoročno uporabnost v jezikoslovnih in jezikovnotehnoloških raziskavah), vključuje povezane nize zaključenih govornih vlog (s čimer zagotovimo možnost morebitne spremembe meril za njihovo segmentacijo na posamezne izjave) in omogoča naknadno razširitev (za potrebe obsežnejših raziskovalnih projektov).

Algoritem je bil implementiran v obliki računalniške skripte v programskem jeziku python, ki kot vhodno datoteko sprejme referenčni korpus v tabelarnem formatu,⁷³ kot izhodno datoteko vrne vzorec v enakem tabelarnem formatu in s pripadajočimi metapodatki, kot parameter pa je ob zagonu potrebna tudi določitev zelene minimalne velikosti izluščenega vzorca (minimalno število pojavnic). Ob zagonu program za vsako besedilo referenčnega korpusa izbere začetek naključne vloge med prvima dvema tretjinama vseh vlog besedila in v vzorčni korpus zajame toliko

⁷³ O tem formatu več v (Erjavec, 2013).

pojavnici, da je njihovo razmerje glede na želeno velikost vzorčenega korpusa enako razmerju vseh pojavnici besedila glede na velikost referenčnega korpusa, pri čemer je predvideno število pojavnici lahko preseženo, dokler vzorčena vloga ni zaključena. Z drugimi besedami to pomeni, da je bil iz vsakega besedila referenčnega korpusa vzorčen **naključni niz ene ali več zaporednih zaključenih govornih vlog** enega ali več govorcev, katerih skupno število pojavnici je **sorazmerno s številom pojavnici vzorčenega besedila glede na skupno velikost referenčnega korpusa**. Edina izjema so besedila javnega informativno-izobraževalnega diskurza, v katerih so zaradi monološke narave govornih dogodkov vloge posameznih govorcev pogosto zelo dolge ali obsegajo celotni govorni dogodek, zato se vzorčenje besedil tega tipa lahko konča zgolj z zaključeno izjavo in ne nujno zaključeno vlogo, to pa je ustrezno označeno tudi v izhodni datoteki.

Z opisanim algoritmom smo tako ob zagonu skripte na korpusu Gos v različici v1.4 z minimalno predvideno velikostjo vzorca 25.000 pojavnici iz korpusa Gos izdelali **vzorčeni korpus velikosti 29.468 pojavnici, kar predstavlja 2,7 % celotnega korpusa Gos**. Kot prikazuje Tabela 3, vzorčeni korpus vključuje enako število različnih govornih dogodkov (287) in podobno razmerje posameznih diskurznihi tipov, tj. 33,5 % besedil javnega informativno-izobraževalnega diskurza, 23 % besedil javnega razvedrilnega diskurza, 28 % nejavnega zasebnega diskurza in 15,5 % nejavnega nezasebnega diskurza, zaradi omejenosti na kratke zaključene segmente posameznih govornih dogodkov pa vzorčeni korpus vsebuje nekoliko manjše število različnih govorcev (606) kot referenčni korpus (1.561). Posamezno besedilo vzorčenega korpusa tako **v povprečju vsebuje 102 pojavnici, 11 izjav, 8 vlog in 2 različna govorca**.

4.1 Dopolnitev sheme za označevanje večbesednih diskurzni označevalcev

tip diskurza	besedila	govorci	vloge ⁷⁴	izjave	pojavnice
javni informativno-izobraževalni	129	263	703	959	9.898
javni razvedrilni	42	78	499	726	6.826
nejavni nezasebni	45	102	425	497	4.535
nejavni zasebni	71	163	833	1.006	8.209
SKUPAJ	287	606	2.460	3.188	29.468

Tabela 3: Velikost in sestava vzorčenega korpusa govorne slovenščine.

Dejstvo, da vzorec vključuje **širok nabor govornih situacij in različnih govorcev**, je za namen kvalitativnega pridobivanja seznama večbesednih diskurzni označevalcev pomembna prednost, saj nam omogoča zajem širokega nabora jezikovnih sredstev, ki ni omejen s specifikami posameznih govornih dogodkov, razmerij med udeleženci ali njihovimi demografskimi lastnostmi, kakršne pogosto zaznamujejo druge sorodne raziskave govornega jezika. Po drugi strani ob izhodiščno majhnem izbranem vzorcu to pomeni, da so vzorčeni segmenti posameznih besedil trenutno razmeroma kratki in manj primerni za raziskave, ki se ukvarjajo z razčlenjevanjem strukture diskurza, zato bi bilo v tem primeru predstavljeni algoritem vzorčenja smiselno metodološko nadgraditi s hkratnim upoštevanjem diskurzne zaključenosti, reprezentativnosti in ročne obvladljivosti.

4.1.1.2 Uvoz v orodje WebAnno

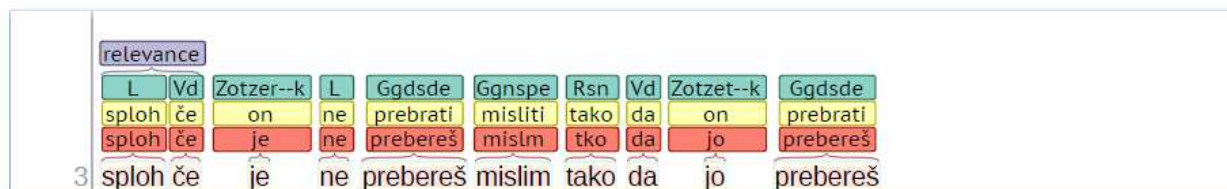
V drugem koraku priprave gradiva za kvalitativno prepoznavo in opis večbesednih diskurzni označevalcev smo vzorčeni korpus uvozili v označevalno orodje WebAnno⁷⁵ (Yimam idr., 2013), prosto dostopno spletno orodje, ki s prilagodljivim in uporabniško prijaznim spletnim vmesnikom omogoča razmeroma samostojno postavitvev in upravljanje raznolikih označevalnih projektov na različnih ravneh jezikoslovnega označevanja. Pred uvozom smo vzorčeni korpus iz izhodiščnega tabelarnega formata pretvorili v format TSV, ki ga zahteva orodje WebAnno, in nastavili

⁷⁴ Med vsemi vzorčenimi vlogami nezaključene vloge informativno-izobraževalnega diskurza predstavljajo zgolj 5,5 % vseh vzorčenih vlog.

⁷⁵ Med več možnimi načini dostopanja do orodja WebAnno smo uporabljali inštalacijo slovenskega konzorcija CLARIN.SI na naslovu <http://nl.ijs.si/webanno>.

4.1 Dopolnitev sheme za označevanje večbesednih diskurznih označevalcev

ustrezne parametre predvidenega označevalnega projekta, kot so določitev nabora besedil, ravni označevanja, njihovih lastnosti in predvidenega nabora oznak. V našem konkretnem primeru to pomeni, da smo poleg privzete ravni jezikoslovne analize (prikaza standardiziranih pojavnic) kot ravni z možnostjo potencialnega urejanja uvozili še pripisane podatke o pogovorni transkripciji, lemi in oblikoskladenjski oznaki JOS, kakršne so bile pojavnicam pripisane že v procesu samodejnega označevanja referenčnega korpusa, dodatno pa je bila za namen naše raziskave poleg teh ravni ustvarjena še raven za označevanje funkcijske kategorije prepoznanih večbesednih diskurznih označevalcev s pripadajočim seznamom 30 možnih oznak, predstavljenih v poglavju 2.2.3.3.1 *Funkcijski opis*. Primer prikaza pojavnic ene izjave z uvoženimi metapodatki prikazuje Slika 2, pri čemer orodje omogoča poljubno vklapljanje in izklapljanje posameznih ravni.



Slika 2: Prikaz označene izjave v orodju WebAnno s pripisom pogovorne oblike (rdeče), leme (rumeno), oblikoskladenjske oznake (zeleno) in funkcije prepoznanega večbesednega diskurznega označevalca sploh če (vijolično).

4.1.1.3 Zamejitev nabora prepoznanih označevalcev

Z opisanim postopkom smo v procesu ročnega označevanja vzorčenega korpusa prepoznali 176 potencialnih večbesednih diskurznih označevalcev, torej strukturno zaključenih večbesednih enot, ki jim je bilo na podlagi teoretičnih izhodišč (poglavje 2.2.3 *Izhodišča raziskave večbesednih diskurznih označevalcev*) in njihovih empiričnih dopolnitev, ki jih podrobneje predstavljamo v naslednjem poglavju (poglavje 4.1.2 *Težavnejša mesta kategorizacije*), mogoče pripisati vsaj eno izmed diskurzni funkcij taksonomije MDMA (poglavje 2.2.3.3.1 *Funkcijski opis*), pri čemer je bilo lahko v primeru funkcijske dvoumnosti ali hkratnega opravljanja več različnih funkcij označevalcu hkrati pripisanih tudi več različnih funkcij.

Pri tem se 51 identificiranih potencialnih označevalcev v celotnem referenčnem korpusu Gos pojavi manj kot 5-krat na milijon pojavnic ali pa se le redko pojavlja v diskurzni rabi, npr.

4.1 Dopolnitev sheme za označevanje večbesednih diskurzni označevalcev

neustaljeni označevalci odprte propozicije (npr. *pa vsega tega, pa tako vse, ali pa kaj tako, ali kaj jaz vem*), označevalci procesov tvorjenja (npr. *kaj sem mislil reči, kaj jaz rečem, kaj se že temu reče*) in ponavljanja (npr. *kot sem že rekla, tako kot si rekla*), narečni označevalci (npr. *pa što tako, kaj je ovo, anche če, višto ker, ma para*), modificirani vezniki (npr. *po tistem ko, glih zato, na koncu ko, pa četudi*) in uvajalne fraze različnih stavčnih tipov (npr. *zgolj za primer, bom drugače vprašal, pazite to, vesta kaj*), pa tudi nekatere idiosinkratične rabe sicer ustaljenih veznikov (npr. *po ta drugi strani pa*). Te enote na podlagi merila ustaljenosti v nadaljnjo analizo niso bile vključene.

Končni seznam prepoznanih večbesednih diskurzni označevalcev v vzorčenem korpusu govorne slovenščine tako obsega **125 enot** s skupno **444 pojavitvami**. Skupaj s podatkom o pogostosti in prepoznanih diskurzni funkcijah v vzorčenem korpusu jih navajamo v Prilogi 1, med njimi pa se v vzorčenem korpusu najpogosteje pojavijo označevalci *tako da* (49 pojavitev), *a ne* (33), *ne vem* (33), *v bistvu* (28), *a veš* (18), *se pravi* (17), *tako kot* (14), *pa to* (13) in *glede na to da* (10), več kot polovica zvez (69) pa se v vlogi diskurznega označevalca v vzorčenem korpusu pojavi zgolj enkrat.

4.1.2 Težavnejša mesta kategorizacije

V procesu ročnega označevanja vzorčenega korpusa se je izkazalo, da **izhodiščna označevalna shema njihovih funkcij omogoča razmeroma hitro in nedvoumno prepoznavo večine izrazov**, ki bi jih lahko umestili v kategorijo večbesednih diskurzni označevalcev. Vendarle pa smo v procesu označevanja identificirali nekaj mejnih kategorij, ki v izhodiščni funkcijski shemi (poglavje 2.2.3.3.1 *Funkcijski opis*) niso bile podrobneje izpostavljene, saj je bila ta prvotno razvita za potrebe prepoznavanja strukturno drugačnega nabora izrazov (povsem gramatikalizirani, večinoma enobesedni označevalci)⁷⁶ v slovenščini manj sorodnih jezikih (angleščini in francoščini), zaradi česar lahko sklepamo, da v drugih aplikacijah sheme spodaj predstavljena vprašanja niso bila deležna tolikšne pozornosti. Čeprav smo nekatera izmed teh

⁷⁶ Med večbesednimi diskurzni označevalci, ki naj bi se pojavljali v analiziranem vzorcu angleškega korpusa intervjujev velikosti približno 28.000 pojavnic, L. Crible (2016) denimo navaja zgolj naslednje večbesedne enote: *as you know, even though, first of all, for example, I guess, I mean, I suppose, if you like, if you will, in fact, in other words, kind of, on the other hand, shall we say, sort of, you know*.

vprašanj že v teoretičnih izhodiščih naslovili s sopostavljanjem izhodiščne funkcijske taksonomije z drugimi domačimi in tujimi funkcijskimi opisi ter z natančnejšo opredelitvijo izbranih strukturnih meril na podlagi dosedanjih slovenističnih razprav (poglavje 2.2.3.3.2 *Strukturna zamejitev*), pa so se v procesu analize dejanskega gradiva nekatere skupine izrazov izkazale za izrazito **težavnejša mesta jezikoslovne kategorizacije** oz. z njo povezane interpretacije, bodisi zaradi pomenskih oz. funkcijskih bodisi zaradi izraznih lastnosti.

4.1.2.1 Funkcijsko problematični izrazi

4.1.2.1.1 Izstopajoče funkcijske podskupine

V opisu funkcij izhodiščne taksonomije smo že izpostavili, da ta v primerjavi z nekaterimi drugimi sorodnimi taksonomijami ponuja nekoliko robustnejši nabor funkcij, zlasti predstavnihi in retoričnih označevalcev, kakršni se najpogosteje pojavljajo tudi v pisnem diskurzu. Tipičen primer je denimo kategorija posledice, ki poleg označevalcev posledice vzroka ali razloga (npr. *zato*) vključuje tudi označevalce posledice pozitivnega pogoja (npr. *potem*) in namena (npr. *da*). V procesu označevanja večbesednih diskurznihi označevalcev v slovenščini se je izkazalo, da so označevalci nekaterih tovrstnih funkcijskih podskupin razmeroma pogosti. Poleg že omenjene kategorije namena oz. cilja, ki je trenutno obravnavana kot del splošnejše kategorije (predstavne oz. semantične) posledice in jo npr. izražajo označevalci *zato da*, *za to da* ali *zaradi tega da* (primer [1]), na podoben način izstopajo še označevalci načina (npr. *tako kot*), primerjave (npr. *tako kot* in *kot da*) in ponazoritve (npr. *na primer* in *a la*), ki se trenutno vsi umeščajo v splošnejšo kategorijo specifikacije (primeri [2]–[4]).

- [1] tiste katere so pa uredili pa ponavadi jih niso uredili **zato da** bi v njih živeli ampak zato da [gap] imajo za okras
- [2] tako da upam da da bo vse to se dalo izpeljat **tako kot** mislimo da bi se moralo
- [3] in so ugotavljali [:voice] [:voice] [:voice] da so naredili res konstantno magnetno polje ne sme biti ne premočno ker potem **kot da** bi bil tam na elektro oni magnetni stimulaciji je bil premočni magneti
- [4] ja pač tam v delovnih so mogli delati **na primer** za [gap]

Poleg naštetih funkcijskih skupin, ki jih izhodiščna shema eksplicitno izpostavlja ali opisuje, a jih načrtno umešča v splošnejše kategorije, pa smo pri analizi večbesednih diskurznihi označevalcev govorjene slovenščine zaznali še dve novi funkciji, ki jih izhodiščna shema ne omenja. Prva je

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihi označevalcev

funkcija označevanja ponavljanja, kakršno med drugim označujejo izrazi, kot so *kot rečeno*, *kot smo rekli*, *saj pravim* ipd., ki bi jo lahko posredno umestili med (sekvenčne) označevalce dodajanja.

[5] kaj naj rečem vesel sem da nisem med njimi ne jim pa **saj pravim** nič ne zavidam

Druga 'nova' zaznana diskurzna funkcija je eksplicitno označevanje urejanja (Shriberg, 1996) oz. označevanje procesov tvorjenja (Verdonik, 2008), kakršno denimo označujejo izrazi *ne vem*, *kaj jaz vem*, *kaj je že* ipd. Izhodiščna shema MDMA te diskurzne funkcije ne predvideva, saj avtorji sheme že v uvodu iz kategorije diskurznihi označevalcev izločajo označevalce urejanja (angl. *editing terms*), ki signalizirajo težave pri priklicu besedišča, izkazujejo nižjo stopnjo gramatikalizacije in vsebujejo sledi propozicijske vsebine (Crible, 2014, str. 5–6) in jih zato, tako kot zapolnjene premore oz. mašila, umeščajo na ločeno raven označevanja označevalcev (ne)tekočnosti v govoru (Crible, 2016). Glede na to, da sami tekočnost govora obravnavamo kot del upravljanja diskurza, smo torej tudi to skupino izrazov umestili med diskurzne označevalce, in sicer kot podskupino označevalcev zapolnjevanja vrzeli, torej označevalcev, ki signalizirajo željo po zadrževanju vloge med načrtovanjem nadaljnjega govora.

[6] ta **kaj je** od eee

[7] recimo **ne vem** eno majico ki jo damo v prej preden jo damo v pralni stroj

[8] pri ustnih harmonikah je to **kaj jaz vem** dvajset trideset evrov ne

Pri nadaljnjih diskurznofunkcijskih analizah slovenskega jezika bi tako veljalo razmisliti o dopolnitvi izhodiščne sheme bodisi z novimi funkcijami bodisi s podrobnejšim opisom posameznih podskupin, v trenutni različici označenega korpusa pa ohranjamo njihovo umestitev znotraj navedenih splošnejših funkcijskih nadkategorij.

4.1.2.1.2 Izrazi epistemske naklonskosti

Funkcije omilitve, tj. govorčevo namerno omiljevanje zavezanosti k veljavnosti izrečenega, ublažitve, tj. preprečevanje ogrožanja dostojanstva, in elipse, tj. označevanja nedoločnosti, odpirajo pomembno vprašanje razmejevanja kategorije diskurznihi označevalcev z drugo funkcijsko sorodno skupino propozicijsko neobveznihi izrazov, izrazi epistemske naklonskosti, s katerimi tvorec izraža svojo oceno o resničnosti propozicije oziroma stopnjo svoje zavezanosti k njeni vsebini (Palmer, 2001). Med širokim naborom izrazov in struktur za izražanje epistemske

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihi označevalcev

naklonskosti v slovenščini (Pisanski Peterlin, 2015c) smo sami med diskurzne označevalce umestili zgolj tiste, pri katerih se funkcija tvorčeve namerne nedoločnosti prepleta s funkcijo zapolnjevanja vrzeli (primeri [9]–[11]), npr. *ali kaj jaz vem, ne vem, bi rekel, na nek način*.

- [9] kaj tako si da bi dala ona kaj noter ne vem ko ima [gap] **kaj jaz vem**
- [10] tu sem ti napisal že v uvodnem delu da predlagam recimo ko daš slovenski prevo [gap] eee prevod ne da vseeno mogoče v oklepaju daš še angleško eem dejansko n [gap] na **na nek način** še enkrat prevedesh nazaj ne
- [11] in **bi rekel** to je ena izmed temeljnih značilnosti sodobne politike

Po drugi strani med diskurzne označevalce nismo uvrstili vseh drugih oblik prav tako izredno pogostih in včasih izrazito gramatikaliziranih izrazov izražanja namerne nedoločnosti, kot so denimo označevalci subjektivnih sodb oz. domnev (primeri [12]–[14]), dokaznosti oz. izvora informacije ([15]–[17]) ali relativizacije ([18]), saj tem ni mogoče pripisati posebne besedilne vloge.

- [12] ker **po moje** je to ful dobra reklama za trgovino
- [13] čakaj mp3 je **meni se zdi**
- [14] predavala gos [gap] bo gospa doktor [name:personal] [name:surname] ki je **kolikor jaz vem** edina v sloveniji ki se poklicno bavi s tem vprašanjem pri umarju urad eee republike slovenije za makroekonomiko če prav če sem prav razložil
- [15] no ampak eee zdaj so znanstveniki dokazali pazite to da testosteroni obliži ki se dajo **kot kaže** kupiti nisem vedel ne rabim testosteroni obliži spodbudijo željo po seksu pri ženskah
- [16] to **kakor vidim** se shrani eee samo naslov v bistvu pa url
- [17] eee iz tega časa so ohranjena še razna železna orodja tudi pri nas grobišča in tudi gradišča **tako imenovana** na hribih
- [18] zdaj če greva klasična polena reči potem je takoj potem je to peč kjer se zdaj te ki so novejša so take da v bistvu eee da v bistvu imate dve **neke vrste** spodnje kurišče kjer u [gap] uplinjate to ta les in potem se v bistvu [gap] ti plini izgoreva

4.1.2.1.3 Stavčni uvajalci premega govora

Med diskurzne označevalce prav tako nismo umestili stavčnih uvajalcev premega govora (primeri [19]–[21]), na katere je v slovenskem prostoru opozorila že D. Zuljan Kumar, ki v analizi briških pogovorov izpostavlja pogosto rabo spremnih stavkov premega govora, ki pripomorejo k ustvarjanju vtisa, da upovedeno dogajanje poteka v tistem trenutku (Zuljan Kumar, 2007, str. 45). Čeprav bi stavčne uvajalce premega govora s funkcijskega vidika lahko umestili med

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihih označevalcev

označevalce navajanja, ki signalizirajo začetek poročanega govora govorca ali govora tretje osebe, pa menimo, da jih je s strukturnega vidika ustrezneje obravnavati kot del vezljivostno nezaključenih struktur, kakršne iz svoje obravnave izločamo, prav tako pa izkazujejo visoko stopnjo skladenjske in oblikoslovne variabilnosti (primer [22]).

- [19] in **sem rekla** zdaj moram pa nekaj zaslužiti ker hočem v španijo hočem v tujino nekaj hočem iti
- [20] in v nekem določenem trenutku so se zakoni nekako sprostili in **so rekli** aha kamioni se pa lahko uvažajo
- [21] ma v glavnem tam se mi usedemo ne ker vidimo tam da je še polno prostega **smo rekli** aha to bomo mi pol počasi ne zaokupirali ne
- [22] verjetno bo sicer so predavatelji v glavnem eee sami izgleda knjige pišejo ne in morajo ... ja ja ... ja in **so kar rekli** zastoj nismo knjig pisali boste kar lepo kupili jih

4.1.2.1.4 Vprašalni pristavki

Ena izmed skupin nepropozicijskih pragmatičnih izrazov, ki jo izhodiščna shema prav tako izloča iz skupine diskurznihih označevalcev, so tako imenovani vprašalni pristavki (angl. *tag questions*), saj jih kljub ustaljeni funkciji uravnavanja govorjene interakcije zaradi svoje oblikoslovne pregibnosti (v angleščini) in propozicijskega pomena ne obravnava kot izrazno povsem gramatikalizirane strukture. Ker sami med diskurzne označevalce prištevamo širši nabor ustaljenih izrazov in ker so vprašalni pristavki v slovenščini z izraznega vidika zaradi glagolske elipse razmeroma ustaljeni, med diskurzne označevalce tako kot G. Redeker (2006, str. 355) umeščamo tudi vprašalne pristavke, v katerih se pogosto prepletajo funkcije zaključevanja vloge, preverjanja pozornosti in/ali izražanja nedoločnosti, npr.:

- [23] [speaker:laughter] kaj ga še kar geekaš **ali kaj** ?
- [24] saj potem si je v bistvu sama kupila ne **ali kako** ?
- [25] in eee predavanje bo trajalo okrog petinštirideset do šestdeset minut eee razprava sledi šele potem **v redu** ?

4.1.2.1.5 Kohezivne prislovne zveze

Kot smo podrobneje izpostavili že v pojasnilu izbranihih strukturnih meril (poglavje 2.2.3.3.2 *Strukturna zamejitev*), kot del diskurznihih označevalcev obravnavamo tudi ustaljena prislovna določila s kohezivno sestavino, ki se s kazanjem v besedilu nazaj ali naprej navezujejo na predhodno izraženo besedilno izraženo propozicijo ali širši besedilni kontekst, npr. *zaradi tega*,

kljub vsemu, na ta način ipd. Toda kljub razmeroma jasno zastavljenim izhodiščem se je izkazalo, da meja med propozicijsko in diskurzno rabo teh izrazov glede na njihove različne stopnje gramatikalizacije ni vedno lahko določljiva, kot prikazujejo spodaj navedeni primeri rabe prislovne zveze *v tem primeru*, kjer ni jasno določljivo, ali zveza deiktično kaže na konkreten primer (prislovna raba, ki jo lahko parafraziramo z izrazom 'v našem primeru') ali pa uvaja posledico predhodno opisane situacije oz. pogoja (diskurzna raba, ki jo lahko parafrazirano z izrazom 'potemtakem').

- [26] in **v tem primeru** povišana trans [gap] eee rotacijska temperatura prevzame vlogo katalizatorja
- [27] če imam **v tem primeru** kot smo rekli pri prvi besedi sova hipoteze eee sova soba kjer bi bila soba druga hipoteza
- [28] ni govora **v tem primeru** o otrocih ne

V zgoraj navedenih primerih smo si tako pri končni odločitvi pogosto pomagali z analizo širšega besedilnega konteksta, tudi onkraj korpusnega vzorca, in parafraznimi testi, v primeru tudi siceršnje dvoumnosti pa smo dali prednost prislovni interpretaciji pred diskurzno. Prav tako smo ne glede na pomen kot diskurzne označevalce obravnavali kataforične časovne prislove tipa *v nadaljevanju, za začetek, na koncu*, ki bodisi vzpostavljajo časovna razmerja s propozicijo, ki sledi, bodisi besedilno enoto, v kateri se pojavijo, metaforično razvrščajo glede na soležne tematske sekvence (Skubic, 1999, str. 224–225).

Med prislovnimi zvezami s kohezivnim pomenom nekoliko izstopata zvezi *še vedno* oz. *še zmeraj*, ki poleg časovnega pomena izražanja nadaljevanja nekega predhodnega dogajanja ('še danes', 'še naprej') hkrati implicirata tudi dopustno negacijo pričakovanega ('kljub vsemu'), npr. v primerih:

- [29] **še vedno** je najbolj razširjeno učinkovito preverjeno in uporabno ugotavljanje sprememb v telesnih odzivih ki so odvisni od eee samodejnega živčnega sistema
- [30] **še vedno** rajši tako v malih zasedbah bolj ne ?

Zaradi visoke stopnje dvoumnosti smo v analizi dali prednost prislovni interpretaciji in teh dveh zvez nismo obravnavali kot diskurzna označevalca.

4.1.2.1.6 Vezniške zveze s kataforično sestavino

Kot smo na podlagi sorodnih slovenističnih razprav napovedali že v teoretičnih izhodiščih, kot večbesedne diskurzne označevalce obravnavamo tudi frazeologirane vezniške zveze, ki jih sestavljajo kombinacije veznika in pred njim stoječega prislova ali členka, med katerimi se v analiziranem gradivu denimo pojavijo modificirani označevalci časovnih razmerij (npr. *takrat ko*, *po tistem ko*), pogoja (npr. *še posebej če*, *sploh če*) in podobno. Kot nekoliko bolj problematične so se izkazale zveze s kazalnim prislovom *tako* in podrednim veznikom *kot*, ki lahko uvajajo opis načina (parafraziran z 'na ta način da'), kot v primeru [31], ali primerjalne skladnosti (parafrazirane s 'prav tako kot'), kot v primeru [32]. Čeprav bi glede na prozodične razlike v izgovoru ene ali druge rabe z vidika skladenjske razstavljalivosti med njima morda veljalo razlikovati, na diskurzni ravni zvezo *tako kot* ne glede na pomen obravnavamo kot diskurzni označevalec (specifikacije), saj je kataforični modifikator (v našem primeru prislovni zaimsek *tako*) v obeh primerih s pomenskega vidika iz rabe izpuštljiv.

[31] tako da upam da da bo vse to se dalo izpeljat **tako kot** mislimo da bi se moralo

[32] ja imela sem take goste **tako kot** imata zdaj vidva tukaj mene

Prav tako na potrebo po natančnejšem premisleku notranje skladenjske strukture kažejo nekatere specifične rabe nekaterih kataforičnih vezniških zvez, v katerih kataforični zaimsek ni ne skladenjsko ne pomensko izpuštljiv (primera [33] in [34]) ali pa se med njim in podrednim veznikom pojavljajo celo druge strukture (primer [35]).

[33] rečeno je samo to da danes in v prihodnjih dneh ne bi želel reči nič ali ravnati **tako da** bi zapravlil možnost da mi uspe sestaviti močno in kompetentno vlado

[34] jaz pa zdaj tule gledam eno položnico in mi je prav zanimivo prizidek si želi
[name:personal] pa me zanima če **zato da** se loči od tašče

[35] t [gap] tudi mogoče tudi mogoče **zaradi tega** če bosta res šla vidva v dvoboj **da** imaš pač ti prednost da lahko ne izbereš način dvoboja ne

4.1.2.1.7 Izrazi odzivanja

Glede na prevladujoči delež dialoških govornih dogodkov se v vzorčenem korpusu govorjene slovenščine zelo pogosto pojavljajo formulaični izrazi, s katerimi se govorci odzivajo na vsebino govora drugih udeležencev v diskurzu, npr. *v redu*, *res je*, *to pa res*, *točno to*, *ma ja*, *čakaj malo*, *a res?* ipd. Kot smo izpostavili že v pregledu raznolikih pristopov k opredelitvi diskurznihih

označevalcev, tradicionalnejše besediloslovne raziskave tem izrazom običajno ne pripisujejo posebne diskurzne vloge, saj ne prispevajo neposredno k smiselni organizaciji besedilno posredovanih sporočil, medtem ko pragmatične raziskave govorjene interakcije po drugi strani pogosto izpostavljajo njihovo pomembno vlogo v upravljanju procesa izmenjave vlog in uravnavanju medosebnih odnosov. Shema MDMA pri tem zavzame sredinsko rešitev, saj med diskurzne označevalce prišteva tudi nekatere izraze odzivanja, a le kadar ti ne nastopajo v vlogi odgovorov na vprašanje oz. zaključenih sporočil, temveč hkrati opravljajo tudi katero izmed drugih (sekvenčnih) diskurznih funkcij, kot sta zaključevanje ali prevzemanje vloge (Crible, 2014, str. 4; 25). Funkciji strinjanja in nestrinjanja se tako v večini primerov kombinirata z različnimi sekvenčnimi funkcijami, pa tudi nasploh sta pri bili pri označevanju angleških in francoskih besedil uporabljeni razmeroma redko (Crible 2016, osebna korespondenca).

Ker tudi sami zagovarjamo besediloslovno opredelitev diskurznih označevalcev, ki diskurzne označevalce opredeljuje z vidika njihove vloge v organizaciji in upravljanju besedilno posredovanih sporočil, smo kot diskurzne označevalce označili zgolj tiste izraze odzivanja, ki se neposredno vpenjajo v *govorčevo* predhodno ali nadaljnje besedilo in torej poleg medosebne opravljajo tudi *besedilno* vlogo. Tak primer je denimo raba izrazov odzivanja v spodnjih dveh primerih, v katerih je izraz *v redu* rabljen (tudi) kot označevalec tematskega prehoda (primer [36]) oz. zaključka (primer [37]).

[36] **v redu** dobro torej to to moramo to je res nujno no to moramo nujno narediti

[37] vse nas malo povleče ven ne te še moremo malo potem še tiste zadnje moči da nekako pridemo k sebi **v redu** ne ... mhm

Vseh ostalih formulaičnih izrazov odzivanja, kot so odgovori na vprašanja (primer [38]), oporni signali (primer [39]), potrjevanje (ne)ustreznosti sogovorčeve vsebine (primera [40] in [41]), (ne)strinjanje s sogovorčevo vsebino ali čustveno odzivanje na sogovorčevo vsebino (primer [42]), torej nismo obravnavali kot del diskurznih označevalcev, saj te obravnavamo kot samostojna sporočila (samostojne diskurzne enote), ki se anaforično nanašajo zgolj na sogovorčev diskurz, ne pa na govorčevo upravljanje lastnega diskurza.

[38] A: trideset sekund toliko kot bo pravih odgovorov toliko bo kart za jutri / zmenjeno ?///
B: **v redu** /// A: tvoja igra tvoj premierni hitri kviz se začena zdaj

- [39] A: kaj je bil na [gap] namen uporabe zatočišč ? /// B: ja da je začasno tam ne. / [...] / da se samo tja zateče / pa da so se skrili pred plenilci /// A: **tako je** varovanje pred vremenskimi pojavi
- [40] A: minus ena /// B: **ne bo držalo** dajva najprej tisto napisati kar si že prvo rekel
- [41] saj ne vem verjetno si sama tudi s [gap] razmišljala o tem ne ja ... **točno to** ja ... seveda ker ne razume ja ... [speaker:laughter] aja okej
- [42] /// tam je pa [name:personal] sedel je rekel čuj daj te bo [name:personal] prišel iskat /// **a res ali kaj?** /// [name:personal] pa ful gleda komu si to rekel? [speaker:laughter] ///

To pomeni, da se tukaj oddaljimo od izhodiščne funkcijske sheme MDMA, saj v svoji ponazoritvi funkcije strinjanja navaja primer, v katerem izraz odzivanja lahko tolmačimo kot samostojno sporočilo (*strinjam se*), ki se v svoji funkciji prav nič ne razlikuje od funkcije, kakršno bi izraz opravljal na začetku ali na koncu stavka, zato ni povsem jasno, kje avtorji vzpostavljajo ločnico med propozicijskim in diskurznihih izražanjem (ne)strinjanja. Tudi raziskava D. Verdonik in sodelavcev (2007) denimo izpostavlja težavnost ločevanja med propozicijsko in diskurzno rabo enobesednega označevalca 'ja'.

Glede na to, da kot diskurzne označevalce obravnavamo zgolj tiste izraze s hkratno funkcijo sekvenčnega upravljanja se seveda postavlja vprašanje smiselnosti funkcijske kategorije (ne)strinjanja, če je ta zgolj drugotna pomenska opredelitev, tako kot to velja tudi za vse druge medosebne funkcije, ki jih predvideva shema MDMA, npr. preverjanja pozornosti, omilitve ali elipse, saj se te večinoma prepletajo s sekvenčnimi funkcijami zapolnjevanja vrzeli in/ali menjave vlog. V prihodnje bi torej prav to skupino označevalcev veljalo ponovno premisliti tako z vidika njihovega razmejevanja z ideacijskimi, retoričnimi in sekvenčnimi rabami kot z vidika razmejevanja z drugimi kategorijami tipično govorjene pragmatične leksike, kot so izrazi odzivanja, naklonskosti, vljudnostni obrazci ipd.

4.1.2.2 Strukturno problematični izrazi

4.1.2.2.1 Eliptični napovedovalci

V vzorčenem korpusu govorjene slovenščine se v vlogi sredstev za organizacijo poteka diskurza pogosto pojavljajo tudi stavčni ali nestavčni (eliptični) uvajalci govornih dejanj, ki jim intonacijsko sledi premor, kakršen bi v pisnem jeziku ustrezal rabi podpičja, npr. *pa še eno vprašanje(:)*, *še ena stvar(:)*, *v nadaljevanju(:)*, *druga zadeva(:)*, *zgolj za primer(:)*, *bom drugače*

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihih označevalcev

vprašal(:), *to pomeni(:)* itd. Te izraze smo prav tako obravnavali kot diskurzne označevalce, pri čemer jih z izjemo izrazov *druga zadeva*, *v nadaljevanju* in *to pomeni* večina ne izkazuje zahtevane stopnje ustaljenosti oz. pogostosti v rabi.

- [43] eee **to pomeni** z tistim minimalnim še potrebnim eee številom ljudi eee z [gap] tudi zaradi tega da so te eee ti stroški čim manjši
- [44] eee **druga zadeva** je majhen priročnik gre kamorkoli hočeš
- [45] **v nadaljevanju** ne vemo kam grejo kupnine od prodanih državnih podjetij

4.1.2.2.2 Razstavljivost kolokacij

V teoretičnih izhodiščih (poglavje 2.2.3.3.2.3 *Kolokacije diskurznihih označevalcev*) smo izpostavili, da se pri interpretaciji (ne)razstavljivosti pogostih sopojavitev diskurznihih označevalcev opiramo na kriterij njihove strukturne in funkcijske samostojnosti. V primeru, da je bila zveza dveh označevalcev skladijsko samostojna, njene vloge v diskurzu pa ni bilo mogoče razstaviti na seštevke funkcij posameznih označevalcev, smo jo torej obravnavali kot primer samostojnega (večbesednega) diskurznehih označevalca, kot npr. označevalca izražanja nestrinjanja oz. odpiranja v spodnjih primerih:

- [46] **no ja** informacije dobivamo seveda od novinarjev ki so tam to so zelo dobri novinarji eee s katerimi sem večkrat sodelovala
- [47] **ja no** saj veš sam kako je z začetki ne začneš z majhnim ne

Če je bilo po drugi strani posameznim gradnikom zveze mogoče pripisati samostojno funkcijo, zveze ne glede na visoko stopnjo formulaičnosti in posledično pogostost rabe nismo obravnavali kot primer večbesednega diskurznehih označevalca. Tipičen primer je denimo zveza *mislim tako*, ki jo sestavljata označevalec reformulacije 'mislim' in označevalec zapolnjevanja vrzeli 'tako', ki se v primeru govorčevega samopopravljanja ali uvajanja podrobnejšega pojasnila pogosto pojavljata v paru (primer [48]). Drugi podobni primeri, ki smo jih prav tako obravnavali kot razstavljive kolokacije, so med drugim tudi zveze *ne vem no*, *ja ne vem*, *ja pač*, *tako pač*, *no ampak*, *no poglejte*, *ja čakajte*, *ja veš kaj*, *ja torej*, *in pol*, *in potem*, *samo po drugi strani*, *okej dobro*, *ma v glavnem*, *a vseeno*, *pa dobro*, *samo veš kaj je*, *no in potem*, *no dobro*, *tako da ja*.

- [48] škoda če je ra [gap] rad bereš eno knjigo pa jo pol **mislim tako** da ti je ena všeč pol jo pa kupiš za ne vem koliko denarja

Kot izrazito težavne so se izkazale tudi zveze s tipično slovenskim označevalcem *pa*, ki se zelo pogosto pojavlja v kombinaciji z drugimi diskurznimi označevalci, kar v svojem pomenskem opisu tega večfunkcijskega označevalca beleži že SSKJ,⁷⁷ prav tako pa vezniške zveze s *pa* pogosto navajajo tudi druge sorodne slovenistične raziskave diskurznofunkcijske leksike.⁷⁸ V vzorčenem korpusu se besedica *pa* pojavi več kot 500-krat, od tega zelo pogosto za drugim diskurznim označevalcem, npr. *ali pa, zdaj pa, pol pa, in pa, potem pa, sicer pa, če pa, zato pa, po drugi strani pa, poleg tega pa, v bistvu pa* itd., redkeje tudi pred njim, npr. *pa potem, pa pol, pa vendarle, pa dobro, pa če, pa četudi* itd.

Za razliko od samostojne rabe označevalca *pa*, v kateri nastopa kot uvajalec dodajanja ali nasprotja, je njegova specifična funkcija v tovrstnih kolokacijah težje določljiva oz. mu lahko pripišemo splošnejšo vlogo poudarjanja diskurzne funkcije soležnega označevalca. Ker *pa* je ta funkcija razmeroma ohlapno opredeljena, tovrstne zveze *pa* izrazito pogoste, smo vezniške zveze z označevalcem *pa*, tako kot druge sorodne raziskave v slovenskem prostoru, obravnavali kot nerazstavljive primere večbesednih diskurznihi označevalcev, zlasti kadar označevalcu *pa* ni bilo mogoče pripisati druge funkcije kot funkcije poudarjanja, npr.:

- [49] dvom zbuja tudi dejstvo da sta bila pravilnika o verski oskrbi sprejeta na hitro in brez javne razprave **po**leg tega **pa** se znova postavlja že staro vprašanje
- [50] eem **sicer** **pa** jaz mislim da še bo več odzivov glede na to da je razstava komaj se začela
- [51] ni prav da samo eden dobi nagrade in da je samo d [gap] eden bogatejši od soseda **zato** **pa** imate na voljo kikiriki vsak dan
- [52] v redu **potem** **pa** jutri prinesete s seboj delovne zvezke bomo s premim govorom nadaljevali

⁷⁷ SSKJ pogosto kombiniranje besedice *pa* z drugimi diskurznofunkcijskimi izrazi denimo izpostavlja v enem izmed pomenov prislova *pa*, kjer ta krepi soležni veznik (s ponazoritvami na zvezah *ali pa, in pa, pa saj, pač pa*), in v več pomenih veznika *pa* – krepitev časovnega prislova (samo pozdravim jih, potem *pa* grem) in poudarjanje nasprotja v pogojnih odvisnikih (*moja bo obveljala, pa če se na glavo postaviš; nikar se ne izmikaj, če pa vem, kje si včeraj hodil; čemu ti bo toliko denarja, ko pa ne moreš vsega porabiti*) – hkrati *pa* so zveze s *pa* izpostavljene tudi v pomenskih opisih in ponazoritvah drugih označevalcev, npr. *vendar (pa)* ali *sicer (pa)*.

⁷⁸ Skubic (1999) npr. omenja naslednje zveze s *pa* z raznovrstnimi funkcijami: *sicer pa, medtem pa, zato pa, lahko pa, drugače pa, lahko pa tudi, ali pa, vendar pa, pa vseeno, pa vendarle, po drugi strani pa*, Balažic Bulc (2009) *pa* npr. zveze, kot so *vendar pa, po drugi strani pa, sicer pa, hkrati pa, ali pa, nato pa* ipd.

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihi označevalcev

Tako kot pri modificiranih veznikih pa bi v prihodnosti veljalo razmisliti, v kolikšni meri pri vzpostavljanju ločnice med večbesednimi diskurznimi označevalci in kolokacijami diskurznihi označevalcev upoštevati tudi merilo skladenjske razstavljivosti, saj se nekatere zveze z označevalcem *pa* res pojavljajo zgolj kot skladenjske nerazstavljive enote (primera [53]–[54]), nekatere pa tudi v zapisu narazen (primeri [55]–[57]), pri čemer je ta raba bistveno manj pogosta:

- [53] prvi primer je sotinski breg in to tukaj gre za drugi primer kjer je investitor eee lokalna eem lokalna skupnost **in pa** seveda j [gap] so je postaja postavljena po vseh standardih ki ustrezajo državni meteorološki eee mreži in je zato lahko vključena v eee državno meteorološko mrežo kot enakovredna postaja meteorološkim postajam ki jih sicer postavlja agencija za okolje
- [54] **ali pa** je skozi minus
- [55] **poleg tega** se **pa** prav tako plačuje televizija pri izbranih teve operaterjih torej jo plačujemo dvojno [primer ni iz SST, temveč GOS]
- [56] **sicer** je **pa** ta ureditev popolnoma enaka kot v prejšnjem zakonu [primer ni iz SST, temveč GOS]
- [57] ne strinjam se z povečane naročnina **vendar** je **pa** tuki še druga plat medalje ker je pa le ta slovenska televizija nacionalnega pomena in mislm da nekaj bi moral dobit denar vendar to brez reklam [primer ni iz SST, temveč GOS]

4.1.2.2.3 Dvodelni vezniki

Glede na to, da se v pričujoči raziskavi osredotočamo zgolj na neprekinjene večbesedne enote, v korpusu nismo označevali ustaljenih sopojavitev razdruženih označevalcev, kot so dvodelni vezniki tipa *ali-ali*, *niti-niti*, *ne samo-ampak tudi*, *če-potem* (Toporišič, 2004, str. 429). Čeprav se zdi ta odločitev glede na dosedanjo obravnavo večbesednih veznikov v slovenskem prostoru nekoliko nenavadna, menimo, da je delno tudi teoretsko upravičljiva, saj bi lahko z vidika kolokabilnosti posameznih označevalcev tovrstno kombiniranje pripisali še bistveno širšemu naboru parov kot le prototipično navajanim dvodelnim veznikom, npr. kombinacijam kataforičnih polarnostnih členkov in anaforičnih dopustnih veznikov tipa '*načeloma a*, *ampak b*' (Skubic, 1999, str. 28), hkrati pa menimo, da je pri njihovem pojasnjevanju ustrežnejše kot o kolokabilnosti označevalcev govoriti o kolokabilnosti posameznih tipov sporočil oz. diskurznihi enot.

Tudi primeri v vzorčenem korpusu namreč potrjujejo, da pogoju, ki ga vpeljuje veznik '*če*', ne sledi nujno posledica z veznikom '*potem*' (primer [58]), izključevalno ločno priredje ni vedno

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihih označevalcev

izraženo z obema veznikoma ([59]–[60]), negacija in dodajalna protivnost v stopnjevalnem priredju nista nujno izraženi z eksplicitnim označevalcem ([61]–[62]), prav tako pa je lahko v govoru prvi ali drugi del dvodelnega veznika besedilno neizražen, kot npr. v primeru kombinacije *po eni strani...po drugi strani* v primerih [63]–[64].

- [58] ker sami **če** bi imeli interes bi že rešili to se pa že predolgo vleče
- [59] iz postaje na postajo motor skozi **ali** pospešuje pol zavira gre energija v nič a ne
- [60] in če bo kdo rekel da ni finančnih sredstev za to da dvignemo ne samo minimalne plače minimalne pokojnine ampak da naredimo sistem tisti sistem ki bo omogočal ponosno življenje vsakega slehernega izmed nas potem ta laže **ali** ne pozna proračuna republike slovenije
- [61] je pa res da tole ni fino samo zaradi izgleda **ampak** se tudi omaka boljše prime na rebraste njoke
- [62] in **ne samo** da je bila ena so prišle dve [gap]
- [63] in da je eee istrabenz eee črpal **po eni strani** iz prodaje nekaterih drugih naložb
- [64] samo **po drugi strani pa** on izredno zrelo misli ne

4.1.2.2.4 Uvajalci eliptičnih pristavkov

Označevalci, ki nastopajo v vlogi besedilnih povezovalcev (konektorjev), včasih poleg vzpostavljanja medstavčnih povezav, pri katerih njihova diskurzna funkcija ni vprašljiva (primer [65]), pojavljajo tudi kot povezovalci skladenjskih enot drugih tipov. Tipičen primer sta priredna veznika *in pa* ter *ali pa*, ki lahko nastopata bodisi v (diskurzni) vlogi povezovalcev dveh stavčno izraženih propozicij (primer [66]) bodisi v (stavčni) vlogi veznikov dveh enakovrednih stavčnih členov (primer [67]).

- [65] zato ker ni džabest **ali pa** ni mobi
- [66] če imamo neko snov ki vsebuje tudi mehurčke **ali pa** prostorčke z zrakom napolnjene tiste prostorčke lahko tudi stisnete zraka ni več notri
- [67] k [gap] ka [gap] kateri so ti verzi recimo **ali pa** te kitice ki vas na to napeljejo da je res ljubezenska ?

Če je ta ločnica razmeroma jasno vzpostavljiva pri vezalnih in ločnih veznikih, pa večji izziv predstavlja kategorizacija drugih tipov označevalcev, kadar uvajajo skladenjske enote brez izraženega glagola, npr. v spodnjih primerih:

- [68] poljaki bodo naslednjo tekmo igrali s san marinom / [pause] / in sicer **prav tako** v sredo doma

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihih označevalcev

- [69] ampak odločila sem se za petnajst tisoč evrov neke nnn **tako kot** v preteklosti
- [70] ti ma [gap] okej razen pač tvojo predstavitev deset do petnajst **pol pa** še pogovora toliko ne
- [71] tako polovičke so tukaj in zaradi tega moraš eno polovičko **potem pa** še eno polovičko dol tako [:voice] dobro [speaker:laughter]
- [72] o spoštovane in spoštovani mi pa gremo naprej sedem je ura in triinštirideset minut nova nagradna igra na frekvencah eee radia capris **in sicer** bogatejši od sosedu
- [73] eee je je to v bistvu eee tudi eee v potrebno imeti dosti izolirano hišo ali je to bolj podobno klasiki ne vem olju drva drva **v bistvu** ogrevanje na drva

Različne teorije diskurznega razčlenjevanja pri tem vprašanju zavzemajo različna stališča, glede na samo opredelitev temeljnih diskurznihih enot (npr. členjenje na stavke, propozicije, zaključena sporočila ipd., glej poglavje 2.2.2.2.1 *Diskurzni označevalci kot povezovalci besedilnih enot*) in tudi glede na izhodiščne predpostavke komplementarne skladenjske teorije (npr. način prepoznave in obravnave glagolske elipse). Menimo, da v kontekstu govorjenega jezika, v katerem govorci zaradi kontekstualnih in kognitivnih dejavnikov povedek izpuščajo bistveno pogostejše kot v pisnem jeziku (Dobrovoljc in Nivre, 2016), členjenje diskurza glede na glagolske stavke ni upravičljivo. Kot smo izpostavili že v teoretičnih izhodiščih, kot temeljno besedilno enoto torej razumemo vsako sporočilno zaključeno enoto besedila, ki jo tipično res predstavljajo stavčno izražene propozicije, lahko pa tudi nižje skladenjsko-pomensko zaključene enote, kot so pristavki, ali daljši besedilni segmenti, kot so tematski sklopi. Vse zgoraj navedene primere ([68]-[73]) torej obravnavamo kot **eliptične (pri)stavke**, njihove uvajalce pa zato prav tako umeščamo med diskurzne označevalce.

Vsekakor pa problematično razmerje med skladenjskim, semantičnim in diskurznim razčlenjevanjem govorjenega jezika ostaja pomemben izziv za prihodnje pristope k modeliranju spontanega govora, kot nazorno prikazuje primer spodnje izjave, v kateri se metadiskurzno upravljanje dogaja na znotrajstavčni ravni, ločnica med stavčnočlenskim priredjem prilastkov in dodajalno ali popravljalno apozicijo pa je zabrisana:

- [74] za njih je zelo zavržno **ali pa** nemoralno **ali pa skratka** katastrofalno da nekdo eee zlorablja otroke (primer je iz Gos-a, ne SST)

4.1.3 Robustna izpeljava funkcij

4.1.3.1 *Problematičnost funkcijskega razdvoumljanja v kontekstu*

Seznam 125 prepoznanih večbesednih diskurznihi označevalcev v vzorčenem korpusu govornje slovenščine in njihovih diskurznihi funkcij glede na označevalno shemo MDMA (Priloga 1) potrjuje, da **številne oblike označevalcev v diskurzu opravljajo več različnih vlog** – bodisi hkrati v danem kontekstu bodisi v različnih diskurznihi kontekstih. Kot smo omenili že v opisu metodologije in predstavitvi rezultatov, večfunkcijskost označevalcev z vidika njihove prepoznavne ni problematična, saj na njej ne utemeljujemo same zamejitve te jezikoslovne kategorije, v katero torej uvrščamo vse izraze, ki jim lahko pripišemo vsaj eno izmed diskurznihi funkcij, bistveno bolj problematična pa je z vidika njihovega natančnega funkcijskega razdvoumljanja v kontekstu.

Za ponazoritev vzemimo označevalec *tako da*, ki se v vzorčenem korpusu govornje slovenščine kot označevalec pojavi 49-krat, in sicer v več različnih vlogah. Kot semantični oz. predstavniki označevalec se uporablja kot označevalec posledice propozicijsko izraženega vzroka oz. razloga (primer [75]), ki bi ga lahko parafrazirali z veznikom 'zato', kot pragmatični označevalec pa najpogosteje nastopa v vlogi retoričnega uvajalca sklepa oz. zaključka (primer [76]), ki bi ga lahko parafrazirali z izrazom 'na podlagi česar lahko rečemo', poleg tega pa tudi kot označevalec predajanja vloge in/ali zapolnjevanja vrzeli (primer [77]), ki se običajno pojavlja na koncu vloge ali kot samostojna izjava, ter povratka po digresiji (primer [78]), v katerem se med poročanje o sklepih petkovega sestanka pevskega zbora vrine še razprava o novi zborovodkinji), ki bi ga lahko parafrazirali z izrazom 'skratka'.

- [75] tako da prihajamo do informacij sama sem pa večkrat bila v gazi **tako da** seveda imam določene eee kontakte in vire da skušam priti do kar se da največ informacij
- [76] absolutno saj so dobili tudi sedem točk eee poslušalcev **tako da** to tudi ni malo ne
- [77] eee če mu je še mama mislim da tamiflu zraven dala samo ga je doma pozabil razumeš [speaker:laughter] / [all:laughter] vse je imel v glavnem vse / **tako da** /// no povej kako sta se še kaj imela
- [78] t da v bistvu samo mene ni pa une [name:personal] tk da bo sedem nas pa vseeno bo pa plus nova ta zborovodkinja // a res mate novo? // nč mi nis povedo! // ja no ... i če v petek smo sestanek meli // kira? // eem // ena [name:personal] iz Mislinje | eem [...] eee sej s [name:personal] se poznata // aja fajn kk se piše pa ne veš slučajn // mmm ja ja // ne // šment // aam // [pause] // **tk da** ja **tk da** v petek smo se zmenli pole

4.1 Dopolnitev sheme za označevanje večbesednih diskurzni označevalcev

Za označevalec *tako da* smo na podlagi funkcijske taksonomije sheme MDMA v vzorčenem korpusu našli kar pet različnih diskurzni funkcij,⁷⁹ ki spadajo v kar tri od štirih različnih funkcijskih skupin (ideacijsko, retorično in sekvenčno), pri čemer pa njihovo določanje ni bilo vedno nedvoumno, zlasti pri določanju meje med namernim predajanjem vloge (funkcija tematskega zapiranja) in željo po njene zadrževanju (funkcija zapolnjevanja vrzeli), kot denimo v primeru [77].

Podobno težavnost funkcijskega razdvajanja denimo izkazuje tudi označevalec v *bistvu*, pri katerem je bila meja med izražanjem retoričnega nasprotja, retoričnega poudarjanja in sekvenčnega zapolnjevanja vrzeli pogosto zabrisana, kot na primer v spodnjem primeru ([79]), kjer bi lahko označevalcu v *bistvu* hkrati pripisali retorično funkcijo uvajanja nasprotja z vsebino predhodnega diskurza (opisom neprofitabilnih kariernih začetkov), retorično funkcijo poudarjanja (protivne diskurzne funkcije označevalca *pa* in poudarjanja presenetljivosti glede na splošno veljavno prepričanje), sekvenčno funkcijo zapolnjevanja vrzeli (načrtovanja ustrezne formulacije) ali celo namerne nedoločnosti (previdne omembe možnosti zaslužka v pogosto nepridobitno usmerjenem kulturnem delovanju).

- [79] [:voice] to je bila ena tako po [gap] polprofesionalna lutkovna skupina ki s [gap] mmm z predstavo kar uspeli in eee takrat sem tudi dejansko prvič videl da pa **v bistvu** mmm nekako eee z d [gap] zadostnim številom predstav eee da pa se tukaj nekaj bi pa tudi finančno dalo kaj pokriti

Nekoliko manj dvoumna je raba izraza v *bistvu* kot uvajalca spremembe ubeseditve, vendar bi tej prav tako lahko hkrati pripisali tudi funkcijo zapolnjevanja vrzeli ob načrtovanju ustrezne formulacije:

- [80] eee je je to v bistvu eee tudi eee v potrebno imeti dosti izolirano hišo ali je to bolj podobno klasiki ne vem olju drva drva **v bistvu** ogrevanje na drva

Kot je razvidno iz zgornjih ponazoritev in številnih podobnih primerov večfunkcijskih označevalcev v vzorčenem korpusu govorjene slovenščine, je torej **shema MDMA zasnovana precej maksimalistično**, saj za natančno določitev predvidenih funkcij zahteva zelo podrobno

⁷⁹ Šesta funkcija specifikacije je pripisana izrazno (naglasno) ločeni obliki modificiranega veznika *tako da* (glej tudi poglavje o veznikih s kataforično sestavino) v naslednjem primeru: *rečeno je samo to da danes in v prihodnjih dneh ne bi želel reči nič ali ravnati tako da bi zapravlil možnost da mi uspe sestaviti močno in kompetentno vlado*.

analizo širšega diskurznege konteksta (za pojasnilo obeh pristopov k pomenskemu opisu funkcijske leksike glej poglavje 2.2.2.3 *Večfunkcijskost*). Pri tem se seveda upravičeno postavlja vprašanje, ali je ob tolikšnem opiranju ne zgolj na besedilni, temveč tudi širši pragmatični kontekst možno pričakovati konsistentno funkcijsko interpretacijo (denimo več različnih strokovnjakov) in v kolikšni meri je sploh upravičeno predpostavljati, da uporabniki jezika med različnimi rabi oz. funkcijami iste oblike pri tvorjenju in razumevanju diskurza dejansko razlikujejo.

Ker metodologija naše raziskave vključuje tako kvalitativni vidik, kjer je potencialno mogoče razdvoumljati tovrstne rabe glede na analizo sobesedila (pa tudi to se je pri analizi dejanskega gradiva izkazalo za težavno), kot kvantitativni vidik, kjer ni mogoče pričakovati tovrstnih kontekstualno občutljivih razdvoumljanj posameznih oblik (npr. določanja vseh možnih diskurznihi funkcij več sto pojavitev izraza *v bistvu* v korpusu Gos), v nadaljevanju predstavljeno maksimalistično tipologijo, ki nam je sicer omogočila razmeroma hitro prepoznavanje izrazov v vlogi diskurznihi označevalcev, dopolnimo še z njeno minimalistično izpeljavo, ki nam obenem omogoča tudi njihovo robustnejšo pomensko klasifikacijo.

4.1.3.2 *Minimalistična funkcijska taksonomija*

V minimalistični izpeljavi zgoraj predstavljene taksonomije funkcij diskurznihi označevalcev izhajamo iz treh predpostavk: da so različne funkcije posamezne oblike diskurznege označevalca med seboj **diahrono povezane**, da se pragmatične rabe s postopno **širitvijo dometa** običajno razvijejo iz semantičnih, ter da je posamezne funkcije mogoče združevati v splošnejše **pomenske skupine**. Če za primer vzamemo zgornji primer označevalca *tako da* torej trdimo, da je temu mogoče pripisati temeljno pomensko jedro, kot je označevanje *rezultata* oz. posledice, glede na različne domete (tipe povezanihi diskurznihi enot) pa se lahko ta posledičnost izraža bodisi na ravni povezovanja dveh propozicij (predstavna funkcija posledice ali namena), dveh besedilnihi sekvenc (retorična funkcija sklepa), dveh govornih vlog (sekvenčna funkcija zaključevanja vloge), dveh tematskih sklopov (sekvenčna funkcija nadaljevanja) ipd. Podobno bi lahko označevalcu *v bistvu* pripisali pomensko jedro označevanja *nasprotja*, ki lahko nato glede na različne domete vzpostavlja pragmatično nasprotje med dvema izjavama (retorična funkcija nasprotja), nasprotje med dvema ubeseditvama (retorična funkcija popravilalne spremembe

ubeseditve), nasprotje med dejanskim in situacijskim védenjem (retorična funkcija poudarjanja),⁸⁰ v pomensko najbolj izpraznjenih rabah pa tudi funkcijo zapolnjevanja vrzeli, kjer poudarjanje novosti dane informacije glede na predhodno védenje pogosto ni več razvidno.

Na podlagi teh predpostavk in funkcijskih lastnosti večbesednih diskurznih označevalcev v vzorčenem korpusu govorjene slovenščine smo zato označevalce združili v **deset jedrnih pomenskih skupin: označevalce razloga, rezultata, nasprotja, pogoja, specifikacije, dodajanja, razvrščanja, dvogovora, nedoločnosti in tvorjenja**. Podrobneje jih predstavimo v nadaljevanju, pri čemer poudarjamo, da je tovrstna delitev trenutno utemeljena zgolj z analizo rabe večbesednih označevalcev, ki bi jo v prihodnje veljalo dopolniti še z analizo funkcijsko analognih posamičnih besed, hkrati pa ponazoritve možnih pragmatičnih izpeljav niso izčrpne, temveč ponazarjalne, saj je dejanski nabor kontekstualno pogojenih funkcij možno utemeljiti šele na obsežnejši analizi gradiva.

V skupino **označevalcev razloga** umeščamo izraze, ki besedilno enoto,⁸¹ v kateri se pojavijo, označujejo kot vzrok, utemeljitev ali pojasnilo druge diskurzne enote. V vzorčenem korpusu govorjene slovenščine so to večbesedni izrazi *glede na to da*, *zaradi tega ker*, *zato ker*, med tipične enobesedne označevalce razloga pa bi lahko uvrstili izraze *ker*, *saj*, *kajti*, *namreč* ipd. Za razliko od semantične oz. predstavne rabe, kjer uvajajo vzrok soležne propozicije (primer [81]), lahko s spremembo dometa v pragmatičnih rabah med drugim razvijejo funkcijo utemeljevanja (primer [82]), pa tudi druge retorične in sekvenčne organizacije diskurza.

⁸⁰ Dodaten argument, da moramo funkcijo retoričnega poudarjanja tolmačiti predvsem kot lastnost označevalcev s specifičnim dometom (izvenbesedilno kazanje na skupno védenje in/ali situacijski kontekst), ne pa kot pomensko lastnost, je tipična raba nekaterih slovenskih členkov kot poudarjalcev ilokucijske sile, ki lahko pripadajo različnim semantičnim skupinam, npr. uvajalcem nasprotja, tako kot v *bistvu* in *vendar* (*Povej vendar!*), uvajalcem posledice, tako kot npr. *torej* (*Kako torej?*), dopustnosti (*ampak rezultat referendumu vendarle lahko predvidimo*) in tako dalje.

⁸¹ Z izrazom **besedilna enota** v nadaljevanju označujemo sporočilno zaključene enote besedila, ki jih tipično predstavljajo stavčno izražene propozicije, lahko pa tudi nižje skladenjsko-pomensko zaključene enote, kot so pristavki, ali daljši besedilni segmenti, kot so tematski sklopi. Z izrazom **diskurzna enota** pa uporabljamo kot nadpomenko za vse možne domete diskurznega označevalca v sporazumevalnem kontekstu, ki poleg sobesedilnega konteksta in z njim povezanega védenja (znotrajbesedilni kontekst), vključuje tudi enciklopedično védenje (kulturološki kontekst), védenje o sporazumevalni situaciji (situacijski kontekst) in védenje o naslovniku (medosebni kontekst). Običajno torej diskurzna enota sovpada z besedilno enoto, a jo moramo pri razlagi pomenskih lastnosti nekaterih označevalcev razumeti tudi širše, saj lahko ti kažejo tudi izven besedila.

4.1 Dopolnitev sheme za označevanje večbesednih diskurzni označevalcev

[81] in eee pre [gap] in zato smo tudi predlagali da **glede na to da** bi radi več stvari uredili pa v zakonu ni podlage da bi bilo dobro **glede na to da** je zakon o izvrševan kazenskih sankcij zdaj v eee v pred [gap] eee predviden za spremembe

[82] mislim da še ne glih ampak eee **glede na to da** eee res je lepo zakaj pa ne

V skupino **označevalcev rezultata** umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot posledico, namen ali sklep druge diskurzne enote. V vzorčenem korpusu govorne slovenščine so to večbesedni izrazi *če ne, kakorkoli že, na ta način, samo da, s tem, se pravi, se pravi da, tako ali tako, tako da, tako ka, to pomeni, to se pravi, v glavnem, v končni fazi, v tem primeru, v vsakem primeru, za to da, zaradi tega, zaradi tega da, zato da, zato pa*, med tipične enobesedne označevalce razloga pa bi lahko uvrstili izraze *zato, torej, skratka* ipd. Za razliko od semantične rabe, kjer uvajajo posledico (primer [83][90]) ali namen (primer [84]) soležne propozicije, vključno s posledico načina (primer [85]), lahko s spremembo dometa v pragmatičnih rabah med drugim razvijejo funkcije sklepanja oz. povzemanja (primer [86]), poudarjanja (primer [87]), (sklepalne) spremembe ubeseditve ([88]), (sklepalne) ponazoritve oz. opisa ([89]), pa tudi sekvenčne organizacije diskurza, kot so menjava ([90]) ali nadaljevanje teme po digresiji ([91]), zaključevanje vloge ali zapolnjevanje vrzeli ([92]) ipd.

[83] čakaj tebi je aaa prejšnje stanovanje bilo majhno pa ti ni bilo všeč **zaradi tega**

[84] in zgolj za primer preko sto tisoč evrov da naša država tudi delavka mure tudi vsi mi **za to da** leto dni stoji plakat v b [gap] na bruseljskem letališču

[85] in nas to zelo zelo jezi ker pa me vemo da smo sposobne a ne [speaker:laughter] in dejansko tudi smo ampak si samo želimo da bi nas poslušali slišali da nam je težko nas razumeli in nam po [gap] **s tem** pomagali a ne nič ne rabimo nobene konkretne pomoči

[86] dobro **se pravi** [name:personal] ti si za quartissimo

[87] v največje povpraševanje je **tako ali tako** po teh strokovnih tehničnih besedilih to po teh splošnih tako niti ne pričakujte da bi dobili

[88] midva z dimom **se pravi** z dimitrijem ruplom sva bila pa najbolj izpostavljena v tistem času ko je izšla sedeminpetdeseta številka nove revije in to je vzbudilo v celi jugoslaviji cel cirkus

[89] vidite da so se stvari nekako zgostile zgostile so se pa relativno nedavno **se pravi** kar dolgo časa v temle vesolju v tej naši okolici naše galaksije je bilo zelo tako živahno življenje veliko stvari je noter padlo veliko stvari je ven špricnilo veliko je bilo tega da je majhna galaksija ena galaksija pojedla drugo

[90] da smo mi že v drugem obdobju ko si pa on že zgradi prvo bivališče iz vej slame in tako naprej **se pravi** kaj je bil na [gap] namen uporabe zatočišč ?

- [91] noro danes **v glavnem** znanstveniki odkrivajo ter preizkušajo različne metode in instrumente
- [92] eee če mu je še mama mislim da tamiflu zraven dala samo ga je doma pozabil razumeš [speaker:laughter] / [all:laughter] vse je imel v glavnem vse / **tako da** /// no povej kako sta se še kaj imela

V skupino **označevalcev nasprotja** umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot drugačno oz. nasprotno neki drugi diskurzni enoti. V vzorčenem korpusu govorne slovenščine so to izrazi *kljub temu da*, *medtem ko*, *medtem ko pa*, *namesto da*, *ne pa da*, *pa vendarle*, *po drugi strani*, *samo da*, *sicer pa*, *tudi če*, *v bistvu*, *v resnici*, *vendar pa*, med tipične enobesedne označevalce razloga pa bi lahko uvrstili izraze *ampak*, *vendar*, *pa*, *čeprav*, *pravzaprav*. Za razliko od semantične rabe, kjer izpostavljajo razliko med lastnostmi dveh soležnih propozicij ([93]) ali med pričakovanim in izraženim ([94]), lahko v pragmatičnih rabah glede na domet med drugim razvijejo funkcije retoričnega nasprotja ([95]), (popravljalne) spremembe ubeseditve ([96]), poudarjanja nasprotja z domnevanim ([97]), kontrastivnega nadaljevanja teme ([98]), poudarjalnega zapolnjevanja vrzeli ([99]) ipd.

- [93] in seveda vlekel sem vprašanja noter in sem gledal samo da izvlečem da popravljam jaz da je tekst pa da jaz popravim kje je treba vejica kje je treba dvopičje **ne pa da** da jaz delam kaj je eee osebek povedek te pa sem ni mi nekako nisem ni [gap] nisem tu jaz bil cel ne
- [94] njih je jasno potrebno pripeti **kljub temu da** statistika da jih je samo sedemdeset eem kakšna bila statistika od sedem ... pripetih otrok mmm tam je res neodgovorna no
- [95] seveda je pohvalno da imamo eee več eee ministrice v tej vladi **vendar pa** jaz sama eem ne ocenjujem ljudi na takšnih položajih po spolu ampak po tem eee kako učinkoviti so eee ali dovolj dobro poznajo resor na katerem so predvsem pa ali znajo v nekem trenutku odreagirati eee pravilno
- [96] eee je je to v bistvu eee tudi eee v potrebno imeti dosti izolirano hišo ali je to bolj podobno klasiki ne vem olju drva drva **v bistvu** ogrevanje na drva
- [97] se pravi imamo **v resnici** ravnotežno situacijo to je tako kot da nimamo nobenega mikrovalovnega gretja
- [98] ja gospod [name:surname] **pa vendarle** ali dopuščate možnost da se je sodnica ki sicer kot vi pravite odloča na podlagi zakona a vseeno tudi subjektivno glede na okoliščine da se je preprosto zmotila in bi lahko recimo nek drug sodnik odločil kako drugače in ga ne bi poslal v hišni pripor ?
- [99] eee teh eee to je **v bistvu** najbolj bom rekel prijazna kar se tiče kolesarja za vzdrževanje ker ni nobenih olja ni nobenega zračnih komor notri tako da v bistvu taka vilica potem če se enkrat na leto pogleda na servisu se tile vodila namažejo se prah obriše se te zadeve pogleda pa je to to

V skupino **označevalcev pogoja** umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot pogoj druge diskurzne enote. V vzorčenem korpusu govorne slovenščine so to denimo izrazi: *edino če*, *sploh če*, *sploh če*, *še posebej če*, *v kolikor*, med tipične enobesedne označevalce pogoja pa bi lahko uvrstili veznik *če*. Za razliko od semantične rabe, kjer uvajajo pozitivni ali negativni pogoj za uresničitev posledice v drugi izjavi ([100]) ali pogojne izjeme ([101]), lahko v pragmatičnih rabah glede na domet med drugim razvijejo funkcijo pragmatičnega pogoja oz. relevantnosti glede na soledno izjavo ([102]).

[100] namreč kot lion eee imaš pravico eee kjerkoli v svetu si poiskati lokalni klub **še posebej če** rabiš pomoč in lahko računaš na pomoč

[101] ne **edino če** starši se trudijo ne ampak ja ja ja

[102] eee za te razmere ki jih imamo danes pa **sploh če** upoštevamo tisti stavek ki sem ga na začetku omenil

V skupino **označevalcev specifikacije**⁸² umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot dodaten opis, podrobnost, ponazoritev ali elaboracijo druge diskurzne enote. V vzorčenem korpusu govorne slovenščine so to izrazi *a la*, *in sicer*, *in to*, *kot da*, *na primer*, *ne da*, *s tem da*, *tako ko*, *tako kot*, *v smislu*, *v smislu da*, med tipične enobesedne označevalce razloga pa bi lahko uvrstili izraze *npr.*, *tj.*, *recimo*. Ta pomenska skupina je izrazito metadiskurzna, zato se v predstavnih oz. semantičnih rabah ne pojavlja, v pragmatičnih rabah pa lahko glede na domet med drugim razvijejo različne funkcije specifikiranja, npr. ponazoritve ([103]), izpostavitve ([104]), primerjave ([105]) ali dopolnitve ([106]).

[103] eee **na primer** eee rokovanjaci od jurčiča ki mi je bilo sploh spet spet mi je zoprno bilo ko sem notri eee am [gap] nekaj slovenščine sem seveda naučil se

[104] pregovor pravi da neumnemu kmetu zraste najdebelejši krompir **in to** na asfaltu ampak kaj pa če ne zraste ?

[105] so take čez cele doline na primer **kot da** bi se upičili tam z otlice kje mimo stomaža viadukt narediti glih tako

[106] no saj ne malo se hecam ampak dejansko bomo bomo tokrat v tem projektu malo bolj sodelovali eem **v smislu** ne samo lektorsko pa akcentov in eem logike stavka ampak

⁸² V izogib napačnega tolmačenja v okviru drugih, prav tako medsebojno različnih opredelitev funkcije pojasnjevanja v slovenskem prostoru (Toporišič, 1984; Gorjanc, 1998; Balažic Bulc, 2009), namesto termina pojasnilo raje uporabljamo izraz specifikacija, ki sugerira dodajanje podrobnosti, ne pa denimo pojasnjevanja vzroka ipd.

bomo probali te like tako ko je eee [name:personal] zdaj že malo nastavlil bomo probali
skoz jezik in način govora [speaker:laughter]

V skupino **označevalcev dodajanja** umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot dodajanje ali razširitev predhodne diskurzne enote. Za razliko od specifikacije, kjer se diskurzni enoti nanašata na isto propozicijo, pri čemer ena predstavlja elaboracijo druge, označevalci dodajanja uvajajo novo propozicijo, ki je prvi sporočilno enakovredna (prim. funkcije konjunkcije, disjunkcije in razširitve v Bunt in Prasad 2016). V vzorčenem korpusu govorne slovenščine so to izrazi *ali pa*, *če ne drugega*, *glj tako*, *in pa*, *ob tem da*, *poleg tega pa*, *prav tako*, *predvsem pa*, *pri tem*, *s tem da*, *sploh pa*, med tipične enobesedne označevalce razloga pa bi lahko uvrstili izraze *in*, *pa*, *tudi*. Za razliko od semantične rabe, kjer povezujejo dva hkratna ([107][114]) ali alternativna dogodka ([108][115]), lahko v pragmatičnih rabah glede na domet med drugim razvijejo funkcije poudarjanja ([109][117]), pojasnjevalne razširitve ([110]–[111]) ali sekvenčnega dodajanja ([112]–[113]).

- [107] medaljo za hrabrost je dobil tudi občan [name:personal] [name:surname] ki je priskočil na pomoč v prometni nesreči in iz gorečega tovornjaka **prav tako** rešil nezvestnega voznika
- [108] zato ker ni džabest **ali pa** ni mobi
- [109] seveda je pohvalno da imamo eee več eee ministric v tej vladi vendar pa jaz sama eem ne ocenjujem ljudi na takšnih položajih po spolu ampak po tem eee kako učinkoviti so eee ali dovolj dobro poznajo resor na katerem so **predvsem pa** ali znajo v nekem trenutku odreagirati eee pravilno
- [110] ja **s tem da** možnosti niso ne za priklop televizije ker ste oddaljeni več kot tri pa pol kilometre od centrale zdaj vidim
- [111] **pri tem** uporabljamo eee poligrafski instrument oziroma uporabljajo policisti in kriminalisti in upoštevajo eee zahteve različnih metod za odkrivanje zavajanja oziroma preverjanja verodostojnih izjav
- [112] prvi primer je sotinski breg in to tukaj gre za drugi primer kjer je investitor eee lokalna eem lokalna skupnost **in pa** seveda j [gap] so je postaja postavljena po vseh standardih ki ustrezajo državni meteorološki eee mreži in je zato lahko vključena v eee državno meteorološko mrežo kot enakovredna postaja meteorološkim postajam ki jih sicer postavlja agencija za okolje
- [113] ja ja ma pazi tam je pa tudi višina notri na tistem viaduktu / **ali pa** ista mmm zadeva je dol če greš med parmo in la spezio

V skupino **označevalcev razvrščanja** umeščamo izraze, ki besedilno enoto, v kateri se pojavijo, označujejo kot prekrivno ali zaporedno glede na drugo diskurzno enoto. V učnem korpusu govorne slovenščine so to izrazi *danes ko*, *druga zadeva*, *medtem ko*, *na koncu*, *pol pa*, *potem*

4.1 Dopolnitev sheme za označevanje večbesednih diskurznihih označevalcev

ko, potem pa, prej ko, sedaj pa, takrat ko, te pa, tedaj pa, v nadaljevanju, za začetek, zdaj pa, že odkar. Za razliko od semantične rabe, kjer označujejo časovno zaporedje ali prekrivanje dveh propozicijsko izraženih dogodkov ([114]–[115]), lahko v pragmatičnih rabah glede na domet med drugim razvijejo funkcije retoričnega sklepa ([116]), sekvenčnega nizanja ([117]), menjave teme ([118]) ipd.

[114] tako [name:personal] in [name:personal] sta povedala kako delate kako nekako začnete z ustvarjanjem ampak **potem ko** je treba to spraviti na oder pa najbrž kar potrebno vaditi ne ?

[115] **že odkar** je [gap] majčken

[116] v redu **potem pa** jutri prinesete s seboj delovne zvezke bomo s premim govorom nadaljevali

[117] eee **druga zadeva** je majhen priročnik gre kamorkoli hočeš

[118] dobro naj bo ljubezenska **zdaj pa** kaj jo pa dela ljubezensko kaj ?

V skupino **označevalcev dvogovora** umeščamo izraze, ki se tipično pojavljajo v govorni interakciji med tvorcem in enim ali več soudeleženci, in sicer besedilno enoto, v kateri se pojavijo, povezujejo z naslovnikom tako, da ga na izrazni ravni eksplicitno nagovarjajo bodisi z vprašanjem bodisi z odzivom. V primerjavi s tipologijo D. Verdonik (2008) ta skupina delno sovпада z interakcijskimi označevalci. V učnem korpusu govorne slovenščine so to denimo izrazi *a ne, a ne da, a veš, a veš kaj, a vidite, ali kaj, ali kaj je, ali kako, ali ne, če dovolite, ja no, kaj je, ma ja, no ja, saj veš, tako torej, v redu, veš kaj, veš kaj je, znaš kaj*. Pri tem lahko glede na domet oz. izgubo propozicijske vsebine poleg vzpostavljanja in uravnavanja medosebnih odnosov, kot so funkcije preverjanja pozornosti ([119]–[121]), (ne)strinjanja ([122]), ohranjanja dostojanstva ([123]), nedoločnosti ([124]) in/ali elipse ([125]), med drugim opravljajo tudi vloge zapolnjevanja vrzeli ([126]), odpiranja ([127]) ali zapiranja besedilnih sekvenc ([128]) ipd.

[119] ja ja pa moraš malo dohajati **a ne da** [speaker:laughter]

[120] ja saj itak kaj boš pa drugega počel **a veš**

[121] saj potem si je v bistvu sama kupila ne **ali kako** ?

[122] **ma ja** to mi hodi gor ma ni lepo da [gap] v elastiki

[123] ja no to je pa dobra novica tako da [speaker:laughter] [gap] da ne bo dolga seansa ja ker imamo pol **saj veš** tisto [gap] in tako naprej ja

[124] ja zdaj n [gap] pravijo da bo še nekaj menda država dala **ali kaj** ne

[125] iz druge neke naše nanizanke ali nadaljevanke **ali kaj je**

[126] ta **kaj je** od eee

[127] **a veš kaj** malega otroka ima

[128] **v redu** dobro torej to to moramo to je res nujno no to moramo nujno narediti

V skupino **označevalcev nedoločnosti** umeščamo izraze, ki se prav tako tipično pojavljajo v govorjeni interakciji, in sicer na izrazni ravni eksplicitno zmanjšujejo sporočilno neposrednost besedilne enote, v kateri se pojavijo, bodisi z namenom zmanjšanja stopnje zavezanosti k veljavnosti izrečenega bodisi z namenom izogibanja navajanja nerelevantnih, občutljivih ali splošno znanih podrobnosti. Po tipologiji D. Verdonik (2008), ta skupina večinoma sovпада s skupino interpretacijskih označevalcev. V učnem korpusu govorjene slovenščine so to izrazi *ali karkoli, ali nekaj takega, ali pa nekaj, in tako, in tako naprej, in to, in vse skupaj, kaj jaz vem, ne vem, ne vem kaj, nekaj takega, pa take, pa take fore, pa take stvari, pa tako, pa tako naprej, pa to, po svoje, prav tako, saj ne vem*. Pri tem lahko glede na domet oz. izgubo propozicijske vsebine poleg vzpostavljanja in uravnavanja medosebnih odnosov, kot sta funkcija nedoločnosti ([129]) in/ali elipse ([130]), med drugim opravljajo tudi vloge zapolnjevanja vrzeli ([131]) ali zapiranja vlog ([132]), pogosto v kombinaciji z označevalci dvogovora ([133]) ipd.

[129] pri ustnih harmonikah je to **kaj jaz vem** dvajset trideset evrov ne

[130] da smo mi že v drugem obdobju ko si pa on že zgradi prvo bivališče iz vej slame **in tako naprej** se pravi kaj je bil na [gap] namen uporabe zatočišč ?

[131] tako da že osemdeset narediš ne lahko ti pa pač zavrnejo diplomu ne pač rečejo **ne vem** plagiat preveč ste **ne vem** niste uporabljali ne

[132] aha eee ja ne mislim naročilo kao bi imeli pa če bi lahko pol to pripisali pa **ne vem** [gap]

[133] petdeset zdaj ne ... ja zdaj smo nekaj ne so nam nekaj dopusta pobrali **pa to ne**

V skupino **označevalcev tvorjenja** umeščamo izraze, ki se na izrazni ravni eksplicitno nanašajo na procese tvorjenja spontanega govora, tj. vsebujejo glagole rekanja in mišljenja, v primerjavi s skupino označevalcev procesov tvorjenja v tipologiji D. Verdonik (2008) pa ne tudi védenja, ki se v naši razvrstitvi umeščajo v pomensko skupino označevalcev nedoločnosti. V vzorčenem korpusu govorjene slovenščine so to izrazi *bi človek rekel, bi rekel, bom rekel, bom rekla, kaj naj rečem, kot rečeno, kot smo rekli, moram reči, recimo temu, saj pravim, tako rekoč*. Pri tem lahko glede na domet oz. izgubo propozicijske vsebine poleg retoričnih funkcij poudarjanja ([134])

opravljajo tudi medosebne funkcije izražanja nedoločnosti ([135]) in/ali sekvenčne vloge zapolnjevanja vrzeli ([136]), nadaljevanja oz. povratka ([137]a) ali ponavljanja ([137]b).

- [134] jaz v bistvu vsakemu odgovorim in tudi **moram reči** eee na to na ta samostojen koncert že dostikrat naletela eee beseda tako da upam da ni samo par tistih ki so mi pisali ampak jih je ne vem koliko jih mora biti
- [135] aha aha aha aha aha aha a [gap] ... eee ba [gap] babure ja ... ja ki jo jaz poznam **tako rekoč** dedca poznam
- [136] ki **bi rekel** ti konfliktni interesi tudi ustvarjajo potem to neko nek vladajoči razred ne iz katerega pravzaprav ko enkrat prideš vanj ne moreš več izpasti
- [137] /// ni jim lušno ne / seveda kakšen si je tudi sam kriv velika večina je verjetno po nekem spletu okoliščin se z [gap] znašla v težavah in eee potem seveda ne z [gap] ne najdejo izhoda / **kaj naj rečem** vesel sem da nisem med njimi ne jim pa **saj pravim** nič ne zavidam ///

Če sklenemo, nam tovrstna robustna izpeljava torej omogoča, da lahko z izjemo nekaterih izrazov (kot je denimo označevalec *medtem ko*, ki lahko označuje tako časovno razvrščanje kot izražanje nasprotja) **večino oblik umestimo v eno samo krovno pomensko kategorijo, ne glede na njeno specifično vlogo v specifičnih rabah**, torej ne glede na to, ali je rabljena kot semantični ali pragmatični povezovalac propozicij, izjav, ubeseditev, besedilnih sekvenc, tematskih sklopov, predpostavljenega vedenja ali situacijskega konteksta, obenem pa so te krovne kategorije dovolj informativne in medsebojno razločljive, da lahko denimo v nadaljnjih analizah podatkov in drugih raziskavah ločujemo med različnimi temeljnimi skupinami označevalcev, kot so označevalci vzpostavljanja medbesedilnih povezav oz. **besedilni konektorji na eni strani** (označevalci rezultata, razloga, pogoja, nasprotja, specifikacije, dodajanja, razvrščanja) in **označevalci procesa interakcije na drugi** (označevalci dvogovora, nedoločnosti, tvorjenja).

Po eni strani predstavljena taksonomija ponuja možnost bolj sistematičnega pristopa k samemu pomenskemu opisu diskurznofunkcijske leksike, v katerem označevalcu najprej določimo pomensko jedro, šele nato pa s sistematično razčlenitvijo konteksta posameznih rab (kjer se kot zelo uporaben kaže parameter dometa) določamo tudi njegove specifične funkcijske uresničitve. Po drugi strani pa je tovrstna robustna pomenska taksonomija še zlasti koristna za namen naše konkretne raziskave, saj v naslednjem koraku začetni seznam večbesednih diskurznihih označevalcev v vzorčenem korpusu govorne slovenščine dopolnimo še z rezultati luščenja teh

4.1 Dopolnitev sheme za označevanje večbesednih diskurznih označevalcev

izrazov iz celotnega korpusa Gos, pri katerem ni mogoče pričakovati podrobne maksimalistične analize diskurznega konteksta vsakokratne rabe posameznega označevalca.

4.2 Metode prepoznavanja večbesednih diskurznihi označevalcev v besedilnih korpusih

V tem poglavju na podlagi presenetljivo slabe raziskanosti vprašanja optimalne metode za pridobivanje nabora diskurznofunkcijskih večbesednih enot (poglavje 4.2.1 *Pregled metod za prepoznavanje večbesednih diskurznihi označevalcev*) kvalitativno prepoznavanje večbesednih diskurznihi označevalcev v vzorčenem korpusu govorne slovenščine dopolnimo še z različnimi kvantitativnimi metodami luščenja teh izrazov iz celotnega korpusa Gos (poglavje 4.2.2 *Priprava seznamov potencialnih večbesednih diskurznihi označevalcev*). Z njihovo primerjavo v poglavju 4.2.3 *Primerjava metod prepoznavanja večbesednih diskurznihi označevalcev* nato skušamo odgovoriti na drugo raziskovalno vprašanje »Katera metoda je najustreznejša za prepoznavanje večbesednih diskurznihi označevalcev v besedilnih korpusih?«.

4.2.1 Pregled metod za prepoznavanje večbesednih diskurznihi označevalcev

4.2.1.1 Kratek pregled metod za prepoznavo diskurznihi označevalcev v besedilnih korpusih

Na ožjem področju raziskav diskurznihi označevalcev se vprašanju izdelave njihovega izčrpnega seznama le **redko namenja posebna pozornost**, saj se ob izraziti funkcijski naravi te jezikoslovne kategorije večina raziskav osredotoča predvsem na vprašanje njihove opredelitve, razmejitve, pomenskega opisa, komunikacijske vloge ali diahronega razvoja, z metodološkega vidika pa se pri tem zadovolji z analizo peščice vnaprej izbranih ponazoritvenih primerov in naknadnim posploševanjem na kategorijo kot celoto. Poskuse izdelave izčrpnih seznamov diskurznihi označevalcev tako najdemo predvsem na področju raziskav diskurznega razčlenjevanja, ki se v svojem opisu načinov strukturiranja zaključenih besedilnih enot oz. vzpostavljanja medbesedilnih povezav opirajo tudi na njihove eksplicitne kazalce na površini.

Seznami diskurznihi označevalcev so znotraj tovrstnih raziskav najpogosteje izdelani v procesu ročnega razčlenjevanja manjših besedilnih korpusov, pri čemer si ti pri prepoznavi že v izhodišču pogosto pomagajo s spiski, nabranimi na podlagi slovarjev, slovnici ali drugih sorodnih raziskav, ali s prevodom seznamov označevalcev iz drugih jezikov. S kombinacijo **vnaprej zbranihi seznamov** in/ali njihovim dopolnjevanjem na podlagi **ročnega pregleda manjših besedilnih korpusov** so denimo nastali najodmevnejši leksikoni angleških (Knott in Dale, 1994; Knott,

1996; Marcu, 1998; Prasad idr., 2008), nemških (Stede, 2002), španskih oz. katalonskih (Alonso, Castellón, in Cor, 2002; Briz Gómez idr., 2008) in francoskih (Roze idr., 2012) diskurznihi označevalcev, z ročnim pregledom korpusa pa je bil izdelan tudi seznam alternativnih leksikalizacij v angleškem korpusu PDTB (Prasad idr., 2010) oz. sekundarnih povezovalcev v češkem korpusu Prague Discourse Treebank (Rysová in Rysová, 2015).

Še **bistveno redkeje pa v literaturi najdemo poskuse samodejne prepoznave** diskurznihi označevalcev (strojnega luščenja), denimo z namenom izdelave njihovega spiska v večjih besedilnih korpusih. E. Lapshinova-Koltunski in K. Kunz (2014a) tako pri izdelavi seznama povezovalnih adjunktov oz. konjunktivov (Halliday in Hasan, 1976) v nemško-angleškem korpusu oblikujeta specifične iskalne pogoje v obliki regularnih izrazov, ki v korpusih iščejo izraze, ki se pojavijo na začetku stavka, so označeni kot vezniki, niso del predmetnih odvisnikov ipd. Podobno tudi orodje X-TRACTOR (Alonso idr., 2002), ki je bilo razvito za luščenje diskurznihi označevalcev v španskih besedilih, pri prepoznavanju potencialnih diskurznihi označevalcev upošteva kompleksen sistem skladenjskih pravil, poleg tega pa še različne druge tipe podatkov, kot so pogostost rabe, dolžina, vzajemna vrednost med posameznimi sestavinami, skladenjska struktura, vsebovanost tipično diskurznihi leksemov (npr. *razlog*, *posledica*) itd., s čimer je bilo kljub obširnemu izhodiščnemu leksikonu več kot 500 potencialnih označevalcev med 100 ročno pregledanimi izluščenimi kandidati denimo prepoznanih tudi 15 povsem novih označevalcev.

Ob tako maloštevilnih razpravah, posvečenih metodološkemu vprašanju pridobivanja seznamov diskurznihi označevalcev v besedilnih korpusih se zdi, da področje raziskav diskurznihi označevalcev inherentno zaznamuje **implicitna predpostavka, da diskurzni označevalci predstavljajo razmeroma majhno, zaprto skupino izrazov, ki so v rabi tako pogosti, da je njihov nabor v jeziku mogoče določiti že s pregledom razmeroma majhnih vzorcev jezikovne rabe**. Področje tako zaznamujejo izrazito kvalitativni pristopi k prepoznavi in opisu teh jezikovnih izrazov, pri premiku fokusa z manjših na obsežnejše besedilne korpusse pa se kvantitativne raziskave ukvarjajo predvsem z razvojem metod samodejnega razdvoumljanja (razločevanja med diskurzno in drugimi rabami določene oblike) ali samodejne funkcijske kategorizacije (razločevanja med funkcijami večfunkcijskega označevalca) vnaprej znanih oblik označevalcev, skoraj povsem neraziskano pa ostaja področje kvantitativnih metod njihove

izhodiščne prepoznave. Čeprav je ob primarni osredotočenosti na enobesedne in visoko gramatikalizirane večbesedne označevalce ta predpostavka do neke mere upravičljiva,⁸³ pa se s **premikom fokusa na bistveno slabše raziskano področje večbesednih enot kaže potreba po njenem ponovnem premisleku**, še zlasti ob upoštevanju dejstva, da v pričujoči raziskavi med večbesedne diskurzne označevalce prištevamo širši nabor strukturno zaključenih večbesednih enot z ustaljeno diskuržno funkcijo kot zgolj tiste najbolj gramatikalizirane oz. pogoste. V nadaljevanju zato na kratko predstavimo še pregled komplementarnega področja kvantitativnih pristopov k prepoznavanju večbesednih enot v besedilnih korpusih, kakršni se uporabljajo na drugih področjih korpusnega jezikoslovja, tudi analize diskurza.

4.2.1.2 *Kratek pregled metod za luščenje večbesednih enot iz besedilnih korpusov*

Kot je to značilno za področje korpusnega jezikoslovja nasploh, je tudi metodologija samodejnega prepoznavanja večbesednih enot neločljivo povezana s statistiko, pri čemer pa posamezne metodološke pristope lahko razdelimo v dve krovni skupini glede na to, katere statistične podatke o jezikovni rabi obravnavajo kot relevantne za iskanje empiričnih odgovorov na vprašanja tvorjenja in razumevanja jezika. Čeprav predstavljena dihotomija zaznamuje področje raziskav večbesednih enot nasploh, pa se pri njihovi predstavitvi osredotočamo predvsem na pristope k luščenju neprekinjenih besednih nizov (ne pa na primer kombinacij s prosto stavo ali leksikogramatičnih vzorcev) iz jezikoslovno neoznačenih korpusov (ne pa na primer iz korpusov s pripisanimi oblikoslovnimi, skladenjskimi, semantičnimi ali diskurznimi podatki).

Na eni strani tako najdemo pristope, ki v ospredje svojega zanimanja postavljajo predvsem podatek o pogostosti in s tem tiste večbesedne enote, ki se v rabi pojavljajo najpogosteje, ne glede na to, ali predstavljajo strukturno (ne)zaključene, pomensko (ne)razstavljive ali psihološko (ne)izstopajoče enote,⁸⁴ saj jih bolj kot njihov kognitivni status zanima njihova vloga pri

⁸³ Tudi s kvalitativnega vidika se namreč odpira vprašanje reprezentativnosti (žanrske, časovne, regionalne) temeljnih korpusov in na njih temelječih seznamih označevalcev.

⁸⁴ Izraz *saliency* (tudi *saliency* in prid. *salient*; pogosto v zvezah *perceptual saliency*, *psychological saliency*, *psycholinguistic saliency*) je eden izmed osrednjih pojmov teorij o formulaični naravi jezika in označuje zaznavno očitnost, opaznost, relevantnost, celovitost ipd. besednih zvez, predvsem v nasprotju z njihovo statistično

prepoznavanju komunikacijskih značilnosti posameznih diskurznihi tipov ali njihove medsebojne primerjave. Z metodološkega vidika tovrstne raziskave torej temeljijo na analizi **najpogostejših besednih nizov**, tj. najpogostejših kombinacij dveh ali več korpusnih pojavnic (večinoma oblik in ne lem), za katere se danes na področju raziskav formulaičnega jezika običajno uporablja izraz *leksikalni skupi* (Biber idr., 1999, 2004), na področju računalniškega jezikoslovja pa izraz besedni *n-gram*, glede na število vsebovanihi besed tudi *dvo-*, *tri-* ali večgram. Pri določanju **frekvenčnega praga** za ločevanje relevantnih od nerelevantnih nizov se različne raziskave odločajo različno, običajno pa ta številka variira med relativno pogostnostjo 10 pojavitev niza na milijon pojavnic (Biber idr., 1999; Simpson-Vlach in Ellis, 2010; Cortes, 2015), 20 pojavitev (Cortes, 2004) ali 40 pojavitev niza na milijon pojavnic (Biber idr., 2004; Biber, 2009), nekoliko redkejša pa so raziskave z nižjim pragom, npr. štirimi pojavitvami niza na milijon pojavnic (McCarthy in Carter, 2006). Pogosto so raziskave najpogostejših besednih nizov omejujejo tudi na nize ene same dolžine, običajno trehi, štirih ali petih besed, zato se pri njihovem luščenju in analizi pogosto niti ne ukvarjajo z vprašanjem njihove medsebojne vsebovanosti, torej dejstvom, da vsaki daljši niz sestavljajo krajši nizi. Redke raziskave, ki hkrati primerjajo pogostost leksikalnih skupov različnih dolžin (O'Donnell, 2010; Adolphs in Carter, 2013), tako namesto preproste frekvence uporabljajo nekoliko **prilagojene metode štetja**, ki pojavitev niza v korpusu zanemarijo, če se ta pojavi kot del daljšega relevantnega niza, za kar se na področju računalniškega jezikoslovja uporablja tudi izraz statistična redukcija podnizov (Lü idr., 2005).

Na drugi strani metodološkega kontinuuma na področju kvantitativne prepoznave večbesednih enot pa lahko umestimo pristope, ki v ospredje svojega zanimanja postavljajo predvsem **funkcijsko in/ali strukturno samostojne enote**, ki jih bolj kot statistična relevantnost zaznamuje status zaznavne celovitosti. Čeprav se pri tem različne raziskave osredotočajo na različne vrste samostojnih večbesednih enot, od pomensko povsem nerazstavljivih (idiomatičnih) do pogostih in pomensko razstavljivih (institucionaliziranihi) besednih zvez, ki niso nujno tudi strukturno celovite, je vsem skupno to, da se pri njihovem prepoznavanju v obsežnih besedilnih

prepoznavnostjo. V disertaciji kot slovensko ustreznico uporabljamo izraz (zaznavno, psihološko, psiholingvistično) *izstopanje*, čeprav se zavedamo, da zaradi konotacij, ki izhajajo iz njegove splošne rabe, ni idealna terminološka izbira.

korpusih pogosto opirajo na t. i. **mere besedne povezovalnosti** (angl. *association measures*), ki kažejo na to, kako tesna je povezava med besedama v dani zvezi (kolokaciji). Za izračun stopnje medbesedne povezovalnosti (kolokabilnosti ali kolokacijske moči) je bilo doslej predlaganih že na desetine različnih statističnih izračunov (za njihovo pregled in primerjavo glej (Wiechmann, 2008; Evert, 2009; Pecina, 2010), ki na različne načine združujejo podatke o pogostosti zveze kot celote, pogostosti posameznih besed, ki jo sestavljajo, ter velikosti korpusa. V primerjavi z razvrščanjem besednih kombinacij po njihovi absolutni pogostosti te mere torej skušajo prepoznati enote, ki niso posledica predvidljive skladijske kombinatorike v jeziku nasploh, temveč je njihovo sojavljanje v paru verjetnostno izstopajoče (pogostejše kot pričakovano). Z vidika naše raziskave, v kateri se sicer prav tako osredotočamo na funkcijsko in strukturno celovite enote, je ena izmed glavnih pomanjkljivosti mer za izračun besedne povezovalnosti torej dejstvo, da je **zvezam, ki vsebujejo (pogoste) funkcijske besede**, zaradi njihove siceršnje pogostosti v jeziku običajno pripisana manjša stopnja kolokacijske povezanosti kot kombinacijam manj pogostih predmetnopomenskih besed.

Prav tako se mere besedne povezovalnosti običajno razvijajo, preizkušajo in evalvirajo za potrebe luščenja **dvobesednih kolokacij, bistveno redkeje pa za potrebe luščenja daljših besednih zvez**. Med njimi za izračune kolokabilnosti tro- ali večbesednih nizov obstajajo različni načini prilagoditev izhodiščnih mer, denimo s pretvorbo v vse možne psevdobigrame in različnimi načini povprečenja njihovih posameznih kolokabilnosti (J. F. da Silva in Lopes, 1999; Wei in Li, 2013), s povprečenjem kolokabilnosti med prvim in zadnjim psevdobigramom (Tadić in Šojat, 2003), s kombinacijo več različnih pristopov (Petrović idr., 2010), z iterativnim ponavljanjem iskanja kolokatorjev (Kilgariff idr., 2012; Gries, 2013), z iskanjem najpogostejšega niza prepoznane dvobesedne kolokacije (Kilgariff idr., 2015), pogosto pa način izpeljave formule niti ni podrobneje pojasnjen (Ramisch idr., 2010).

Tudi v tej skupini metodoloških pristopov je **problem medsebojne vsebovanosti** krajših in daljših nizov običajno namenjene manj pozornosti, saj se večina raziskav z njo bodisi ne ukvarja (Petrović idr., 2010; Ramisch idr., 2010) bodisi se zaradi tega osredotočajo zgolj na enote ene same dolžine, običajno dvobesedne kolokacije, ali enote, daljše od dveh besed (Simpson-Vlach in Ellis, 2010). Med statističnimi metodami se za ta namen najpogosteje uporablja algoritem lokalnega maksimuma (J. Da Silva idr., 1999; Wei in Li, 2013), ki kolokacijsko moč danega n-

grama primerja s kolokacijsko močjo njegovih podnizov ali nadnizov. Če je kolokacijska moč danega n-grama večja od kolokacijske moči obeh podnizov in obeh nadnizov, je niz prepoznan kot večbesedna enota, vendar to obenem pomeni, da je med nizoma dveh zaporednih dolžin (npr. *zaradi tega* in *zaradi tega ker*) s tovrstno metodo vedno lahko izbran le en. Tudi S. Th. Gries (2013) v pregledu stanja in perspektiv na področju izračunov besedne povezovalnosti zato ugotavlja, da je njihov razvoj za potrebe luščenja leksikalnih enot poljubne dolžine ostaja eden ključnih izzivov prihodnosti, prav tako pa tudi razprava o (ne)ustreznosti metodoloških pristopov, ki temeljijo zgolj na pogostosti, na eni strani in pristopov, ki temeljijo na različnih izračunih besedne povezovalnosti, na drugi še traja (Biber, 2009; Bybee, 2010; Hanks, 2010; Gries, 2012).

V korpusnih raziskavah **večbesednih enot slovenskega jezika** se je doslej običajno uporabljala predvsem metoda luščenja kolokacij na podlagi izračunov besedne povezovalnosti, denimo za potrebe pridobivanja leksikografsko relevantnih večbesednih enot različnih tipov (Kosem idr., 2013; Gantar idr., 2016), avtomatske izdelave leksikona večbesednih enot (Ljubešić idr., 2015), prepoznavanja terminoloških zvez (Vintar, 2010; Logar idr., 2014) ali ugotavljanja leksikalnih specifik spletne komunikacije (Pollak, 2015a). Čeprav so bile ob začetkih korpusnega jezikoslovja v slovenskem prostoru opravljene krajše primerjave različnih načinov izračunov besedne povezovalnosti za potrebe pridobivanja stalnih zvez v slovenščini (Gorjanc in Krek, 2001; Gorjanc, 2005; Gantar, 2007),⁸⁵ se danes metodološkemu vprašanju izbire najustrenejše mere namenja precej manj pozornosti. Večina naštetih raziskav namreč uporablja mero logDice (Rychlý, 2008), ki jo kot privzeto mero za prepoznavanje kolokacij ponuja leksikografsko orodje SketchEngine, v katerem se nahaja večina referenčnih slovenskih korpusov (Erjavec, 2013), pri njeni evalvaciji za raziskave večbesednih enot v slovenščini pa se ukvarjajo predvsem z vprašanjem določanja minimalnega pragu kolokabilnosti glede na besedno vrsto in frekvenčno skupino leme (Gantar idr., 2016). Ker se je pri primerjavi avtomatsko in ročno izdelanih gesel

⁸⁵ Vse tri raziskave, ki v resnici primerjajo zgolj nekaj deset najpogostejših kolokatorjev nekaj izbranih mer, se osredotočajo predvsem na primerjavo uporabnosti mere vzajemne vrednosti (MI) in kubične vzajemne vrednosti (MI3), z ugotovitvijo, da slednja daje boljše rezultate za leksikološko analizo večbesednih enot v slovenščini. Gantar (Gantar, 2007, str. 46–48) v primerjavo dodatno vključi še mero LL, za katero na podlagi analize najpogostejših kolokatorjev leme *zimski* v korpusu FidaPLUS ugotavlja, da v primerjavo z vrednostjo MI3 ne prinaša bistvenih razlik.

Leksikalne baze za slovenščino (Gantar idr., 2016) izkazalo, da so leksikografi v ročno izdelanih geslih dajali večji poudarek podatku o pogostosti kot pa zgolj podatku o razvrščanju glede na stopnjo kolokabilnosti, načrtovalci novega kolokacijskega slovarja slovenskega jezika pri naboru kolokacij, ki ustrezajo predvidenim skladenjskim vzorcem, upoštevajo tako kolokatorje, razvrščene po pogostosti, kot kolokatorje, razvrščene glede na mero logDice (Gantar idr., 2015), kar potrjuje, da tudi v slovenskem prostoru **vprašanje optimalne metode za luščenje statistično izstopajočih večbesednih enot ostaja odprto**. Pri tem izpostavimo še, da so se vse zgoraj omenjene raziskave osredotočale predvsem na analizo predmetnopomenskih leksikalnih enot, **z diskurznimi in drugimi funkcijskimi večbesednimi enotami pa se doslej slovensko korpusno jezikoslovje ni ukvarjalo**.

4.2.1.3 Izbrana metoda pridobivanja seznama večbesednih diskurznihi označevalcev

V slovenskem prostoru doslej še ni bilo sistematične gradivno utemeljene raziskave diskurznihi drugih funkcijskih večbesednih enot, na podlagi katere bi lahko oblikovali razmeroma izčrpen in dosleden nabor večbesednih diskurznihi označevalcev, kot so opredeljeni znotraj pričujoče raziskave. Ob hkratnem razmeroma slabo raziskanem vprašanju njihovih slovničnih in distribucijskih lastnosti v širšem jezikoslovnem prostoru nasploh je zato na tej stopnji razvoja področja bolj kot opiranje na eno samo metodo (npr. ročna prepoznavna proti samodejno luščenje; luščenje pogostih nizov proti luščenje kolokacij) smiselna izbira izrazito induktivnega pristopa, ki v analizo gradiva vstopa **brez predpostavk o lastnostih večbesednih označevalcev na eni strani in najustreznejši metodi njihovega prepoznavanja na drugi**. Poleg relevantnosti za pričujočo raziskavo je izdelava čim bolj izčrpnega in doslednega nabora večbesednih diskurznihi označevalcev še toliko pomembnejša z vidika njihovega pomenskega in izraznega opisa v okviru nadaljnjih raziskav, s katerimi bi lahko razvili prilagojene metode njihovega prepoznavanja in nenazadnje pripomogli tudi k razmisleku o ustreznosti njihove trenutne opredelitve. Na podlagi teh izhodišč smo zato v pričujoči raziskavi seznam potencialnih večbesednih diskurznihi označevalcev, pridobljen na podlagi ročne analize reprezentativnega vzorca korpusa Gos, skušali dopolniti še **z unijo več različnih kvantitativnih metod** za pridobivanje potencialnih večbesednih diskurznihi označevalcev iz celotnega korpusa Gos.

Za ta namen smo najprej izdelali seznam vseh nizov **dveh ali več standardiziranih pojavnic**, ki se v referenčnem korpusu govorne slovenščine Gos pojavijo **vsaj 5-krat na milijon pojavnic**, in sicer pri **vsaj dveh različnih govornikih**. Pri tem smo luščenje nizov omejili na mejo **znotraj izjave**, kot relevantne pojavnice pa upoštevali samo **besedne pojavnice** (vključno z opornimi signali tipa *ja*, *mhm*), ne pa zapolnjenih premorov oz. mašil (*eee*, *eem*, *mmm*, *nnn*, *mm*, *een*, *nn*) in nejezikovnih pojavnic, kot so intonančne oznake (?/), oznake prekrivnega govora (...), premori (*[pause]*), oznake nerazumljivih ali nedokončanih besednih fragmentov (*[gap]*), oznake glasovnih in neglasovnih zvokov (*[voice]*, *[incident]*, *[speaker:laughter]*, *[audience:laughter]*, *[all:laughter]*). Prav tako med jezikovne pojavnice nismo vključili anonimiziranih lastnih imen (*[name:personal]*, *[name:surname]*, *[name:organisation]*, *[name:place]*, *[name:phone]*, *[name:address]*).

Za dobljeni seznam nizov smo nato v korpusu pridobili **več ločenih statističnih vrednosti**. Po vzoru metod, ki večbesedne enote razvrščajo z vidika njihove pogostosti, smo nizom pripisali podatek o njihovi **absolutni pogostosti**, dodatno pa tudi podatek o reducirani oz. **prilagojeni pogostosti**, ki upošteva njihovo medsebojno vsebovanost (po algoritmu (O'Donnell, 2010)). Po vzoru metod, ki večbesedne enote razvrščajo z vidika kolokabilnosti njihovih gradnikov, smo za izluščene nize nato izračunali še kolokacijsko moč. Za ta namen smo uporabili več uveljavljenih statističnih mer na področju korpusnega jezikoslovja (Manning in Schütze, 1999; Evert, 2009), ki temeljijo na različnih matematičnih principih in posledično dajejo tudi medsebojno različne rezultate, tj. **izračun vzajemne vrednosti**, angl. mutual information (Church in Hanks, 1989), **kubični izračun vzajemne vrednosti** (Oakes, 1998), **izračun t-vrednosti** (angl. t-test ali t-score), **izračun logaritma verjetnosti**, angl. log likelihood ali G^2 (Dunning, 1993), **Diceov koeficient** (Dice, 1945) in **vrednost χ^2** . Pri njihovi prilagoditvi za enote, daljše od dveh besed, smo med obstoječimi predlogi razširitev uporabili **metodo pravične porazdelitve** (J. F. da Silva in Lopes, 1999), ki je v primerjavi z drugimi zgoraj naštetimi metodami razmeroma preprosta za prenos v prakso, z univerzalnim načinom posplošitve pa omogoča konsistentno izpeljavo za poljuben nabor mer in ne zgolj za eno ali nekaj izbranih, kot so denimo izračun vzajemne vrednosti (Wei in Li, 2013) ali zgolj preprostejše statistične mere (Ramisch idr., 2010).

Pred predstavitvijo primerjave vseh osmih načinov rangiranja nizov (absolutna pogostost, prilagojena pogostost, MI, MI³, LL, Dice, t-test, χ^2) z vidika priklica večbesednih diskurznihi označevalci te v naslednjem razdelku najprej še podrobeneje opišemo.

4.2.2 Priprava seznamov potencialnih večbesednih diskurznihi označevalci

4.2.2.1 Luščenje in običajno štetje besednih nizov

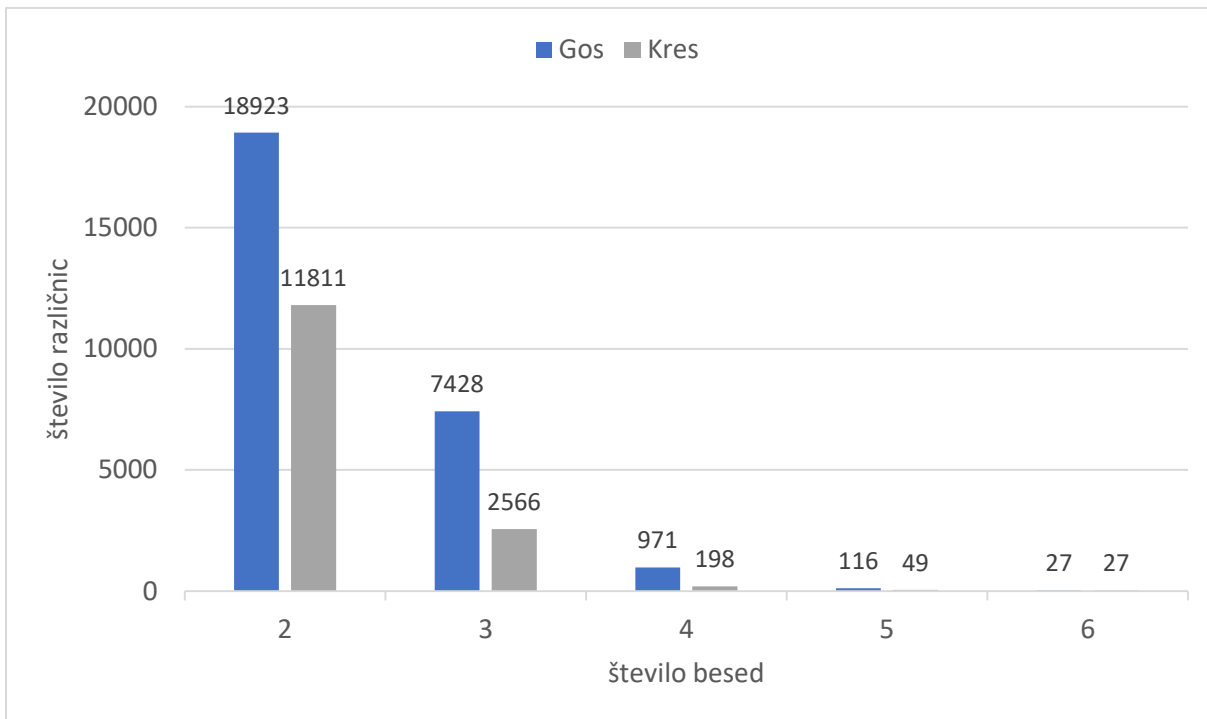
Kot je razvidno iz podatkov v Tabeli 4, smo z luščenjem vseh dvo- ali večbesednih nizov besednih standardiziranih besednih pojavnic, ki se znotraj meja ene govorne izjave pojavijo vsaj petkrat na milijon pojavnic (pri čemer smo kot velikost korpusa upoštevali število besednih pojavnic, tj. 1.005.455) in pri vsaj dveh različnih govornicah, iz korpusa Gos s pomočjo za ta namen razvite računalniške skripte pridobili 27.495 različnih besednih nizov z dvema ali več besednimi pojavnicami. Med njimi prevladujejo nizi dolžine dveh besed, precej manj pogosti pa so nizi dolžine pet ali več besed. Ker se je v podrobnejši analizi konkordanc nizov dolžine sedem ali več besed izkazalo, da se med njimi pojavljajo zgolj nizi, ki niso rabljeni kot diskurzni označevalci (npr. nizi opornih signalov tipa *ja ja ja ja ja ja ja ja ja*, izseki radijskih napovednikov tipa *v najboljšem glasbenem miksu za vso slovenijo* ali ponavljanja v kontekstu šolskega diskurza – *prijatelj navadno sedi v isti klopi in gleda*), smo nadaljnjo raziskavo omejili na preučevanje nizov **do dolžine šest besed**, med katerimi se denimo pojavi označevalec odprte propozicije *pa ne vem kaj še vse*.

4.2 Metode prepoznavanja večbesednih diskurznihi označevalcev v besedilnih korpusih

Dolžina niza	Število različnic	Število pojavitev	Povp. št. govorcev
2 besedi	18.923	417.283	15
3 besede	7.428	86.971	10
4 besede	971	8.929	7
5 besed	116	1.193	6
6 besed	27	334	4
7 besed	14	140	3
8 besed	9	59	3
9 besed	4	26	2
10 besed	2	14	2
11 besed	1	7	2
Skupaj	27.495	514.956	13

Tabela 4: Število večbesednih nizov z minimalno relativno pogostostjo 5/milj. in vsaj dvema različnima govorcema v korpusu Gos.

Graf na Sliki 3 dodatno prikazuje še njihovo primerjavo s pogostostjo formulaičnih nizov v korpusu pisne slovenščine, ki smo jih z enako metodo izluščili iz korpusa Kres, pri čemer smo namesto pogoja pojavljanja niza pri vsaj dveh različnih govoricah uporabili pogoj pojavljanja v vsaj dveh izmed skupno 21.456 besedil, za izračun relativnega frekvenčnega praga 5 pojavitev pa smo upoštevali skupno število besednih pojavnic brez ločil (tj. 99.831.145 pojavnic) oz. minimalno absolutno pogostost 499 pojavitev v celotnem korpusu Kres. Vidimo lahko, da tudi za slovenščino rezultati dokazujejo izrazito formulaičnost govornega diskurza v primerjavi s pisnim sporazumevanjem in potrjuje potrebo bo podrobnejših analizah večbesednih enot v govorni slovenščini. Korpus Gos namreč vsebuje skoraj enkrat več formulaičnih besednih nizov kot korpus Kres (511.917 proti 263.258 pojavitev na milijon pojavnic), pri čemer je razlika v formulaičnosti obeh tipov diskurza še očitnejša, če primerjamo število *različnih* besednih nizov po posameznih dolžinah, saj vidimo, da z izjemo nizov dolžine 6 besed govorni diskurz vsebuje več različnih pogostih nizov kot pisni, pri tri- in štiribesednih nizih celo do 3-krat oziroma 5-krat več.



Slika 3: Primerjava števila različnic za besedne nize z minimalno relativno pogostostjo 5 pojavitev na milijon pojavnic, dolžine 2-6 besed v korpusih Gos in Kres.

4.2.2.2 Prilagojeno štetje besednih nizov

Natančnejši vpogled na seznam najpogostejših n-gramov za standardizirane oblike v korpusu GOS v Prilogi 2 kaže, da so določeni večbesedni nizi v govorjeni rabi izrazito pogosti (npr. *ja ja, to je, ne vem, da je, tako da* itd.), tudi v primerjavi z enobesednimi pojavnicami, saj se denimo le 42 besed v korpusu Gos pojavlja pogostejše kot najpogostejša besedna zveza *to je*. Vendarle pa tovrstni frekvenčni spiski, kakršna ponujajo različna specializirana korpusna orodja, kot so kfNgram,⁸⁶ N-Gram Phrase Extractor,⁸⁷ N-gram Extraction Tool⁸⁸ ali mwe toolkit,⁸⁹ kot tudi večina zmogljivejših orodij za splošno korpusno analizo, npr. SketchEngine⁹⁰, WordSmith⁹¹,

⁸⁶ <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>

⁸⁷ http://lextutor.ca/n_gram/

⁸⁸ <http://homepages.inf.ed.ac.uk/lzhang10/ngram.html>

⁸⁹ <http://mwetoolkit.sourceforge.net/>

⁹⁰ <http://www.sketchengine.co.uk/>

⁹¹ <http://www.lexically.net/wordsmith/index.html>

NooJ⁹² ali AntConc⁹³ ne dajejo zadovoljivega odgovora na vprašanje, kako pogost je določen niz v primerjavi z drugimi besedami ali besednimi nizi, saj pri njihovem štetju ne upoštevajo medsebojne vsebovanosti, torej dejstva, da je vsak besedni niz (n-gram) sestavljen iz krajših nizov (n-1-gramov).

Za ponazoritev tega problema vzemimo dvobesedni niz *glede na*, ki se v kar 58 % vseh pojavitev (178 od skupno 309 pojavitev) v korpusu Gos pojavlja kot del daljšega besednega niza *glede na to*. To pomeni, da bi bilo na frekvenčnem seznamu večbesednih enot v korpusu Gos niz *glede na to* ustrezneje navajati pred nizom *glede na*, saj se ta izven te besedne zveze pojavlja manj pogosto kot znotraj nje. Po drugi strani bi morali na enak način tudi pri izračunu pogostosti besednega niza *glede na to* upoštevati, da se tudi sam pojavlja kot del pogostih daljših besednih nizov, npr. v 69 % (122 od 178 pojavitev) kot del niza *glede na to da*, ta pa se denimo včasih pojavi kot del nadrejenega besednega niza *ne glede na to da* (15 od 122 pojavitev).

Podobno se potreba po drugačnem štetju korpusnih pojavitev pojavlja tudi pri analizah frekvenčnih seznamov besednih nizov z enako dolžino. Za trigram *glede na to* in *na to da* vpogled v korpus denimo razkrije, da se v kar 122 pojavitvah od skupno 178 (*glede na to*) oz. 170 pojavitev (*na to da*) pojavita kot del istega štirigrama *glede na to da*, pri čemer sta oba trigram uvrščena na 304. oz. 322. mesto seznama najpogostejših nizov, štirigram pa šele na 485. mesto. To pomeni, da bi morali za realnejšo sliko razmerij v rabi pogostost obeh trigramov prilagoditi tako, da so štete zgolj njune samostojne pojavitve izven tega ali katerega drugega pogostega štirigrama, štirigram *glede na to da* pa bi moral biti na skupnem frekvenčnem seznamu besedišča v korpusu uvrščen precej višje od obeh trigramov, ki ju vsebuje.

Za tovrstno hkratno upoštevanje medsebojno vsebovanih nizov različnih dolžin se v tuji literaturi najpogosteje uporablja izraz statistična redukcija podnizov. Med različnimi predlaganimi metodami (npr. Cavnar idr., 1994; Lü idr., 2005; Nagao in Mori, 1994) smo za osnovo luščenja besednih nizov iz korpusa govorne slovenščine preizkusili algoritem za izdelavo t. i.

⁹² <http://www.nooj4nlp.net/>

⁹³ <http://www.laurenceanthony.net/software/antconc/>

prilagojenega frekvenčnega seznama, angl. *adjusted frequency list* (O'Donnell, 2010), ki je bil razvit prav za potrebe jezikoslovnih korpusnih analiz in je že bil preizkušen pri raziskavi besednih nizov v govorjeni angleščini (Adolphs in Carter, 2013). Če implementacijo te metode v obliki računalniške skripte v programu python na kratko povzamemo, ta v predhodno indeksiranem korpusu, v katerem ima vsaka pojavnica pripisan svoj unikatni številčni indeks, za vsak relevantni niz (n-gram), ki smo ga izluščili v prvem koraku, opisanem v prejšnjem poglavju, shrani podatek o njegovih konkretnih indeksih (mestih pojavljanja v korpusu), in nato preveri, ali se v besedilu pojavi kot del daljšega relevantnega niza (n+1-grama). Če to drži, se iz seznama vseh pojavitev danega n-grama ta pojavitev odstrani, s čimer se njegova končna pogostost zmanjša oz. ustrezno prilagodi.

Če ta iterativni postopek ponazorimo na primeru nizov, ki vsebujejo besedo *zaradi*, se torej v prvi iteraciji prilagajanja, v kateri se posameznim besedam zmanjšujejo frekvence glede na vsebovanost v dvobesednih nizih, pojavnici *zaradi* odštejejo vse tiste pojavitve, kjer nastopa kot del pogostega bigrama (npr. *zaradi tega*). V drugi iteraciji so se nato bigramom (npr. *zaradi tega*) odštejejo vse tiste pojavitve, kjer ti nastopajo kot del pogostega trigrama (npr. *zaradi tega ker*, *zaradi tega da*, *in zaradi tega*), in tako naprej vse do zadnje iteracije prilagajanja frekvenc nizov dolžine pet besed glede na pojavljanje v nizih maksimalne dolžine šest besed. Med štetimi pojavitvami besede *zaradi* tako ostanejo samo še tiste, kjer se ta ne pojavlja kot del besednega niza *zaradi tega*, temveč samostojno (npr. *to je prav kajti svetlobo potrebujemo zaradi marsičesa*), med pojavitvami niza *zaradi tega* so štete samo tiste, kjer se ta ne pojavlja kot del nizov *zaradi tega ker* ipd. (npr. *ali je zaradi tega ta višinski pas poseljen*) in tako dalje.

S tovrstno metodo štetja iz korpusa Gos pridobimo nekoliko drugačen seznam najpogostejših nizov kot z običajnim štetjem, ki ne upošteva medsebojne vsebovanosti pogostih nizov. Prva razlika med seznamoma, ki jo povzemajo podatki v Tabeli 5, je ta, da je v primerjavi z začetnim frekvenčnim seznamom prilagojeni frekvenčni seznam krajši, saj vsebuje **manjše število različnic** oz. potencialnih večbesednih enot (25.261 v primerjavi s 27.465), kar pomeni, da je bila v postopku prilagajanja pogostost tistih n-gramov, ki so se vedno pojavljali zgolj kot del daljših besednih nizov in nikoli kot samostojne enote besedišča, zreducirana na manj kot pet pojavitev. Analiza najpogostejših dvobesednih nizov na seznamu 981 besednih nizov, ki jim je bila v postopku prilagajanja frekvenca zreducirana celo na nič pojavitev, denimo razkriva, da med njimi

4.2 Metode prepoznavanja večbesednih diskurznihih označevalcev v besedilnih korpusih

prednjačijo gradniki predložnih zvez (npr. *po/na eni strani*, v *zvezi z*, na *kakšen način*, v *smislu da*) in drugih stalnih zvez (npr. *se mi/ti zdi*, *dame in gospodje*) ter stavčnih vzorcev (npr. *je bilo še*, *če se ne motim*).

Poleg realnejšega seznama različnic v korpusu pa prilagojeni frekvenčni seznam prinaša tudi **ustreznejše štetje njihovih pojavitev** (pogostosti), saj upošteva, da lahko določeno mesto v korpusu zaseda zgolj ena enota besedišča (n-gram).⁹⁴ Kot prikazujejo podatki v Tabeli 5, imajo nizi na prilagojenem besednem seznamu v povprečju pričakovano manjše število pojavitev, zanimivo pa je, da se je pri tem zmanjšala predvsem pogostost dvobesednih nizov (22 proti 15 pojavitev na različnico), pri daljših nizih pa je razlika v pogostosti bistveno manjša, kar pomeni, da se v korpusu veliko dvobesednih nizov pojavlja kot del pogostih tribesednih nizov, medtem ko slednji največkrat predstavljajo samostojne enote besedišča, ne pa delov daljših (štiribesednih) nizov. V primerjavi z neprilagojenim frekvenčnim štetjem je tako slika distribucije večbesednih nizov glede na njihovo dolžino realnejša, saj relativizira prepad med pogostostjo dvobesednih in daljših besednih nizov, kakršnega sugerira neprilagojeno štetje.

št. besed	začetni frekvenčni seznam za nize z vsaj 5 pojavitvami			prilagojeni frekvenčni seznam za nize z vsaj 5 pojavitvami		
	različnic	pojavnic	poj./razl.	različnic	pojavnic	poj./razl.
2	18.923	417.283	22,1	17.444	268.637	15,4
3	7.428	86.971	11,7	6.872	70.665	10,3
4	971	8.929	9,2	839	6.961	8,3
5	116	1.193	10,3	79	676	8,56
6	27	334	12,4	27	334	12,4
Skupaj	27.465	514.710	18,7	25.261	347.273	13,7

Tabela 5: Primerjava štetja pojavnic in različnic za n-grame v korpusu Gos pred in po implementaciji algoritma za izdelavo prilagojenega frekvenčnega seznama.

⁹⁴ Čeprav niz ni štet, kadar se pojavlja kot del pogostega daljšega niza, pa to ne pomeni, da lahko vsak indeks v korpusu pripada zgolj eni enoti besedišča (besedi ali besednemu nizu). Smisel prilagojenega frekvenčnega seznama je namreč predvsem to, da reduciramo pogostost krajših nizov, kadar se pojavijo kot del daljših nizov, ne pa štetje, ki bi vsako pojavnico v korpusu pripisalo zgolj eni enoti. V nizu [ne [zaradi tega ker] je] niz 'zaradi tega ker' sicer ni bil štet, ker se pojavi znotraj niza 'ne zaradi tega ker', vendar algoritem ne skuša razdvoumiti med hkratnima pojavitvama nizov 'ne zaradi tega ker' in 'zaradi tega ker je', ki oba presegata želeni frekvenčni prag.

Bistveno bolj kot primerjava števila izluščenih nizov in števila njihovih pojavitev pa je za namen naše raziskave pomembno dejstvo, da oba vidika vplivata tudi na **drugačno razvrščanje** besednih nizov po pogostosti. Primerjava prvih 100 besednih nizov na začetnem in prilagojenem frekvenčnem seznamu (Priloga 3) namreč kaže, da je 73 % enot enakih na obeh seznamih, spremenilo pa se je njihovo zaporedje, saj je bil v procesu prilagajanja pogostosti zmanjšana frekvenca tistim besedam, ki se pojavljajo tudi kot del pogostih daljših nizov, s čimer na vrh seznama prodirajo tisti besedni nizi, ki večinoma nastopajo v poljubnih besednih kombinacijah in torej delujejo kot samostojne enote besedišča. Med njimi prevladujejo zlasti izrazi z diskurzno ali pragmatično vlogo (npr. *dobro jutro, ali kaj, tako je, in potem, pa tako, zdaj pa, a veš, se pravi, zato ker*).

Med 27 enotami, ki po prilagajanju frekvence niso več del seznama najpogostejših 100 besednih nizov korpusa Gos, kakršnega smo pridobili z običajnim štetjem nizov, najdemo predvsem kombinacije s pomožnimi glagoli (npr. *ne bi, kako je, bi bilo*), kazalnimi zaimki (npr. *to pa, ja to, to da*) in drugimi funkcijskimi besedami (npr. *da ne, se mi, se pa*). Med 27 enotami, ki so po drugi strani postale del najpogostejših 100 enot šele po prilagoditvi frekvence, prav tako prevladujejo nepropozicijske večbesedne enote (npr. *a res, ja no, ne samo, na primer, v glavnem, in tako naprej, ali ne, hvala lepa*), ponavljanja (npr. *mhm mhm, aha aha, ja ja ja ja, v v, na na*), pa tudi stavčni fragmenti različnih tipov (npr. *tudi v, so se, ki ga, so bili*).

4.2.2.3 Mere kolokabilnosti

Kot lahko razberemo iz seznama najpogostejših besednih nizov v korpusu Gos (Prilogi 2 in 3) pa ne glede na način njihovega štetja med njimi vendarle **še vedno prevladujejo strukturno nezaključeni nizi** s pomožnimi glagoli, vezniki, členki in drugimi pogostimi funkcijskimi besedami, ki jim težko pripišemo status psihološko izstopajočih samostojnih enot ali, v primeru naše konkretne raziskave, status diskurznihi označevalcev. V tretjem koraku smo zato za vse relevantne nize, izluščene v prvem koraku (poglavje 4.2.2.1 *Luščenje in običajno štetje besednih nizov*) izračunali še stopnjo povezovalnosti med vsebovanimi besedami, ki naj bi z izračunom verjetnosti sopojava itve posameznih besed v dani kombinaciji glede na verjetnost pojava itve besed v korpusu nasploh na vrh seznama potiskala izstopajoče besedne kombinacije.

Kot smo omenili že v splošnem opisu izbranega metodološkega postopka, smo za ta namen **izbrali šest različnih mer povezovalnosti**, ki na tak ali drugačen način sopostavljajo informacije o pogostosti besednega niza, pogostosti posameznih vsebovanih besed, včasih pa tudi informacijo o sami velikosti korpusa, in sicer izračun vzajemne vrednosti (MI), izračun kubične vzajemne vrednosti (MI³), izračun logaritma verjetja (LL), izračun koeficienta Dice ter izračun mer χ^2 in t-test. Za njihovo posplošitev z dvobesednih enot na večbesedne enote poljubne dolžine smo uporabili metodo pravične porazdelitve (J. F. da Silva in Lopes, 1999), ki v svojem bistvu deluje tako, da niz poljubne dolžine najprej pretvori v vse možne bigrame oz. psevdobigrame (npr. petbesedni niz *abcd* v psevdobigrame *a*bcde*, *ab*cde*, *abc*de* in *abcd*e*, pri čemer zvezdica označuje mesto preloma na dva soležna niza), nato pa pri izračunu kolokacijske moči celotnega niza kot frekvenco prve enote (Avx) vzame aritmetično povprečje frekvenc vseh možnih levih bigramov (npr. *f(a)*, *f(ab)*, *f(abc)* in *f(abcd)*), kot frekvenco druge enote (Avy) aritmetično povprečje frekvenc vseh desnih bigramov (npr. *f(bcde)*, *f(cde)*, *f(de)* in *f(e)*), kot produkt verjetnosti obeh enot (Avp) pa aritmetično povprečje produktov verjetnosti vsakega para desnega in levega bigrama (npr. *p(a)·p(bcde)*, *p(ab)·p(cde)*, *p(abc)·p(de)* in *p(abcd)·p(e)*).

Seznam izhodiščnih mer in njihovih prilagoditev za večbesedne enote poljubne dolžine na podlagi ponazoritve in izpeljave enačb v članku (J. F. da Silva in Lopes, 1999) prikazuje Tabela 7, pri čemer N označuje število vseh pojavnic v korpusu, *f(w₁...w_n)* pogostost niza dolžine n besed v celotnem korpusu in *p(w₁...w_n)* verjetnost pojavitve tega niza v korpusu, tj. količnik razmerja med pogostostjo in velikostjo korpusa (*p(w₁...w_n) = f(w₁...w_n) / N*). Enačbe za izračun vseh treh omenjenih aritmetičnih povprečij (Avp, Avx, Avy) so podane v Tabeli 6.

A	$Avx = \frac{1}{n-1} \sum_{i=1}^{i=n-1} f(w_1 \dots w_i)$
B	$Avy = \frac{1}{n-1} \sum_{i=2}^{i=n} f(w_i \dots w_n)$
C	$Avp = \frac{1}{n-1} \sum_{i=1}^{i=n-1} p(w_1 \dots w_i) \cdot p(w_{i+1} \dots w_n) = \frac{1}{n-1} \sum_{i=1}^{i=n-1} \frac{f(w_1 \dots w_i) \cdot f(w_{i+1} \dots w_n)}{N^2}$

Tabela 6: Načini povprečenja statističnih podatkov prve in druge besede za nize dolžine n besed.

	MI ⁹⁵	
	izhodiščna mera	prilagoditev za nize poljubne dolžine
1	$MI(x, y) = \log \frac{p(x, y)}{p(x) \cdot p(y)} = \log \frac{f(x, y) \cdot N}{f(x) \cdot f(y)}$	$MI(w_1 \dots w_n) = \log_2 \frac{p(w_1 \dots w_n)}{Avp}$ $= \log_2 \frac{f(w_1 \dots w_n)}{Avp \cdot N}$
	MI ³	
	izhodiščna mera ⁹⁶	prilagoditev za nize poljubne dolžine
2	$MI3(x, y) = \log_2 \frac{f(x, y)^3 \cdot N}{f(x) \cdot f(y)}$	$MI3(w_1 \dots w_n) = \log_2 \frac{f(w_1 \dots w_n)^3}{Avp \cdot N}$
	χ^2	
	izhodiščna mera	prilagoditev za nize poljubne dolžine
3	$\chi^2(x, y) = \frac{[f(x, y) \cdot N - f(x) \cdot f(y)]^2}{f(x) \cdot f(y)(N - f(x)) \cdot (N - f(y))}$	$\chi^2(w_1 \dots w_n) = \frac{[f(w_1 \dots w_n) \cdot N - Avp]^2}{Avp \cdot (N - Avx) \cdot (N - Avy)}$ popravljen prilagoditev: ⁹⁷

⁹⁵ V dejanskem prenosu enačbe za izračun vrednosti MI in MI3 v računalniško skripto smo namesto desetiškega logaritma uporabili logaritem z osnovo 2, ki je v korpusnem jezikoslovju bolj uveljavljen (glej npr. SketchEngine, Wei in Li 2013), vendar pa ta odločitev ne vpliva na same rezultate, saj je razvrščanje nizov ne glede na način logaritmiranja enako.

⁹⁶ Enačba za izračun mere MI³, ki ni del izhodiščnega članka (J. F. da Silva in Lopes, 1999), je povzeta po: <https://www.sketchengine.co.uk/wp-content/uploads/ske-stat.pdf>

⁹⁷ Za razliko od vseh drugih mer članek pri prilagoditvi mere χ^2 kot Avp navede zgolj aritmetično sredino vseh zmnožkov levih in desnih frekvenc, tj. $f(x)f(y)$, ne pa zmnožkov njihovih verjetnosti, tj. delitve z vrednostjo $f(x)f(y)/N^2$, kot to prikazuje vrstica C v Tabeli 6. Zavaljo konsistentnosti smo sami tudi v tej enačbi uporabili enako izpeljavo vrednosti Avp kot v vseh drugih, ustreznost naše odločitve pa dodatno potrjuje tudi dejstvo, da bi po prilagoditvi, ki jo predlaga članek, vrednost χ^2 označevalca *zaradi tega da* (glej primer v nadaljevanju) znašala 36857761343, če pa za njen izračun uporabimo popravljen enačbo, ta vrednost znaša 0,036. Pri tem poudarimo, da različni načini izračuna vplivajo zgolj na pripisano vrednost (rezultat enačbe), ne pa na način razvrščanja nizov, saj razmerja v njihovi kolokabilnosti ne glede na izbiro ene ali druge enačbe ostajajo nespremenjena.

4.2 Metode prepoznavanja večbesednih diskurzivnih označevalcev v besedilnih korpusih

		$\chi^2(w_1 \dots w_n) = \frac{[f(w_1 \dots w_n) \cdot N - Avp \cdot N^2]^2}{Avp \cdot (N - Avx) \cdot (N - Avy)}$
LL		
	izhodiščna mera	prilagoditev za nize poljubne dolžine
4	$\begin{aligned} \loglike(x, y) &= 2 \cdot (\logl(p1, k1, n1) \\ &\quad + \logl(p2, k2, n2) \\ &\quad - \logl(p, k1, n1) \\ &\quad - \logl(p, k2, n2)) \end{aligned}$ pri čemer $\logl(P, K, M) = K \cdot \ln(P) + (M - K) \cdot \ln(1 - P)$ in $k1 = f(x, y) \quad k2 = f(x) - f(x, y)$ $n1 = f(y) \quad n2 = N - f(y)$ $p1 = \frac{f(x, y)}{f(y)} \quad p2 = \frac{f(x) - f(x, y)}{N - f(y)}$ $p = \frac{f(x)}{N}$	$\begin{aligned} \loglike(w_1 \dots w_n) &= 2 \cdot (\logl(pf1, kf1, nf1) \\ &\quad + \logl(pf2, kf2, nf2) \\ &\quad - \logl(pf, kf1, nf1) \\ &\quad - \logl(pf, kf2, nf2)) \end{aligned}$ pri čemer $\logl(P, K, M) = K \cdot \ln(P) + (M - K) \cdot \ln(1 - P)$ in $kf1 = f(w_1 \dots w_n) \quad kf2 = Avx - f(w_1 \dots w_n)$ $nf1 = Avy \quad nf2 = N - Avy$ $pf1 = \frac{f(w_1 \dots w_n)}{Avy} \quad pf2 = \frac{Avx - f(w_1 \dots w_n)}{N - Avy}$ $pf = \frac{Avx}{N}^{98}$
Dice		
	izhodiščna mera	prilagoditev za nize poljubne dolžine
5	$Dice(x, y) = \frac{2 \cdot f(x, y)}{f(x) + f(y)}$	$Dice(w_1 \dots w_n) = \frac{2 \cdot f(w_1 \dots w_n)}{Avx + Avy}$
t-test		

⁹⁸ Vrednost pf je v članku izpeljana napačno, saj v navedeni izpeljavi $pf = (kf1 + kf2)/N = Avx$ imenovalci N ni bil prenesen v naslednji korak.

	izhodiščna mera ⁹⁹	prilagoditev za nize poljubne dolžine
6	$t_score(x, y) = \frac{f(x, y) - p(x, y)}{\sqrt{f(x, y)}}$ $= \frac{f(x, y) - \frac{f(x) \cdot f(y)}{N}}{\sqrt{f(x, y)}}$	$t_score(w_1 \dots w_n) = \frac{f(w_1 \dots w_n) - Avp \cdot N}{\sqrt{f(w_1 \dots w_n)}}$

Tabela 7: Seznam uporabljenih mer povezovalnosti in njihova generalizacija za večbesedne enote poljubne dolžine (dolžine n besed).

Za lažjo predstavo izračun stopnje povezovalnosti ponazorimo na primeru večbesednega diskurznega označevalca *zaradi tega ker*, ki se v korpusu Gos (velikosti 1.005.455 pojavnic) pojavi 187-krat in ga lahko razdelimo v dva psevdobigrama: '*zaradi*' + '*tega ker*' ter '*zaradi tega*' + '*ker*'. Povprečje vseh možnih levih bigramov (Av_x) izračunamo z aritmetičnim povprečjem seštevka pogostosti leve enote prvega psevdobigrama, tj. '*zaradi*' (783), in leve enote drugega psevdobigrama, tj. '*zaradi tega*' (371), torej $Av_x = (783 + 371) / 2 = 577$. Povprečje desnih bigramov (Av_y) dobimo tako, da izračunamo aritmetično povprečje seštevka pogostosti desne enote prvega bigrama, tj. '*tega ker*' (193), in desne enote drugega psevdobigrama, tj. '*ker*' (4.779), torej $Av_y = (193 + 4.779) / 2 = 2.486$. Povprečje produktov verjetnosti levih in desnih bigramov dobimo tako, da izračunamo aritmetično povprečje seštevka produkta verjetnosti leve in desne enote prvega bigrama, tj. '*zaradi*' in '*tega ker*', torej $(783 / 1.005.455) \cdot (193 / 1.005.455) = 1,49484E-07$, ter produkta verjetnosti leve in desne enote psevdobigrama '*zaradi tega*' in '*ker*', torej $(371 / 1.005.455) \cdot (4.779 / 1.005.455) = 1,75382E-06$, s čimer torej dobimo povprečje produktov $Av_p = (1,49484E-07 + 1,75382E-06) / 2 = 9,51653E-07$. Ko te vrednosti nato vnesemo v zgornje enačbe, za niz '*zaradi tega ker*' dobimo podatek o kolokacijski moči, in sicer: $MI \approx 7,61$, $\chi^2 \approx 0,036$, $LL \approx$, $Dice \approx 0,122$, $t_score \approx 13,6$, $DP \approx 0,0748$.

Tovrstni izračun s pomočjo računalniške skripte ponovimo še za vse ostale besedne nize izhodiščnega seznama, da dobimo šest različnih načinov njihovega razvrščanja glede na stopnjo povezanosti posameznih besed. Primerjava 20 najmočnejših oziroma najvišje uvrščenih kolokacij

⁹⁹ Enačba za izračun vrednosti t-test, ki ni del izhodiščnega članka (J. F. da Silva in Lopes, 1999) je povzeta po: <https://www.sketchengine.co.uk/wp-content/uploads/ske-stat.pdf>

za vsako izmed šestih načinov izračuna te lastnosti v Tabeli 8 potrjuje, da tovrstne mere kot statistično izstopajoče nize predlagajo drugačen nabor najrelevantnejših nizov kot običajni ali prilagojen seznam pogostnosti (Tabela 6), razvrstitve nizov glede na stopnjo kolokabilnosti pa se med seboj tudi razlikujejo, pri čemer mere MI, X2 in Dice favorizirajo manj pogoste predmetnopomenske zveze, kot npr. lastna imena, mere MI3, LL in t-test pa pogoste funkcijske zveze, med katerimi se poleg pragmatičnih formul in stavčnih fragmentov pojavljajo tudi nekateri diskurzni označevalci (podčrtano in odebeljeno).

	MI	frek.	MI ³	frek.	χ^2	frek.	LL	frek.	Dice	frek.	t-test	frek
1	savage garden	5	<u>ne vem</u>	2743	robbie williams	7	<u>ne vem</u>	2743	mata hari	13	<u>to je</u>	3853
2	nelly furtado	5	<u>to je</u>	3853	mata hari	13	<u>v bistvu</u>	985	bruce springsteen	11	<u>ne vem</u>	2743
3	gnarls barkley	5	ja ja	3420	bruce springsteen	11	<u>to je</u>	3853	robbie williams	7	ja ja	3420
4	cyndi lauper	5	dobro jutro	440	joe cocker	5	je bilo	1681	joe cocker	5	je bilo	1681
5	joe cocker	5	<u>v bistvu</u>	985	cyndi lauper	5	<u>v redu</u>	833	cyndi lauper	5	<u>a ne</u>	1836
6	oton župančič	5	jaz sem	1316	gnarls barkley	5	jaz sem	1316	gnarls barkley	5	<u>tako da</u>	1804
7	true blood	5	hvala lepa	225	nelly furtado	5	ja ja	3420	nelly furtado	5	da je	2697
8	sean connery	5	je bilo	1681	savage garden	5	<u>se pravi</u>	832	savage garden	5	jaz sem	1316
9	žabice in rege	6	<u>v redu</u>	833	premierni hitri kviz	14	<u>a ne</u>	1836	premierni hitri kviz	14	da bi	1438
10	robbie williams	7	<u>zaradi tega</u>	371	žabice in rege	6	<u>a veš</u>	839	lindsay wonn	12	<u>ali pa</u>	1358
11	proud mary	6	<u>a veš</u>	839	lindsay wonn	12	<u>tako da</u>	1804	žabice in rege	6	je bil	1176
12	fabienne sutter	6	<u>a ne</u>	1836	denise carbon	11	je bil	1176	denise carbon	11	da se	1539
13	amy macdonald	6	dober večer	143	freddie mercury	9	dobro jutro	440	freddie mercury	9	<u>v bistvu</u>	985
14	šmarje pri jelšah	5	<u>tako da</u>	1804	majdo juvan	8	mislim da	1038	majdo juvan	8	mislim da	1038
15	monte carlo	5	<u>se pravi</u>	832	tesa worley	8	da bi	1438	tesa worley	8	je bila	1031
16	whitney houston	6	da je	2697	fabienne sutter	6	<u>ali pa</u>	1358	fabienne sutter	6	ja ja ja	999
17	onkološkem inštitutu	6	javnem sektorju	33	proud mary	6	<u>s tem</u>	597	proud mary	6	<u>v redu</u>	833
18	taina barioz	5	<u>s tem</u>	597	amy macdonald	6	je bila	1031	amy macdonald	6	<u>a veš</u>	839
19	nadia fanchini	5	<u>zato ker</u>	508	jan plestenjak	17	<u>zato ker</u>	508	jan plestenjak	17	<u>se pravi</u>	832
20	andrea fischbacher	5	lep pozdrav	73	oton župančič	5	<u>zaradi tega</u>	371	oton župančič	5	je to	1867

Tabela 8: Primerjava 20 najvišje uvrščenih besednih nizov glede na izračun besedne povezovalnosti z merami MI, MI³, χ^2 , LL, Dice in t-test.

Prikazani spiski peščice najvišje uvrščenih večbesednih enot glede na različne statistične vrednosti, kot sta pogostost pojavljanja ali kolokacijska moč vsebovanih besed, so torej že nakazali, da različne statistične metode za prepoznavanje večbesednih enot dajejo različne sezname potencialnih kandidatov, v nadaljevanju pa njihovo dejansko primerljivost in primernost za luščenje diskurznofunkcijske leksike ovrednotimo še z analizo večjega vzorca, in sicer za 2.000 najvišje uvrščenih enot vsakega izmed spiskov.

4.2.2.4 Prekrivnost seznamov najvišje uvrščenih kandidatov

Primerjava prekrivnosti 2.000 najvišje uvrščenih večbesednih enot v Tabeli 9 kaže, da je **razlika v primerljivosti predlaganih kandidatov precejšnja**, od kar 1.654 (82,7 %) prekrivnih kandidatov med vrednostma MI in χ^2 do zgolj 53 (2,65 %) prekrivnih kandidatov med vrednostjo MI in prilagojeno pogostostjo (PP). Čeprav se tovrstne primerjave za vsak posamezni par statističnih vrednosti precej razlikujejo, jih vendarle lahko glede na podobnost umestimo v nekaj komplementarnih skupin. V prvo denimo spadajo običajni seznam pogostosti (OP), prilagojeni seznam pogostosti (PP) in mera t-test, ki se precej razlikujejo od druge skupine mer z medsebojno razmeroma sorodnimi spiski, tj. merami MI, Dice in χ^2 . V slednjo skupino se delno umešča tudi mera MI³, ki pa ima za razliko od drugih treh mer precej visok delež prekrivnih kandidatov tudi z merami iz prve skupine. Kot tretja skupina izrazito izstopa mera LL, ki si sicer z vsako izmed ostalih mer deli vsaj 20 % prekrivnih kandidatov, a hkrati z nobeno ne več kot 51 %.

Unikatnost mere LL potrjuje tudi primerjava deleža večbesednih enot, ki se pojavljajo zgolj na enem izmed spiskov, saj se kar 41 % (819) kandidatov na seznamu 2.000 enot z najvišjo vrednostjo LL ne pojavlja na nobenem izmed drugih sedmih seznamov. Po drugi strani je mera MI³ s tega vidika precej manj informativna, saj se vsi njeni kandidati pojavijo tudi na vsaj enem izmed drugih spiskov, zlasti na seznamih mer χ^2 in t-test. Zanimivo je v celotnem naboru **5.813 potencialnih kandidatov večbesednih enot** (tj. uniji 2.000 najvišje uvrščenih kandidatov na vsakem izmed osmih seznamov) le 41 takih, ki se pojavljajo na vseh osmih seznamih, pri čemer se na vseh osmih seznamih med najvišje 1.000 uvrščenimi kandidati presenetljivo pojavita zgolj dve besedni zvezi, *lep pozdrav* in *dober večer*, vse ostale potencialne večbesedne enote pa vsaj na enem izmed seznamov ne presegajo praga prvih 1.000 uvrščenih enot.

4.2 Metode prepoznavanja večbesednih diskurznihih označevalcev v besedilnih korpusih

	OP	PP	MI	MI ³	χ^2	LL	Dice	t-test	Unikaten
OP		<u>1650</u>	<u>63</u>	716	284	659	352	<u>1214</u>	123
PP	1650		<u>53</u>	655	255	597	320	<u>1160</u>	182
MI	63	53		1217	<u>1654</u>	409	<u>1450</u>	<u>193</u>	271
MI ³	716	655	1217		<u>1532</u>	<u>1014</u>	<u>1480</u>	935	0
χ^2	284	255	1654	1532		702	<u>1581</u>	497	16
LL	659	597	409	1014	702		726	904	819
Dice	352	320	1450	1480	1581	726		541	171
t-test	1214	1160	193	935	497	904	541		350

Tabela 9: Delež prekrivnih kandidatov na seznamih 2.000 najvišje uvrščenih večbesednih enot za vsakega izmed različnih statističnih izračunov: OP = običajna pogostost, PP = prilagojena pogostost. V zadnjem stolpcu je navedeno število kolokacij, ki se pojavljajo zgolj na seznamu dane mere.

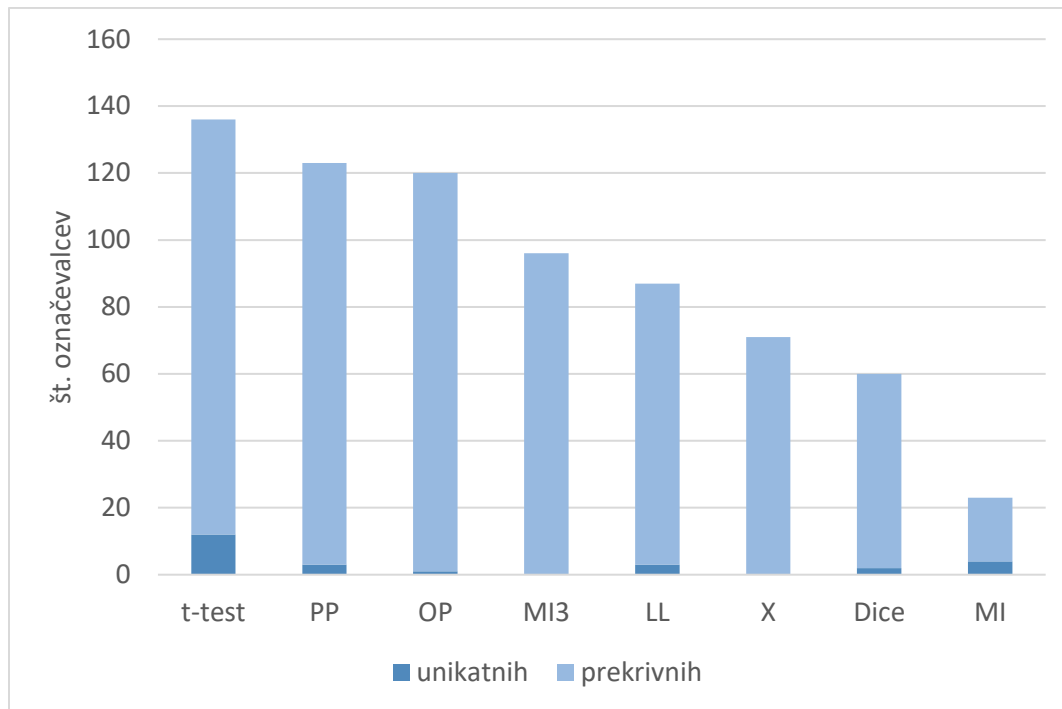
Predstavljena hitra primerjava prekrivnosti seznamov potencialnih večbesednih enot, kakršne dobimo z različnimi statističnimi izračuni njihove ustaljenosti, zelo jasno dokazuje, da njihova **izbira za potrebe pridobivanja seznamov večbesednih enot ni trivialna metodološka odločitev**, saj lahko različne mere kažejo povsem drugačno sliko najrelevantnejših večbesednih enot (Evert, 2009; Gries, 2010). Čeprav tako slovenske kot tuje raziskave večbesednih enot le redko argumentirajo izbiro uporabljene mere in se običajno zadovoljijo z analizo kolokabilno najmočnejših enot ene same mere, npr. MI³ (Gantar, 2007) ali logDice (Logar idr., 2014; Ljubešić idr., 2015; Pollak, 2015a; Gantar idr., 2016), pričujoča primerjava kaže, da njihova izbira za potrebe luščenja besednih kolokacij v korpusih slovenskega jezika zahteva skrbnejši začetni premislek, prav tako pa je za realno oceno njihove komplementarnosti oz. prekrivnosti potrebno primerjati daljši seznam potencialnih kandidatov kot le nekaj deset najvišje uvrščenih enot (prim. Tabeli 8 in 9). V primeru uporabe statističnih mer za pridobivanje seznamov slabše raziskanih skupin večbesednih enot, kot so diskurzni označevalci, pa ti rezultati obenem tudi empirično potrjujejo **upravičenost analize unije rezultatov več različnih statističnih metod**, ki jo opišemo in ovrednotimo v nadaljevanju.

4.2.3 Primerjava metod prepoznavanja večbesednih diskurznihi označevalcev

4.2.3.1 Primerjava statističnih metod

V prvem koraku analize izbranihi metode prepoznavanja večbesednih diskurznihi označevalcev v korpusu govornje slovenščine smo želeli izvedeti, katera izmed izbranihi metod statističnega razvrščanja večbesednih enot daje najboljše rezultate pri pridobivanju seznama predstavnikov obravnavane kategorije. Splošna primerjava skupnega števila prepoznanihi diskurznihi označevalcev med 2.000 najvišje uvrščenimi enotami vsake izmed mer (Slika 4) kaže, da smo **največ diskurznihi označevalcev prepoznali¹⁰⁰ s pomočjo mere t-test** (136), ki daje precej podobne rezultate kot razvrščanje nizov glede na pogostost, ne glede na to, ali smo za izračun njihove pogostosti uporabili običajno štetje (120) ali prilagojeno štetje (123). Bistveno manj označevalcev je bilo prepoznanihi z merami MI³ (96), LL (87), X² (71) in Dice (60), daleč najslabše rezultate pa daje mera MI (23 prepoznanihi označevalcev med 2.000 pregledanimi enotami). Z najboljšo statistično metodo t-test smo tako pridobili 78,6 % vseh **173 označevalcev**, prepoznanihi s kvantitativnimi metodami razvrščanja kolokacij, ki jih sicer v celoti skupaj s podatkom o temeljni funkcijski skupini ter pripadajočih statističnih vrednostih navajamo v Prilogi 4. Med 37 kvantitativno prepoznanimi označevalci, ki glede na mero t-test niso bili uvrščeni med najmočnejših 2.000 kolokacij, smo jih največ pridobili s kombinacijo pregleda seznama najpogostejših nizov z običajnim oz. prilagojenim štetjem (18 večinoma prekrivnih označevalcev) ter merami MI (14 označevalcev), LL (3 označevalci) in Dice (2 označevalca), meri X² in MI³ pa ne prinašata novih informacij, saj na teh dveh seznamih ni unikatnih označevalcev, ki jih ne bi pokrivala že druge mere.

¹⁰⁰ Na podlagi pregleda prvih 100 naključnih korpusnih konkordanc smo vsako izmed večbesednih enot kategorizirali glede na to, ali se v rabi pojavlja kot diskurzni označevalec ali ne, pri čemer smo niz kot relevanten prepoznali tudi, če se v rabi pojavlja tako v diskurzni kot drugih vlogah.

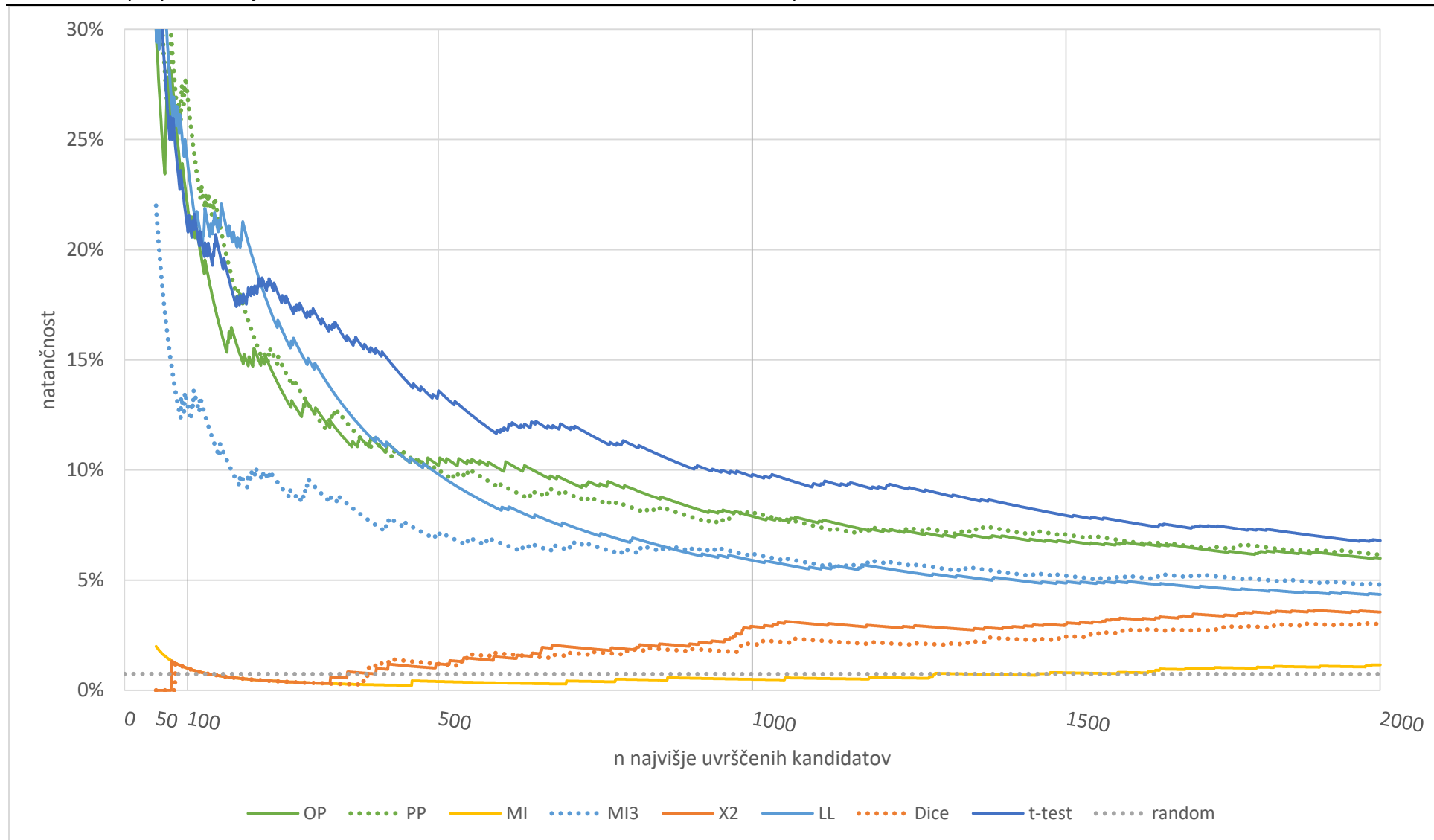


Slika 4: Število diskurznihih označevalcev med 2.000 najvišje uvrščenimi kandidati izbrane statistične mere.

Zgornja primerjava števila prepoznanihih diskurznihih označevalcev na arbitrarno izbranem seznamu najvišje 2.000 uvrščenih enot nadalje odpira vprašanje, kakšna je natančnost posamezne izmed mer ne glede na izbrani interval pregledanihih enot. Rezultate prikazujemo v obliki t. i. grafa natančnosti (angl. *precision plot*) na Sliki 5, ki prikazuje **delež diskurznihih označevalcev za poljubno število najvišje uvrščenih enot** (do izbranega praga 2.000 enot). Pri tem kot deveto mero vključujemo tudi prikaz natančnosti naključnega izbora kandidatov, torej delež (znanihi) diskurznihih označevalcev na celotnem seznamu nizov (tj. 204 od 27.465 nizov oz. 0,74 %).¹⁰¹ Evalvacija metod je namreč smiselna le, če jih primerjamo z naključnim naborom kandidatov, saj je mera povezovalnosti uporabna le v toliko, kolikor daje pomembno boljše rezultate, kot bi jih dobili s pregledovanjem naključnihih kandidatov (Evert, 2009).

¹⁰¹ Kot število vseh znanihih diskurznihih označevalcev jemljemo unijo označevalcev, ki so bili prepoznani bodisi z analizo statističnihih seznamov (173) bodisi z ročno analizo vzorčenega korpusa (125), skupaj 204 različni diskurznihih označevalci (glej tudi poglavje 4.2.3.2. Primerjava statističnihih metod in ročne prepoznavne).

4.2 Metode prepoznavanja večbesednih diskurznih označevalcev v besedilnih korpusih



Slika 5: Primerjava natančnosti prepoznavanja večbesednih diskurznih označevalcev v korpusu Gos za različne mere, tj. primerjava deleža večbesednih diskurznih označevalcev na seznamih n najvišje uvrščenih večbesednih enot vsake izmed mer.

Graf natančnosti za različne mere razkriva več presenetljivih ugotovitev. Med splošnimi ugotovitvami primerjave izstopa predvsem dejstvo, da **se rezultati primerjave natančnosti metod razlikujejo glede na izbrani interval analiziranih nizov**, saj primerjava deleža diskurznihi označevalcev v označenem intervalu prvih 100 enot (tj. 100 statistično najbolj izstopajočih nizov) denimo kaže povsem druga razmerja kot primerjava deleža diskurznihi označevalcev na (realnejšem) večjem vzorcu, kot je npr. prvih 2.000 enot (Slika 4). Če bi denimo pod drobnogled vzeli le prvih 100 enot vsakega seznama, bi na seznamih najpogostejših nizov (OP), nizov s prilagojenim štetjem (PP), nizov z največjo mero t-test in LL našli primerljivo število označevalcev, tj. 21 (t-test), 22 (OP), 24 (LL) oz. 27 (PP) vseh statistično relevantnih enot,¹⁰² medtem ko na večjem vzorcu enot mera LL daje bistveno slabše rezultate kot druge tri mere (Slika 5). To potrjuje sklep, ki smo ga izpostavili že ob splošni primerljivosti posameznih statističnih mer, da o ustreznosti izbire posamezne mere za opazovani tip večbesednih enot, jezik, velikost korpusa ipd. ne moremo sklepati zgolj na podlagi primerjave peščice najvišje uvrščenih kandidatov (prim. Tabela 9) oz. je izbira najprimernejše mere odvisna tudi od nameravanega števila analiziranih kandidatov. V nadaljevanju kot orientacijsko točko primerjave jemljemo označeni interval 1.000 enot, saj razmerja med merami po tem prelomu ostajajo večinoma nespremenjena.

Nadaljnja podrobnejša analizi celotnega grafa natančnosti potrjuje začetno trditev, da med vsemi merami **največjo natančnost pri prepoznavanju (večbesednih) diskurznihi označevalcev dosega mera t-test**, pri kateri denimo v označenem intervalu 1.000 najvišje uvrščenih enot (enot z najmočnejše povezanimi besedami) diskurznihi označevalci predstavljajo 9,8 % vseh enot, medtem ko natančnost analize najpogostejših nizov (OP) brez posebnih izračunov notranje povezovalnosti pri istem pragu dosega 7,9 %, kar z drugimi besedami pomeni, da smo s pregledom najvišje uvrščenih 1.000 kandidatov glede na t-test pridobili 98, s pregledom seznama po pogostosti pa 79 diskurznihi označevalcev. Pri prepoznavanju večbesednih diskurznihi označevalcev izračun kolokabilnosti nizov z mero t-test torej daje boljše rezultate kot zgolj njihovo razvrščanje po pogosti. Čeprav se na seznamu nizov glede na njihovo prilagojeno štetje

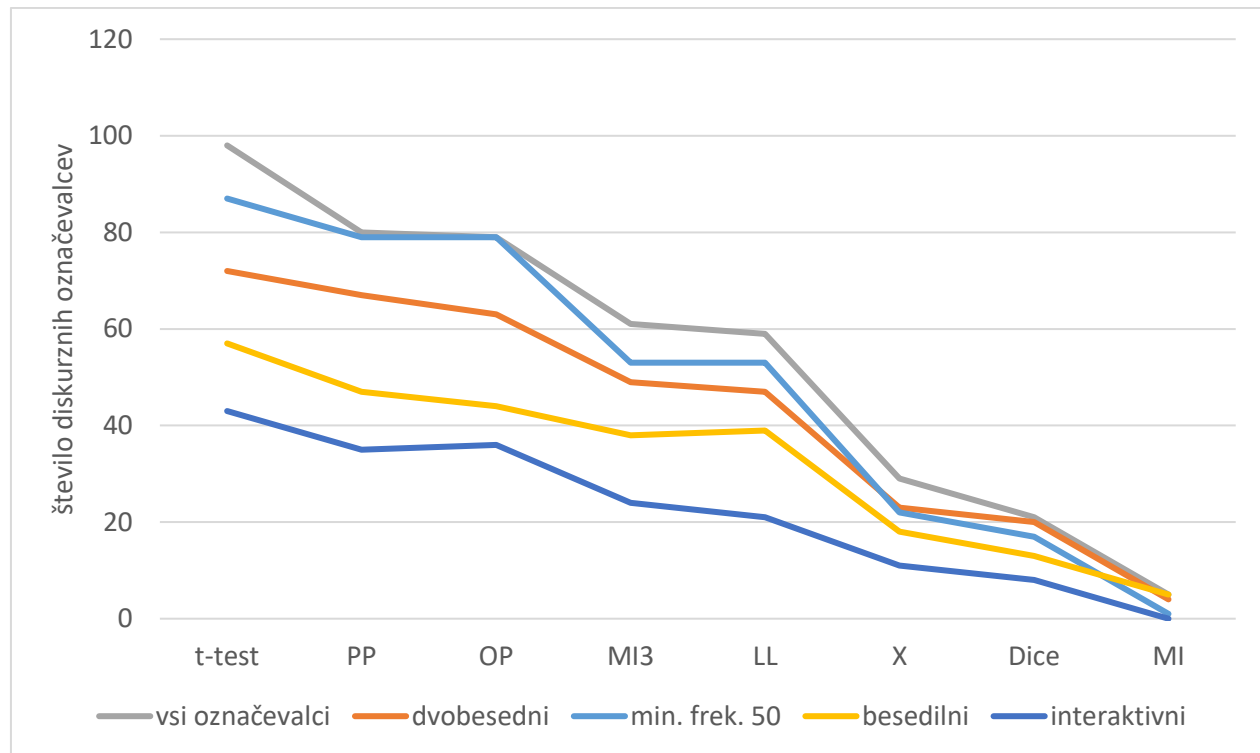
¹⁰² Glede na ugotovljeno maloštevilnost teh jezikovnih sredstev to število označevalcev obenem predstavlja tudi presenetljivo velik delež (statistično relevantnih) večbesednih enot.

(PP) med prvimi 100 kandidati pojavlja nekoliko več diskurznihi označevalcev kot na enakem seznamu nizov, razvrščenih po absolutni pogostosti (OP), pa na večjem vzorcu pregledanih kandidatov metoda prilagojenega štetja pogostosti zanimivo ne ponuja bistvene izboljšave glede na običajno štetje.

Nekoliko slabše rezultate pri prepoznavanju večbesednih diskurznihi označevalcev dajeta meri MI3 in LL (6,1 % oz. 5,9 % diskurznihi označevalcev med prvih 1.000 kandidatov), pri čemer je pri manjšem intervalu prvih sto enot bistveno bolj natančna mera LL (13 % proti 24 %). Izrazito slabše so se pri natančnosti prepoznavanja diskurznihi označevalcev uvrstile mere X2, Dice in MI, za katere smo že pri primerjavi prekrivnosti kandidatov nasploh ugotovili precejšnjo mero podobnosti. Med prvimi 1.000 kandidati mer X2 in Dice se je tako (v nasprotju z natančnostjo, o kakršni priča primerjava števila diskurznihi označevalcev za interval 2.000 enot na Sliki 4) pojavilo le 29 oz. 21 diskurznihi označevalcev, med prvimi 1.000 kandidati mere MI pa le 5 označevalcev. To pomeni, da je **mera MI z vidika pridobivanja večbesednih diskurznihi označevalcev povsem neuporabna**, saj je njena natančnost primerljiva s številom diskurznihi označevalcev, ki bi ga dobili, če med vsemi nizi pregledali enako število povsem naključno izbranihi kandidatov.

Pri tem poudarimo še, da rezultati primerjave niso odvisni od izbranega načina posplošitve statističnihi mer za izračun kolokabilnosti enot, od izbranega frekvenčnega pragu ali od posamezne funkcijske skupine, saj se povsem enaka razmerja kažejo tudi, če deleže diskurznihi označevalcev na seznamih posameznih mer zgolj za dvobesedne enote, zgolj za večbesedne enote z 10 ali več pojavitvami ali zgolj za označevalce medbesedilnih povezav oz. besedilne konektorje na eni strani (tj. označevalci dodajanja, nasprotja, specifikacije, razloga, razvrščanja in rezultata) in označevalce interakcije na drugi (tj. označevalci dvogovora, nedoločnosti in procesov tvorjenja).¹⁰³

¹⁰³ Označevalca *in to* ter *prav tako*, ki spadata tako med besedilne konektorje kot označevalce interakcije, sta šteta dvakrat.



Slika 6: Primerjava števila diskurznihih označevalcev na seznamu vseh 1.000 najvišje rangiranihih večbesednih enot za vse označevalce, dvobesedne označevalce, označevalce s 50 ali več pojavitvami, za besedilne označevalce in interaktivne označevalce, glede na osem različnih mer razvrščanja.

Tako nizka natančnost mere MI pri identifikaciji diskurznofunkcijskih večbesednih enot po eni strani ni presenetljiva, saj je bilo v literaturi že večkrat izpostavljeno, da ta mera favorizira redke dogodke (Oakes, 1998; Evert, 2009; Gries, 2010), torej idiosinkratične kombinacije besed, ki se pojavljajo zgolj v dani kombinaciji, redko pa izven nje, ne glede na siceršnjo pogostost celotnega niza (Oakes 1998, Evert 2009), kar velja tudi za luščenje večbesednih enot v slovenščini (Gorjanc in Krek, 2001; Gorjanc, 2005; Gantar, 2007). Že hitra analiza leksikalne zgradbe prepoznanihih diskurznihih označevalcev (Priloga 5) namreč razkriva, da ti izrazi v slovenščini izkazujejo ravno obratne lastnosti, saj **večinoma vsebujejo tudi sicer pogosto rabljene funkcijske besede**, kot so vezniki ali členki (npr. *a, ali, če, da, ne, kot, in, pa*), predlogi (npr. *pri, s, pred, zaradi*), zaimki (npr. *kaj, kako, neki, tako, to*) in pomožni glagoli (npr. *bi, bom, je*), pa tudi med označevalci s predmetnopomenskimi sestavinami prevladujejo pogosto rabljeni glagoli védenja, rekanja ali mišljenja (npr. *biti, misliti, reči, vedeti*) ali abstraktni samostalniki (*človek, stališče, razlog, način, vidik*). Ker **kar 198 od 204 znanihih označevalcev vsebuje vsaj eno funkcijsko besedo**, torej ni presenetljivo, da smo z mero MI med 2.000 pregledanimi kandidati našli zgolj 23 diskurznihih

označevalcev,¹⁰⁴ saj je bila prednost pri razvrščanju dana idiosinkratičnim kombinacijam polnopomenskih besed, kot so lastna imena (npr. *savage garden*, *oton župančič*), terminološke zveze (*kaznivih dejanj*, *električnega toka*, *natrijev klorid*) in druge predmetnopomenske stalne besedne zveze.

Po drugi strani pa je tako slab rezultat vendarle presenetljiv zato, ker se **prav mera MI pogosto uporablja kot referenčna mera za izračunov besedne povezovalnosti** v raziskavah formulaičnega jezika. Biber (2009) v svoji razpravi o ustreznosti uporabe mer besedne povezovalnosti v raziskavah formulaičnega jezika denimo primerja seznam najvišje rangiranih večbesednih enot, ki ga dobi z luščenjem najpogostejšim štiribesednih enot (imenuje jih večbesedni leksikalni skupi), s seznamom najvišje razvrščenih večbesednih enot, ki ga dobi z izračun vzajemne informacije MI (imenuje jih večbesedne leksikalne kolokacije). Ugotovi, da seznam večbesednih kolokacij vsebuje večinoma terminološke predmetnopomenske izraze, ki jih sestavljajo leksikalne/predmetnopomenske besede, seznam večbesednih skupov pa vsebuje nize, ki vključujejo tako leksikalne kot slovnične besede, na podlagi česar sklene, da kolokacijska moč ni enaka formulaičnemu statusu, zato meni, da za raziskave formulaičnih nizov mere povezovalnosti niso uporabne. Po drugi strani Wei in Li (2013) v svoji evalvacije nove metode za luščenje neprekinjenih frazeoloških nizov iz akademskih korpusov kot referenčno mero uporabljata zgolj mero MI, Ellis s sodelavci (2008) pa mero MI uporablja kot eno izmed spremenljivk v prilagojeni metodi za luščenje formulaičnih izrazov v akademskem diskurzu za potrebe njihovega poučevanja, pri čemer mednje poleg referenčnih izrazov in izrazov naklonskosti umeščajo tudi izraze za organizacijo diskurza (Simpson-Vlach in Ellis, 2010). Čeprav tudi naša raziskava potrjuje, da za priklic diskurznihih označevalcev, ene izmed pogostih funkcijskih podskupin formulaičnih izrazov, običajno štetje po pogostosti daje povsem zadovoljive rezultate, pa vendarle rezultati kažejo, da doprinos mer besedne povezovalnosti, kot je mera t-test, ni zanemarljiv.

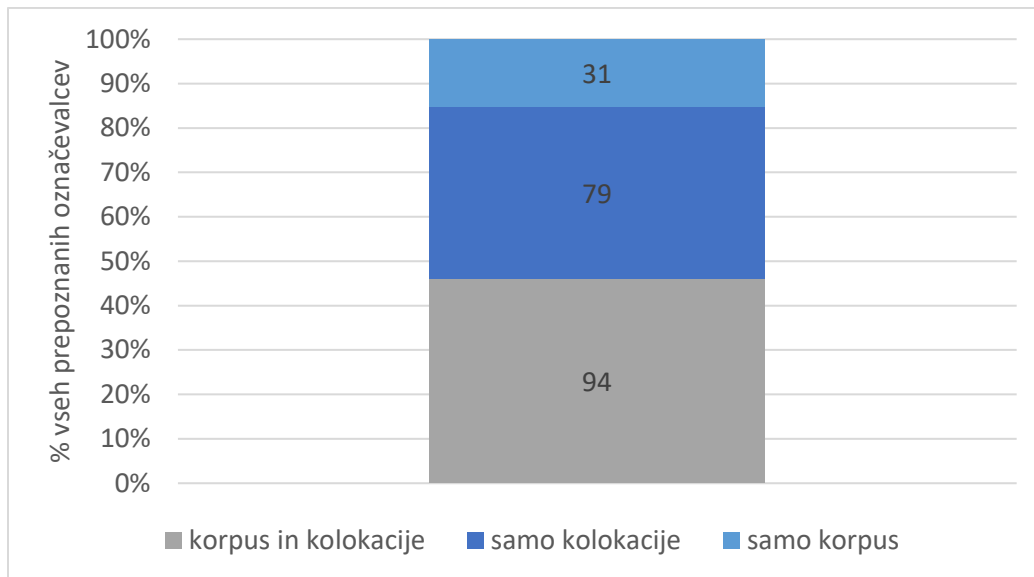
¹⁰⁴ Večbesedni diskurzni označevalci, prepoznani z mero MI: *alla fine*, *kljub vsemu*, *konec koncev*, *prvo kot prvo*, *kljub temu*, *moram priznati*, *kar se tega tiče*, *kako se že reče*, *veš kaj je fora*, *moram reči*, *od takrat naprej*, *med drugim*, *ali pa kaj podobnega*, *ali pa kaj takega*, *s tega stališča*, *s tega vidika*, *po drugi strani*, *še enkrat poudarjam*, *kaj sem hotela reči*, *zaradi tega*, *pri čemer*, *kako naj rečem*, *bi človek rekel*.

4.2.3.2 Primerjava statističnih metod in ročne prepoznavne

V zadnjem koraku primerjave različnih metod prepoznavanja večbesednih diskurznihi označevalcev v besedilnih korpusih smo seznam 173 označevalcev, pridobljen z različnimi načini statističnega razvrščanja nizov (kvantitativni pristop), primerjali še z izhodiščnim seznamom 125 označevalcev, prepoznanih v procesu ročne analize vzorčenega korpusa (kvalitativni pristop).¹⁰⁵

Primerjava obeh seznamov večbesednih diskurznihi označevalcev v korpusu govorne slovenščine razkriva, da izbrani krovni metodi prinašata med seboj različna seznama večbesednih diskurznihi označevalcev govorne slovenščine. Med **skupno 204 različnimi označevalci**, ki so bili identificirani bodisi kvalitativno bodisi kvantitativno, se jih je namreč le 94 (46,1 %) pojavilo tako v analiziranem vzorcu korpusa kot na enem izmed statističnih seznamov kolokacij, medtem ko se 79 označevalcev (38,7 % vseh znanih označevalcev) pojavlja zgolj na seznamu kolokacij, 31 označevalcev (15,2 %) pa smo identificirali zgolj z ročno analizo vzorčenega korpusa (Slika 7). Z drugimi besedami to pomeni, da se **kvalitativna in kvantitativna metoda medsebojno dopolnjujeta, pri čemer pa smo izčrpnjši seznam diskurznihi označevalcev vendarle dobili na podlagi analize statistično najrelevantnejših enot celotnega korpusa**, kar dokazuje neupravičenost (implicitne) predpostavke, da diskurzni označevalci tvorijo maloštevilno skupino izrazov, ki jih je zaradi njihove pogoste rabe mogoče identificirati že s kvalitativno analizo razmeroma majhnih in bolj ali manj reprezentativnih vzorcev realne komunikacije.

¹⁰⁵ Tako v kvalitativnem kot kvantitativnem pridobivanju seznama označevalcev vsak standardizirani zapis obravnavamo kot samostojno obliko in jih v skladu z radikalnim korpusnim pristopom ne skušamo združevati, npr. v primeru nedoslednih transkripcij (npr. *po pa – pol pa, se prav – se pravi*), pregibnih (npr. *bi rekel – bi rekla, a veš – a veste*), narečnih (npr. *tako da – tako ka*) ali drugih tipov bližnjih (variantnih) parov.



Slika 7: Delež diskurznihih označevalcev, pridobljenih z metodo ročnega pregleda reprezentativnega vzorca korpusa in/ali analize statistično najbolj izstopajočih kolokacij.

Če si podrobneje ogledamo seznam 79 označevalcev, ki so bili prepoznani šele na podlagi analize seznamov najpogostejših oz. kolokacijsko najmočnejših enot (Tabela 10), vidimo, da se na njem pojavljajo označevalci vseh pomenskih skupin razen označevalcev razloga, med njimi **tudi taki z visoko stopnjo gramatikalizacije in pogosto diskurzno rabo**, npr. *drugače pa*, *na neki način*, *ali pa kaj*, *zdaj ko*, *kaj je že*, *konec koncev* itd. Povprečna pogostost oblik kvantitativno identificiranih označevalcev sicer znaša 53 pojavitev na milijon pojavnic, kar potrjuje, da razlika v uporabnosti obeh metod ni zgolj posledica dejstva, da med diskurzne označevalce prištevamo širši nabor izrazov kot zgolj najbolj gramatikalizirane oz. pogosto rabljene označevalce.

Funkcijska skupina	Različnic	Večbesedni diskurzni označevalci
označevalci nedoločnosti	19	<i>ali nekaj, ali pa kaj, ali pa kaj podobnega, ali pa kaj takega, bolj tako, čisto tako, in podobno, in tako dalje, in vse, jaz ne vem, kaj pa jaz vem, kaj takega, na neki način, na vem, ne znam, pa ne vem kaj, pa vse, pa vse to, zdaj ne vem</i>
označevalci razvrščanja	14	<i>alla fine, na začetku, ob koncu, od takrat naprej, po pa, po tem, pol ko, pred tem, prvo kot prvo, še nekaj, v nadaljevanju pa, v tem času, za konec, zdaj ko</i>
označevalci tvorjenja	11	<i>bi rekla, bi rekli, bomo rekli, kaj sem hotela reči, kako bi rekel, kako naj rečem, kako se reče, kako se že reče, lahko rečem, moram priznati, še enkrat poudarjam</i>
označevalci rezultata	8	<i>glede na to, konec koncev, na tak način, ne glede na to, se prav,¹⁰⁶ toliko da, v tem smislu, za to</i>
označevalci dodajanja	8	<i>če že, glede tega, kar se tega tiče, ob tem, poleg tega, pri čemer, s tega stališča, s tega vidika</i>
označevalci nasprotja	5	<i>drugače pa, kljub temu, kljub vsemu, pač pa, samo pa</i>
označevalci pogoja	5	<i>če že, razen če, recimo če, recimo da, samo če</i>
označevalci specifikacije	2	<i>med drugim, tako kakor</i>
označevalci dvogovora	1	<i>a veste</i>

Tabela 10: Seznam 79 večbesednih diskurzni označevalcev, ki so bili prepoznani na podlagi analize kolokacij in se v vzorčenem korpusu govorjene slovenščine ne pojavljajo.

Čeprav je analiza primerljivosti posameznih statističnih mer v poglavju 4.2.2.4 *Prekrivnost seznamov najvišje uvrščenih kandidatov* sugerirala, da utegnemo z unijo izbire več komplementarnih mer zaradi njihove različnosti priklicati precej izčrpen nabor večbesednih enot v slovenščini, pa se presenetljivo 31 (15,2 %) označevalcev ne pojavi med vrhnjih 2.000 enot na nobenem izmed osmih načinov njihovega razvrščanja. Kot prikazuje njihov seznam v Tabeli 11, s pripisom najvišje uvrstitve glede na vseh osem načinov razvrščanja, bi tudi ob izbiri nižjega

¹⁰⁶ Opomba glede nekonsistentnih transkripcij (*se pravi proti se prav, ne vem proti na vem, po pa proti pol pa*). Združili samo obravnavo na nek način in na neki način.

4.2 Metode prepoznavanja večbesednih diskurznihi označevalcev v besedilnih korpusih

pragu, npr. analizi prvih 3.000 kandidatov, 19 označevalcev (9,3 % vseh znanih označevalcev) ostalo statistično nezaznanih.

	označevalec	najvišja uvrstitev		označevalec	najvišja uvrstitev		označevalec	najvišja uvrstitev
1	<i>recimo temu</i>	2.012 (t-test)	12	<i>kakorkoli že</i>	2.789 (t-test)	23	<i>še posebej če</i>	5.332 (MI)
2	<i>namesto da</i>	2.028 (t-test)	13	<i>edino če</i>	3.175 (t-test)	24	<i>tako torej</i>	5.415 (PP)
3	<i>v končni fazi</i>	2.034 (t-test)	14	<i>ali karkoli</i>	3.442 (t-test)	25	<i>pa take stvari</i>	6.400 (t-test)
4	<i>prej ko</i>	2.070 (t-test)	15	<i>v kolikor</i>	3.718 (t-test)	26	<i>poleg tega pa</i>	6.717 (t-test)
5	<i>če ne drugega</i>	2.073 (X2)	16	<i>pa take fore</i>	3.728 (MI)	27	<i>in vse skupaj</i>	6.746 (t-test)
6	<i>ali pa nekaj</i>	2.247 (t-test)	17	<i>če dovolite</i>	3.743 (X2)	28	<i>a la</i>	7.696 (t-test)
7	<i>v smislu da</i>	2.298 (LL)	18	<i>sploh pa</i>	3.932 (OP)	29	<i>ob tem da</i>	9.758 (MI)
8	<i>pa vendarle</i>	2.482 (t-test)	19	<i>sploh če</i>	4.099 (PP)	30	<i>danes ko</i>	11.889 (Dice)
9	<i>a vidite</i>	2.563 (t-test)	20	<i>zaradi tega da</i>	4.402 (PP)	31	<i>medtem ko pa</i>	13.039 (MI)
10	<i>tedaj pa</i>	2.565 (t-test)	21	<i>kot smo rekli</i>	4.543 (MI)			
11	<i>kaj naj rečem</i>	2.637 (X2)	22	<i>že odkar</i>	4.831 (MI)			

Tabela 11: Seznam večbesednih diskurznihi označevalcev, ki so bili prepoznani na podlagi ročne analize vzorčenega korpusa in se ne pojavljajo na seznamu 2.000 najvišje uvrščenih enot na nobenem izmed osmih načinov statističnega razvrščanja besednih nizov.

Če ugotovitve rezultatov analiz znotraj poglavja 4.2 *Metode prepoznavanja večbesednih diskurznihi označevalcev v besedilnih korpusih* strnjeno povzamemo, smo torej s primerjavo več različnih metod za prepoznavanje večbesednih diskurznihi označevalcev ugotovili, da ti izrazi v slovenščini predstavljajo razmeroma **maloštevilno skupino stalnih besednih zvez** v slovenščini, pri pridobivanju njihovega izčrpnega nabora pa je prevladujočo metodo ročnega pregledovanja reprezentativnih vzorcev jezikovne rabe smiselno **dopolnjevati ali celo nadomestiti s kvantitativnimi metodami luščenja večbesednih enot iz celotnega korpusa**, kot je denimo luščenje statistično najrelevantnejših večbesednih enot, saj te omogočajo hitrejši in obsežnejši priklic opazovanega tipa večbesednih enot. Pri tem smo ugotovili, da je tako kot pri drugih tipih stalnih besednih zvez razvrščanje po absolutni pogostosti smiselno dopolniti tudi z izračuni besedne povezovalnosti, saj nekoliko **večjo natančnost pri prepoznavanju večbesednih diskurznihi označevalcev kot razvrščanje po pogostosti daje mera t-test**. Druge mere besedne

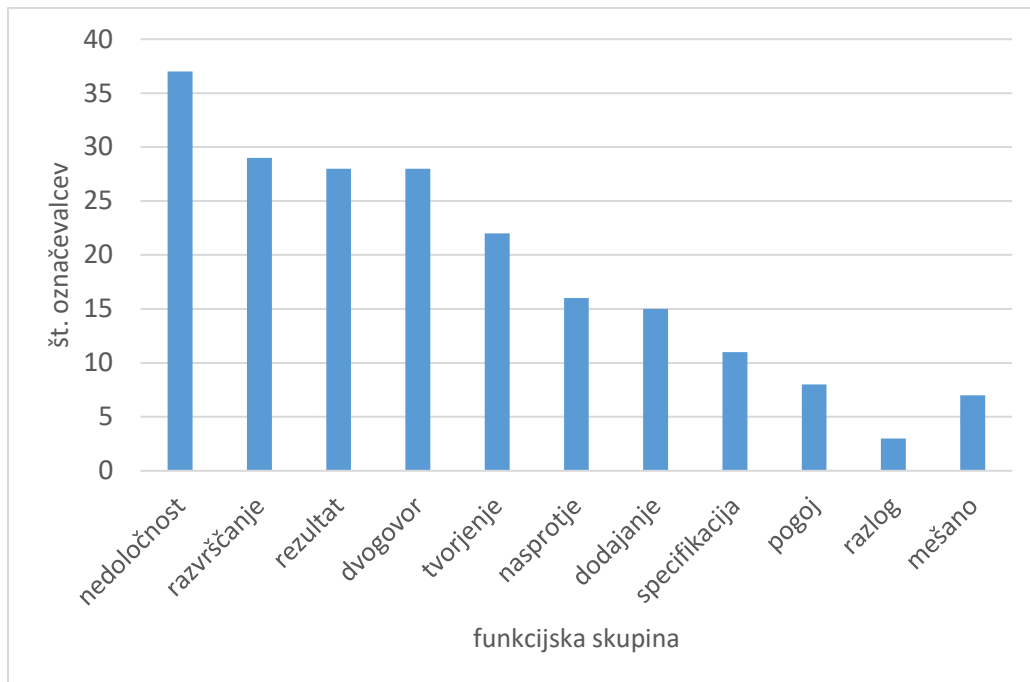
povezovalnosti so se izkazale za manj uporabne, še zlasti mera MI, ki za luščenje diskurznofunkcijske leksike v slovenščini ni primerna. Poleg relevantnosti za nadaljnje raziskave večbesednih diskurznihi označevalcev (npr. morebitno širjenje njihovega nabora z nadaljnjim pregledom rezultatov mere t-test onkraj 2.000 najvišje uvrščenih enot) so ti rezultati, objavljeni tudi v (Dobrovoljc, 2017), pomembni tudi za nadaljnje raziskave drugih tipov besednih zvez s podobno leksikalno zgradbo, kot so formulaični odgovori, vljudnostni obrazci, vzkliki in drugi izrazi čustvene naklonskosti, predložne zveze itd., ki prav tako pogosto vsebujejo tudi sicer pogosto rabljeno funkcijsko besedišče.

4.3 Seznam večbesednih diskurznih označevalcev govorjene slovenščine

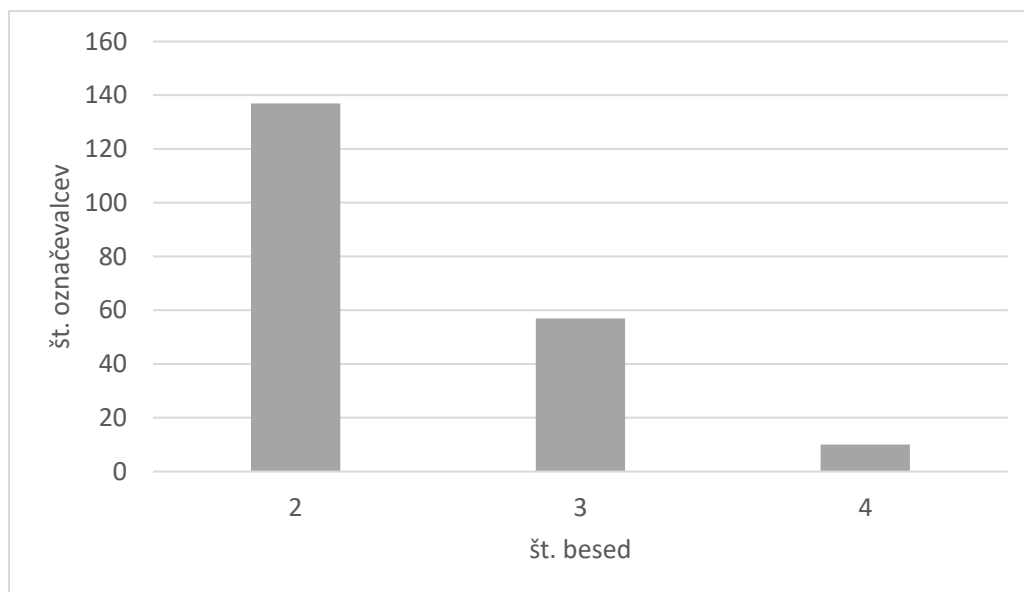
V tem poglavju na tretje raziskovalno vprašanje »Kateri večbesedni diskurzni označevalci se rabijo v govorjeni slovenščini?« odgovarjamo s predstavitvijo unije nabora označevalcev, ki so bili v korpusu Gos prepoznani bodisi z ročnim pregledom reprezentativnega vzorca korpusa Gos (poglavje 4.1 *Dopolnitev sheme za označevanje večbesednih diskurznih označevalcev*) bodisi s pregledom statistično najbolj izstopajočih večbesednih enot celotnega korpusa Gos (poglavje 4.2 *Metode prepoznavanja večbesednih diskurznih označevalcev v besedilnih korpusih*).

Z obema krovnama metodama smo torej v korpusu Gos identificirali skupno **204 večbesedne diskurzne označevalce**, ki jih s pripisom jedrne funkcijske skupine prikazujemo v Tabeli 12 spodaj, z drugimi podatki, zbranimi v okviru nadaljnje raziskave (pogostost rabe v korpusih Gos in Kres, stopnja dvoumnosti, stopnja tipičnosti) pa jih navajamo tudi v Prilogi 5. Kot strnjeno prikazuje graf na Sliki 8, smo največ različnih označevalcev identificirali znotraj funkcijskih skupin označevalcev nedoločnosti (37), razvrščanja (29), rezultata (28) in dvogovora (28), najmanj pa je označevalcev pogoja (8) in razloga (3). Z vidika dolžine prevladujejo označevalci dolžine dveh besed (137), med prepoznanimi označevalci pa ni izrazov, daljših od štirih besed (Slika 9).

4.3 Seznam večbesednih diskurzni označevalcev govorne slovenščine



Slika 8: Večbesedni diskurzni označevalci glede na funkcijsko skupino.¹⁰⁷



Slika 9: Večbesedni diskurzni označevalci glede na dolžino.

¹⁰⁷ V funkcijsko skupino *mešano* uvrščamo oblike označevalcev, ki se umeščajo v več različnih funkcijskih skupin: *prav tako* (dodajanje, nedoločnost), *s tem da* (dodajanje, specifikacija), *medtem ko* (razvrščanje, nasprotje), *in to* (nedoločnost, specifikacija), *če že* (dodajanje, pogoj), *glah tako* (dodajanje, rezultat), *samo da* (rezultat, nasprotje).

	OZNAČEVALEC	SKUPINA		OZNAČEVALEC	SKUPINA		OZNAČEVALEC	SKUPINA
1	a la	specifikacija	69	kljub temu da	nasprotje	137	s tem da	dodajanje, specifikacija
2	a ne	dvogovor	70	kljub vsemu	nasprotje	138	saj ne vem	nedoločnost
3	a ne da	dvogovor	71	konec koncev	rezultat	139	saj pravim	tvorjenje
4	a veste	dvogovor	72	kot da	specifikacija	140	saj veste	dvogovor
5	a veš	dvogovor	73	kot rečeno	tvorjenje	141	saj veš	dvogovor
6	a veš kaj	dvogovor	74	kot smo rekli	tvorjenje	142	samo če	pogoj
7	a vidite	dvogovor	75	lahko rečem	tvorjenje	143	samo da	rezultat, nasprotje
8	ali kaj	dvogovor	76	ma ja	dvogovor	144	samo pa	nasprotje
9	ali kaj je	dvogovor	77	med drugim	specifikacija	145	se prav	rezultat
10	ali kako	dvogovor	78	medtem ko	razvrščanje, nasprotje	146	se pravi	rezultat
11	ali karkoli	nedoločnost	79	medtem ko pa	nasprotje	147	se pravi da	rezultat
12	ali ne	dvogovor	80	moram priznati	tvorjenje	148	sedaj pa	razvrščanje
13	ali nekaj	nedoločnost	81	moram reči	tvorjenje	149	sicer pa	nasprotje
14	ali nekaj takega	nedoločnost	82	na koncu	razvrščanje	150	sploh če	pogoj
15	ali pa	dodajanje	83	na neki način	nedoločnost	151	sploh pa	dodajanje
16	ali pa kaj	nedoločnost	84	na primer	specifikacija	152	še enkrat poudarjam	tvorjenje
17	ali pa kaj podobnega	nedoločnost	85	na ta način	rezultat	153	še nekaj	razvrščanje

18	ali pa kaj takega	nedoločnost	86	na tak način	rezultat	154	še posebej če	pogoj
19	ali pa nekaj	nedoločnost	87	na vem	nedoločnost	155	tako ali tako	rezultat
20	alla fine	razvrščanje	88	na začetku	razvrščanje	156	tako da	rezultat
21	bi človek rekel	tvorjenje	89	namesto da	nasprotje	157	tako ka	rezultat
22	bi rekel	tvorjenje	90	ne da	specifikacija	158	tako kakor	specifikacija
23	bi rekla	tvorjenje	91	ne glede na to	rezultat	159	tako ko	specifikacija
24	bi rekli	tvorjenje	92	ne pa da	nasprotje	160	tako kot	specifikacija
25	bolj tako	nedoločnost	93	ne vem	nedoločnost	161	tako rekoč	tvorjenje
26	bom rekel	tvorjenje	94	ne vem kaj	nedoločnost	162	tako torej	dvogovor
27	bom rekla	tvorjenje	95	ne znam	nedoločnost	163	takrat ko	razvrščanje
28	bomo rekli	tvorjenje	96	nekaj takega	nedoločnost	164	te pa	razvrščanje
29	čakaj malo	dvogovor	97	no ja	dvogovor	165	tedaj pa	razvrščanje
30	če dovolite	dvogovor	98	ob koncu	razvrščanje	166	ti povem	dvogovor
31	če ne	rezultat	99	ob tem	dodajanje	167	to pomeni	rezultat
32	če ne drugega	dodajanje	100	ob tem da	dodajanje	168	to se pravi	rezultat
33	če že	dodajanje, pogoj	101	od takrat naprej	razvrščanje	169	toliko da	rezultat
34	čisto tako	nedoločnost	102	pa ne vem kaj	nedoločnost	170	tudi če	nasprotje
35	danes ko	razvrščanje	103	pa take	nedoločnost	171	v bistvu	nasprotje

36	druga zadeva	razvrščanje	104	pa take fore	nedoločnost	172	v glavnem	rezultat
37	drugače pa	nasprotje	105	pa take stvari	nedoločnost	173	v kolikor	pogoj
38	edino če	pogoj	106	pa tako	nedoločnost	174	v končni fazi	rezultat
39	glede na to	rezultat	107	pa tako naprej	nedoločnost	175	v nadaljevanju	razvrščanje
40	glede na to da	razlog	108	pa to	nedoločnost	176	v nadaljevanju pa	razvrščanje
41	glede tega	dodajanje	109	pa vendarle	nasprotje	177	v redu	dvogovor
42	glih tako	dodajanje, rezultat	110	pa vse	nedoločnost	178	v resnici	nasprotje
43	in pa	dodajanje	111	pa vse to	nedoločnost	179	v smislu	specifikacija
44	in podobno	nedoločnost	112	pač pa	nasprotje	180	v smislu da	specifikacija
45	in sicer	specifikacija	113	po drugi strani	nasprotje	181	v tem času	razvrščanje
46	in tako	nedoločnost	114	po pa	razvrščanje	182	v tem primeru	rezultat
47	in tako dalje	nedoločnost	115	po svoje	nedoločnost	183	v tem smislu	rezultat
48	in tako naprej	nedoločnost	116	po tem	razvrščanje	184	v vsakem primeru	rezultat
49	in to	nedoločnost, specifikacija	117	pol ko	razvrščanje	185	vendar pa	nasprotje
50	in vse	nedoločnost	118	pol pa	razvrščanje	186	veste kaj	dvogovor
51	in vse skupaj	nedoločnost	119	poleg tega	dodajanje	187	veš kaj	dvogovor
52	ja no	dvogovor	120	poleg tega pa	dodajanje	188	veš kaj je	dvogovor
53	jaz ne vem	nedoločnost	121	potem ko	razvrščanje	189	veš kaj je fora	dvogovor

54	<i>kaj jaz vem</i>	nedoločnost	122	<i>potem pa</i>	razvrščanje	190	<i>za konec</i>	razvrščanje
55	<i>kaj je</i>	dvogovor	123	<i>prav tako</i>	dodajanje, nedoločnost	191	<i>za to</i>	rezultat
56	<i>kaj je že</i>	dvogovor	124	<i>pred tem</i>	razvrščanje	192	<i>za to da</i>	rezultat
57	<i>kaj naj rečem</i>	tvorjenje	125	<i>predvsem pa</i>	dodajanje	193	<i>za začetek</i>	razvrščanje
58	<i>kaj pa jaz vem</i>	nedoločnost	126	<i>prej ko</i>	razvrščanje	194	<i>zaradi tega</i>	rezultat
59	<i>kaj sem hotela reči</i>	tvorjenje	127	<i>pri čemer</i>	dodajanje	195	<i>zaradi tega da</i>	rezultat
60	<i>kaj takega</i>	nedoločnost	128	<i>pri tem</i>	dodajanje	196	<i>zaradi tega ker</i>	razlog
61	<i>kaj že</i>	dvogovor	129	<i>prvo kot prvo</i>	razvrščanje	197	<i>zato da</i>	rezultat
62	<i>kako bi rekel</i>	tvorjenje	130	<i>razen če</i>	pogoj	198	<i>zato ker</i>	razlog
63	<i>kako naj rečem</i>	tvorjenje	131	<i>recimo če</i>	pogoj	199	<i>zato pa</i>	rezultat
64	<i>kako se reče</i>	tvorjenje	132	<i>recimo da</i>	pogoj	200	<i>zdaj ko</i>	razvrščanje
65	<i>kako se že reče</i>	tvorjenje	133	<i>recimo temu</i>	tvorjenje	201	<i>zdaj ne vem</i>	nedoločnost
66	<i>kakorkoli že</i>	rezultat	134	<i>s tega stališča</i>	dodajanje	202	<i>zdaj pa</i>	razvrščanje
67	<i>kar se tega tiče</i>	dodajanje	135	<i>s tega vidika</i>	dodajanje	203	<i>znaš kaj</i>	dvogovor
68	<i>kljub temu</i>	nasprotje	136	<i>s tem</i>	rezultat	204	<i>že odkar</i>	razvrščanje

Tabela 12: Seznam prepoznanih večbesednih diskurzivnih označevalcev govorne slovenščine.

4.4 Tipičnost za govorjeni jezik

V tem poglavju bomo odgovorili na četrto raziskovalno vprašanje »Kateri večbesedni diskurzni označevalci se tipično rabijo predvsem v govorjeni rabi?«, pri čemer po opisu izbrane metode izračuna pogostosti diskurzne rabe (poglavje 4.4.1 *Izračun pogostosti diskurzne rabe*) v prvem koraku primerjamo splošno pogostost rabe večbesednih diskurzni označevalcev v govornem korpusu Gos glede na pisni korpus Kres (poglavje 4.4.2 *Pogostost rabe v govorjeni in pisni slovenščini*), v drugem koraku pa z vidika njihove ključnosti za korpus Gos glede na korpus Kres razvrstimo še posamezne označevalce in s tem določimo končni nabor tipično govorjenih večbesednih diskurzni označevalcev (poglavje 4.4.3 *Tipično govorjeni večbesedni diskurzni označevalci*).

4.4.1 Izračun pogostosti diskurzne rabe

Pri označevalcih, ki se v jezikovni rabi prevladujoče uporabljajo zgolj kot diskurzni označevalci (npr. *kljub temu da*), izračun pogostosti rabe predstavlja preprost metodološki korak, kjer za dano obliko niza pridobimo njeno pogostost v opazovanem (Gos) in primerjalnem (Kres) korpusu, pri čemer glede na ortografske razlike med obema korpusoma v korpusu pisne slovenščine Kres nismo razlikovali med različnimi načini zapisi glede na velikost črk ali stavo ločil¹⁰⁸ – nize '*kljub temu, da*', '*Kljub temu da*' in '*kljub temu da*' torej vse obravnavamo kot pojavitve diskurznega označevalca *kljub temu da*.

Bolj problematični pa so nizi, ki se lahko v rabi pojavljajo tako kot diskurzni označevalci kot v drugih vlogah, denimo kot del propozicijske vsebine, ki je lahko z označevalcem diahrono povezana (npr. *ne vem ali v glavnem*) ali pa je zgolj naključna (npr. *takó da* proti *táko da*). V literaturi se za ta tip pojava običajno uporablja izraz **dvoumnost**, ki pa ga ne smemo enačiti z večfunkcijskostjo (glej poglavje 2.2.2.3 *Večfunkcijskost*), tj. pojavom, ko lahko posamezni diskurzni označevalec v hkrati ali v različnih kontekstih opravlja različne diskurzne vloge (npr.

¹⁰⁸ Stava ločil je zaradi visoke stopnje gramatikalizacije v pisni rabi namreč precej neustaljena, predvsem pri konektorjih s kataforično sestavino (npr. *ob tem da*) in modificiranih konektorjih (npr. *še posebej če*, *že odkar*, *danes ko*).

označevalec *medtem ko* v vlogi uvajalca nasprotja in/ali časovnih razmerij). Čeprav tudi večfunkcijski označevalci z vidika strojnega razločevanja predstavljajo velik izziv (glej npr. Meyer idr. 2011), se v nadaljevanju osredotočimo predvsem na raziskave, ki se ukvarjajo z razdvoumljanjem diskurznihi označevalcev v besedilnih korpusih, s kakršnimi bi lahko dobili podatek o dejanski pogostosti rabe posameznih diskurznihi označevalcev v korpusih Gos in Kres.

4.4.1.1 Kratek pregled metod za razdvoumljanje diskurznihi označevalcev

Kot povzema Stede (2014), lahko metode za razdvoumljanje diskurznihi označevalcev razdelimo v štiri krovne skupine glede na stopnjo jezikoslovne označenosti korpusov: (i) metode, ki se opirajo zgolj na površinsko besedilo, (ii) metode, ki se opirajo na pripisane oblikoskladenjske oznake, (iii) metode, ki se opirajo na skladenjsko razčlenbo besedila, in (iv) metode, ki se opirajo na besedilo s pripisanimi semantičnimi oz. pragmatičnimi interpretacijami. Primer prve skupine je denimo D. Marcu (1998, 2000), ki je diskurzne označevalce v angleščini prepoznaval z oblikovanjem različnih pravil (regularnih izrazov), ki med drugim iščejo izraze, ki se pojavljajo na začetku stavka (njihova stava v angleščini je namreč bolj predvidljiva kot v slovenščini) ali v kombinacijami z določenimi ločili (npr. *Besides, ...*), s čimer je pri razdvoumljanju diskurznihi označevalcev denimo dosega f-vrednost 85 % (skupna vrednost za natančnost in priklic). Podobno Hirschberg in Litman (1993) pri razdvoumljanju označevalcev tematske organizacije v angleškem govoru, kot so *now, well, incidentally, by the way, that reminds me, anyway, in any case*, uporabljata pravila, ki upoštevajo prozodične lastnosti, kot sta (ne)poudarjenost in nizka intonacije, pa tudi prozodične lastnosti, ki so bile v transkripcijah označene z ločil, kot so vejice, pike, vezaji in oznake za odstavke. V drugo skupino lahko umestimo raziskavi Schneider in Stede (2012), ki se pri razdvoumljanju označevalcev v nemščini opirata na avtomatsko pripisani podatek o besedni vrsti, vendar ta metoda učinkovita le za zelo majhno število označevalcev (9 od 41 dvoumnih označevalcev), prav tako bistvene izboljšave ne prinaša podatek o besedni vrsti besed v okolici.

Med metodami tretje skupine, ki temeljijo na razdvoumljanju na podlagi skladenjskih informacij, sta Pitler in Nenkova (2009) označevalce v skladenjsko razčlenjenem korpusu Penn Discourse Treebank denimo prepoznavali na podlagi informacij o njihovi oblikoskladenjski oznaki, oblikoskladenjski oznaki nadrejenega elementa (ki npr. ločuje med samostalniško in stavčno

vlogo veznika *and*), oblikoskladenjsko oznako levega in desnega soležnega elementa (ki npr. ločuje med homonimnimi stavčnimi vezniki in predlogi), itd. S to metodo in hkratnim upoštevanjem same oblike označevalca so bili diskurzni označevalci prepoznani s 95,04-odstotno natančnostjo (f-vrednost 92,28 %). Med sorodne skladenjske pristope lahko umestimo še raziskavi Hutchinson (2003), ki diskurzne označevalce prepozna s skladenjskim razčlenjevanjem spletnih besedil (povedi, ki vsebujejo oblike potencialnih diskurzni označevalcev), ter Faiz in Mercer (2013), ki uporabljata še kompleksnejšo kombinacijo površinskih in skladenjskih značilk.

Čeprav Stede (2014, str. 134) izpostavlja, da doslej še ni bilo raziskav, ki bi se pri prepoznavanju diskurzni označevalcev (besedilnih konektorjev) opirale na pomenske lastnosti sobesedilnega konteksta in/ali pragmatične značilke (četrta skupina), bi mednje lahko prištevali raziskavo (Popescu-Belis in Zufferey, 2011), ki se sicer ukvarja zgolj z razdvoumljanje označevalcev oz. prislovov *like* in *well* v govorni angleščini. Avtorja pri prepoznavanju diskurzne rabe s pomočjo strojnega učenja upoštevata leksikalne lastnosti (besede v okolici), prozodične lastnosti (dolžina, premori pred in po izgovorjenem označevalcu) oz. položaj potencialnega označevalca v izjavi in sociolingvistične lastnosti govorca, kot so spol, starost, stopnja izobrazbe ipd. s čimer za označevalec *like* dosežeta točnost 84,5 %, za označevalec *well* pa 97,5 %. Pri tem ugotavljata, da najboljše rezultate dajejo leksikalne lastnosti besed v okolici, sociolingvistične lastnosti pa vsaj pri razdvoumljanju oz. prepoznavanju obravnavanih označevalcev ne igrata bistvene vloge. Eno pomembnih spoznanj njune raziskave je tudi dejstvo, da se rezultati oz. vpliv posameznih značilk za vsakega izmed obeh obravnavanih označevalcev med seboj razlikujejo, kar tudi empirično potrjuje, da diskurzni označevalci tvorijo zelo nehomogeno skupino jezikovnih izrazov, ki je ni nujno mogoče preučevati z enakim metodološkim aparatom.

4.4.1.2 Metoda sklepanja po vzorcu

Ne glede na to, ali za prepoznavanje uporabljajo vnaprej napisana pravila ali metode strojnega učenja, vse zgoraj predstavljene metode **predpostavljajo obstoj korpusa, ki je na razmeroma zanesljiv način** (najpogosteje ročno) **označen na več različnih jezikoslovnih ravneh**, bodisi z informacijami o prozodičnih lastnostih besedila, ločili, morfološki kategorijami ali skladenjskimi razčlembami. Za razliko od korpusov, uporabljenih v zgornjih raziskavah, **korpus**

Gos teh informacij ne vsebuje. Ročno so bile obdelane zgolj segmentacija, tokenizacija in transkripcija na dveh različnih nivojih, metapodatki o lemah in oblikoskladenjskih oznakah pa so bili korpusu pripisani avtomatsko, natančnost uporabljenega označevalnika (Grčar idr., 2012) pa za govorjeni jezik še ni bila ustrezno evalvirana. Glede na obstoječo stopnjo označenosti korpusa Gos torej ni mogoče avtomatsko ugotoviti, ali je niz 'a ne' v izjavi

[138] strpnost to smo tud **a ne** sprejemanje drugačnih
rabljen kot diskurzni označevalec (z interpretacijo: *strpnost, to smo tudi, a ne, sprejemanje drugačnih*) ali kot kombinacija protivnega veznika in nikalnega členka, ki uvaja stavke z glagolsko elipso (z interpretacijo: *strpnost, to smo tudi, a ne sprejemanje drugačnih*). Tako kot brez pripisanih prozodičnih informacij, oblikoslovnih oznak ali skladenjske razčlembe o ustrezni interpretaciji ne more sklepati človeški interpret tudi od strojnih mehanizmov torej ne moremo pričakovati, da bi to nalogo opravili razmeroma zanesljivo.

Poleg okrnjene razpoložljive infrastrukture pa je naša raziskava specifična tudi glede na svoj namen. Medtem ko se zgornje raziskave ukvarjajo z vprašanjem, kako čim natančneje prepoznati diskurzne označevalce v konkretnih pojavitvah, pa za potrebe pridobivanja *pogostosti* njihove diskurzne rabe za nas podatek o konkretnih rabah ni relevanten. Namesto pisanja kompleksnih pravil za prepoznavanje posamičnih pojavitev **izrazno in distribucijsko medsebojno zelo raznolikih označevalcev** v konkretnih besedilih, ki bi se lahko opirala na bolj ali manj zanesljive podatke o položaju, lemi in morfoloških lastnostih označevalca ali besed v njegovi okolici, smo z namenom pridobivanja podatka o deležu diskurzne rabe uporabili bistveno preprostejšo metodo, ki jo bomo v nadaljevanju imenovali **metoda sklepanja po vzorcu**.

O podatku o dejanskem številu pojavitev diskurznega označevalca smo namreč sklepali na podlagi deleža diskurznih rab v pregledu naključnega vzorca vseh pojavitev dane oblike. Za ponazoritev vzemimo analizo deleža diskurznih rab zgoraj omenjenega označevalca 'a ne' v korpusu Kres, v katerem lahko natančnost metode sklepanja po vzorcu dodatno primerjamo tudi z metodo razdvoumljanja na podlagi ločil. Oblika 'a ne' se ne glede na velikost črk v korpusu Kres skupno pojavi 4.434-krat (36,8-krat na milijon pojavnic). Če v orodju noSketchEngine s funkcijo

izbire naključnega vzorca za analizo izberemo 100 naključnih konkordanc,¹⁰⁹ ugotovimo, da se niz 'a ne' v vzorcu 100 pojavitev 23-krat pojavi v vlogi diskurznega označevalca (primeri [139]-[141]), 77-krat pa v drugih vlogah, denimo kot kombinacija protivnega veznika in eliptičnega členka (primer [142]), kot kombinacija vprašalnega in nikalnega členka (primer [143]) ali v drugih redkih kombinacijah (primer [144]).

[139] Malo je kislega notri, **a ne**?

[140] Naslov je dvoumen, mal' zavede, **a ne**, drugače pa so slike lepa.

[141] Seveda je nedolzen, tako kot sta Pahor pa Golobic pa Bavcar itd. pošteni pa jih noben ne preganja **a ne**.

[142] Blizu so, **a ne** vedo, kako bi izigrali zadnji adut.

[143] **A ne** dobim lubčka za dobro jutro?

[144] Sonja Javornik je zatrdila, da na Kanalu **A ne** dela čistk.

Z metodo sklepanja po vzorcu bi torej sklepali, da se tudi v celotnem korpusu Kres označevalec 'a ne' pojavi v 23 % primerov, tj. v 1.020 primerih od vseh 4.434 pojavitev niza. To posplošitev nato preverimo še z razdvoumljanjem na podlagi predvidljivih lastnosti sobesedilnega konteksta, za kar smo oblikovali preprost regularni izraz (spodaj), ki v korpusu Kres poišče vse pojavitve niza standardiziranih pojavnic 'a' in 'ne' znotraj ene povedi, ki mu sledi poljubno ločilo, pri čemer v drugem koraku izločimo še (31) nizov, ki jim za tem ločilom sledi pojavnica *da* (*A ne, da očka ni čisto pri sebi?*).

```
[lc="a"] [lc="ne"] [tag="U"] within <s/>
```

Skupno tako s kombinacijo obeh iskalnih pogojev tako dobimo 1,029 primerov¹¹⁰ oz. 23,2 % vseh pojavitev niza 'a ne', med katerimi se vsaj med prvimi 100 pregledanimi konkordancami ne pojavi noben primer lažnega zadetka, v katerem bi bil niz rabljen v propozicijski ali kakršnikoli drugi nediskurzni vlogi. S predstavljeno metodo sklepanja po vzorcu smo torej dobili **povsem zanesljivo informacijo o deležu diskurzne rabe** niza v korpusu, ki je vsaj pri prepoznavanju

¹⁰⁹ Seznam 100 naključnih konkordanc niza 'a ne' v korpusu Kres: <http://goo.gl/PRm5yK>

¹¹⁰ Seznam pojavitev niza 'a ne' v korpusu Kres, ki ustrezajo opisanemu iskalnemu pogoju: <http://goo.gl/jNO7ML>

niza 'a ne' celo natančnejša kot nekatere izmed zgoraj predstavljenih kompleksnejših metod razdvoumljanja označevalcev v kontekstu.¹¹¹

Analiza konkordanc vzorca nam je obenem omogočila, da smo ročno razdvoumili primere, kjer se en označevalec (npr. *se pravi*) pojavlja kot del drugega označevalca (npr. *to se pravi, se pravi da, to se pravi da*). V takih primerih smo kot primer diskurzne rabe niza šteli zgolj pojavitve, kjer se označevalec pojavlja kot samostojna strukturna enota in ne kot del drugega diskurznega označevalca. Enako velja za vse medsebojno vsebovane kombinacije označevalcev, kot so pari osnovnih in modificiranih označevalcev (npr. *zato ker* in *še zlasti zato ker*), ali razširitve s členkom *pa* (npr. *poleg tega* in *poleg tega pa*). Pripisani podatek o pogostosti diskurzne rabe opazovanega niza nam torej pove, **kolikokrat se v korpusu kot samostojni diskurzni označevalec pojavi točno ta oblika**, podatek o pogostosti drugih rab pa tako lahko združuje tako propozicijske (npr. *poleg tega človeka*) ali besednozvezne vezniške rabe (npr. *hiša in pa miza*) kot rabe znotraj drugega označevalca (npr. *poleg tega pa*).

4.4.2 Pogostost rabe v govorjeni in pisni slovenščini

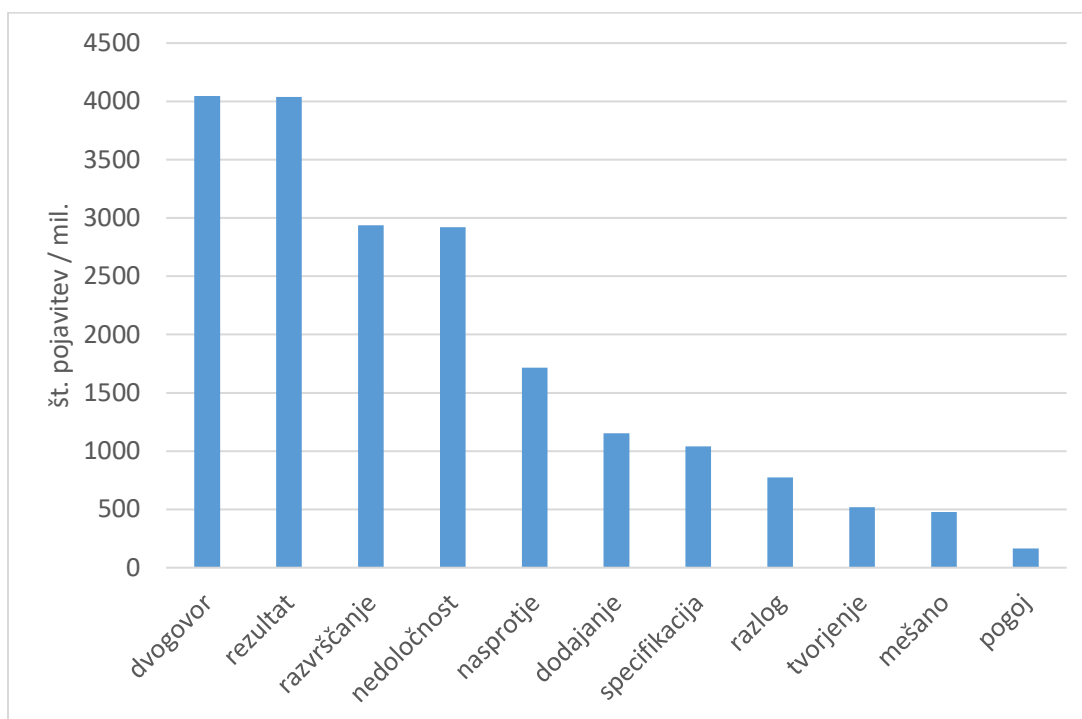
4.4.2.1 Pogostost rabe v korpusu Gos

Ugotovljeno pogostost pojavljanja posameznih diskurzni označevalcev v korpusu Gos prikazujemo v Prilogi 5, kjer poleg obravnavanega niza navajamo število vseh pojavitev oblike v korpusu, delež prepoznanih diskurzni rab v analiziranem vzorcu in oceno števila pojavitev diskurzni rab s sklepanjem na celoten korpus. Skupno se 204 obravnavani označevalci v korpusu Gos pojavijo 19.784-krat (19.709-krat na milijon pojavnic), kar pomeni, da se **posamezni označevalec v korpusu Gos povprečno pojavi kar 97-krat**. Ta podatek že kaže na izredno pogosto rabo teh stalnih besednih zvez v govorjenem jeziku, saj se denimo med 27.495 izluščenimi nizi (poglavje 4.2.2.1 *Luščenje in običajno štetje besednih nizov*) v korpusu Gos pogosteje (98-krat ali več) pojavi le 636 oz. 2,3 % vseh nizov.

¹¹¹ Pri tem smo ne glede na siceršnjo pogostost oblike in velikost analiziranega korpusa izbrali enotno velikost vzorca (100 analiziranih konkordanc), saj menimo, da ta za namen naše raziskave daje dovolj dobro oceno približne pogostosti, kot je razvidno iz primera analize označevalca 'a ne' v obsežnem korpusu Kres. Več kot 100-krat se v korpusu Gos sicer pojavi le 63, več kot 500-krat pa le 18 od skupno 204 obravnavanih nizov.

4.4 Tipičnost za govorjeni jezik

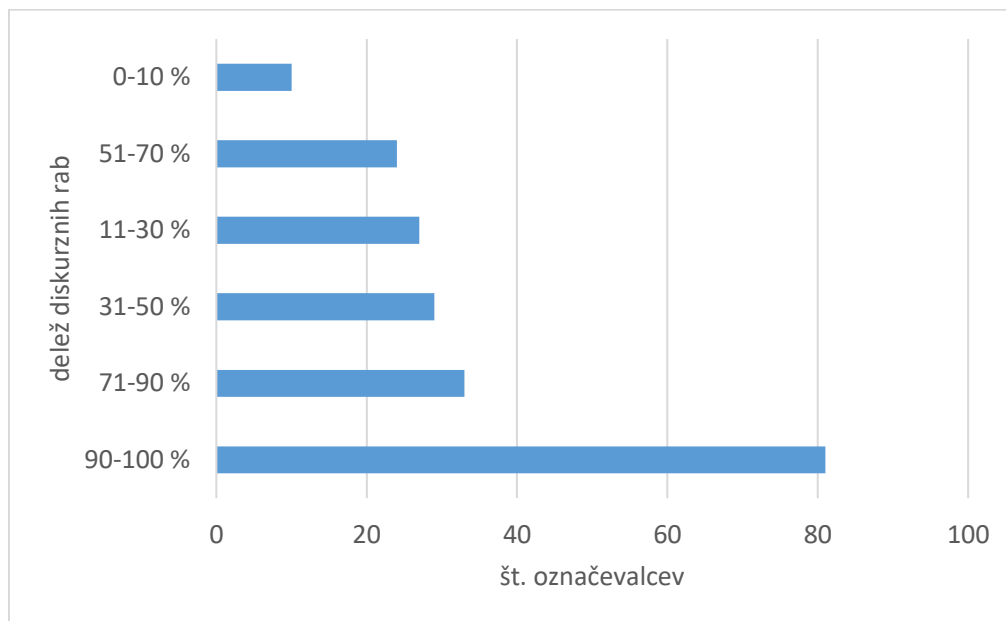
Če si podrobneje ogledamo še pogostost rabe označevalcev glede na jedrno funkcijsko skupino (Slika 10), vidimo, da je v govoru **res pogosta raba interakcijskih označevalcev**, kot so označevalci dvogovora (med njimi najpogosteje *a ne*, *a veš*, *ali kaj*, *ja no*, *v redu*) in označevalci nedoločnosti (*ne vem*, *pa to*, *in tako naprej*, *na neki način*, *pa tako*), **vendar prav tako izstopajo tudi nekatere skupine konektorskih označevalcev**, zlasti označevalci rezultata (*tako da*, *se pravi*, *v glavnem*, *zaradi tega*, *zato da*) in razvrščanja (*pol pa*, *zdaj pa*, *potem pa*, *na koncu*, *takrat ko*).



Slika 10: Pogostost rabe večbesednih diskurznihih označevalcev v korpusu Gos glede na funkcijsko skupino.

Za nazornejšo sliko visoke stopnje dvoumnosti (večbesednih) diskurznihih označevalcev v slovenščini smo analizirane nize nadalje razdelili v pet odstotkovnih skupin glede na delež prepoznanihih diskurznihih rab, od nizov, ki se v rabi večinoma pojavljajo zgolj kot diskurzni označevalci (v več kot 90 % vseh primerov), npr. *tako da*, *v bistvu*, *zdaj pa*, *potem pa*, *ja no*, *na primer*, *in tako naprej*, *in sicer*, *na neki način*, *v resnici*, *zdaj ko* ipd., do nizov, ki se v tej vlogi pojavljajo zelo redko (v manj kot 10 % primerov), npr. *in to*, *za to*, *ne da*, *ne vem kaj*, *kaj takega*, *kaj je*, *še nekaj*, *ali ne* ipd. Prikaz števila nizov glede na stopnjo dvoumnosti na Sliki 11 kaže, da se med (večinoma) nedvoumne označevalce umeščata slabi dve petini (39,7 %) vseh prepoznanihih

označevalcev, medtem ko se **preostale tri petine označevalcev v korpusu pojavljajo tudi v drugih, nediskurznih rabah.**



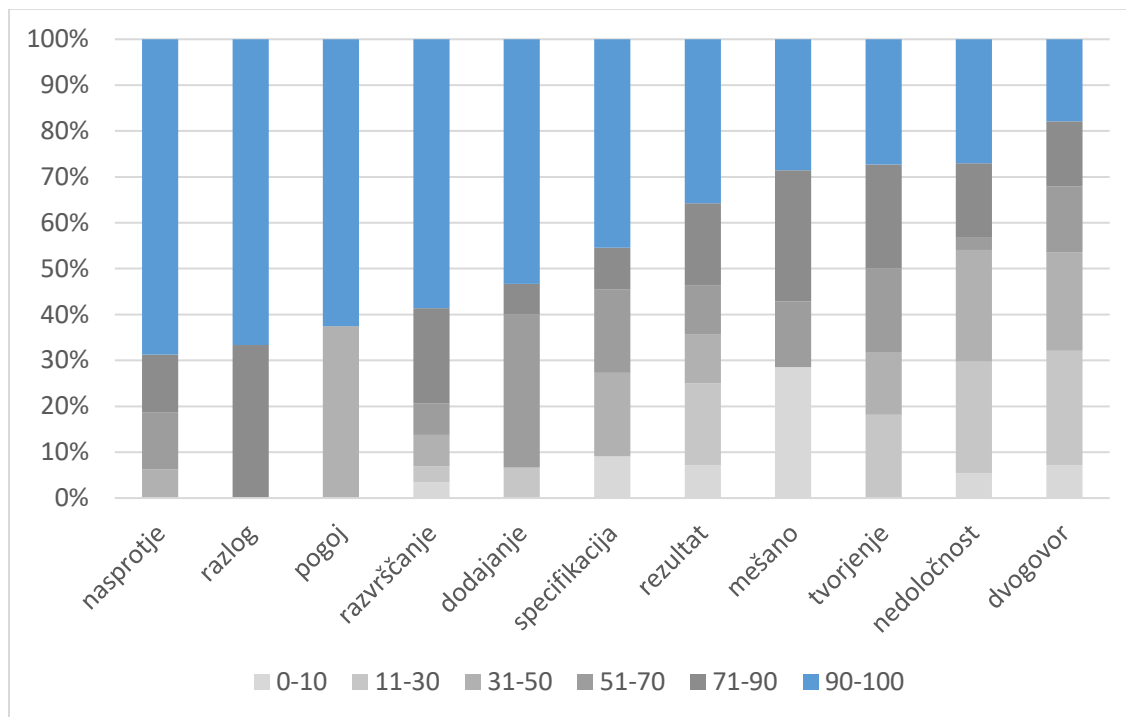
Slika 11: Stopnja dvoumnosti večbesednih diskurzni označevalcev govorjene slovenščine s prikazom števila označevalcev glede na delež diskurznih rab med vsemi pojavitvami niza v korpusu Gos.

Dokler za slovenščino niso podrobneje raziskani še enobesedni diskurzni označevalci (v širšem smislu), ta podatek težko ovrednotimo, saj se tudi podatki o deležu dvoumnosti diskurzni označevalcev v drugih jezikih med seboj precej razlikujejo: medtem ko Pitler in Nenkova (Pitler in Nenkova, 2009) poročata, da je v angleškem korpusu PDTB le 11 % označevalcev razmeroma nedvoumni (le 11 od 100 obravnavanih označevalcev se v več kot 90 % vseh pojavitev pojavlja zgolj v vlogi diskurzni označevalcev), Dipper in Stede (2006) po drugi strani ugotavljata, da je med 170 vezniki v leksikonu nemških besedilnih konektorjev DiMLex kar 129 (76 %) takih, ki se v rabi pojavljajo zgolj kot diskurzni označevalci.

Tudi pri podrobnejši analizi stopnje dvoumnosti glede na posamezne funkcijske skupine (Slika 12) se izkaže, da je stopnja dvoumnosti precej odvisna od same funkcijske skupine oz. oblik, ki jo sestavljajo, **pri čemer so označevalci interakcije, kot so označevalci tvorjenja, nedoločnosti in dvogovora, bistveno bolj dvoumni od besedilnih konektorjev**, kar je z diahronega vidika tudi pričakovano, saj izkazujejo nižjo stopnjo gramatikalizacije (pregibnost, sledovi propozicijske vsebine). Za vse tipe označevalcev pa visoka stopnja dvoumnosti večbesednih diskurzni

4.4 Tipičnost za govorni jezik

označevalcev v slovenščini zagotovo predstavlja **velik jezikovnotehnološki izziv prihodnosti**, zlasti za potrebe kvantitativnih korpusnih raziskav njihove distribucije v obsežnejših besedilnih korpusih, kot je naša.



Slika 12: Stopnja dvoumnosti večbesednih diskurzivnih označevalcev govorne slovenščine po posameznih funkcijskih skupinah s prikazom deleža označevalcev glede na delež diskurzivnih rab med vsemi pojavitvami niza v korpusu Gos.

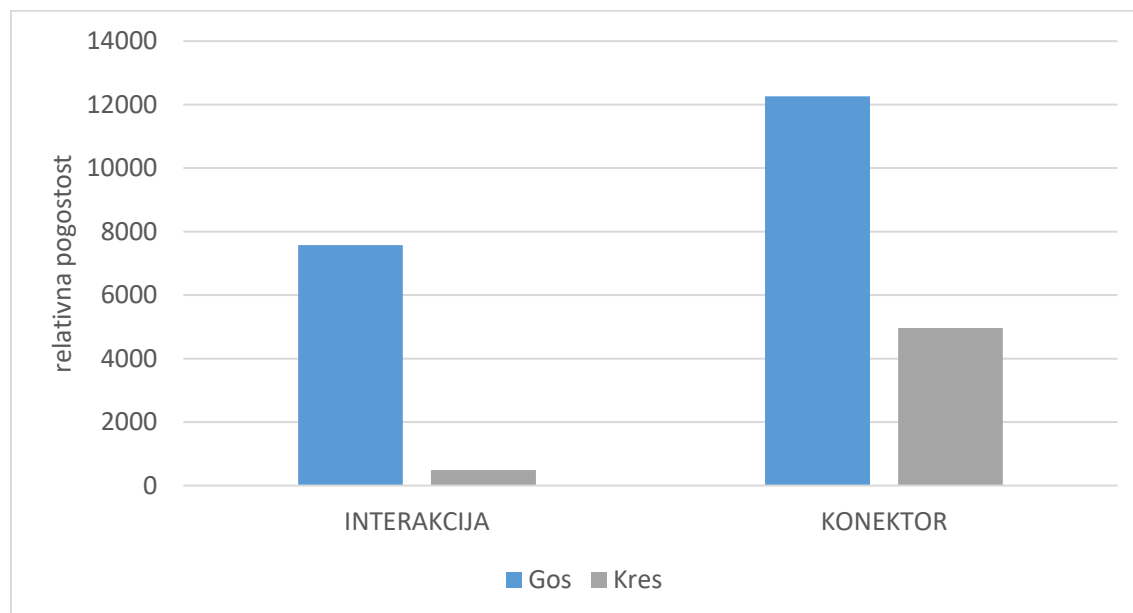
4.4.2.2 Pogostost rabe v korpusu Kres

Z enako metodo smo nato za vse označevalce pridobili še podatek o pogostosti rabe v korpusu pisne slovenščine Kres.¹¹² Skupno smo v korpusu Kres našli 518.340 pojavitev obravnavanih diskurzivnih označevalcev oz. 5.192 pojavitev na milijon pojavnic. V primerjavi z 19.709 pojavitvami na milijon pojavnic v korpusu Gos to pomeni, da se prepoznani večbesedni diskurzivni označevalci govornenega jezika **v govorni komunikaciji v povprečju rabijo skoraj štirikrat pogosteje kot v pisnih besedilih**. Čeprav nam tudi na tem mestu za celotno sliko primerjave

¹¹² Pri pridobivanju pogostosti v korpusu Kres kot variant nismo obravnavali morebitnih variantnih nestandardnih zapisov (npr. variante *zarad tega* označevalca *zaradi tega*), saj predpostavljamo, da so te v korpusu Kres redke, hkrati pa jih ni mogoče sistematično predvideti.

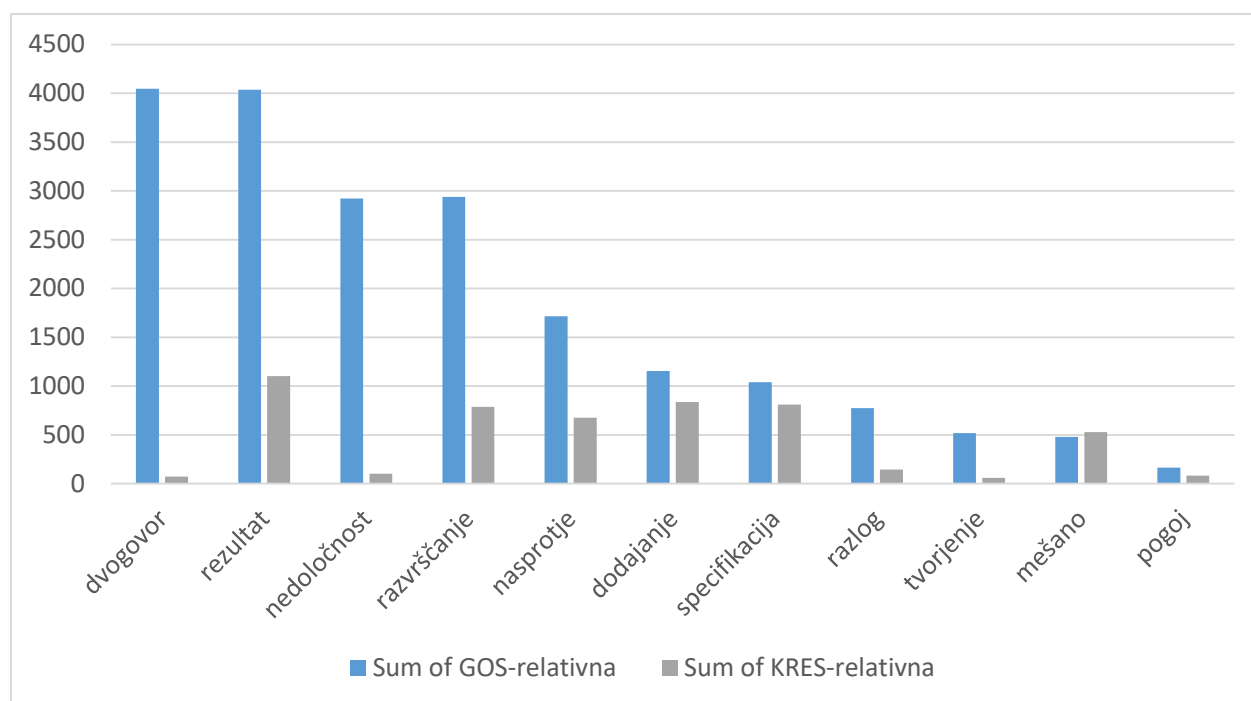
rabe diskurznofunkcijske leksike med govorjeno in pisno komunikacijo umanjka podatek o rabi funkcijsko analognih enobesednih označevalcih in hkrati tudi o morebitnih specifičnih označevalcih pisnega jezika, pa ta podatek vendarle sugerira sklep, **da je eksplicitno upravljanje diskurza bistveno pomembnejši mehanizem v govorjeni kot v pisni komunikaciji.**

Če v nadaljevanju rabo označevalcev primerjamo še glede na dve krovni funkcijski skupini – označevalce interakcije (dvogovor, nedoločnost, tvorjenje) na eni strani in besedilne konektorje (rezultat, razvrščanje, nasprotje, dodajanje, specifikacija, razlog, pogoj) na drugi – primerjava (Slika 13) razkriva tri ključne ugotovitve. Prva je (pričakovana) ugotovitev, da med raznolikimi funkcijami diskurznega upravljanja v govorjenem diskurzu glede na pisni diskurz dejansko **izstopajo tiste, povezane s specifičnimi dejavniki spontano govorjene komunikacije**, kot so uravnavanje medosebnih odnosov, relativizacija izrečenega in vzpostavljanje tekočnosti govora, ki so zaradi drugačnih okoliščin pisnega razumevanja manj pomembne. Tudi sicer se večina zaznanih primerov rabe označevalcev interakcije, kot so npr. izrazi *a ne, ne vem, a veš, kot rečeno, veš kaj, bom rekel* ipd., v korpusu Kres pojavlja v premem govoru leposlovnih besedil, prepisih sej državnega zbora in spletnih besedilih (besedilnozvrstna kategorija Drugo).



Slika 13: Primerjava pogostosti rabe večbesednih diskurznihih označevalcev v korpusih Gos in Kres.

Nekoliko bolj presenetljiva je druga ugotovitev (Slika 13), da v govorjeni komunikaciji po pogostosti rabe ne izstopajo le označevalci interakcije, vezani na specifične situacijske dejavnike govorjenega sporazumevanja, kot sta prisotnost sogovornika in pritisk tvorjenja v realnem času, temveč tudi označevalci drugih funkcijskih tipov, saj je **tudi raba besedilnih povezovalcev (opomenjevalnih konektorjev) v govorjenem jeziku v povprečju 2,5-krat pogostejša kot v pisnem jeziku**. Podrobnejša primerjava posameznih funkcijskih skupin na Sliki 14 potrjuje, da to velja prav za vse skupine besedilnih konektorjev, še zlasti pa izstopa pogosta raba označevalcev razloga (5,3-krat pogostejša raba), označevalcev razvrščanja (3,7-krat pogostejša raba), in označevalcev rezultata (3,7-krat pogostejša raba).



Slika 14: Primerjava pogostosti rabe večbesednih diskurzivnih označevalcev v korpusih Gos in Kres za posamezne funkcijske skupine

Med besedilnimi konektorji, pri katerih je razlika v pogostosti rabe med govorjenim in pisnim diskurzom največja, resda prevladujejo taki, ki lahko **poleg vzpostavljanja semantičnih povezav s širjenjem dometa razvijejo tudi različne vrste pragmatičnih, tj. retoričnih in/ali sekvenčnih funkcij**, kot so označevalci *tako da*, *v bistvu*, *ali pa*, *pol pa*, *se pravi*, presenetljivo pa podatki kažejo, da govorci v govorjenem jeziku pogosteje uporabljajo **tudi funkcijsko enovite označevalce**, za katere bi težko trdili, da so v govoru pogosteje rabljeni zaradi tega, ker bi tam

opravljali drugačne funkcije kot v pisnem jeziku, npr. označevalec specifikacije *v smislu da* (166-krat pogostejša raba v korpusu Gos kot v korpusu Kres), označevalec sklepa ali specifikacije *to se pravi* (28-krat pogostejša raba v korpusu Gos kot v korpusu Kres), označevalec dodajanja *in pa* (19-krat pogostejša raba v korpusu Gos kot v korpusu Kres), označevalec rezultata *glede na to* (18-krat pogostejša raba v korpusu Gos kot v korpusu Kres), označevalec namena *zaradi tega da* (15-krat pogostejša raba v korpusu Gos kot v korpusu Kres) ipd.

Vse tri ugotovitve torej potrjujejo **pomen eksplicitnega upravljanja diskurza v govornem jeziku**, pri čemer pa v primerjavi s pisnim jezikom v govoru ne izstopa zgolj raba jezikovnih sredstev, vezanih na specifične okoliščine tvorjenja spontane interakcije, temveč se govorci pogosteje kot v pisni komunikaciji poslužujejo tudi sredstev za vzpostavljanje smiselne povezanosti posredovanih sporočil.

4.4.3 Tipično govorni večbesedni diskurzni označevalci

4.4.3.1 Pregled metod za izračun ključnosti

V zadnjem koraku analize večbesednih diskurzni označevalcev v korpusu govorne slovenščine smo nato skušali ugotoviti še, kateri konkretni označevalci se tipično uporabljajo predvsem v govornem jeziku. Za ta namen se v metodologiji korpusnega jezikoslovja običajno uporabljajo t. i. **izračuni ključnosti** (angl. *keyness*), ki pogostost pojavljanja določenega jezikovnega pojava v opazovanem korpusu (v našem primeru korpusu Gos) primerjajo s pogostostjo pojavljanja v referenčnem korpusu (v našem primeru korpusu Kres) in na podlagi tega obravnavane pojave razvrščajo glede na njihovo nenavadno visoko pogostost pojavljanja v opazovanem korpusu (Scott, 1997). Tako kot se v korpusnem jezikoslovju še naprej preizprašuje sam teoretski koncept tipičnosti oz. ključnosti (Bondi in Scott, 2010; Rayson, 2013), obstajajo tudi številne različne metode za njeno prepoznavanje, ki jih lahko v splošnem razdelimo na dve skupini.

Na eni strani najdemo metode, ki izhajajo iz klasičnih statističnih metod za **izračun statistične pomembnosti** (angl. *significance test statistics*), ki povedo, kako pomembna oz. izstopajoča je razlika v pogostosti rabe določenega jezikovnega pojava v opazovanem korpusu glede na referenčni korpus. Čeprav se je včasih za ta namen uporabljala tudi mera χ^2 (Scott, 1997), so poznejše raziskave pokazale, da je med različnimi merami za izračun ključnosti ustrežnejša mera logaritma verjetnosti (LL), saj je manj občutljiva na velikost primerjanih korpusov in samo

pogostost opazovanih nizov (A. Baron idr., 2009). Tako kot druge mere za izračun statistične pomembnosti se vrednost logaritma verjetnosti običajno pretvarja v vrednost p ,¹¹³ ki nam pove, kakšna je verjetnost, da je do razlike v pogostosti rabe prišlo po naključju (Scott, 2010, str. 48), oziroma povedano drugače, »manjša, kot je vrednost p , večja je verjetnost, da pogosto pojavljanje določene besede v opazovanem korpusu ni zgolj naključje, temveč rezultat avtorjeve zavestne ali nezavestne izbire, da to besedo uporablja pogosteje« (Baker, 2006, str. 126).

Na drugi strani najdemo metode z **izračuni velikosti učinka** (angl. *effect size statistics*), ki nam namesto pomembnosti razlike povedo, kako velika je dejansko razlika v pogostosti rabe med enim in drugim korpusom. Za razliko od metod za izračun pomembnosti, kjer za izračun ključnosti prevladuje raba mere LL, so bile med merami za izračun velikosti učinka doslej predlagane številne različne mere (Rayson, 2008b), npr. mera %DIFF (Gabrielatos in Marchi, 2012), mera BIC (A. Wilson, 2013), velikost učinka za LL (Johnston idr., 2006), razmerje obetov, angl. *odds ratio* (Gries, 2014, 2015), razmerje logaritma, angl. *log ratio* (Hardie, 2014), razmerje relativnih frekvenc (Kilgariff, 2001, 2012; Gries, 2010) itd. Pri tem se tudi pri teh merah kot komplementarna informacija pri analizi najvišje uvrščenih ključnih jezikovnih pojavov najpogosteje upošteva tudi podatek o statistični pomembnosti razlike, kot denimo v zadnji različici programa WordSmith¹¹⁴ (Scott, 2016), kjer se pri izračunu ključnosti upoštevata tako razmerje logaritma (log ratio) kot logaritem razmerja verjetnosti (LL). Čeprav Gabrielatos in Marchi (2012; 2013) na podlagi podrobne analize 100 najbolj ključnih besed glede na razvrščanje z mero statistične pomembnosti (LL) in razvrščanje z mero velikosti učinka (%DIFF) trdita, da so testi statistične pomembnosti za luščenje ključnega besedišča manj primerni kot mere za izračun velikosti učinka, Pojanapunya in Todd (2016) v svoji izčrpni primerjavi mer LL in OR¹¹⁵ ugotavljata, da je namesto (ne)ustreznosti obeh krovnih metod ustrezneje govoriti o njuni komplementarnosti, saj z različnimi seznamom ključnih besed služita tudi različnim

¹¹³ Vrednost LL 3,84 pomeni $p \leq 0,05$, vrednost LL 6,63 pomeni $p \leq 0,01$, vrednost LL 10,83 pomeni $p \leq 0,001$, vrednost LL 15,13 pomeni $p \leq 0,0001$, povzeto po (Rayson, 2008b).

¹¹⁴ http://www.lexically.net/downloads/version7/HTML/index.html?keywords_calculate_info.htm

¹¹⁵ Čeprav se avtorja med več obstoječimi merami velikosti učinka odločita za mero razmerij obetov (OR), poudarjata, da zelo podobne rezultate (tj. identičen seznam 500 najbolj ključnih besed) kažejo tudi druge sorodne mere, kot so %DIFF, relativno tveganje, razmerje logaritma, TF-IDF (Pojanapunya in Todd, 2016, str. 16).

raziskovalnim ciljem. Medtem ko mera statistične pomembnosti LL izpostavlja besede, ki so v opazovanem korpusu rabljene razmeroma pogosto (žanrska specifičnost), mera velikosti učinka OR izpostavlja predvsem specializirano besedišče danega korpusa (terminološka specifičnost).

Tudi v slovenskem prostoru so se doslej uporabljale tako mere za izračun statistične pomembnosti (Puc in Erjavec, 2006; Kosem in Verdonik, 2012; Pollak in Bo, 2014) kot mere za izračun velikosti učinka (Zwitter Vitez in Fišer, 2015a; Verdonik in Maučec, 2016), pri čemer je bila izbira mere največkrat utemeljena z možnostmi in omejitvami izbranega orodja za korpusno analizo (npr. WordSmith na eni in SketchEngine na drugi strani), manj pa z vidika njene ustreznosti glede na raziskovalne cilje.

Ker želimo v pričujoči raziskavi identificirati predvsem označevalce, ki so tipični za govorjeno komunikacijo, ne pa zgolj označevalce, ki se tipično uporabljajo zgolj v govorjenem jeziku, smo tako kot sorodne diskurznoanalitične korpusne raziskave (McEnery idr., 2006; Simpson-Vlach in Ellis, 2010) tudi sami za razvrstitev označevalcev glede na tipičnost za korpus Gos uporabili **mero za izračun statistične pomembnosti LL**, ki pove, kako signifikantna je razlika v relativni pogostosti rabe posameznega označevalca med obema korpusoma (Dunning, 1993; Rayson, 2008a). Zanj smo uporabili poenostavljeno enačbo, kakršna je predstavljena v (Rayson, 2013), ki upošteva absolutno pogostost pojavljanja označevalca v opazovanem korpusu (a), v našem primeru korpusu Gos, absolutno pogostost pojavljanja označevalca v referenčnem korpusu (b), v našem primeru korpusu Kres, ter siceršnjo velikost obeh korpusov, opazovanega (C) in referenčnega (D).

$$LL = 2 \cdot \left(a \cdot \ln \frac{a}{E1} + b \cdot \ln \frac{b}{E2} \right), \text{ pri čemer } E1 = \frac{C \cdot (a+b)}{C+D} \text{ in } E2 = \frac{D \cdot (a+b)}{C+D}$$

4.4.3.2 Seznam tipično govorjenih večbesednih diskurznihih označevalcev

Statistično pomembno pogostejšo rabo (vrednost LL nad kritično mejo 6,63, ki ustreza vrednosti p , manjši ali enaki kot 0,01) izkazuje kar 174 od 204 označevalcev, kar pomeni, da se **kar 85 %**

vseh obravnavanih označevalcev statistično gledano tipičneje uporablja v govorni kot v pisni rabi.¹¹⁶

Seznam 30 najbolj tipičnih označevalcev govornega jezika (Tabela 13) potrjuje predhodne ugotovitve, da med njimi prevladujejo predvsem označevalci dvogovora (7 označevalcev) in označevalci nedoločnosti (6), npr. *a ne, ne vem, a veš, ja no, pa to, ali kaj, pa tako* itd., s katerimi govorniki uravnavajo stik z naslovnikom in/ali pod pritiskom tvorjenja v realnem času ohranjajo tekočnost svojega govora. Sledijo jim označevalci rezultata (6), kot npr. izrazi *se pravi, tako da, v glavnem*, s katerimi govorniki poudarjajo uvajanje sklepa, pojasnila, povzetka ali tematskega prehoda, in označevalci razvrščanja (4), kot npr. izrazi *pol pa, zdaj pa, pol ko*, s katerimi govorniki označujejo sosledje dogodkov, o katerih pripovedujejo, napovedujejo prehode med tematskimi sekvencami ali uvajajo sklep (*pol pa, potem pa*). Med drugimi tipično govornimi označevalci izstopajo še označevalec nasprotja v *bistvu*, s katerim govorniki poudarjajo ilokucijsko silo sporočanega in nasprotje s pričakovanim, napovedujejo spremembo ubeseditve in/ali ohranjajo tekočnost govora in (glej podrobnejšo analizo v poglavju 4.1.3.1 *Problematičnost funkcijskega razdvajanja v kontekstu*), označevalec tvorjenja *saj pravim*, s katerim govorci poudarjajo ponovno izpostavitve že znanega mnenja, ter nekoliko presenetljivo tudi tipična besedilna konektorja *zato ker* in *zaradi tega ker*, s katerimi govorniki bodisi vzpostavljajo vzročno-posledično povezavo med dvema pripovedovanima dogodkoma bodisi uvajajo pojasnilo za predhodno govorno dejanje.¹¹⁷

¹¹⁶ Preostalih 30 označevalcev s statistično nepomembno razliko v rabi in/ali pogostejšo rabo v korpusu Kres: *ne pa da* (22 pojavitev na milijon pojavnic v korpusu Gos / 12 pojavitev na milijon pojavnic v korpusu Kres), *pred tem* (19/32), *še nekaj* (9/3), *tako torej* (4/1), *kot da* (136/164), *namesto da* (23/35), *v tem primeru* (78/62), *kljub vsemu* (16/25), *sicer pa* (81/98), *ob tem da* (6/3), *v nadaljevanju* (49/61), *tako rekoč* (30/38), *za to da* (17/12), *razen če* (25/29), *sploh pa* (22/17), *ne glede na to* (5/3), *glede tega* (13/17), *danes ko* (8/6), *zato pa* (39/34), *s tega vidika* (8/11), *še posebej če* (7/9), *samo če* (16/13), *kakorkoli že* (15/12), *ali ne* (7/5), *v tem času* (41/44), *ob koncu* (8/7), *po tem* (9/8), *v resnici* (108/111), *kljub temu da* (63/32), *po svoje* (13/13).

¹¹⁷ Za primerjavo navedimo še, da se z alternativnim izračunom ključnosti označevalcev glede na razmerje obetov (angl. *odds ratio*) na seznam 30 najvišje uvrščenih označevalcev (Priloga 5) umeščajo predvsem označevalci, ki se v korpusu Kres ne pojavijo ali se v njem pojavijo zelo redko, npr. *pa tako* (110 pojavitev v korpusu Gos, 0 pojavitev v korpusu Kres), *in tako* (66, 0), *ali pa kaj takega* (41, 1), *na vem* (26, 0), *ne znam* (24, 0), *ali nekaj* (20, 0), *in vse* (20, 0), *ne vem kaj* (17, 0), *znaš kaj* (16, 0), *veš kaj je* (15, 0), *tako ka* (52, 4), *ali pa nekaj* (10, 0), *ali kaj je* (10, 1), *kaj takega* (10, 0), *kaj je* (9, 0) itd. Če torej izračun logaritma verjetnosti (LL) izpostavlja najpogostejše, tipične govorne označevalce, izračun razmerja obetov (OR) izpostavlja predvsem najunikatnejše, tipično govorne

4.4 Tipičnost za govorjeni jezik

	označevalec	skupina	Abs. pog. Gos	Abs. pog. Kres	Ključnost (LL)
1	<i>a ne</i>	dvogovor	1.726	1.038	12.267,9
2	<i>ne vem</i>	nedoločnost	1.399	310	11.280,7
3	<i>a veš</i>	dvogovor	705	111	5.848,4
4	<i>pol pa</i>	razvrščanje	714	901	4.379,7
5	<i>v bistvu</i>	nasprotje	985	4.920	3.852,6
6	<i>ja no</i>	dvogovor	328	142	2.449,8
7	<i>se pravi</i>	rezultat	666	3.856	2.433,1
8	<i>tako da</i>	rezultat	1.804	42.181	2.414,9
9	<i>pa to</i>	nedoločnost	339	240	2.345,1
10	<i>ali kaj</i>	dvogovor	345	509	2.035,4
11	<i>zdaj pa</i>	razvrščanje	693	8.081	1.700,6
12	<i>v glavnem</i>	rezultat	193	379	1.051,6
13	<i>pa tako</i>	nedoločnost	110	1	1.000,4
14	<i>v redu</i>	dvogovor	183	536	883,2
15	<i>ali pa</i>	dodajanje	815	23.052	866,2
16	<i>veš kaj</i>	dvogovor	176	563	821,4
17	<i>saj pravim</i>	tvorjenje	119	89	811,8
18	<i>zato ker</i>	razlog	498	9.923	787,4
19	<i>in tako naprej</i>	nedoločnost	235	1.820	741,1
20	<i>tako kot</i>	specifikacija	315	5.039	606,8
21	<i>in tako</i>	nedoločnost	66	1	593,5
22	<i>ma ja</i>	dvogovor	91	90	589,6
23	<i>to se pravi</i>	rezultat	126	454	564,0

označevalce. Če seznam glede na razmerje obetov nadalje zamejimo samo na enote, ki se v korpusu Gos pojavijo vsaj 40-krat, se med najvišje 30 uvrščenimi enotami pojavi 20 enot, ki jih med 30 ključnih izrazov umešča tudi mera LL, pri čemer se na prvih treh mestih pojavljajo označevalci *pa tako*, *in tako* in *ali pa kaj takega*.

4.4 Tipičnost za govorni jezik

24	<i>zaradi tega ker</i>	razlog	170	1.220	559,4
25	<i>drugače pa</i>	nasprotje	164	1.533	462,5
26	<i>tako ka</i>	rezultat	52	4	450,5
27	<i>pol ko</i>	razvrščanje	65	61	425,0
28	<i>bi rekel</i>	tvorjenje	89	291	414,1
29	<i>potem pa</i>	razvrščanje	507	17.678	397,2
30	<i>ali pa kaj takega</i>	nedoločnost	41	1	368,4

Tabela 13: 30 najbolj tipičnih diskurzni označevalcev v korpusu Gos glede na korpus Kres na podlagi izračuna ključnosti s pomočjo logaritma verjetnosti (LL).

Če povzamemo, smo torej s primerjavo pogostosti rabe večbesednih diskurzni označevalcev govornega jezika v govornem in pisnem diskurzu ugotovili, da se v govoru bistveno pogosteje kot v pisni rabi uporabljajo predvsem označevalci, vezani na specifične situacijske okoliščine govorne interakcije (prisotnost naslovnika, tvorjenje v realnem času), kar **potrjuje prvo hipotezo pričujoče disertacije**, da se v govoru tipično uporabljajo predvsem diskurzni označevalci interakcije. Vendarle pa ta **sklep drži le delno**, saj tako primerjava pogostosti rabe označevalcev glede na posamezne funkcijske skupine (poglavje 4.4.2.2 Pogostost rabe v korpusu Kres) kot seznam najbolj tipičnih označevalcev (poglavje 4.4.3.2 Seznam tipično govornih večbesednih diskurzni označevalcev) kažeta, da je **za govor tipična funkcija diskurznega upravljanja nasploh**, saj so poleg označevalcev interakcije v govornem diskurzu pogosteje rabljeni tudi drugi tipi diskurznofunkcijske leksike, kot so semantični in/ali pragmatični besedilni povezovalci oz. konektorji. Med statistično izstopajoče označevalce govornega jezika se tako umešča zelo **širok in heterogen nabor izrazov**, ki jih ni mogoče opazovati kot funkcijsko homogeno množico.

5 Večbesedni diskurzni označevalci govornega jezika v uporabniških spletnih vsebinah

V tem poglavju predstavljamo rezultate analize rabe najbolj tipičnih večbesednih diskurzni označevalcev govornega jezika v korpusu uporabniških spletnih vsebin Janes glede na govorni korpus Gos in pisni korpus Kres, da bi odgovorili na raziskovalna vprašanja »Kako pogosto se tipično govorni večbesedni diskurzni označevalci rabijo v uporabniških spletnih vsebinah?« (poglavje 5.1 *Pogostost rabe v uporabniških spletnih vsebinah*), »Ali je pogostost rabe tipično govornih večbesednih diskurzni označevalcev odvisna od tipa uporabniških spletnih vsebin?« (poglavje 5.2 *Raba glede na tip uporabniških spletnih vsebin*) in »Ali tipično govorni večbesedni diskurzni označevalci v uporabniških spletnih vsebinah opravljajo enako vlogo kot v govorni rabi?« (poglavje 5.3 *Značilnosti rabe v uporabniških spletnih vsebinah*).

Glede na zelo širok in funkcijsko heterogen nabor izrazov, ki so bili kot tipično govorni prepoznani v poglavju 4.4.3.2 *Seznam tipično govornih večbesednih diskurzni označevalcev*, pa smo se pri tem v nadaljnji analizi rabe tipično govornih večbesednih diskurzni označevalcev govorne slovenščine v uporabniških spletnih vsebinah namesto kvantitativne distribucijske analize celotne množice prepoznanih tipično govornih označevalcev osredotočili na **študijo primera treh najbolj tipičnih označevalcev, tj. označevalcev *a ne*, *a veš* in *ne vem***, ki omogoča sistematičnejšo in metodološko zanesljivejšo raziskavo.¹¹⁸ Glede na splošno neraziskanost prehajanja (diskurzno)funkcijske leksike in govorne v spletno slovenščino namreč menimo, da lahko analiza rabe izbranih treh označevalcev ponudi relevantne odgovore na v uvodu predstavljene hipoteze ter izhodišča za nadaljnje raziskave, saj vsi trije označevalci spadajo v najbolj tipični skupini govornih označevalcev, tj. označevalce dvogovora in/ali nedoločnosti, so visoko gramatikalizirani in v govoru pogosto rabljeni, hkrati pa so med seboj izrazno in funkcijsko dovolj različni, da osmišljajo tudi medsebojno primerjavo (v skladu z metodološkimi koraki, opisanimi v poglavju 3.3 *Struktura raziskave*).

¹¹⁸ Z metodološkega vidika je še zlasti problematično dejstvo, da sta statistično najbolj izstopajoči skupini označevalcev nedoločnosti in dvogovora obenem tudi izrazno najbolj dvoumni (Slika 11).

5.1 Pogostost rabe v uporabniških spletnih vsebinah

V primerjavi s pisno rabo drugih jezikovnih zvrsti je ena glavnih izraznih značilnosti uporabniških spletnih vsebin oz. računalniško posredovane komunikacije nasploh predvsem izstopajoča pogostost nestandardnih jezikovnih prvin, ki na leksikalni ravni poleg izbire nestandardnega besedišča (npr. *glih* namesto *ravno*) ali pregibnih oblik (npr. *sve* namesto *sva*) vključuje zlasti različne oblike nestandardnega zapisa, najpogosteje pod vplivom posebnosti v izgovoru (npr. *useen* namesto *vseeno*). Za razliko od korpusa govorjene slovenščine Gos, ki na pogovorni ravni prepisa izkazuje podobno mero variantnosti¹¹⁹ in v katerem so bile standardizirane ustreznice pogovornim oblikam pripisane ročno, je bil proces dodajanja normalizirane ravni zapisa (prevod iz izvornih, zapisanih oblik v neznaznamovane, standardne oblike) povsem avtomatiziran. Ker program za normalizacijo nestandardnih zapisovalnih posebnosti spletne slovenščine, uporabljen v korpusu Janes (Ljubešić idr., 2014), trenutno dosega omejeno natančnost od 76,3 % za neznano besedišče do 85,6 % za znano besedišče,¹²⁰ smo v začetnem koraku pridobivanja podatka o pogostosti rabe obravnavanih označevalcev v korpusu Janes glede na njihovo maloštevilnost **najprej skušali identificirati vse različice njihovega izvirnega zapisa**, da bi se tako izognili opiranju na normalizirano raven zapisa.

Pri identifikaciji zapisovalnih posebnosti smo se pri tem oprli na tipologijo označevanja nestandardnih prvin, ki nastaja v okviru raziskave regionalne jezikovne variantnosti besedil v korpusu Janes (Čibej, v pripravi). Med številnimi predlaganimi kategorijami znotraj krovnih kategorij nestandardnega besedišča (npr. *zihr*, *omg*), zapisovalnih različic pogostega besedišča (npr. *tolk* za *toliko*, *jst* za *jaz*), alternativnih zapisov (npr. *useen* za *vseeno*, *biu* za *bil*), črkovnih redukcij (npr. *kradl* za *kradli*), nestandardnega oblikoslovja (npr. *šou* za *šel*), pogostih transformacij (npr. *neki* za *nekaj*) in drugih posebnosti (npr. zapis skupaj ali narazen), smo za

¹¹⁹ Med 61.706 različnimi standardiziranimi pojavnici korpusa Gos se jih kar 8.973 (14,5 %) pojavi z vsaj dvema različnima pogovornima zapisoma, npr. pogovorne oblike *noter*, *not*, *notr*, *nuot*, *neutr*, *nut* itd. za standardizirano obliko *noter*.

¹²⁰ Omejeno natančnost normalizacije v korpusu Janes potrjuje tudi pilotna analiza zapisovalnih variant obravnavanih označevalcev. Iskanje po normaliziranem nizu *a ne* denimo poleg relevantnih variant *a ne*, *a neee*, *a neeee* ipd. prepozna tudi konkordance z nizi *apriori ne*, *a nevem*, *a neighbour* ipd., po drugi strani pa pogosta oblika zapisa skupaj *ane* v korpusu nikoli ni normalizirana v razdružen zapis *a ne*.

identifikacijo zapisovalnih posebnosti obravnavanih izrazov *a ne*, *a veš* in *ne vem* kot relevantne prepoznali zlasti kategorije, povezane z alternativnim zapisom soglasnikov (npr. *ne wem*), izpustom končnega samoglasnika (npr. *a n*), izpustom nekončnega nepoudarjenega samoglasnika (npr. *ne vm*) in variantnim zapisom skupaj (npr. *aveš*). Poleg teh kategorij smo na podlagi komplementarnih raziskav kot relevanten pojav prav tako prepoznali preglas samoglasnika *e* v *i* (npr. *a wiš*), zapis z angleškimi digrafi (npr. *a vesh*), spremembo šumnika v sičnik (npr. *a ves*) in pojav multiplikacije posamičnih črk (npr. *a neee*). Na podlagi tega smo nato z oblikovanjem več različnih hevrističnih iskalnih pogojev v iskalnem jeziku CQL preko konkordančnika noSketchEngine¹²¹ pridobili razmeroma izčrpen seznam zapisovalnih variant posameznih označevalcev (45 za niz *a veš*, 44 za niz *ne vem* in 38 za niz *a ne*, če združimo zapise, ki se razlikujejo zgolj v velikosti črk). Pričakovano se vsi trije nizi najpogosteje pojavljajo v svoji standardni varianti, vendarle pa statistično zelo izstopajo tudi nekatere nestandardne variante – standardna oblika *ne vem* se v korpusu pojavi samo 6-krat pogosteje kot nestandardni zapis *nevem*, zapis *a ne* pa samo dvakrat pogosteje kot nestandardni zapis *ane*). Ker je veliko zapisovalnih variant rabljenih zelo redko, smo kot relevantne v nadaljnji analizi upoštevali samo tiste z vsaj pet pojavitvami, tj. **17 različnih zapisov niza *ne vem*** (*ne vem*, *nevem*, *nwm*, *nvm*, *newem*, *nvem*, *nwem*, *neum*, *ne wem*, *num*, *n vem*, *ne vm*, *nevm*, *ne veeeem*, *ne um*, *neuem*, *nevemm*), **14 različnih zapisov niza *a ne*** (*a ne*, *ane*, *anne*, *a n*, *aneee*, *a neee*, *aaa ne*, *anee*, *a neeee*, *aneeee*, *aaaa ne*, *aneeeee*, *aa ne*, *aaaaa ne*) in **11 različnih zapisov niza *a veš*** (*a veš*, *a ves*, *aveš*, *aves*, *a viš*, *a wiš*, *a veeš*, *a veeeeš*, *a veeeeš*, *aveeeeš*, *a wes*).

V drugem koraku smo na podlagi identificiranega nabora možnih zapisov za vsakega izmed označevalcev nato pridobili seznam vseh njegovih konkordanc, pri čemer smo za izračun pogostosti diskurzne rabe tudi v korpusu Janes za razdvoumljanje posameznih oblik uporabili metodo sklepanja po vzorcu (glej poglavje 4.4.1.2 *Metoda sklepanja po vzorcu*). Ker se v tem sklopu raziskave osredotočamo na zelo majhen nabor označevalcev, v nadaljevanju pa nas bo

¹²¹ Iskalni pogoj "[Nn]+([Ee]+)?" "([Vv]+[Ww]+[Uu]+)([Ee]+[Ii]+)?[Mm]+" denimo poišče vse pojavitve dveh pojavnic, v katerem prvo pojavnico sestavlja poljubno dolg niz črke n, ki mu sledi nič ali več pojavitev črke e (npr. *ne*, *NE*, *n*, *nne*, *neee*), drugo pojavnico pa sestavlja poljubno dolg niz črk v, w ali u, ki mu sledi nič ali več pojavitev črk e ali i ter poljubno dolg niz črke m (npr. *vem*, *VEM*, *wem*, *veeem*, *vemm*, *vim*, *vm*). S tem v korpusu prepoznamo 38 različnih načinov zapisa niza *ne vem* brez upoštevanja zapisovanja skupaj.

5.1 Pogostost rabe v uporabniških spletnih vsebinah

zanimala tudi pogostost njihove rabe glede na posamezne tipe uporabniških besedil, smo namesto 100 naključnih konkordanc celotnega korpusa Janes v tej raziskavi analizirali 100 naključnih konkordanc vsakega izmed šestih krovnih podkorpusev (tviti, forumi, blogi, komentarji na bloge, komentarji na novice in pogovorne strani Wikipedije), torej skupno **600 konkordanc** za vsakega izmed treh označevalcev. Izračunane podatke o pogostosti znotraj posameznih podkorpusev smo nato sešteli v krovni podatek o pogostosti pojavljanja v celotnem korpusu Janes.

Kot prikazujejo pridobljeni podatki v Tabeli 14, se tako kot v govorjeni rabi med tipično govorjenimi označevalci v uporabniških spletnih vsebinah najpogosteje pojavlja označevalec *a ne* (152 pojavitev na milijon pojavnic), približno enkrat redkeje se uporablja označevalec *ne vem* (83 pojavitev na milijon), najredkeje pa se uporablja označevalec *a veš* (36 pojavitev na milijon).¹²² To dokazuje, da tipično govorjeni diskurzni označevalci **z nezanemarljivo pogostostjo prehajajo tudi v spletno rabo**, na podlagi česar lahko **ovržemo drugo hipotezo raziskave**, da tipično govorjeni diskurzni označevalci v uporabniške spletne vsebine ne prehajajo, kakršna je bila sugerirana tudi v nekaterih starejših raziskavah.¹²³

označevalec	pogostost oblike	ocenjena pogostost označevalca	relativna pogostost označevalca
<i>a ne</i>	38.808	26.567	151,6 poj./milj.
<i>ne vem</i>	111.002	12.844	73,3 poj./milj.
<i>a veš</i>	10.028	6.252	35,7 poj./milj.

Tabela 14: Pogostost rabe treh najbolj tipično govorjenih diskurznih označevalcev v korpusu Janes.

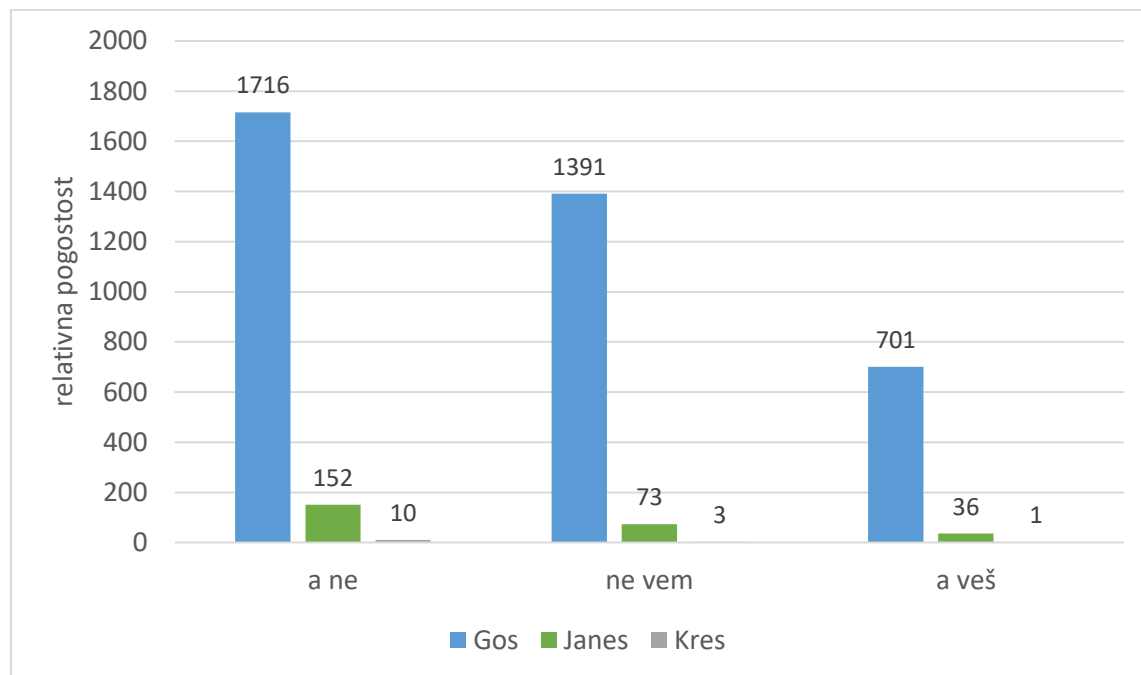
Nadaljnja primerjava z relativno pogostostjo rabe v korpusu govorjene slovenščine Gos in korpusu pisne slovenščine Kres (poglavje 4.4.2 *Pogostost rabe v govorjeni in pisni slovenščini*) na

¹²² Tako kot v korpusu Kres kot merilo velikosti korpusa Janes jemljemo število besednih pojavnic (vseh pojavnic razen ločil), tj. 175.244.406 pojavnic.

¹²³ D. Crystal (2011, str. 22) v razmisleku o možnostih raziskav prepleta govorjenega in spletnega jezika denimo ugotavlja: »To take an example: an important feature of informal conversation is the use of reaction signals, comment clauses, and tag questions (such as *mhm*, *you know* and *isn't it*), which give the listener the option of providing feedback. In a situation where this feedback is missing, will such features continue to be used, or will they be adapted in some way? My impression is that they are generally absent, but we need descriptive studies. [...] Some writers have suggested that the lack of these features is one of the reasons why so many Internet interactions are misperceived as abrupt, cold, distant, or antagonistic.«

5.1 Pogostost rabe v uporabniških spletnih vsebinah

Slika 15 kaže, da so tipično govorjeni diskurzni **označevalci v uporabniških spletnih vsebinah vendarle rabljeni izrazito manj pogosto kot v govoru**, saj se vsi trije analizirani označevalci v korpusu Janes pojavljajo precej bolj poredkoma kot v korpusu Gos, in sicer od 11-krat (*a ne*) do 20-krat (*a veš*) redkeje. To potrjuje, da je pogosta raba analiziranih diskurzni označevalcev v govoru **predvsem posledica tistih okoliščin tvorjenja govorjenega diskurza, ki so specifične zgolj za ta tip komunikacije**, zlasti tvorjenje v realnem času (prava sinhronost komunikacije) in deljenem prostoru (dejanska prisotnost naslovnika), manj pa druge okoliščine, ki jih izpostavljajo nekatere študije diskurzni označevalcev v govoru, kot so jasna struktura udeležencev, odnosov med njimi ter dialoška narava komunikacije, saj so te vsaj v določeni meri značilne tudi za komunikacijo uporabniških spletnih vsebin. Da pa njihov potencialni vpliv na rabo označevalcev vendarle ni nezanemarljiv, nakazuje pogostejša raba označevalcev v spletni pisni rabi glede na pisno rabo tradicionalnih medijev, saj so vsi trije označevalci v korpusu Janes rabljeni od 15-krat (*a ne*) do 32-krat (*a veš*) **pogosteje kot v korpusu Kres**.



Slika 15: Primerjava relativne pogostosti najbolj tipično govorjenih diskurzni označevalcev v korpusu govorjene slovenščine Gos, korpusu pisne slovenščine Kres in korpusu uporabniških spletnih vsebin Janes.

Sklenemo torej lahko, da se tipično govorjeni diskurzni označevalci za razliko od zelo redke rabe v tradicionalni pisni komunikaciji pojavljajo tudi v uporabniško pisni komunikaciji na spletu, a

vendarle bistveno redkeje kot v spontanem govoru, kar potrjuje, da je njihova raba v govorjeni komunikaciji pogojena predvsem s specifičnimi prostorskimi in kognitivnimi značilnostmi spontanega govora, ki jih ni mogoče pripisati drugim modalnostim. Dejstvo, da E. Fox Tree (2015) v svoji raziskavi rabe nekaterih tipično govorjenih diskurznih označevalcev v spletnih klepetalnicah ugotavlja, da sta označevalca *well* in *oh* tam rabljena približno enako pogosto kot v spontanem pogovoru (mi pa za vse tri označevalce ugotavljamo bistveno večji prepad), obenem sugerira, da je delež prehajanja leksikalnih prvin govorjenega jezika v računalniško posredovano komunikacijo močno odvisen od tipa tovrstne komunikacije (npr. spletne klepetalnice na eni in uporabniške spletne vsebine na drugi strani).

5.2 Raba glede na tip uporabniških spletnih vsebin

Kot smo izpostavili že v pregledu literature (poglavje 2.3.2 *Značilnosti uporabniških spletnih vsebin*), uporabniških spletnih vsebin kljub nekaterim skupnim značilnostim, kot so vezanost na razmeroma nove spletne platforme, podobna družbena vloga spodbujanja izmenjav in sodelovanja ali laičnost avtorjev vsebin, ni mogoče obravnavati kot enotne jezikovne zvrsti, saj na jezikovno rabo znotraj teh platform vplivajo številni tehnološki in družbeni dejavniki. V pričujočem poglavju nas zato zanima, ali lahko kot relevantno okoliščino, ki vpliva na stopnjo prehajanja tipično govorjenih diskurznihih označevalcev, upoštevamo tudi tip uporabniških besedil in z njim povezane dejavnike. Pri tem v korpusu Janes ločimo med šestimi tipi spletnih vsebin: besedila na družbenem omrežju Twitter, besedila na spletnih forumih, blogovski zapisi, komentarji na blogovske zapise, komentarji na spletne novice ter pogovorne strani slovenske Wikipedije.

Kot prikazuje primerjava skupne relativne pogostosti vseh treh označevalcev v posameznih besedilnih tipih na Sliki 16, med posameznimi besedilnimi tipi **dejansko prihaja do razlik** v pogostosti rabe analiziranih označevalcev. Ti so najpogostejše rabljeni na družbenem omrežju Twitter (323 pojavitev na milijon pojavnic), nekoliko manj pogosto na spletnih forumih ter v komentarjih na blogovske zapise in novice, bistveno redkeje pa so rabljeni na pogovornih straneh Wikipedije in v samih blogovskih zapisih. Kot dejavnik, ki najbolj vpliva na splošno rabo diskurznihih označevalcev v spletnih besedilih, se torej kaže predvsem **stopnja interaktivnosti oz. dialoščnosti** posameznih besedilnih tipov oz. platform, na katerih se ti pojavljajo, kar **potrjuje tretjo hipotezo naše disertacije**, da se diskurzni označevalci v uporabniških spletni komunikaciji rabijo predvsem v tipih s podobnimi sporočanjskimi okoliščinami.

Najbolj dinamična je zagotovo izmenjava besedil na družbenem omrežju Twitter, kjer uporabniki z namenom jedrnate izmenjave sporočil tvorijo kratka besedila in se pogosto zapletejo v diskusijo (naslovnik je največkrat konkreten). Prav tako je izmenjava mnenj več uporabnikov o isti temi eden izmed glavnih namenov spletnih forumov, kjer lahko posamezni uporabniki bodisi replicirajo na temo oz. vprašanje, ki ga je izpostavil prvi uporabnik v nizu, bodisi se odzivajo na mnenja drugih uporabnikov, daljše dialoške diskusije, kot jih denimo omogoča Twitter, pa so običajno sankcionirane s strani administratorjev. Podobno velja tudi za komentarje na statične

vsebine, kot so novice ali blogovski zapisi, ki sicer s tehnološkega vidika omogočajo podoben način izmenjave mnenj, vendar so v primerjavi s forumi bolj kot diskusiji (vstopanju v dialog z drugimi uporabniki) namenjeni podajanju enostranskih odzivov na izvorno statično vsebino. Blogovski zapisi so torej statična vsebina, brez jasno identificiranega naslovnika, zato jih v primerjavi z vsemi drugimi zgoraj naštetimi tipi ne moremo obravnavati kot odziv, temveč pobudo.

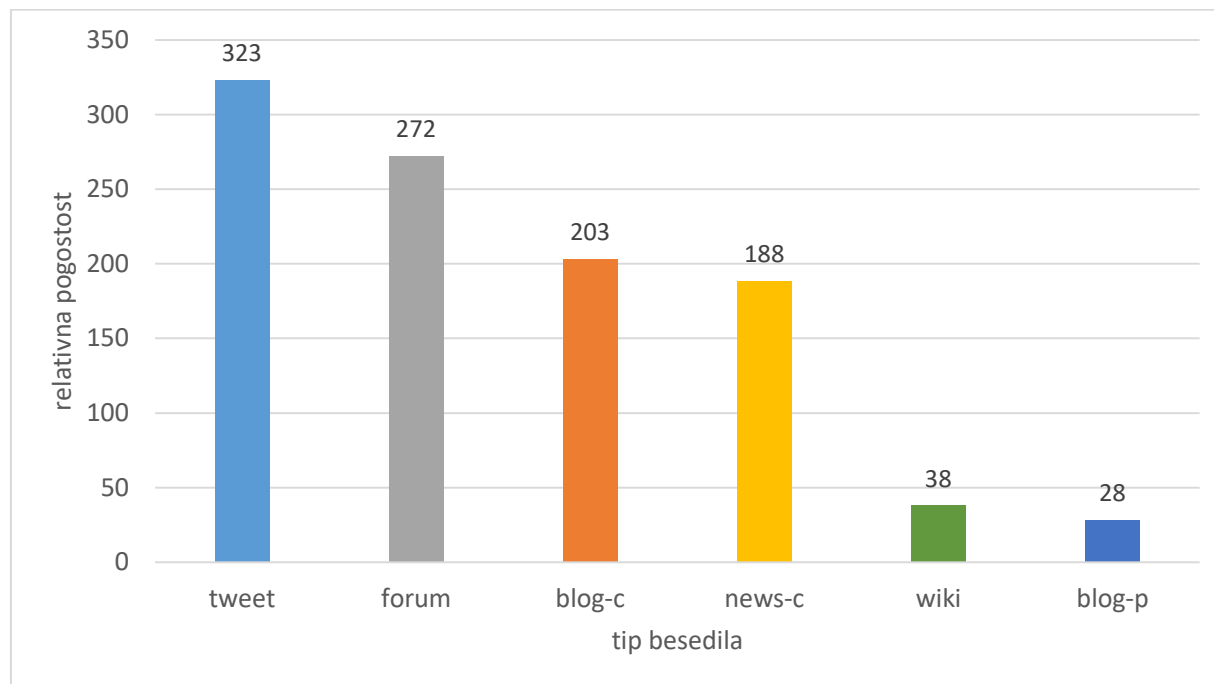
Če bi torej blogovske zapise v primerjavi z **dialoškimi** (twitter) in **diskusijskimi** besedili (forumi, komentarji) lahko umestili v tretjo skupino **statičnih** besedil in s tem kontinuumom od bolj proti manj interaktivnim besedilnim tipom pojasnili tudi upad rabe diskurznihih označevalcev, pa to predpostavko pod vprašaj postavlja **presenetljivo nizka raba diskurznihih označevalcev na pogovornih straneh Wikipedije**. Te so namreč prav tako namenjene izmenjavi mnenj med uporabniki, in sicer bodisi izražanju predlogov glede izboljšave vsebin posameznih geselskih člankov (geselske pogovorne strani)¹²⁴ bodisi pošiljanju sporočil posameznemu uporabniku Wikipedije (uporabniške pogovorne strani),¹²⁵ najpogosteje glede njegovih/njenih geselskih člankov ali (ne)spoštovanja pravil urejanja. Kot je razvidno iz obeh navedenih primerov, se v obeh podtipih uporabniških strani Wikipedije pogosto razvije diskusija v obliki več replik dveh ali več uporabnikov, podobna tistim na spletnih forumih. Pri pojasnjevanju izrazito redkejši rabe diskurznihih označevalcev bi zato poleg stopnje interaktivnosti (ki je vsekakor bližje dialoškosti kot statičnosti) veljalo upoštevati tudi druge specifične komunikacije na pogovornih straneh slovenske Wikipedije, kot so struktura udeležencev (majhna skupnost), odnosi med njimi

¹²⁴ Za primer glej diskusije na pogovorni strani članka Slovenska osamosvojitvena vojna: https://sl.wikipedia.org/wiki/Pogovor:Slovenska_osamosvojitvena_vojna.

¹²⁵ Za primer glej diskusije na strani uporabnika Eleassar: https://sl.wikipedia.org/wiki/Uporabni%C5%A1ki_pogovor:Eleassar.

5.2 Raba glede na tip uporabniških spletnih vsebin

(spoštljivost), namen platforme (izobraževanje), tehnološke omejitve (platforma ne spodbuja diskusij), obstojnost sporočil (dolgotrajno viden arhiv) itd.



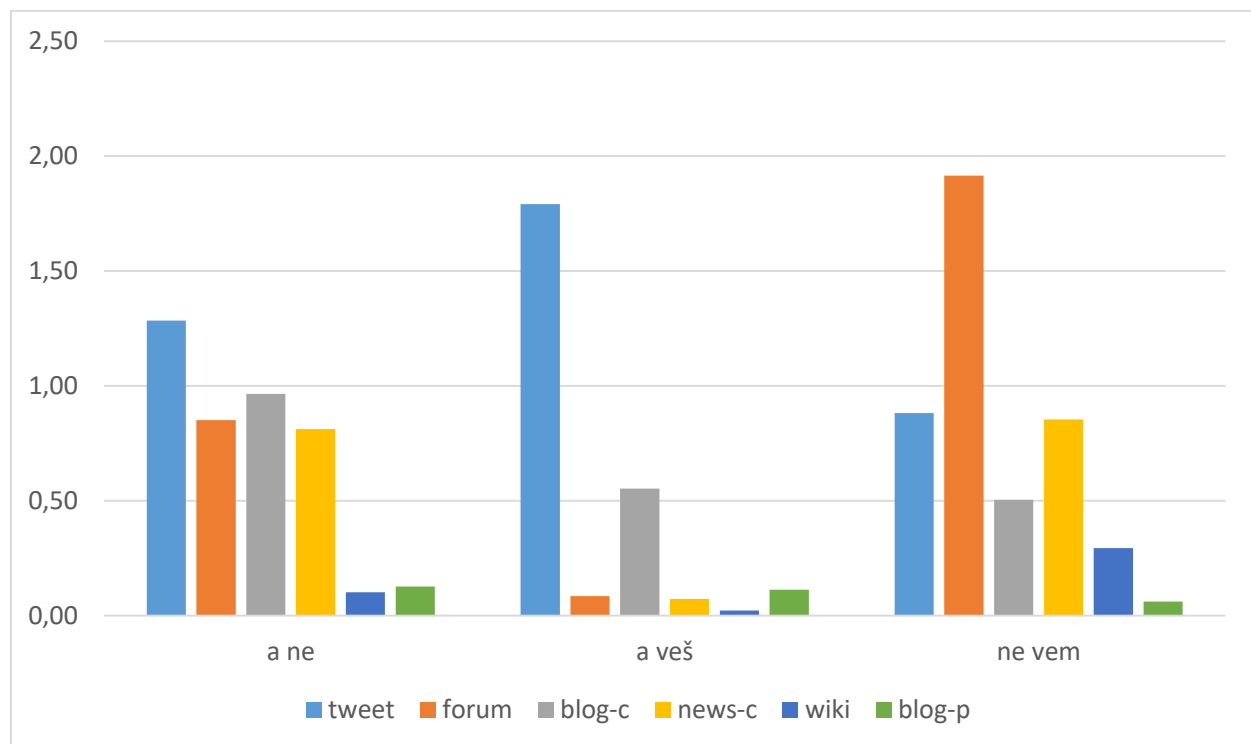
Slika 16: Primerjava relativne pogostosti najbolj tipično govorjenih večbesednih diskurzivnih označevalcev v šestih besedilnih tipih korpusa Janes.

Glede na izrazno in funkcijsko raznolikost vseh treh obravnavanih označevalcev nas je v naslednjem koraku zanimala še primerjava njihove individualne besedilnotipske distribucije. Ker korpus Janes ni besedilnozvrstno uravnotežen (vsi besedilni tipi niso zastopani z enakim številom pojavnic), hkrati pa se označevalci pojavljajo z različno pogostostjo, primerjavo v nadaljevanju ponazarjamo z izračunom t. i. relativne besedilnotipske pogostosti (angl. *relative text type frequency*), ki nam pove, kolikokrat bolj ali manj pogosta je raba označevalca v posameznem besedilnem tipu v primerjavi z njegovo rabo v celotnem korpusu.¹²⁶ Prikaz distribucije po posameznih besedilnih tipih za vsakega izmed označevalcev na Sliki 18 potrjuje, da kljub enotni

¹²⁶ Na podlagi formule, opisane na strani <https://www.sketchengine.co.uk/rel>, relativno besedilnotipsko pogostost izračunamo tako, da delež pojavitev v posameznem podkorpusu glede na vse pojavitve v korpusu (npr. 5768 pojavitev označevalca *a* veš v podkorpusu tvitov od skupno 6252 pojavitev v celotnem korpusu Janes, tj. 92,26 %) delimo z relativno velikostjo posameznega besedilnega tipa (npr. 51,52 % za podkorpus tvitov), pri čemer nam dobljeni količnik (v našem primeru 1,79) pove, kolikokrat pogostejša oziroma redkejša je raba določenega izraza v opazovanem podkorpusu glede na rabo v celotnem korpusu.

5.2 Raba glede na tip uporabniških spletnih vsebin

korelaciji med pogostejšo rabo v interaktivnejših besedilih (npr. tvitih) na eni strani in redkejšo rabo v statičnejših besedilih (npr. blogovskih zapisih) na drugi, **posamezni označevalci kažejo različne tendence v rabi glede na posamezne besedilne tipe**. Pri tem je zanimivo, da za vse tri označevalce nekateri tipi izkazujejo izrazito redko rabo (npr. 500-krat redkejša raba označevalca *a veš* na pogovornih straneh Wikipedije v primerjavi z rabo v uporabniških spletnih vsebinah nasploh), v nobenem izmed tipov pa se označevalci ne pojavljajo izredno pogosto, saj tudi najbolj izstopajoče pogosta raba označevalca *ne vem* v besedilih slovenskih forumov ne predstavlja več kot 1,9-kratnika povprečne rabe tega označevalca v korpusu nasploh.



Slika 17: Primerjava relativne besedilnotipske pogostosti posameznih označevalcev (po označevalcih).

Najbolj enakomerno rabo glede na različne tipe uporabniških spletnih vsebin izkazuje označevalec dvogovora *a ne*, s katerimi govorci nagovarjajo naslovnika, preverjajo njegovo pozornost, mu predajajo vlogo, poudarjajo pravkar izrečeno in zapolnjuje vrzeli. Ta označevalec se najpogosteje (1,3-krat pogosteje kot v korpusu nasploh) pojavlja na Twitterju, nekoliko redkejša in približno enakomerna pa je raba tega označevalca v spletnih forumih ter komentarjih na novice in bloge. Zelo redko se ta označevalec pojavlja na pogovornih straneh Wikipedije ali besedilih blogov (približno 10-krat manj pogosto kot v korpusu nasploh), kar lahko delno

pojasnimo s tem, da tvorci pri pisanju blogov redko nagovarjajo (konkretnega) naslovnika, saj so namenjeni splošnejši oz. širši zainteresirani javnosti.

Precej bolj raznolika je besedilnotipska distribucija označevalca *a veš*, ki se prav tako umešča med označevalce dvogovora (za preverjanje pozornosti, zapolnjevanje vrzeli, poudarjanje vsebine izrečenega, odpiranje tematskih sekvenc itd.), a se od označevalca *a ne* loči predvsem po tem, da z glagolsko obliko v drugi osebi ednine na izrazni ravni nagovarja točno določenega naslovnika, s katerim ima tvorec obenem vzpostavljen razmeroma neformalen odnos. Ni torej presenetljivo, da se ta označevalec najpogosteje rabi na družbenem omrežju Twitter (1,8-krat pogosteje kot v korpusu nasploh), na katerem uporabniki sami zamejujejo svoj krog sledilcev oz. naslovnikov posameznih tvitov, medtem ko tvorci besedil na drugih tipih platform običajno naslavljajo širok krog naslovnikov ali anonimne naslovnike, ki jih sicer ne poznajo.¹²⁷ Izjema so komentarji na blogovske zapise, v katerih se avtorji komentarjev (vključno s tvorcem izvirnega zapisa) v diskusiji pogosto tikajo (tudi z rabo označevalca *a veš*), kar lahko delno pojasnimo s tem, da so bili v korpus Janes zaradi tehničnih dejavnikov zajeti samo blogi na portalih publishwall.si in rrtv.si, ki imajo razmeroma majhno oz. zaprto skupino uporabnikov v primerjavi z nekaterimi drugimi blogerskimi portali s širšim razponom bralcev in komentatorjev.

Če sta oba označevalca dvogovora najbolj tipično rabljena na družbenem omrežju Twitter, manj tipična pa je njuna raba za druge tipe uporabniških spletnih vsebin, pa se označevalec nedoločnosti *ne vem* najpogosteje uporablja v besedilih spletnih forumov (1,9-krat pogosteje kot v korpusu nasploh). Glede na to, da označevalec poleg sekvencijskih funkcij zapolnjevanja vrzeli, odpiranja ali zapiranja vlog z vidika svojega pomenskega jedra pogosto opravlja tudi vlogo epistemskega omiljevalca, s katerim govorci relativizirajo svojo zavezanost k veljavnosti izrečenega (omiljevanje, ublažitev), je njegova pogosta raba na forumih na nek način pričakovana, saj je ta med vsemi drugimi tipi uporabniških spletnih vsebin edini eksplicitno namenjen izmenjavi uporabniških *mnenj*, ne pa drugim tipom dialoških izmenjav ali odzivov, kot je to značilno za komunikacijo na tviterju ali v spletnih komentarjih, kjer se ta označevalec rabi

¹²⁷ To velja tudi za pogovorne strani na Wikipediji, za katere Michelizza (2015) sicer ugotavlja, da naj bi jih zaznamoval razmeroma domač diskurz in vsesplošno tikanje uporabnikov.

nekoliko redkeje. Tako kot pri drugih označevalcih tudi *ne vem* ni pogost v blogovskih zapisih, v primerjavi z označevalcema dvogovora pa je pogostejše rabljen na pogovornih straneh Wikipedije, ki so prav tako namenjene izmenjavi mnenj glede rešitev za doseganja skupnega cilja, tj. izdelave kvalitetnega geselskega članka (Michelizza, 2015).

Pričujoče poglavje lahko torej sklenemo s krovno ugotovitvijo, da je pogostost spletne rabe tipično govornih večbesednih diskurzni označevalcev **odvisna tudi od tipa** uporabniških spletnih vsebin. Kot dejavnik, ki najbolj vpliva na delež rabe označevalcev, se pri tem kaže **stopnja interaktivnosti** oz. dialoški posameznih besedilnih tipov (bolj kot medij spodbuja dinamično izmenjavo besedil med uporabniki, pogostejša je raba diskurzni označevalcev), vendar podrobnejša primerjava distribucije posameznih označevalcev kaže, da je njihova raba odvisna tudi od funkcijskega razpona posameznega označevalca in njegovega **sovpadanja s specifičnimi družbenimi dejavniki posameznih platform**, kot so struktura udeležencev, odnosi med njimi in sporazumevalni namen.

5.3 Značilnosti rabe v uporabniških spletnih vsebinah

V zadnjem koraku raziskave najbolj tipično govorjenih večbesednih diskurznihih označevalcev smo po analizi kvantitativnih kazalnikov njihove rabe na spletu s kvalitativnim pregledom konkretnih primerov rabe v kontekstu želeli izvedeti še, ali obravnavani diskurzni označevalci v spletni komunikaciji opravljajo podobno vlogo kot v spontanem govoru. Kot referenčni nabor primerov rabe v obeh modalnostih smo pri tem obravnavali primere iz naključnih vzorcev, izdelanih za potrebe razdvoumljenega štetja v korpusu Gos (poglavje 4.4.2.1 *Pogostost rabe v korpusu Gos*) in korpusu Janes (poglavje 5.1 *Pogostost rabe v uporabniških spletnih vsebinah*), za vse tri označevalce pa jih navajamo v Prilogah 6–12.

5.3.1 Raba označevalca *a ne*

5.3.1.1 Raba označevalca *a ne* v korpusu Gos

Najbolj tipični večbesedni diskurzni označevalec govorjenega jezika *a ne* se v analiziranih primerih rabe korpusa Gos (Priloga 6) pojavlja v dveh tipičnih vlogah. Najpogostejša je raba označevalca v funkciji **zapolnjevalca vrzeli**, s katerim govorec označuje, poudarja oz. zapolnjuje prehode med deli besedila, kot so izjave, stavki ali nižje pomenonosne enote (primeri [145]–[147]). Pri tem tvorec na semantični ravni z rabo označevalca dvogovora obenem eksplicitno vzdržuje stik z naslovnikom, preverja njegovo pozornost, razumevanje ali strinjanje, vendar ne z namenom predajanja govorne vloge. Na izrazni ravni ta funkcija najpogosteje sovпада s padajočo intonacijo, najbolj značilno pa se uporablja v javnem informativno-izobraževalnem govoru (predavanja ali predstavitve) in pripovedih znotraj različnih tipov zasebne komunikacije.

[145] in da ma tud sedež na prevajalstvu in tko naprej **a ne** | k pol lahk dejansko od sam [gap]
oddelka za prevajalstvo tud financira sofinancira pa take fore a ne //

[146] zdej jasno ključno temeljno vprašanje v kakšnem jeziku je to potekalo **a ne** se na žalost ne
ve zarad tega ker ni ohranjenih zapisov |

[147] skoraj zaščitni znak **a ne** Mehike //

Po drugi strani se označevalec *a ne* v govorjeni rabi uporablja tudi kot **usmerjevalec poteka interakcije**, s katerim tvorec zaključuje svojo vlogo in s pozivom k izmenjavi mnenj vlogo predaja naslovniku (primer [148]). Izrazno to funkcijo zaznamuje vprašalna intonacija in tipično pojavljanje na koncu izjave, vendar ne nujno, saj lahko označevalcu sledi še navajanje dodatnih

pojasnil v obliki apozicije (primer [149]), v primeru neodzivnosti naslovnika pa tudi tvorčev odgovor na lastno vprašanje, kot v primeru učiteljičinega odgovora na vprašanje, naslovljeno učencem (primer [150]). Pri tem je vprašalni pristavek *a ne* v spontanem govoru rabljen bistveno redkeje kot intonančno nezaznamovani zapolnjevalec vrzeli (približno osmina vseh pojavitev).

[148] je blo prejšn Sladko vince bol **a ne** ? ... ja ja // meni ... o sladko vince šnici omakce to pa mam rad //

[149] sej roke lahko uporabm **a ne** drugo roko |

[150] kaj je včasih obsegala Kranjska // današnjo Gorenjsko // današnjo Gorenjsko pa samo današnjo Gorenjsko | pa še [gap] | Štajersko // nekej Štajerske Dolenjsko pa nekej Notranjske **a ne** ja | eee povejte mi in kdo je tisti ki je živel na Kranjskem in je bil zelo dober po junaštvih? //

5.3.1.2 Raba označevalca *a ne* v korpusu Janes

Tudi primere rabe označevalca *a ne* v korpusu uporabniških spletnih vsebin Janes (Priloga 10) lahko v splošnem razdelimo v enaki krovni skupini – povedne rabe na eni (primer [151]) in vprašalne rabe (primer [152]) na drugi strani – vendar vsaka izkazuje nekaj pomembnih razlik glede na rabo v korpusu Gos.

[151] Pa Avstrici tudi, a ne, ki so krampasti kot Švabi, pa Italjani **a ne** , ki v vsaki dobri mrhi iščejo svojo mamico.

[152] @anzet Drzi . A bo sam 1 sezona **a ne** ? 8 delov ?

Najbolj opazna razlika med obema modalnostma je zagotovo ta, da je povedna raba označevalca v primerjavi s spontanim govorom, kjer ta predstavlja veliko večino vseh pojavitev označevalca *a ne*, v uporabniški spletni komunikaciji **precej manj pogosta** kot vprašalna raba. To je na nek način tudi pričakovano, saj v uporabniški spletni komunikaciji, ki ne glede na hitrost izmenjave sporočil med tvorcem in naslovnikom nikoli ne poteka v realnem času, kot je to značilno za spontani govor,¹²⁸ težko govorimo o kognitivni vlogi zapolnjevanja vrzeli med načrtovanjem nadaljnjega besedila. Ker tvorci besedil na spletu nadzirajo, kdaj in v kakšni obliki bo njihovo sporočilo posredovano naslovniku, označevalca *a ne* v povednih rabah torej opravlja **druge vloge kot predvsem vlogo mašila v procesu načrtovanja** (kot bi jo na primer lahko pripisali

¹²⁸ Glej tudi pojasnilo razlike med sinhronim in kvazisinhronim sporazumevanjem v poglavju 2.3.2 *Značilnosti uporabniških spletnih vsebin*.

govorjeni rabi v primerih [146] in [147]). Toda njihova natančna opredelitev nikakor ni enostavna – kot skupno funkcijsko lastnost lahko sicer prepoznamo tvorčevo namero po vzpostavljanje stika z naslovnikom oz. naslovniki, vendar v nekaterih primerih označevalec deluje predvsem kot **poudarjalec splošne veljavnosti** soležne trditve, ki bi jo lahko parafrazirali z »saj se vsi strinjamo« ali »saj vsi vemo« (primera [153] in [154]), tudi v pokroviteljskih oz. sarkastičnih rabah (primer [155]), spet v drugih pa se zdi, da je označevalec rabljen predvsem z namenom dodatne omilitve oz. **relativizacije izrečenega** (primera [156] in [157]), ki že prehaja iz trdilne v vprašalno rabo in bi jo lahko parafrazirali z »saj se vsi strinjate?«.

- [153] Pa kaj je s tem srcem ? - bela cesta , važno je da imaš srce ? Ja sej ga mam vsi **a ne** , samo eni ga ne matrajo pa obremenjujejo z ljubeznijo , pa srečo , pa veseljem pa žalostjo . Eni mu enostavno ukažejo .
- [154] je pa tko **ane** : če ne znaš nič s telefonom počet , kupi iOS (iphone)
- [155] Ne , saj , kot sem rekel zgoraj , ti načeloma ni potrebno popravljati članka , če je » Center « res » center « , **a ne** .
- [156] Sej **ane** o kromu bo treba se globoko premislit
- [157] Huh ja bo letos treba tudi neki nardit , ce bo folk za oz bo udeležba kr **ane** , pol bi lahko naredili mogoče mal vec kot lansko leto .

Tudi pri rabi vprašalnega pristavka *a ne*, ki mu pogosto sledi vprašaj, ni pa nujno, uporabniška spletna komunikacija izkazuje nekaj specifik glede na rabo v spontanem govoru. Obema modalnostma je skupno to, da je vprašalni pristavek rabljen kot sredstvo za preverjanja naslovnikovega strinjanja z izrečenim in posledičnim odpiranjem dialoga z naslovnikom (predajanje vloge). Toda v primerjavi z govorno rabo, kjer tvorec z rabo označevalca naslavlja konkretnega naslovnika oz. več njih, ki se lahko na vprašanje odzove v realnem času (glej primere zgoraj), v uporabniških spletnih vsebinah tvorci naslavlajo **več tipov naslovnikov**. V prvo skupino lahko umestimo naslavljanje **enega ali več konkretnih posameznikov**, tj. drugih uporabnikov izbrane spletne platforme, pri čemer tak individualistični nagovor najpogosteje spremlja omemba konkretnega uporabniškega imena oz. psevdonima (primeri [158]–[160]), včasih pa lahko naslovnika razberemo šele iz konteksta, kot v primeru [161], v katerem komentator blogovskega zapisa nagovarja njegovega avtorja.

- [158] @mihavete sam me boš pelu en krog poj, **ane** ?
- [159] Se mam česa veselit , **a ne** @lisjakm @tmoven @ErikRenko @n_e_j_k_o @jure_mastnak @sZlatko

[160] jazorit , si ti mal preprost **ane** , kultura tebi ni prioriteta , a ?

[161] Ojdip je konstrukt ne fakt, to ve vsak otrok, želja ni retrogardna, marveč je parcialna, zaradi sistema nenehnih restrikcij in dominacij - saj si bral Deleuza, **a ne** ?

V drugo skupino lahko uvrstimo primere rabe, v katerih tvorec z rabo označevalca *a ne* podobno kot v povednih rabah naslavlja **splošnega naslovnika**, tj. vse druge uporabnike z dostopom do tvorčevih vsebin (sledilci, drugi sodelujoči in nesodelujoči uporabniki, anonimni bralci itd.). Tudi tukaj odvisno od sobesedilnega konteksta tvorci z rabo označevalca bodisi z relativizacijo izrečenega dejansko pozivajo k izmenjavi mnenj (primera [162] in [163]) bodisi predvsem poudarjajo samoumevnost izrečenega, dialog pa odpirajo le navidezno, tj. ne pričakujejo neposrednega odgovora (primera [164] in [165]).

[162] Torej , rečemo lahko , da lahko vemo kaj ni naključnost , nikakor in nikoli ne moremo naključnosti potrditi , **a ne** ?

[163] Ja, mame prenesejo vse, **ane** ?

[164] Komaj čakam, da Rusi vrnejo sankcije in nehajo uvažat zdravila iz zahoda . Ko bo cela Dolenjska na zavodu, država pa v buli, kako bomo šele takrat kričali - Živela Amerika in EU, **a ne** ?

[165] Pariz je pa malo bližje kot Sirija in Irak , **a ne** ?

V prav posebno, tretjo skupino pa spadajo primeri, v katerih tvorci naslavlajo konkretne **posameznike, ki v razpravi niso udeleženi**, kot so javne osebnosti. To ironično naslavljanje odsotnega udeleženca je značilno zlasti v komentarjih novic, v katerih so ti posamezniki omenjeni (primera [166] in [167]) ali ob omembi tem, ki jih javnost s to osebo običajno povezuje (primer [168]).

[166] to je pa tko k je kitarist za selektorja **ane** Čeferin .

[167] Ko mački stopiš na rep zacvili **ane** Ferlinc ?

[168] Pa kozlička za potenco , pa jagnje za izpričevanje nedolžnosti , pa odojka , da ne bo zrasel v svinjo , ki bi delala z nami kot z » namiljenim mehurjem « (karkoli to že je , **a ne** , g. Štuhec ?

5.3.2 Raba označevalca *a veš*

5.3.2.1 Raba označevalca *a veš* v korpusu Gos

Tudi drugi najbolj tipični večbesedni označevalec govorjenega jezika *a veš* se umešča med označevalce dvogovora, s katerimi govorci na semantični ravni vzpostavljajo in vzdržujejo stik z

naslovnikom, pri čemer sta glede na glagolsko obliko druge osebe ednine tvorec in naslovnik v tem primeru običajno v neformalnem, prijateljskem odnosu (ni presenetljivo, da se med analiziranimi primeri rabe v korpusu Gos niti en ne pojavlja v informativno-izobraževalnem javnem govoru). V vseh analiziranih primerih rabe v korpusu Gos (Priloga 7) se označevalec *a veš* s semantičnega vidika pojavlja kot sredstvo za preverjanje razumevanja in pozornosti naslovnika, z besedilnega vidika pa bi njegove raznolike mikrofunkcije lahko razvrstili v dve krovni skupini.

Na eni strani najdemo rabe, kjer označevalec *a veš* uvaja novo besedilno enoto v obliki dodatnega pojasnila, argumentacije ali spremembe ubeseditve (primera [169] in [170]) in se običajno pojavlja na začetku izjav, stavkov ali drugih pomenonosnih enot (kaže v besedilu **naprej**), včasih pa to pojasnilo, podobno kot pri označevalcih odprte propozicije, ob zadovoljivi povratni informaciji naslovnika niti ni eksplicitno izpeljano (primer [171]).

[169] ja kok deleč ga pol spustiš? // **a veš** a si upaš it do njega kako to zgleda kaj rečeš preden se ... ja ja ja //

[170] lepe sprehajalne sam **a veš** namest da bi bla tkole naravnost ma take mini stopničke tak tak tak tak tak tak

[171] sej se prilagajaš ampak tko **a veš** mislm s [gap] //

V drugo skupino pa lahko uvrstimo rabe označevalca na koncu stavka, v katerih označevalec zaključuje in poudarja predhodno trditev (jo povezuje z naslovnikom in preverja njegovo strinjanje), ne implicira pa njenega nujnega nadaljevanja (kaže v besedilu **nazaj**), kot v primerih [172]–[174].

[172] odzuni sem odzuni mam sam še fasado za prebarvat **a veš** // pa tud sem tlakovce lan na nov po celem dvorišču dala kurac pa so se tam nek pogreznal pa tam nek je bula pa t [gap] zdej a veš //

[173] rastejo rastejo sam treba jih je nabirat **a veš** // aja ... ja ja ja //

[174] tu jest ne morem pozabet sej pravm eee zmer sem rekla pr nabenmu ne b bla rada pr smrti ti pa sem bla pr očet pr mami in pr teti [name:personal] ne morem pozabit tega **a veš** // ja no vidiš //

Meja med obema rabama ni vedno nedvoumno določljiva, a ju lahko posplošimo s skupno lastnostjo hkratne vloge členjenja in povezovanja besedila (bodisi v smeri naprej ali nazaj) ter njegovega usmerjenega povezovanja z naslovnikom. Poleg teh dveh vlog pa lahko tudi v primeru označevalca *a veš* govorimo o hkratni vlogi zapolnjevalca vrzeli (vzpostavljalca tekočnosti

besedila), ki tvorcu omogoča daljši načrtovalni čas, zlasti kadar se označevalec pojavlja med več zaporednimi segmenti tvorčevega diskurza (prva skupina, primer [175]).

[175] ja sam mmm tko je **a veš** Bled je ful turističen ful bl turističen je |//

5.3.2.2 *Raba označevalca a veš v korpusu Janes*

Tudi v korpusu Janes se v analiziranih primerih rabe (Priloga 11) označevalec *a veš* uporablja tako v vlogi uvajalca nove vsebine oz. pojasnila (primera [176]) kot v vlogi poudarjalca predhodne vsebine (primera [178] in [179]).

[176] ja , tudi takšni morate obstajati :) . **A veš** , potem več ostane za nas , ki nam je mar :) .

[177] Imaš prav , **a veš** , zgodovino pišejo vedno tisti , za katerimi ostanejo različni pisni dokumenti v takšni ali drugačni obliki (pa naj bo to na papirju , leseni tabli , glini , kamnu ...

[178] Ob takem vremenu je pika na i to , da naročiš kurilca . Pol si ful hepi , **a veš** .

[179] Lahko ' naštepaš ' 8.036.0221.383 in eno škrbino , če želiš in če zmoreš - meni je načeloma vseeno , saj niso moji kalifornijski strežniki . **A veš** .

Raba obeh je približno enako pogosta, zanimivo pa je, da se med njimi prav tako pojavljajo primeri, ki jih v kontekstu govorne rabe običajno tolmačili kot zapolnjevalca vrzeli (primer [180]), njihova raba v asinhroni pisni komunikaciji pa nakazuje, da so tudi med posameznimi pomensko nezaključenimi besedilnimi segmenti označevalci rabljeni hoté, da naslovniku izpostavijo in poudarijo soležno vsebino. Po drugi strani med analiziranimi konkordancami razen v primerih ironičnega posnemanja govorne interakcije (primer [181]) nismo našli primerov, v katerih bi se tvorec z apeliranjem na naslovnikovo védenje izognil podrobnejši opredelitvi vsebine (kot denimo v primeru [171] zgoraj), kar potencialno nakazuje, da so tovrstne rabe predvsem posledica kognitivnih omejitev tvorjenja v realnem času, ne pa namernega izogibanja konkretizaciji vsebine.

[180] @dzaden Dear god pa tega nisem vidla ! Ampak 2013 **aveš** , to je še dodiplomc , ne moreš verjet , kako se kvaliteta potem spremeni . Not . :D

[181] @vanfranco Hvala lepa . Joj , pa ravno 1.12. , ko imam ... tisto ... **a veš** , no , nedelja , pa ... joj ... ne morem . * zbeži * @loudandwicked

Med primeri rabe označevalca *a veš* v uporabniških spletnih vsebinah smo sicer zaznali dve izstopajoči posebnosti. Prva je – podobno, kot smo to izpostavili že pri analizi označevalca *a ne* – dejstvo, da z označevalcem *a veš* uporabniki na spletu nagovarjajo **raznolik nabor naslovnikov**,

čepprav se ta oblika v spontanem govoru običajno uporablja zgolj v dialoški interakciji med tvorcem in konkretnim naslovnikom, s katerim je tvorec vzpostavil prijateljski odnos. V spletni komunikaciji uporabniki označevalca *a veš* namreč ne uporabljajo zgolj pri naslavljanju znanih konkretnih uporabnikov (npr. prijateljev ali znancev), kot v primeru [182], temveč tudi **neznane konkretne uporabnike**, ki jih osebno ali poimensko ne poznajo, a jih kljub temu tikajo (primer [183]). To namreč ni nujno tudi znak domačnosti med obema uporabnikoma, ki bi jo ta dva razvila v predhodni komunikaciji (kot je to predpostavljeno v govorjeni interakciji), saj se zdi tikanje med statusno enakovrednimi uporabniki (npr. razpravljalci na forumu, komentatorji, tviteraši, ki niso javne osebnosti) prevladujoča oblika naslavljanja na spletu.¹²⁹

[182] @E_nk_a **A veš** , ko si imela letos r. d. in sem ti poslala neprimerno pesem ? No , zdaj se mi zdi , da sm našla ta pravo : <http://t.co/4cyePPL2jJ>

[183] Sam **a ves** SF010 , v Olimpiji se da tko fino dnar oprat pa po raznoraznih zepih poskrit .. Poleg nagovarjanja konkretnih znanih ali neznanih uporabnikov pa se označevalec *a veš* kljub svoji izrazni zaznamovanosti v uporabniški spletni komunikaciji uporablja tudi pri naslavljanju neopredeljenega splošnega naslovnika, denimo v blogovskih zapisih, v katerih se avtorji povezujejo z **neidentificiranim neznanim bralcem** (primer [184]), čepprav se v tem besedilnem tipu označevalec *a veš* sicer uporablja predvsem v poročanem govoru (primer [185]), kar potrjuje, da so tipično govorjeni diskurzni označevalci pomembno **slogovno sredstvo pri ustvarjanju učinka pristnosti poročanega govornega dogodka**.

[184] Kot se da predvideti že iz tega , da ni že v začetku poročila JUHEJ ! in JUHUHU ! , nisem dobil niti svežnja knjig kamoli 800 ih evriča . Jebaj ga zdaj , saj bi bilo tudi malo , (maaaaalo , **a veš**) pretenciozno pričakovat , da se prvič udeležiš ene take reči in kar zmagaš , jel ? Bom pa za drugič vedel ... Hvala vsem , ki ste me bodrili s fizično prisotnostjo [...]

[185] " **A veš** , tamladi bodo pa dojenčka dobili , drug mesec se mi zdi , da bo , " mi je zaupno povedala . Da je prav tako , da mora iti življenje dalje , je še dodala .

¹²⁹ V slovenskem prostoru sistematčne raziskave rabe glagolskih oblik tikanja/vikanja v dialoškem sporazumevanju spletnih uporabnikov doslej sicer še ni bilo, vendar že preprosta primerjava pogostosti zvez *a veš* in *a veste*, ki lahko nastopata v podobnih skladenjskih in diskurznih vlogah, kaže, da oblika v drugi osebi ednine močno prevladuje (9.339 proti 873 pojavitev v korpusu Janes).

Kot drugo posebnost rabe označevalca *a veš* v uporabniških spletnih vsebinah pa lahko izpostavimo prav posebno obliko naslavljanja generičnega naslovnika, v kateri se označevalec pojavlja v **ustaljenih (frazelogiziranih) skladenjskih zvezah** tipa *a veš* + samostalniška zveza v imenovalniku (primeri [186]–[190]), najpogosteje z določilom ali jedrom v obliki kazalnega zaimka, ter (bistveno redkeje) *a veš* + časovni odvisnik (primer [182] zgoraj).¹³⁰

- [186] **A veš** uno.... , ko ti je vseeno, ce imajo v kavarni dobro kavo, šteje le hiter wifi.
- [187] **A veš** tisto, k si na obisku in se ti neki strese, pa upaš da te noben ni vidu. No to.
- [188] **A veš** tist filling, k maš še milijon snovi za predelat in se morš skor u temi učit, ker ni elektrike :/ No, to!
- [189] **A veš** un,k je sin od unga,k je brat od sosede,k so tam živel, kjer je bla moja sodelovka,k je sin z uno poročen #mati #IdontknowIdontcare

Ta tip zvez, s katerim uporabnik uvaja kratko enostavno pripoved o nekem izkušenem dogajanju, stanju ali občutenju, ki je bodisi zelo splošno znano (npr. primer [186]) bodisi zelo nepričakovano (npr. primer [188]), se najpogosteje pojavlja na družbenem omrežju Twitter, kjer predstavlja veliko večino vseh pojavitev označevalca *a veš*. Kot slogovni efekt je rabljen tudi v drugih besedilnih tipih, kot prikazuje spodnji iztržek blogovskega zapisa (primer [190]), ki sicer predstavlja več kot četrtino vseh analiziranih primerov rabe v podkorpusu blogovskih zapisov.

- [190] A veš ono , ko pride tak moment v dnevu , ko se nasmejiš do srca ? **A veš** ono , ko dobiš filing , da si preprosto srečen , pa četudi se to sliši klišejsko in pocukrano na kubik ? A veš ono , ko opustiš strah za bodoče bivanje v študentskem domu zaradi bojazni pred nemirom ? **A veš** ono , ko kličeš svojo staro mamó , spremeniš ton glasu , pokriješ slušalko , v fori rečeš » O , Cirila ! « , ona pa ti odzdravi » O , Marinka ! « in ne pogrunta , da kliče njen vnuk (brez slušnih aparatov je bila izjemoma) , potem pa se še na polno režiš dalje , ko se pretvarjaš , da si pač Marinka in ob tem uživaš ? A veš ono , ko si poln

¹³⁰ Ta tip zvez sicer odpira pomembno vprašanje razmejevanja zveze *a veš* v vlogi samostojnega označevalca brez vezljivostnih dopolnil (diskurzna raba) na eni strani in vezljive glagolske zveze *a veš* (propozicijska raba) na drugi strani. Tako kot nekateri drugi označevalci z glagolsko sestavino je namreč niz *a veš* gramatikaliziran v nekatere ustaljene skladenjske vzorce, kot so denimo zveze s predmetnimi odvisniki (npr. *a veš kako ...*, *a veš kaj ...*, *a veš da ...*), ki sicer odpirajo vezljivostna mesta, a ne uvajajo nujno dejanskega vprašanja (*a veš kako sem utrujena*). V pričujoči raziskavi smo ločnico med diskurzno in propozicijsko rabo vzpostavili glede na vezljivostno dopolnjevanje glagola in zveze, ki mu sledi, kar pomeni, da zveze s predmetnimi odvisniki ne glede na stopnjo gramatikalizacije niso obravnavane kot primeri diskurzne rabe, zveze, ki jih izpostavljamo (npr. *a veš tisto, ko ...*) pa zaradi neujemanja med sklonom ali odvisniškim tipom, ki ga predvideva glagol, in določilom, ki mu sledi, vendarle obravnavamo kot primere diskurzne rabe označevalca *a veš*. Ta ločnica je seveda le ena izmed možnih rešitev arbitrarnega razmejevanja kompleksnega kontinuuma med propozicijsko in diskurzno rabo.

ciljev in planov za prihodnost , pa četudi so malce preveč idealizirani ? **A veš ono** , ko spoznaš , da je sreča stvar odločitve ? [...]

Čeprav smo se med analiziranimi konkordancami pojavlja en sam tak primer rabe, bi med frazeologizirane zveze z označevalcem *a veš* lahko uvrstili tudi zvezo *a veš kao*, s katero uporabniki z ironičnim posmehom relativizirajo resničnost trditve pred ali za njo (primer [191]) in se v korpusu Gos ne pojavlja.

[191] zakaj ste pa zbrisali moj komentar ?? ... aja , pardon , že vem ... vi ste ja nedemokratski rdeči kajtnig ... ups ... ma pravico vsak do mnenja a ne Mladina ? **aveš** kao ...

5.3.3 Raba označevalca *ne vem*

5.3.3.1 Raba označevalca *ne vem* v korpusu Gos

Označevalec nedoločnosti *ne vem* se s semantičnega vidika uvršča med označevalce nedoločnosti, s katerimi tvorec relativizira svojo stopnjo zavezanosti glede natančnosti posredovanih informacij ali neomajnosti izraženega mnenja. Vendarle pa lahko z besedilnega vidika njegovo vlogo v diskurzu glede na analizirane primere rabe v korpusu Gos (Priloga 8) členimo še na tri podskupine. V prvo se umeščajo primeri, v katerih tvorec označevalec rabi **sredi** svoje izjave ali vloge, pred popravkom ali spremembo načrtovane ubeseditve (primer [192]), pred navedbo primera (primer [193]) ali specifikacijo podrobnosti (primer [194]), pri čemer med slednjimi v korpusu Gos prevladuje predvsem raba pred konkretizacijo količinskih oz. številčnih podrobnosti (primer [195]). V vseh tovrstnih primerih rabe se epistemska funkcija nedoločnosti tudi najbolj izrazito prepleta z besedilno funkcijo vzdrževanja tekočnosti govora (zapolnjevanja vrzeli).

[192] a kaj misliš eee mmm kak nate gledajo punce kot eee južnjaka ali a mmm **ne vem** ali majo e [gap] etikete pa ne vem [gap] ... še |//

[193] zej če vidva **ne vem** da je enmu res ne vem kaj se je zgodil ne potem ja drgač pa ponavad ne vem al ignorirava al pa |//

[194] eem tisti senzorji pa dajo na izhodu ne dajo linearen signal ampak je **ne vem** pač nelinearen a ne |

[195] in s [gap] tud smo pazli da recimo pozimi že s temi klasičnimi kolektorji na sončn dan se je voda zagrela na ne vem dvajset stopinj al pa trenutno je se je zagrela celo na **ne vem** trideset petintrideset |//

V drugo skupino lahko glede na prevladujočo pozicijo in z njo povezano funkcijo umestimo rabe označevalca na **začetku** govorne vloge ali posamezne tematske sekvence, v kateri govorec

nakaže prevzem vloge oziroma nadaljevanja diskurza, obenem pa sledečo vsebino relativizira, jo omili kot zgolj nezavezujoče mnenje, primer ali osebno izkušnjo, kot je razvidno iz primerov [196]–[198].

- [196] ja sej pravm de [gap] // **ne vem** se mi zdi de sta de je [name:personal] dol pr [name:surname] //
- [197] eni so rekli ja jz nimam nobenga problema pa kej čem pisat tku // **ne vem** jz bi šla eee za prvi nivo je važnu sej ostalu bos [gap] bojo pol //
- [198] ja [gap] ... ja razumem ja // **ne vem** z biomaso je tako biomaso imate vi ze tri vrste bi reku tri vrste al mate klasična polena //

V tretjo skupino pa se umeščajo primeri rabe označevalca ob **zaključku** tvorčeve vloge (npr. [199] in [200]), s katero tvorec (vnovič) poudari subjektivnost in relativnost pravkar povedanega, vlogo pa z namenom izmenjave mnenj preda sogovorniku.

- [199] k majo tisto igrišče k so mel otvoritu se mi je zdev dost mirno no **ne vem** // sej je //
- [200] se lahko poročajo pa to otroke pa to **ne vem** // jah ja to nega kaj se to samo eee katoliška vera nema //

5.3.3.2 Raba označevalca *ne vem* v korpusu Janes

Vse tri položaje oz. vloge označevalca *ne vem*¹³¹ smo zaznali tudi v analiziranih primerih rabe v korpusu Janes (Priloga 12), vendar s precej drugačno distribucijo. Če v spontanem govoru prevladuje raba označevalca kot napovedovalca priklica ustrezne vsebine ali ubeseditve (primeri [192]–[195] znotraj prve zgoraj izpostavljene skupine), kakršna se običajno povezuje s položajem na sredini izjave, so tovrstni primeri v spletni komunikaciji **zelo redki**, bolj kot zapolnjevanje vrzeli pri priklicu pa je poudarjena njihova semantična vloga omiljevanja vsebine, ki sledi (primera [201] in [202]).

- [201] Uporništvo je tisto , s katerim so se razvijale generacije in generacije . A , današnja , **ne vem** , upam , da se motim , a se mi dozdeva , da le redki ohranjajo uporništvo v svojem srcu .

¹³¹ Zelo podobno prepletanje epistemske in sekvenčne funkcije v govornjeni rabi francoskega označevalca *je ne sais pas* izpostavlja tudi raziskava (Pekarek Doehler, 2016), ki ugotavlja, da se v francoščini ta označevalec v položajih sredi govorne vloge uporablja kot sredstvo za omiljevanje, zadrževanje vloge in signaliziranje približka, v položajih na začetku vloge opozarja na nepričakovano oz. neustrezno (non-fitted) repliko, v položajih na koncu izjave pa signalizira zaključevanje vloge.

[202] @Skravzlana Orto . Plesat . Kar je navadno fino . Ampak **ne vem** , čudna sem . :)

Kot posebnost uporabniških spletnih vsebin lahko izpostavimo tudi dejstvo, da je ta vloga označevalca pogosto rabljena **ironično**, kot v primerih [203] in [204], v katerih uporabnika z rabo zveze *ne vem* poudarjata samoumevnost trditve ali sklepa, ki sledi.

[203] K sveti Hijacinti se bojda moli , ko je nekdo v nevarnosti utopitve . Kaj pa recimo ... **ne vem** ... mu greš pomagat ?

[204] @pengovsky Se spomniš suma korupcije pri zaposlitvi izraelskega kirurga v korist ... **ne vem** , pacientov ? No , to .

Bistveno bolj pogosto se v uporabniških vsebinah pojavlja raba označevalca na začetku stavkov oz. izjav, pri čemer se za razliko od spontanega govora, kjer je ta položaj pogosto povezan tudi s prevzemanjem vloge govorca (glej primere [196]–[198] zgoraj in [205]–[206] spodaj), v uporabniških spletnih vsebinah prevladuje predvsem **raba na začetku sklepnih trditev**, ki se običajno pojavljajo proti koncu govorčevega besedila oz. vloge. Ko torej uporabnik predstavi svoje mnenje, v izogib konfliktu z drugimi uporabniki končni sklep omili z rabo označevalca (pogosto tudi v kombinaciji z označevalcem *no*), s čimer obenem tudi spodbuja nadaljnjo izmenjavo mnenj (primera [207] in [208]).

[205] _ris , **ne vem** , si ti res prepričan , da smo se res odbili od dna ?

[206] **Ne vem** ko sem delal pri dostavi prek študenta , sem imel kombi brez turbine mal manj kot 90 konjev , brez težav je šel čez 100.

[207] Stoji že 2 leti in nič ne kaže da bi blo kaj narobe , streha je gor pločevinasta Hosekra katere teža je okoli 7 kg / m² , tak da **ne vem** no , meni se vse skupaj zdi še predimenzionirano in ne poddimenzionirano .

[208] Ampak Janez se gotovo ne jezi tjavdan . **Ne vem** , sam o osebnih imenih nimam virov in se nikoli pravzaprav nisem ukvarjal z njimi .

Zelo podoben preplet funkcij hkratnega **zaključevanja diskurzne enote, odpiranja nadaljnje razprave in omiljevanja tvorčeve zavezanosti k veljavnosti** pravkar izrečenega lahko pripišemo tudi primerom, v katerih uporabniki označevalec *ne vem* rabijo na koncu svoje vloge oz. besedila (primera [209] in [210]).

[209] n meni j enjegov odgovor tudi bolj logičen - če gre res za estrogenom podobnim snovem - hormoni . Znano je da jo lahko zatremo z KT , nosečnost , dojenje **ne vem** no

[210] Tale tekma v Domžalah pa vse bolj smrdi , ker tako felerji , kartoni in penali ... **Ne vem** .

Če analizo rabe treh najbolj tipičnih diskurznihih označevalcev govorjenega jezika v uporabniški spletni komunikaciji strnjeno povzamemo, primerjava z njihovo rabo v spontanem govoru potrjuje, da so ti izrazi v govoru rabljeni predvsem kot sredstvo za zapolnjevanje vrzeli oz. segmentacijo diskurza med kognitivnim načrtovanjem nadaljnjega govora, v uporabniških spletnih vsebinah pa ob odsotnosti pritiska tvorjenja v realnem času izstopajo predvsem druge, **semantično motivirane vloge**. Kot v sorodni analizi rabe zapolnjenih premorov v nemških tvitih ugotavlja I. Rehbein (2015), gre torej za izraze z nepropozicijskim pomenom, s katerimi tvorci svojim besedilom namerno dodajajo določen podtekst oz. dodatna subtilna sporočila. Pri tem naša raziskava kaže, da so ta sporočila analiziranih treh označevalcev predvsem želja po **vzpostavljanju stika z naslovniki, preverjanje strinjanja ali omiljevanje izrečenega**, ki se hkrati prepletajo tudi z besedilnimi funkcijami **poudarjanja, prevzemanja ali predajanja vloge**. Vsi trije označevalci se na spletu pojavljajo tudi v nekaterih **novih kontekstih rabe**, kot so nagovarjanje neznanega oz. generičnega naslovnika, doseganje humornega učinka (ironija) ali izražanje negativnega (pokroviteljskega) odnosa, vstopajo pa tudi v nekatere nove frazeološke zveze.

6 Razprava

Kljub razmeroma specifičnemu končnemu cilju predstavljene raziskave, analizi rabe treh najbolj tipičnih večbesednih diskurznihi označevalcev govorne slovenščine v besedilih slovenskih spletnih uporabnikov, smo v procesu njene uresničitve podrobneje raziskali več vidikov jezikoslovnega opisa kategorije diskurznihi označevalcev. Ti prinašajo pomembne ugotovitve za nadaljnje raziskave na področjih analize diskurza, diskurznofunkcijske leksike, večbesednih enot, kontrastivnih zvrstnih raziskav in metodologije korpusnega jezikoslovja, a obenem odpirajo tudi nekaj pomembnih epistemoloških vprašanj.

Kot najbolj pereče vprašanje zagotovo izstopa **problematičnost same opredelitve kategorije diskurznihi označevalcev**, za katero smo v pregledu literature ugotovili, da so bili skladno s težavno opredelitvijo samega koncepta diskurza zanjo doslej predlagane številne različne zamejitve glede na (ne)izbrani teoretski model, raziskovalne cilje, preučevano gradivo in različne načine sopostavljanja pomenskih in izraznih prepoznavnih lastnosti. Govorimo lahko torej o nekakšnem paradoksu diskurznihi označevalcev, saj so po eni strani ta jezikovna sredstva v jeziku izrazito pogosta in vseprisotna, po drugi strani pa njihov dosleden in izčrpen opis še danes predstavlja enega najtežje rešljivih izzivov sodobnega jezikoslovja, tudi v slovenskem prostoru. Zdi se, da bo tega mogoče doseči šele, ko bodo ti neposredno vpeti tudi v širše uveljavljeno teorijo diskurza, ki bi zadovoljiveje kot obstoječe pojasnjevala socio-kognitivne mehanizme tvorjenja in razumevanja diskurza, tako v pisni kot govorni rabi. Dokler torej v jezikoslovju ni konsenza glede tega, kako pristopati k opisu jezikovne rabe onkraj stavčnih propozicij, v katerih govorniki pripovedujejo o tem, kdo je kaj storil komu kako, kje, kdaj in zakaj, je vsakršna nepropozicijska jezikoslovna kategorija, vključno s kategorijo diskurznihi označevalcev, obsojena na določeno mero **arbitrarnosti**. Vsak poskus njene zamejitve je tako le poskus, a hkrati nujno potreben korak na poti k empirično utemeljenemu razumevanju strategij diskurznega povezovanja, strukturiranja ali upravljanja.

Tudi opredelitev diskurznihi označevalcev, iz kakršne smo izhajali znotraj te raziskave, je torej **le ena izmed možnih opredelitev**, a je bila prav zaradi tega v primerjavi s predhodnimi sorodnimi raziskavami, ki so jo utemeljevale s sklicevanjem na metodološko težko prenosljive modele diskurza, ponazoritvami s peščico izbranihi prototipičnih primerov ali nejasnimi razmejitvami

glede na druge skupine nepropozicijske leksike, zastavljena **teoretsko razmeroma nevtrarno, široko in z opiranjem na metodološko čim bolj objektivizirana merila**. Med slednjimi izstopata predvsem razumevanje diskurza kot tvorčeve koherentne besedilne produkcije, ki omogoča ločevanje glede na vse ostale kategorije nepropozicijske oz. pragmatične leksike, in pa zavračanje predpostavke, da uporabniki jezika pri tvorjenju diskurza sistematično razlikujejo med različnimi dometi večfunkcijskih oblik, kot so povezovanje propozicij, izjav, besedilnih sekvenc, zunajbesedilnega konteksta ipd., ali različnimi strategijami diskurznega upravljanja, kot sta objektivno in subjektivno povezovanje, na kakršne so se opirale nekatere predhodne raziskave diskurznih označevalcev v slovenščini.

Kljub tovrstni široki, besediloslovno usmerjeni opredelitvi diskurznih označevalcev, ki sicer sovпада s trenutnimi trendi na področju aplikativne analize diskurza, pa se je pri prenosu teoretičnih meril, oblikovanih na podlagi sinteze pregleda širokega nabora sorodnih raziskav, na konkretno gradivo izkazalo, da je te **nujno dopolnjevati s spoznanji empirične evalvacije**, pri čemer se je v predstavljeni analizi vzorčenega segmenta korpusa govorne slovenščine poleg potrebe po natančnejši opredelitvi ločnice med diskurzni označevalci in **drugimi pragmatičnimi jezikovnimi sredstvi v govoru**, kot so izrazi odzivanja ali naklonskosti, izkazala predvsem potreba po natančnejši opredelitvi izraznih meril, ki ustaljene strukturno zaključene večbesedne diskurzne označevalce ločuje od drugih **funkcijsko ekvivalentnih formulaičnih nizov** v slovenščini.

Proces označevanja (večbesednih) diskurznih označevalcev v avtentičnem gradivu pa je razkril še eno neizogibno protislovje: če je za nedvoumno in dosledno *prepoznavo* teh jezikovnih sredstev nujna čim natančnejša specifikacija diskurznih funkcij, kakršno smo povzeli po taksonomiji MDMA in nam je omogočila odmik od metodološko težko upravičljivega prepoznavanja 'izrazov v vlogi diskurza' k identifikaciji konkretnih strategij diskurznega upravljanja in z njimi povezanih jezikovnih sredstev, po drugi strani tovrstno drobljenje funkcij otežuje njihov nedvoumen in dosleden *opis*, torej funkcijsko razvrščanje in razdvoumljanje v kontekstu. Zdi se torej, da so **maksimalistični pristopi k analizi funkcijske leksike**, ki členitev pomena posameznih oblik pogojujejo s podrobno analizo sobesedilnega in sporazumevalnega konteksta, **z aplikativnega vidika manj uporabni**, hkrati pa težavnost njihovega razdvoumljanja odpira tudi pomislek, ali je tovrstno drobljenje pri modeliranju jezikovne rabe potemtakem sploh teoretsko upravičljivo, če

tudi (poklicni) jezikovni uporabniki med različnimi kontekstualnimi dometi oblik težko razlikujejo ali pa morajo zanje oblikovati zelo podrobne označevalne protokole. Kot alternativo smo v raziskavi zato oblikovali novo funkcijsko taksonomijo, ki temelji na minimalistični predpostavki, da lahko vsakemu označevalcu običajno pripišemo en sam jedrni pomen oz. funkcijo, glede na stopnjo diahronega razvoja oblike pa se ta v rabi lahko pojavlja v različnih kontekstih, dometih oz. maksimalistično prepoznanih funkcijah. Menimo, da ta dvostopenjski pristop, v katerem se označevalcu najprej določi semantično jedro, šele nato pa različne funkcije v odvisnosti od dometa, predstavlja tudi dobro izhodišče za nadaljnji sistematičnejši opis besedilnopovezovalne leksike v slovenščini, ki je bila doslej glede na različne načine kombiniranja izraznih in pomenskih lastnosti bolj ali manj dosledno obravnavana kot del veznikov, členkov, prislovov in/ali medmetov.

Drugi pomemben vidik, ki ga načenja predstavljena raziskava, pa je **relativizacija ločnice med enobesednimi in večbesednimi diskurznimi označevalci**. Čeprav se diskurzni označevalci v svojem bistvu vzpostavljajo kot funkcijska jezikoslovna kategorija, torej kot skupina jezikovnih sredstev, ki jim je skupna predvsem določena vloga v sporazumevanju, ne pa nujno tudi izrazne lastnosti (kot je to denimo značilno za druge besednovrstne kategorije), v njihovih raziskavah le **redko zasledimo onomaziološke pristope**, ki bi se spraševali, kateri izrazi v jeziku opravljajo določeno (npr. diskuržno) funkcijo. Nekoliko ironično namesto tega tako na tujem kot pri nas še vedno prevladuje izrazito semaziološki pristop, ki se osredotoča na analizo peščice vnaprej izbranih, najpogostejših ali drugače izstopajočih diskurznih označevalcev, ugotovitve pa naknadno posplošuje na kategorijo ali diskurz kot celoto. Razumljivo so se tovrstne raziskave doslej večinoma omejevale predvsem na enobesedne ali najbolj gramatikalizirane večbesedne označevalce, pri njihovem prepoznavanju pa niso bile obremenjene z metodami njihovega priklica, zato so tovrstni sezname označevalcev najpogosteje nastajali na podlagi naključne izbire, paberkovanja po drugih virih ali analize majhnih korpusov. Širok nabor več kot 200 različnih večbesednih diskurznih označevalcev, ki smo jih znotraj predstavljene raziskave z različnimi metodami prepoznali v referenčnem korpusu govorjene slovenščine, smotrnost tovrstnih pristopov tako postavlja v novo luč, saj dokazuje, da povsem **enake funkcije v jeziku opravljajo tudi številni ustaljeni večbesedni izrazi**, za njihov čim bolj izčrpen priklic pa je kvalitativne

metode **smiselno dopolnjevati tudi s kvantitativnimi metodami**, kot je luščenje nizov ali kolokacij, ki s časovnega vidika dajejo tudi bistveno boljše rezultate.

Nenazadnje pomen onomaziološkega pristopa potrjuje že naša raziskava, saj smo z naivnim, induktivnim pristopom k prepoznavanju tipično govorjenih diskurzni označevalcev, v katerem smo najprej skušali identificirati vse diskurzne označevalce v govoru, šele nato pa s pomočjo statističnih metod izluščili tiste najbolj tipične, razkrili več zanimivih ugotovitev. Prepoznani diskurzni označevalci nasploh se tako kar nekajkrat pogostejše pojavljajo v govorjeni kot v pisni rabi, hkrati pa se med najbolj tipično govorjenimi označevalci ne pojavljajo zgolj označevalci interakcije, vezani na specifične situacijske okoliščine tvorjenja spontanega govora, kot sta tvorjenje v realnem času in navzočnost naslovnika, kot bi morda intuitivno predpostavljali, temveč tudi besedilni konektorji, kakršni se pojavljajo tudi v pisni rabi, vključno s tistimi, ki nikoli niso razvili posebnih pragmatičnih vlog in v diskurzu nastopajo zgolj kot semantični povezovalci oz. vezniki. To dokazuje, da si govorci **v spontanem govoru aktivneje, tj. s pogostejšo rabo eksplicitnih jezikovnih sredstev, prizadevajo za koherentno podajanje vsebine**, tako z vidika smiselne povezanosti posredovanih sporočil kot z vidika njihove vpetosti v širši sporazumevalni kontekst. Čeprav so za natančno določitev vzrokov za ugotovljene razlike v rabi diskurzni označevalcev v obeh modalnostih potrebne nadaljnje podrobnejše raziskave, naši rezultati sugerirajo, da govorci z njihovo rabo glede na omejitve sprotnega tvorjenja in sprejemanja predvsem **kognitivno razbremenjujejo** tako sebe (npr. vzdržujejo tekočnost govora med načrtovanjem njegovega nadaljevanja) kot naslovnika (npr. mu lajšajo razumevanje, izpostavljajo poudarke, usmerjajo izmenjavo).

Vendarle pa trditev, da so diskurzni označevalci, še zlasti tisti, ki se tipično pojavljajo predvsem v spontanem govoru, zgolj jezikovna kompenzacija pritiska procesiranja v realnem času, nekoliko relativizirajo ugotovitve sklepnega dela naše raziskave. Pokazali smo namreč, da se tudi v pisni komunikaciji uporabnikov spleta druge generacije najbolj tipično govorjeni večbesedni diskurzni označevalci dejansko pojavljajo bistveno redkeje kot v govoru, kar potrjuje, da je pogosta raba analiziranih diskurzni označevalcev v govoru **predvsem posledica specifičnih okoliščin tvorjenja govorjenega diskurza**. Toda po drugi strani je raba teh označevalcev na spletu vendarle prisotna in obenem tudi pogostejša kot v tradicionalni pisni komunikaciji, podrobnejša primerjava njihovih funkcij v govorjeni in spletni komunikaciji pa kaže, da v spletni

rabi izstopajo predvsem **funkcije, ki izhajajo iz njihovega semantičnega jedra** (npr. nagovarjanje naslovnika, preverjanje strinjanja ali omiljevanje izrečenega). Govorimo lahko torej o določeni stopnji **leksikalizacije** teh izrazov, pri čemer pa je zanimivo tudi to, da ti kljub razmeroma kratkemu obstoju tega medija na uporabniških spletnih platformah že **razvijajo nove kontekste rabe**, kot sta npr. naslavljanje neznanega naslovnika ali vstopanje v frazeološke zveze, ki jih prej ni bilo mogoče zaznati ne v govorjeni ne v pisni komunikaciji. Raba tipično govorjenih (večbesednih) diskurznihih označevalcev v uporabniških spletnih vsebinah tako dokazuje, da to niso zgolj redundantna, pomensko izpraznjena 'mašila', temveč jih spletni uporabniki prepoznavajo kot pomemben gradnik posredovanih sporočil, tudi v kontekstih, kjer je njihova vsebina prostorsko omejena.

Pri tem pa eno ključnih vprašanj ostaja, **v kolikšni meri v kontekstu uporabniške spletne komunikacije sploh lahko govorimo o konvencionalizirani jezikovni rabi** teh ali drugih tipično govorjenih jezikovnih sredstev. Če izhajamo iz predpostavke, da obstaja le en jezikovni kontinuum, ki ga ni mogoče ločevati na ločena sistema pisnega ali govorjenega jezika, lahko dihotomijo med značilnostmi govorjene in pisne rabe pripišemo predvsem dejstvu, da so konvencije pisne rabe, ki delno odražajo različnost tudi samih namenov in okoliščin med obema tipoma sporazumevanja (Verdonik, 2015a, str. 395), jezikovnim uporabnikom običajno priučene skozi proces izobraževanja oz. socializacije. Preklapljanje iz govorjene v pisno rabo je torej formalno priučeno, utrjeno in hkrati tudi družbeno sankcionirano. Toda v primerjavi s starejšimi zvrstmi pisne jezikovne rabe je medij uporabniških spletnih platform tako nov, da težko govorimo o obstoju standardiziranih konvencij ali njihovem sistematičnemu usvajanju. Jezikovne izbire spletnih uporabnikov so tako zelo raznolike, individualizirane in le redko družbeno sankcionirane. Čeprav naša raziskava kaže, da raba vseh treh tipično govorjenih diskurznihih označevalcev govorjenega jezika **kaže določene tendence glede na različne tipe uporabniških spletnih vsebin**, kot sta pogosta raba označevalcev v dialoških tvitih na eni strani in redka raba v statičnih blogovskih zapisih na drugi, razlike v primerjavi rabe posameznih označevalcev, še zlasti pa nenavadno redka raba vseh treh označevalcev na (dialoških) pogovornih straneh Wikipedije, odpirajo prostor za nadaljnje raziskave dejavnikov prehajanja tipično govorjenih prvin v spletno rabo, ki bi poleg besedilnih tipov raziskale tudi vpliv drugih okoliščin, zlasti demografskih **značilnosti samih uporabnikov**.

7 Zaključek in nadaljnje raziskave

V doktorski raziskavi z naslovom *Leksikalne prvine govornega jezika v uporabniških spletnih vsebinah: primer večbesednih diskurznihi označevalci* smo raziskali več vidikov rabe večbesednih diskurznihi označevalci, ki smo jih na podlagi pregleda sorodnih tujih in domačih raziskav opredelili kot propozicijsko neobvezne kombinacije dveh ali več zaporednih grafemskih besed z ustaljeno funkcijo diskurznega povezovanja, strukturiranja ali upravljanja, z namenom čim bolj objektivizirane in dosledne prepoznave v avtentičnem gradivu pa smo izhodiščna funkcijska in izrazna merila dopolnili še s podrobnejšo obravnavo težavnejših mest tovrstne jezikoslovne kategorizacije, ki smo jih zaznali v procesu ročnega označevanja vzorčenega korpusa govorne slovenščine. S primerjavo rezultatov tega procesa in več izbranih kvantitativnih metod luščenja večbesednih nizov iz celotnega korpusa govorne slovenščine smo ugotovili, da je pri prepoznavanju diskurznofunkcijskih večbesednih enot v besedilnih korpusih smiselno običajne, kvalitativne metode dopolnjevati tudi s kvantitativnimi, zlasti luščenjem najpogostejših nizov in izračunom njihove kolokabilnosti glede na mero t-test. Z unijo vseh metod smo tako v korpusu govorne slovenščine prepoznali 204 različne večbesedne diskurzne označevalce, med katerimi so bili kot tipično govorni prepoznani predvsem označevalci procesa govorne interakcije, a tudi nekateri besedilni povezovalci. Analiza rabe treh najbolj tipično govornih večbesednih diskurznihi označevalci v uporabniških spletnih vsebinah je potrdila, da so ti tudi v tej obliki pisnega sporazumevanja rabljeni bistveno redkeje kot v spontanem govoru, vendarle pa njihova raba statistično ni zanemarljiva, zlasti v besedilnih tipih s poudarjeno interaktivno oz. dialoško izmenjavo uporabniških sporočil. Primerjava specifičnih funkcij teh izrazov v govorni in spletni rabi je potrdila, da se ti na spletu pojavljajo v semantično in ne kognitivno motiviranih funkcijah ter v nekaterih povsem novih kontekstih rabe.

Čeprav tudi obravnavana raziskovalna vprašanja odpirajo **številne možnosti nadaljnjih raziskav**, kot so evalvacija oblikovane označevalne sheme s primerjavo odločitev več neodvisnih označevalci, primerjava z metodami luščenja na podlagi strojnega učenja, dopolnitev z luščenjem diskurznihi označevalci pisnega jezika in analiza spletne rabe širšega nabora tipično govornih diskurznihi označevalci, menimo, da bi te prioriteto veljalo dopolniti predvsem s **prenosom izdelanih teoretičnih in metodoloških izhodišč na funkcijsko analogne enobesedne označevalce**, ki so bili doslej v slovenistiki obravnavani le parcialno. Rezultate pričujoče

raziskave namreč lahko dokončno zanesljivo potrdimo šele s hkratno in metodološko enotno obravnavo vseh izraznih in funkcijskih podskupin diskurznihi označevalcev, ne glede na onomaziološko redundantno merilo eno- oz. večbesednosti. To še posebej velja za tiste ugotovitve, kjer večbesednost težko upravičujemo kot relevantno spremenljivko, ki bi jo veljalo obravnavati ločeno od enobesednosti, kot so kontrastivne primerjave pogostosti in strategij rabe diskurznihi označevalcev v govorjeni, pisni ali spletni rabi.

Glede na trenutne smernice slovenskega in tujega uporabnega jezikoslovja pa še večji potencial nadaljnjih raziskav znotraj področij onkraj analize diskurza predstavljajo **novi jezikovni viri**, ki so bili posredno ali neposredno izdelani v okviru pričujoče raziskave. **Vzorčeni korpus govorne slovenščine**, ki smo ga izdelali z namenom kvalitativne prepoznave večbesednih označevalcev, tako predstavlja prvo reprezentativno podmnožico referenčnega korpusa govorne slovenščine, ki je bila v okviru ločene raziskave (Dobrovoljc in Nivre 2016) poleg predstavljenega označevanja diskurznihi označevalcev ročno označena tudi na nižjih jezikoslovnih ravneh (lematizacija, oblikoslovje, skladnja), njihov preplet pa odpira številne možnosti nadaljnjih empiričnih jezikoslovnih raziskav posebnosti govorne komunikacije ali jezikovnotehnoloških vidikov njenega modeliranja in procesiranja.

Prav tako dragoceni so tudi izluščeni **sezname najpogostejših besednih nizov** v govorjeni slovenščini, razvrščenih glede na različne uveljavljene metode korpusnega jezikoslovja, saj ponujajo obširno gradivno osnovo za razmeroma slabo raziskano področje stalnih besednih zvez v govorjeni slovenščini, od širših raziskav formulaičnega jezika do specifičnih raziskav tipično govornih večbesednih leksikalnih enot. Med slednjimi se z metodološkega vidika kot najbolj zanimivo odpira področje nadaljnjih primerjav uporabnosti posameznih kolokacijskih mer pri prepoznavanju drugih vrst nepropozicijskih (funkcijskih) večbesednih enot, saj so bile te tako v slovenskem kot tujem jezikoslovju doslej v primerjavi s polnopomenskimi zvezami večinoma zapostavljene. Prav dejstvo, da se je tudi pri priklicu diskurznihi označevalcev kot najuporabnejša metoda izkazala ena izmed mer kolokabilnosti, namreč dokazuje neupravičenost predpostavke, da so za luščenje nepropozicijskih večbesednih enot zaradi vsebovanosti pogostih funkcijskih besed mere kolokabilnosti neuporabne. Nenazadnje koristen doprinos k nadaljnjemu razvoju slovenističnega korpusnega jezikoslovja predstavlja tudi izdelano **programsko orodje za**

luščenje in razvrščanje besednih nizov, ki ga je na enak način mogoče prenesti tudi na druge referenčne korpusne slovenskega jezika.

Končni **seznam več kot dvesto različnih večbesednih izrazov**, ki v jeziku lahko opravljajo tudi določene diskurzne oz. besedilne vloge, predstavlja enega najizčrpnjših spiskov diskurznofunkcijskih večbesednih enot v (govorjeni) slovenščini doslej, pomembno dopolnjuje spiske drugih sorodnih raziskav in je kot tak uporaben na več področjih uporabnega jezikoslovja. Kljub dejstvu, da rojeni govorci rabe diskurznih označevalcev, zlasti tistih, ki se tipično uporabljajo predvsem v govornem jeziku, običajno ne uzaveščaj, jo jemljejo kot samoumevno ali celo redundantno, ti izrazi predstavljajo enega najpomembnejših kazalcev jezikovne zmožnosti, zato jih je smiselno vključevati v bodoče slovarje, slovnice in učbenike, namenjene (tudi) poučevanju slovenščine kot drugega/tujega jezika.

Poleg leksikografije in jezikovne didaktike pa lahko posredno nastali **leksikon večbesednih diskurznih označevalcev**, ki poleg seznama standardiziranih oblik vključuje tudi vse ostale distribucijske in slovnične podatke, zbrane v okviru te in zgoraj omenjene komplementarne raziskave (npr. podatek o funkcijskih uresničitvah oblike v govoru, pogovornih variantah v korpusu Gos, zapisovalnih variantah v korpusu Kres, pogostosti (diskurzne) rabe in stopnji dvoumnosti v obeh korpusih, stopnji tipičnosti za govor, osnovni obliki, besednovrstni zgradbi, skladijski strukturi in skladijski vlogi), pomembno pripomore tudi k razvoju orodij za računalniško procesiranje (govorjene) slovenščine. Analogno z drugimi učenci slovenščine mora namreč tudi računalnik za uspešno razumevanje ali generiranje besedil v pisni, govorni ali spletni slovenščini kljub njihovi sekundarnosti glede na propozicijsko vsebino tovrstne izraze ustrezno prepoznati in ovrednotiti. Morda bo namreč prav trenutek, ko bodo računalniki ustrezno uzavestili rabo diskurznih označevalcev, zlasti tistih tipično govornih, pomensko najbolj redundantnih 'mašil', predstavljal prelomni trenutek v razvoju umetne inteligence, ključni preskok od izumetničenega k naravnemu.

Povzetek

Pregled tujih in domačih raziskav diskurznihi označevalci in sorodnih diskurznofunkcijskih jezikoslovnih kategorij potrjuje, da se pri zamejitvi krovne prepoznavne funkcije teh jezikovnih sredstev od funkcij drugih propozicijsko neobveznih izrazov različni raziskovalci strinjajo, da je temeljna prepoznavna lastnost diskurznihi označevalci njihova povezovalnost, a pri tem ponujajo različne opredelitve glede na razumevanje pojma diskurz in z njim povezanega dometa označevalci (od označevalci medbesedilnih povezav do označevalci nejezikovnih vidikov sporazumevalnega procesa), glede na potrebo po razlikovanju med semantičnim in pragmatičnim upravljanjem diskurza ter tudi glede na način samega razlikovanja med semantiko in pragmatiko. Poleg funkcijskih razhajanj, ki pogosto sovpadajo tudi s tipom preučevanega gradiva in različnimi načini sopostavljanja pomenskih in izraznih prepoznavnih lastnosti, podpodročje raziskav večbesednih diskurznihi označevalci dodatno zaznamuje predvsem vprašanje njihove strukturne celovitosti in ustaljenosti, zlasti pod vplivom raziskav formulaičnega jezika, ki kažejo, da podobne funkcije kot jedrna skupina gramatikaliziranihi diskurznofunkcijskih stalnih besednih zvez opravljajo tudi druge, skladenjsko bolj ali manj zaključene sintagme, ki so v rabi nezanemarljivo pogoste.

V želji po konsolidaciji izkazane teoretske razdrobljenosti smo v skladu s trenutnimi smernicami na področju aplikativne analize diskurza večbesedne diskurzne označevalci v naši raziskavi opredelili kot propozicijsko neobvezne in strukturno zaključene pogoste kombinacije dveh ali več zaporednih grafemskih besed s prepoznavno funkcijo diskurznega povezovanja, strukturiranja ali upravljanja, s čimer se naša opredelitev diskurznihi označevalci vzpostavlja kot nadpomenka dosedanjih slovenističnih terminov, kot so opomenjevalni konektorji, členki s kohezivno vlogo, pragmatični povezovalci, členkovni besedilni povezovalci, govorni signali, kazalci stopenj besedilne zgradbe in diskurzni označevalci (v ožjem smislu). Kljub razmeroma podrobno razdelanim izhodiščnim funkcijskim in izraznim merilom za prepoznavanje večbesednih diskurznihi označevalci na podlagi sorodnih tujih in domačih raziskav pa so se pri njihovem prenosu na avtentično jezikovno gradivo v procesu označevanju vzorčenega korpusa govorjene slovenščine nekatere skupine izrazov izkazale kot izrazito težavne z vidika nedvoumne jezikoslovne kategorizacije. S pomenskega oz. funkcijskega vidika tako izstopa predvsem

vprašanje razmejitve (večbesednih) diskurznihi označevalci glede na druge skupine nepropozicijskih izrazov, kot so izrazi epistemske naklonskosti, izrazi odzivanja in drugi tipično govorjeni pragmatični izrazi, z izraznega oz. strukturnega vidika pa po drugi strani izstopata predvsem vprašanji objektivnega določanja ustaljenosti in minimalne diskurzne enote.

Na podlagi dopolnjenih izhodiščnih meril smo tako v procesu označevanja vzorčenega korpusa prepoznali 125 večbesednih diskurznihi označevalci, ki smo jih v drugem koraku raziskave na podlagi istih meril dopolnili še z avtomatskih luščenjem večbesednih diskurznihi označevalci iz celotnega korpusa Gos, pri čemer smo glede na maloštevilnost raziskav, povezanih z določanjem optimalne metode luščenja diskurznofunkcijskih večbesednih enot, za ta namen uporabili osem različnih statističnih metod za razvrščanje dvo- ali večbesednih nizov, tj. razvrščanje z običajnim in prilagojenim štetjem ter razvrščanje glede na šest uveljavljenih mer povezovalnosti (MI, MI³, LL, Dice, χ^2 , t-test). Rezultati luščenja so potrdili, da ti izrazi v slovenščini predstavljajo razmeroma maloštevilno, a izrazito pogosto skupino večbesednih enot, vendarle pa je pri pridobivanju njihovega izčrpnega nabora prevladujočo metodo ročnega pregledovanja majhnih vzorcev jezikovne rabe smiselno dopolnjevati s kvantitativnimi metodami luščenja iz obsežnejših besedilnih korpusov, saj smo na ta način v korpusu govorjene slovenščine prepoznali 79 novih označevalci. Pri tem so se med uporabljenimi statističnimi metodami za priklic večbesednih diskurznihi označevalci kot najbolj učinkoviti izkazali mera t-test in oba načina razvrščanja glede na pogostost, ostale mere povezovalnosti pa so za namen prepoznavanja teh jezikovnih sredstev manj uporabne.

Z unijo kvalitativnih in kvantitativnih metod smo tako v korpusu Gos identificirali 204 večbesedne diskurzne označevalce govorjene slovenščine, med katerimi se jih glede na rezultate statistične primerjave z rabo v referenčnem korpusu pisne slovenščine Kres velika večina tipično pojavlja predvsem v govorjenem diskurzu. Med najbolj tipično govorjenimi označevalci pričakovano prevladujejo pomensko izpraznjeni označevalci, vezani na specifične okoliščine govorjene interakcije, kot sta prisotnost naslovnika ali tvorjenje v realnem času, vendarle pa se med njimi pojavljajo tudi označevalci drugih funkcijskih skupin, kot so označevalci semantičnih ali pragmatičnih medbesedilnih povezav (besedilni povezovalci oz. konektorji). To dokazuje, da je za govorjeni jezik tipična funkcija diskurznega upravljanja nasploh, med statistično

izstopajoče, tipično govorjene večbesedne diskurzne označevalce pa se tako umešča zelo dolg in raznolik nabor izrazov, ki ga ni mogoče opazovati kot funkcijsko ali izrazno homogeno množico.

Končna analiza rabe treh najbolj tipično govorjenih večbesednih diskurznih označevalcev, tj. označevalcev *a ne*, *a veš* in *ne vem*, v uporabniških spletnih vsebinah je potrdila, da so ti na spletu rabljeni bistveno redkeje kot v spontanem govoru, vendarle pa njihova raba statistično ni nezanemarljiva in je izrazito pogostejša tudi v primerjavi z rabo v drugih, starejših oblikah pisne komunikacije. V skladu z dejstvom, da uporabniških spletnih vsebin ni mogoče obravnavati kot enotne jezikovne zvrsti, saj na jezikovno rabo znotraj teh platform računalniško posredovane komunikacije vplivajo številni tehnološki in družbeni dejavniki, tudi naša raziskava potrjuje, da je raba diskurznih označevalcev odvisna od samega tipa uporabniških spletnih vsebin, kot so tviti, forumske izmenjave, blogovski zapisi ali komentarji. Kot dejavnik, ki najbolj vpliva na pogostost rabe tipično govorjenih označevalcev, se pri tem kaže stopnja interaktivnosti oz. dialošкости posameznih besedilnih tipov (bolj kot medij spodbuja dinamično izmenjavo besedil med uporabniki, pogostejša je raba tipično govorjenih diskurznih označevalcev), primerjava distribucije posameznih označevalcev pa obenem razkriva, da je njihova raba odvisna tudi od njihovega funkcijskega razpona in sovpadanja s specifičnimi družbenimi dejavniki posameznih spletnih platform, kot so struktura udeležencev, odnosi med udeleženci in sporazumevalni namen.

Kot sugerira že sama primerjava pogostosti rabe treh najbolj tipično govorjenih večbesednih diskurznih označevalcev v govorjeni in spletni slovenščini, tudi podrobnejša primerjava njihovih konkretnih funkcij v obeh tipih jezikovne rabe potrjuje, da so ti izrazi v govoru rabljeni predvsem kot sredstvo za zapolnjevanje vrzeli oz. segmentacijo diskurza med kognitivnim načrtovanjem nadaljnega govora. Ob odsotnosti tovrstnega pritiska tvorjenja v realnem času v uporabniških spletnih vsebinah po drugi strani izstopajo predvsem druge, semantično motivirane vloge, kot so vzpostavljjanje stika z naslovnikom, preverjanje strinjanja ali omiljevanje izrečenega, ki se hkrati prepletajo tudi z besedilnimi funkcijami poudarjanja ter prevzemanja ali predajanja vloge. Vsi trije analizirani označevalci se na spletu pojavljajo tudi v nekaterih novih kontekstih rabe, kot so nagovarjanje neznanega oz. generičnega naslovnika, raba za doseganje humornega učinka ali izražanje pokroviteljskega odnosa, vstopajo pa tudi v nekatere nove frazeološke zveze.

Summary

An overview of research on discourse markers and related discourse-functional linguistic categories shows a general consensus on the fact that indexicality can be considered the fundamental feature that distinguishes these linguistic devices from other types of non-propositional expressions. Nevertheless, many differing definitions of this indexical function have been proposed, depending on the underlying theory of discourse segmentation and the related discourse marker scope (from markers of textual coherence to markers of non-linguistic aspects of communication processes), the presupposed distinction between semantic and pragmatic discourse management, and the differentiation between semantics and pragmatics in general. In addition to function-based discrepancies, often coinciding with the type of data under investigation and the divergent interpretations of the form-function relationship, the subfield of research on multi-word discourse markers has further been characterised by questions related to their structural completeness and fixedness, especially under the influence of research on formulaic language demonstrating that discourse-marking functions, hitherto usually attributed to a few prototypical highly grammaticalized fixed expressions, are also performed by other types of frequently-used multi-word expressions.

To overcome this heterogeneity of theory-driven approaches, in line with the recent tendencies in applied discourse analysis, this research adopts a relatively theory-neutral approach, in which multi-word discourse markers are defined as any recurring structurally complete non-propositional combination of two or more orthographic words with a recognizable function of discourse relating, structuring or managing, thus introducing an umbrella category for a wide range of discourse-related expressions that have already been explored in Slovenian linguistics, such as connectors, cohesive particles, pragmatic connectives, text-connective particles, speech signals, text-structure indicators and discourse markers (in the narrow sense). However, despite a relatively elaborated set of functional and formal criteria for multi-word discourse marker identification based on previous related research, their application to authentic data, i.e. the multi-word discourse marker annotation in the sampled section of the Gos reference corpus of spoken Slovenian, revealed several challenging categorisation issues. From the semantic perspective, they mostly include questions related to unambiguous delimitation of (multi-word) discourse markers from other types of non-propositional expressions, such as expressions of epistemic

modality, feedback signals and other speech-specific pragmatic expressions; whereas from the formal perspective, these issues include questions related to the notions of fixedness and discourse segmentation.

Based on the enhanced annotation guidelines, 125 multi-word discourse markers have been identified in the sampled corpus, however, this list was further enlarged in the second stage of our research, in which the same identification criteria were used for semi-automatic extraction of multi-word discourse markers from the Gos reference corpus of spoken Slovene. Given the scarce amount of research on discourse-related multi-word expression extraction methods, eight different statistical methods were used for this task, i.e. ranking by normal and adjusted frequencies, and ranking by six general association measures (MI, MI³, LL, Dice, χ^2 , t-test). The extraction results have confirmed that multi-word discourse markers in spoken Slovenian represent a relatively closed group of frequently used expression. Nevertheless, the fact that as many as 79 new multi-word discourse have been identified in this process suggests that the predominant qualitative methods based on annotation of small language samples should be complemented with quantitative extraction methods based on larger corpora. In particular, best results in multi-word discourse marker identification in Gos have been achieved by using the t-test association measure and both frequency counts, whereas other association measures have proven to be less appropriate for this specific task.

By combining qualitative and quantitative identification methods, 204 multi-word discourse markers have been identified in the Gos reference corpus of spoken Slovene, among which the majority typically occurs in spoken interaction, as demonstrated by keyness calculations in relation to the Kres reference corpus of written Slovene. As expected, the list of the most typical spoken discourse markers mainly includes markers related to the specific circumstances of spontaneous speech production, such as the presence of the addressee or online planning, however, other types of markers also occur in the list, such as markers of semantic and pragmatic discourse relations (discourse connectives). This confirms the importance of explicit discourse management in spontaneous speech in general, leading to a long and diverse list of speech-specific markers, which cannot be considered a semantically or formally homogenous class.

The subsequent analysis of the three most typical multi-word discourse markers occurring in spoken Slovenian (*a ne, a veš, ne vem*) in the Janes reference corpus of user-generated content has confirmed they are used much less often than in spontaneous speech, however, their frequency of usage in user-generated content is nevertheless significant, especially in comparison with other types of written communication. In line with the fact that user-generated content constitutes a heterogeneous linguistic modality, influenced by various technical and social factors, our research confirms that the usage of discourse markers also depends on the type of user-generated content, such as tweets, forum exchanges, blog posts or comment. In particular, the degree of interactivity or dialogue-like communication can be seen as the most influential factor in discourse marker usage (the more a platform encourages a dynamic exchange between users, the more frequent the usage of typically spoken discourse markers is). In addition to that, the comparison of text-type distribution analysis for each of the three markers has shown that their usage also depends on their semantic characteristics and the specific social features of each platform, such as participation structure, participant relations and purpose.

As has already been suggested by the general comparison of their frequency of usage, an in-depth analysis of the usage of the three markers in both corpora confirms that, in speech, typically spoken multi-word discourse markers are mainly used as markers of speech production and online planning, signalling the speaker's intention to retain, take or give turn. Given the absence of such cognitive pressure in asynchronous user-generated communication, however, the usage of discourse markers in this modality is mainly motivated by their semantic potential (e.g. to relate to the addressee, check for agreement or alleviate assertion), intertwined with discourse-structuring functions, such as emphasising or turn taking/giving. All three markers also exhibit some new types of usage in user-generated content, such as addressing an unknown/generic addressee, producing a humorous effect, patronizing, and entering into new types of fixed expressions.

Bibliografija

- Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., in King, J. (2011). How can you say such things?!?: recognizing disagreement in informal political argument. V: *Proceedings of the Workshop on Languages in Social Media*, str. 2–11.
- Adolphs, S., in Carter, R. (2013). *Spoken Corpus Linguistics: From Monomodal to Multimodal*. London, New York: Routledge.
- Afantenos, S., Asher, N., Benamara, F., Bras, M., Fabre, C., Ho-Dac, M., ... Vieu, L. (2012a). An empirical resource for discovering cognitive principles of discourse organisation: the ANNODIS corpus. V: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, str. 2727–2734.
- Afantenos, S., Asher, N., Benamara, F., Cadilhac, A., Cedric, D., Denis, P., ... Vieu, L. (2012b). Modelling Strategic Conversation: model, annotation design and corpus. V: *Proceedings of the 16th Workshop on the Semantics and Pragmatics of Dialogue (SeineDial)*, str. 167–168.
- Aijmer, K. (1996). *Conversational Routines in English: Convention and Creativity*. London, New York: Addison Wesley Longman.
- Aijmer, K. (1997). I think—an English modal particle. V: T. Swan in O. Jansen Westvik (ur.), *Modality in Germanic languages. Historical and Comparative Perspectives*, str. 1–47. De Gruyter Mouton.
- Aijmer, K. (2002). *English Discourse Particles*. Amsterdam: John Benjamins Publishing.
- Aijmer, K. (2013). Analyzing modal adverbs as modal particles and discourse markers. V: L. Degand, B. Corneille in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and Description*, str. 91–106. John Benjamins Publishing.
- Aijmer, K., Foolen, A., in Simon-Vandenberg, A.-M. (2006). Pragmatic markers in translation: a methodological proposal. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 102–114. Elsevier.
- Aijmer, K., in Simon-Vandenberg, A.-M. (2011). Pragmatic markers. V: J. Zienkowski, J. O. Östman in J. Verschueren (ur.), *Discursive Pragmatics*, str. 223–247. John Benjamins Publishing.
- Aitchison, J., in Lewis, D. M. (2003). *New Media Language*. Routledge.
- Allen, K., Carenini, G., in Ng, R. T. (2014). Detecting Disagreement in Conversations using Pseudo-Monologic Rhetorical Structure. V: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, str. 1169–1180.

- Alonso, L., Castellón, I., in Cor, L. P. (2002a). Lexicón Computacional de Marcadores del Discurso. *Procesamiento Del Lenguaje Natural*, 29, str. 239–246.
- Alonso, L., Castellón, I., in Padró, L. (2002b). X-T RACTOR : A Tool For Extracting Discourse Markers. V: *Proceedings of the LREC 2002 workshop on Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*, str. 100–105.
- Al-Saif, A., in Markert, K. (2010). The Leeds Arabic Discourse Treebank: Annotating Discourse Connectives for Arabic. V: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, str. 2046–2053.
- Anscombe, J. C., in Ducrot, O. (1983). *L'argumentation dans la langue*. Bruselj: Pierre Mardaga.
- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Springer Netherlands.
- Asher, N., in Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.
- Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.
- Balažic Bulc, T. (2009a). Slovarski opisi pomensko občutljivih kategorij : konektorji in SSKJ. V: M. Stabej (ur.), *Obdobja 28: Infrastruktura slovenščine in slovenistike*, str. 33–39. Ljubljana: Znanstvena založba FF UL in Center za slovenščino kot drugi/tuji jezik.
- Balažic Bulc, T. (2009b). *Torej, namreč, zato ... o konektorjih: Raba in funkcija konektorjev v slovenskem in hrvaškem jezikoslovnem diskurzu*. Ljubljana: Filozofska fakulteta.
- Balažic Bulc, T. (2010). Pozicija konektorjev v makrostrukuri znanstvenega članka. V: Š. Vintar (ur.), *Slovenske Korpusne Raziskave*, str. 54–71. Ljubljana: Znanstvena založba FF UL.
- Balažic Bulc, T. (2015). Členek v slovenskem jezikoslovju in slovarju. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve*, str. 590–637. Ljubljana: Znanstvena založba FF UL.
- Balažic Bulc, T., in Gorjanc, V. (2015). The Position of Connectors in Slovene and Croatian Student Academic Writing: a Corpus-Based Approach. V: S. Starc, C. Jones in A. Maiorani (ur.), *Meaning Making in Text: Multimodal and Multilingual Functional Perspectives*, str. 51–71. London: Palgrave Macmillan.
- Baron, A., Rayson, P., in Archer, D. (2009). Word frequency and keyword statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies*, 20(1), str. 41–67.
- Baron, N. S. (1998). Letters by phone or speech by other means: the linguistics of email. *Language & Communication*, 18(2), str. 133–170.

- Baron, N. S. (2008). *Always On: Language in an Online and Mobile World*. Oxford University Press.
- Baron, N. S. (2010). Discourse Structures in Instant Messaging : The Case of Utterance Breaks. *Language@Internet*, 7, članek 4.
- Baym, N. K. (1995). The emergence of community in computer-mediated communication. V: S. G. Jones (ur.), *CyberSociety: Computer-mediated communication and community*, str. 38–163. Sage Publications.
- Bazzanella, C. (2006). Discourse markers in Italian: towards a ‘compositional’ meaning. V: K. Fischer (ur.), *Approaches to Discourse Particles* (str. 449–464). Elsevier.
- Beißwenger, M. (2007). Corpora zur computervermittelten (internetbasierten) Kommunikation. *Zeitschrift Für Germanistische Linguistik*, 35, str. 496–503.
- Beißwenger, M., Bartsch, S., Evert, S., in Würzner, K.-M. (2016). EmpiriST 2015: A Shared Task on the Automatic Linguistic Annotation of Computer-Mediated Communication, Social Media and Web Corpora. V: *Proceedings of the 10th Web as Corpus Workshop (WAC-X)*, str. 44–56.
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge University Press.
- Biber, D. (2006). *University Language: A Corpus-based Study of Spoken and Written Registers*. John Benjamins Publishing.
- 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), str. 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), str. 371–405.
- Biber, D., Johansson, S., Leech, G., in Conrad, S. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- Bieswanger, M. (2013). Micro-linguistic structural features of computer-mediated communication. V: S. C. Herring, D. Stein in T. Virtanen (ur.), *Pragmatics of Computer-Mediated Communication*. Berlin, Boston: De Gruyter Mouton.
- Blakemore, D. (1987). *Semantic Constraints on Relevance*. Blackwell Publishing.
- Blakemore, D. (1996). Are Apposition Markers Discourse Markers?. *Journal of Linguistics*, 32(2), str. 325–347.

- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge University Press.
- Blakemore, D. (2006). Divisions of labour: The analysis of parentheticals. *Lingua*, 116(10), str. 1670–1687.
- Blakemore, D. (2008). Discourse Markers. V: L. R. Horn in G. Ward (ur.), *The Handbook of Pragmatics*, str. 221–240. Oxford, UK: Blackwell Publishing.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum*, 1(1), str. 1–14.
- Bolly, C., Crible, L., Degand, L., in Uygun, D. (2014). Towards a Model for Discourse Marker Annotation in spoken French: From potential to feature-based discourse markers. *Pragmatic Markers, Discourse Markers and Modal Particles: What Do We Know and Where Do We Go from Here?*
- Bolly, C., Crible, L., Degand, L., in Uygun-Distexhe, D. (2015). MDMA. Un modèle pour l'identification et l'annotation des marqueurs discursifs « potentiels » en contexte. *Discours*, 16.
- Bondi, M. (2013). Connectives. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- Bondi, M., in Scott, M. (ur.). (2010). *Keyness in Texts*. Amsterdam: John Benjamins Publishing.
- Bordería, S. P. (2006). A functional approach to the study of discourse markers. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 77–100. Elsevier.
- Bordería, S. P. (2008). Do discourse markers exist? On the treatment of discourse markers in Relevance Theory. *Journal of Pragmatics*, 40(8), str. 1411–1434.
- Bou-Franch, P., Lorenzo-Dus, N., in Blitvich, P. G.-C. (2012). Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence. *Journal of Computer-Mediated Communication*, 17(4), str. 501–521.
- Brinton, L. J. (2008). *The comment clause in English*. Cambridge University Press.
- Brinton, L. J. (2010). The Development of I mean: Implications for the Study of Historical Pragmatics. V: S. M. Fitzmaurice and I. Taavitsainen (ur.), *Methods in Historical Pragmatics*, 37–80. Berlin: Mouton de Gruyter.
- Brinton, L. J., in Traugott, E. C. (2005). *Lexicalization and Language Change*. Cambridge University Press.
- Briz Gómez, A., Pons Bordería, S., in Portolés Lázaro, J. (2008). Diccionario de partículas discursivas del español (DPDE): www.dpde.es.

- Bross, F. (2012). German modal particles and the common ground. *Helikon. A Multidisciplinary Online Journal*, 2, str. 182–209.
- Buch-Kromann, M., in Korzen, I. (2010). The unified annotation of syntax and discourse in the Copenhagen Dependency Treebanks. V: *Proceedings of the Fourth Linguistic Annotation Workshop (LAW IV '10)*, str. 127–131.
- Bunt, H., Alexandersson, J., Choe, J., Fang, A. C., Hasida, K., Petukhova, V., ... Traum, D. (2012). ISO 24617-2 : A semantically-based standard for dialogue annotation. V: *Proceedings of LREC 2012*, str. 430–437.
- Bunt, H., in Prasad, R. (2016). Core Concepts for the Annotation of Discourse Relations. V: *Proceedings of the LREC 2016 Workshop 'ISA-12: 12th Joint ACL - ISO Workshop on the Interoperable Semantic Annotation'*, str. 45–55.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge University Press.
- Callaway, C. B., in Lester, J. C. (2002). Narrative prose generation. *Artificial Intelligence*, 139(2), str. 213–252.
- Candido Jr., A., Maziero, E., Gasperin, C., Pardo, T. A. S., Specia, L., in Aluísio, S. M. (2009). Supporting the Adaptation of Texts for Poor Literacy Readers : a Text Simplification Editor for Brazilian Portuguese. V: *Proceedings of the NAACL HLT Workshop on Innovative Use of NLP for Building Educational Applications*, str. 34–42.
- Cardoso, P. C., Maziero, E. G., Jorge, M. L., Seno, E. M., Di Felipo, A., Rino, L. H., ... Pardo, T. A. (2011). CSTNews -A Discourse-Annotated Corpus for Single and Multi-Document Summarization of News Texts in Brazilian Portuguese. V: *Proceedings of the 3rd RST Brazilian Meeting*, str. 88–105.
- Carlson, L., Marcu, D., in Okurowski, M. E. (2003). Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. V: J. van Kuppevelt in W. Smith (ur.), *Current Directions in Discourse and Dialogue*, str. 85–112. Kluwer Academic Publishers.
- Cavnar, W. B., Trenkle, J. M., in Mi, A. A. (1994). N-Gram-Based Text Categorization. V: *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, str. 161–175.
- Chafe, W. (1980). The deployment of consciousness in the construction of narrative. V: W. Chafe (ur.), *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, str. 9–50. Norwood: Ablex.
- Chafe, W., in Tannen, D. (1987). The Relation Between Written and Spoken Language. *Annual Review of Anthropology*, 16(1), str. 383–407.

- Chaudron, C., in Richards, J. C. (1986). The Effect of Discourse Markers on the Comprehension of Lectures. *Applied Linguistics*, 7(2), str. 113–127.
- Cherny, L. (1999). *Conversation and community: chat in a virtual world*. Stanford: CSLI Publications.
- Church, K. W., in Hanks, P. (1989). Word association nouns, Mutual Information, and lexicography. V: *Proceedings of the 27th Annual Conference of the Association for Computational Linguistics*, str. 22–29.
- Clark, H. H., in Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(84), str. 73–111.
- Condon, S. L., in Čech, C. G. (2010). Discourse Management in Three Modalities. *Language@Internet*, 7, str. 1–32.
- Conklin, K., in Schmitt, N. (2012). The Processing of Formulaic Language. *Annual Review of Applied Linguistics*, 32, str. 45–61.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), str. 397–423.
- 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*, str. 197–216.
- Coulmas, F. (1979). On the sociolinguistic relevance of routine formulae. *Journal of Pragmatics*, 3(3–4), str. 239–266.
- Crible, L. (2014). Identifying and describing discourse markers in spoken corpora: Annotation protocol v.8. Université catholique de Louvain.
- Crible, L. (2016). Towards an operational category of discourse markers: A definition and its model. V: A. Sanso and C. Fedriani (ur.), *Discourse markers, Pragmatics Markers and Modal Particles: New Perspectives* (v tisku). John Benjamins.
- Crible, L., in Zufferey, S. (2015). Using a unified taxonomy to annotate discourse markers in speech and writing. V: *Proceedings of the 11th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-11)*, str. 14–22.
- Croft, W. (2000). Parts of speech as language universals and as language-particular categories. V: P. M. Vogel in B. Comrie (ur.), *Approaches to the typology of word classes*, str. 65–102. Berlin: Mouton de Gruyter.
- Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.

- Crystal, D. (2006). *Language and the Internet*. Cambridge University Press.
- Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Routledge.
- Cuenca, M.-J. (2013). The fuzzy boundaries between discourse marking and modal marking. V: L. Degand, B. Cornillie in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and Description*, str. 191–216. John Benjamins.
- Cuenca, M.-J., in Marín, M.-J. (2012). Discourse markers and modality in spoken Catalan: The case of (és) clar. *Journal of Pragmatics*, 44(15), str. 2211–2225.
- da Cunha, I., Torres-Moreno, J.-M., in Sierra, G. (2011). On the development of the RST Spanish treebank. V: *Proceedings of the Fifth Law Workshop (LAW V)*, str. 1–10)
- da Silva, J. F., in Lopes, G. P. (1999). A Local Maxima method and a Fair Dispersion Normalization for extracting multi-word units from corpora. V: *Proceedings of the 6th Meeting on the Mathematics of Language*, str. 369–381.
- Danlos, L., Antolinos-Basso, D., Braud, C., in Roze, C. (2012). Vers le FDTB : French Discourse Tree Bank. V: *Actes de la conférence conjointe JEP-TALN-RECITAL 2012, volume 2: TALN*, str. 471–478.
- Degand, L., Cornillie, B., in Pietrandrea, P. (ur.). (2013a). *Discourse Markers and Modal Particles: Categorization and description*. Amsterdam: John Benjamins Publishing.
- Degand, L., Cornillie, B., in Pietrandrea, P. (2013b). Modal particles and discourse markers: Two sides of the same coin? Introduction. V: L. Degand, B. Cornillie, in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and description*, str. 1–18. Amsterdam: John Benjamins Publishing.
- Degand, L., in Evers-Vermeul, J. (2015). Grammaticalization or pragmaticalization of discourse markers?: More than a terminological issue. *Journal of Historical Pragmatics*, 16(1), str. 59–85.
- Degand, L., Lefèvre, N., in Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design*, 1(1), str. 39–51.
- Degand, L., in Sanders, T. (2002). The impact of relational markers on expository text comprehension in L1 and L2. *Reading and Writing*, 15 (7), str. 739–757.
- Degand, L., in Simon, A. C. (2009). On identifying basic discourse units in speech: theoretical and empirical issues. *Discours*, 4.
- Degand, L., in Simon-Vandenberghe, A.-M. (2011). Introduction: Grammaticalization and (inter)subjectification of discourse markers. *Linguistics*, 49(2), str. 287–294.

- Degand, L., in van Bergen, G. (2016). Discourse markers as turn-transition devices: Evidence from speech and instant messaging. *Discourse Processes*, 15(4), str. 1–25.
- Dehé, N., in Kavalova, Y. (ur.). (2007). *Parentheticals*. John Benjamins Publishing.
- Dehe, N., in Wichmann, A. (2010). Sentence-initial I think (that) and I believe (that): Prosodic evidence for use as main clause, comment clause and discourse marker. *Studies in Language*, 34(1), str. 36–74.
- Dér, C. (2010). On the status of discourse markers. *Acta Linguistica Hungarica*, 57(1), str. 3–28.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3), str. 297–302.
- Diewald, G. (2006). Discourse particles and modal particles as grammatical elements. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 403–426. Elsevier.
- Diewald, G. (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics*, 49(2), str. 365–390.
- Diewald, G. (2013). Same same but different: modal particles and discourse particles and the art (and purpose) of categorization. V: L. Degand, B. Corneille in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and Description*, str. 409–446. Amsterdam: John Benjamins Publishing.
- van Dijk, T. A. (1979). Pragmatic Connectives. *Journal of Pragmatics*, 3, 447–456.
- Dipper, S., in Stede, M. (2006). Disambiguating potential connectives. V: *Proceedings of KONVENS* (Vol. 6), str. 167–173.
- Dittmar, N. (2013). Transcription. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- 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), str. 551–582.
- Dobrovoljc, K., Erjavec, T., in Krek, S. (2015a). Leksikon besednih oblik Sloleks in smernice njegovega razvoja. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve*, str. 80–105. Ljubljana: Znanstvena založba Filozofske fakultete.
- Dobrovoljc, K., Krek, S., Holozan, P., Erjavec, T., in Romih, M. (2015b). Morphological lexicon Sloleks 1.2. Centre for Language Resources and Technologies, University of Ljubljana.

- Dobrovoljc, K., in Nivre, J. (2016). The Universal Dependencies Treebank of Spoken Slovenian. V: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, str. 1566–1573.
- Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, 19(1), str. 61–74.
- Eckle-Kohler, J., Kluge, R., in Gurevych, I. (2015). On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse. V: *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, str. 2236–2242.
- Ellis, N. C., Simpson-Vlach, R., in Maynard, C. (2008). Formulaic Language in Native and Second Language Speakers: Psycholinguistics, Corpus Linguistics, and TESOL. *TESOL Quarterly*, 42(3), str. 375–396.
- Erjavec, T. (2013). Korpusi in konkordančniki na strežniku nl.ijs.si. *Slovenščina 2.0*, 1(1), str. 24–49.
- Erjavec, T., Fišer, D., Krek, S., in Ledinek, N. (2010). The JOS Linguistically Tagged Corpus of Slovene. V: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, str. 1806–1809.
- 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*, str. 49–53.
- Erman, B. (2001). Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics*, 33(9), str. 1337–1359.
- Erman, B., in Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), str. 29–62.
- Evert, S. (2009). Corpora and collocations. V: A. Lüdeling in M. Kytö (ur.), *Corpus Linguistics. An International Handbook* (Vol. 1), str. 1212–1248. De Gruyter Mouton.
- Ferrara, K., Brunner, H., in Whittemore, G. (1991). Interactive Written Discourse as an Emergent Register. *Written Communication*, 8(1), str. 8–34.
- Filliettaz, L., in Roulet, E. (2002). The Geneva Model of discourse analysis: an interactionist and modular approach to discourse organization. *Discourse Studies*, 4(3), str. 369–393.
- Firth, J. (1957). *Papers in linguistics 1934-1951*. Oxford University Press.
- Fischer, K. (2000). *From Cognitive Semantics to Lexical Pragmatics: The Functional Polysemy of Discourse Particles*. Walter de Gruyter.

- Fischer, K. (ur.). (2006a). *Approaches to discourse particles*. Amsterdam: Elsevier.
- Fischer, K. (2006b). Frames, constructions, and invariant meanings: the functional polysemy of discourse particles. V: K. Fischer (ur.), *Approaches to discourse particles*, str. 427–448. Amsterdam: Elsevier.
- Fischer, K. (2006c). Towards an Understanding of the Spectrum of Approaches to Discourse Particles: Introduction to the Volume. V: K. Fischer (ur.), *Approaches to discourse particles*, str. 1–20. Amsterdam: Elsevier.
- Fischer, K. (2010). Beyond the sentence: Constructions, frames and spoken interaction. *Constructions and Frames*, 2(2), str. 1–28.
- Fischer, K. (2013). Discourse Markers. V: *The Encyclopedia of Applied Linguistics*, str. 1743–1748. Oxford, UK: Blackwell Publishing.
- Fischer, K. (2014). Discourse Markers. V: K. P. Schneider and A. Barron (ur.), *Pragmatics of Discourse*, str. 271–294. Berlin: De Gruyter Mouton.
- Fischer, K., in Alm, M. (2013). A radical Construction Grammar perspective on the modal particle-discourse particle distinction. V: L. Degand, B. Cornillie, in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and description*, str. 47–88. Amsterdam: John Benjamins Publishing.
- Fišer, D., Erjavec, T., Čibej, J., in Ljubešić, N. (2015). Gradnja in analiza korpusa spletne slovenščine JANES. V: M. Smolej (ur.), *Simpozij Obdobja 34: Slovnica in slovar – aktualni jezikovni opis*, str. 217–223.
- Fišer, D., Erjavec, T., in Ljubešić, N. (2016). JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0*, 4(2), str. 67–100.
- Flowerdew, J., in Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, 17(4), str. 435–458.
- Fox Tree, J. E. (2015). Discourse markers in writing. *Discourse Studies*, 17(1), str. 64–82.
- Frank-Job, B. (2006). A dynamic-interactional approach to discourse markers. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 395–413. Amsterdam: Elsevier.
- Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14(3), str. 383–398.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), str. 167–190.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, str. 931–952.
- Fraser, B. (2006). Towards a theory of discourse markers. V: K. Fischer (ur.), *Approaches to discourse particles*, str. 189–204. Amsterdam: Elsevier.

- Fraser, B. (2009). An Account of Discourse Markers. *International Review of Pragmatics*, 1, str. 293–320.
- Fraser, B. (2011). The sequencing of contrastive discourse markers in English. *Baltic Journal of English Language, Literature and Culture*, 1, str. 29–35.
- Fraser, B. (2013a). Combinations of Contrastive Discourse Markers in English. *International Review of Pragmatics*, 5(2), str. 318–340.
- Fraser, H. (2013b). Spoken Word Recognition. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- Fried, M., in Östman, J. O. (2005). Construction Grammar and spoken language: The case of pragmatic particles. *Journal of Pragmatics*, 37(11), str. 1752–1778.
- Gabrielatos, C. (2013). *Keyword Analysis*. Dubrovnik Fall School in Linguistic Methods. Centre for Advanced Academic Studies, Dubrovnik, Hrvatska.
- Gabrielatos, C., in Marchi, A. (2012). *Keyness: Appropriate metrics and practical issues*. CADS International Conference, Bologna, Italija.
- Gantar, P. (2007). *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC ZRC SAZU.
- Gantar, P., Kosem, I., in Krek, S. (2016). Discovering Automated Lexicography: The Case of the Slovene Lexical Database. *International Journal of Lexicography*, 29(2), str. 200–225.
- Gantar, P., Krek, S., Kosem, I., in Gorjanc, V. (2015). Collocation dictionary for Slovene: challenge for automatic extraction of data and crowdsourcing. V: *Proceedings of Europhras 2015. Computerised and corpus-based approaches to phraseology: Monolingual and multilingual perspectives*, str. 87–89.
- Garcia, A. C., in Baker Jacobs, J. (1999). The Eyes of the Beholder: Understanding the Turn-Taking System in Quasi-Synchronous Computer-Mediated Communication. *Research on Language and Social Interaction*, 32(4), str. 337–367.
- Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., ... Smith, N. A. (2011). Part-of-Speech Tagging for Twitter: Annotation, Features, and Experiments. V: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Shortpapers*, str. 42–47.
- Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.
- Gonzalez, M. (2005). Pragmatic markers and discourse coherence relations in English and Catalan oral narrative. *Discourse Studies*, 7(1), str. 53–86.

- Gorjanc, V. (1998). Konektorji v slovničnem opisu znanstvenega besedila. *Slavistična revija*, 46(4), str. 367–387.
- Gorjanc, V. (2005). *Uvod v korpusno jezikoslovje*. Domžale: Izolit.
- Gorjanc, V., in Krek, S. (2001). A Corpus-Based Dictionary Database as the Source for Compiling Slovene-X Dictionary. V: *Proceedings of 6th Conference on Computational Lexicography and Corpus Research: »Computational Lexicography and New EU Languages«*, str. 41–47.
- Gray, B., in Biber, D. (2013). Lexical frames in academic prose and conversation. *International Journal of Corpus Linguistics*, 18(1), str. 109–136.
- Grčar, M., Krek, S., in Dobrovoljc, K. (2012). Obeliks : statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. V: T. Erjavec in J. Žganec Gros (ur.), *Zbornik Osme Konference Jezikovne Tehnologije*, str. 89–94.
- Gries, S. T. (2010). Useful statistics for corpus linguistics. V: S. Aquilino in A. Moisés (ur.), *A mosaic of corpus linguistics: selected approaches*, str. 269–291. Frankfurt am Main: Peter Lang.
- Gries, S. T. (2012). Frequencies, probabilities, and association measures in usage-/exemplar-based linguistics: some necessary clarification. *Studies in Language*, 11(3), str. 477–510.
- Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next. *International Journal of Corpus Linguistics*, 18(1), str. 137–166.
- Gries, S. T. (2014). Frequency tables. Tests, effect sizes, and explorations. V: D. Glynn in J. A. Robinson (ur.), *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, str. 365–389. John Benjamins Publishing.
- Gries, S. T. (2015). Quantitative designs and statistical techniques. V: D. Biber in R. Reppen (ur.), *The Cambridge Handbook of English Corpus Linguistics*, str. 50–71. Cambridge: Cambridge University Press.
- Habernal, I., in Gurevych, I. (2016). Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, 43(1), (v tisku).
- Halliday, M. A. K. (1985). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. A. K., in Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Halliday, M. A. K., in Matthiessen, C. M. I. M. (2014). *Halliday's Introduction to Functional Grammar (Fourth Edition)*. Routledge.

- Hancock, J. T. (2009). Digital deception: Why, when and how people lie online. V: A. N. Joinson, K. Y. A. McKenna, T. Postmes, in U.-D. Reips (ur.), *Oxford Handbook of Internet Psychology*. Oxford University Press.
- Hanks, P. (2010). Nine issues in metaphor theory and analysis. *International Journal of Corpus Linguistics*, 15(1), str. 133–150.
- Hansen, M.-B. M. (1998a). *The Function of Discourse Particles: A Study with Special Reference to Spoken Standard French*. Amsterdam: John Benjamins Publishing.
- Hansen, M.-B. M. (1998b). The semantic status of discourse markers*. *Lingua*, 104, str. 235–260.
- Hansen, M.-B. M. (2006). A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French toujours). V: K. Fischer (ur.): *Approaches to discourse particles*, str. 21–41. Elsevier.
- Hardie, A. (2014). *Log Ratio – an informal introduction*. <http://cass.lancs.ac.uk/?p=1133>
- Harrison, S., in Allton, D. (2013). Apologies in email discussions. V: S. C. Herring, D. Stein in T. Virtanen (ur.), *Pragmatics of Computer-Mediated Communication*, str. 315–338. Berlin, Boston: de Gruyter.
- Heeman, P. A., in Allen, J. F. (1999). Speech Repairs, Intonational Phrases and Discourse Markers: Modeling Speakers' Utternaces in Spoken Dialogue. *Computational Linguistics*, 25(4), str. 527–571.
- Heine, B. (2013). On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics*, 51(6), str. 1205–1247.
- Hellermann, J., in Vergun, A. (2007). Language which is not taught: The discourse marker use of beginning adult learners of English. *Journal of Pragmatics*, 39(1), str. 157–179.
- Herring, S. C. (1996). Two variants of an electronic message schema. V: S. C. Herring (ur.), *Computer-Mediated Communication: Linguistic, social, and cross-cultural perspectives*, str. 81–106. John Benjamins Publishing.
- Herring, S. C. (1999). Interactional Coherence in CMC. *Journal of Computer-Mediated Communication*, 4(4).
- Herring, S. C. (2001). Computer-Mediated Discourse. V: D. Schiffrin, D. Tannen in H. Hamilton (ur.), *The Handbook of Discourse Analysis*, str. 612–634. Blackwell Oxford.
- Herring, S. C. (2007). A Faceted Classification Scheme for Computer - Mediated Discourse. *Discourse*, 2007(1), str. 1–37.

- Herring, S. C. (2012a). Discourse in Web2.0: Familiar, Reconfigured and Emergent. V: D. Tannen in A. M. Tester (ur.), *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and New Media*, str. 1–29. Washington, DC: Georgetown University Press.
- Herring, S. C. (2012b). Grammar and Electronic Communication. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- Herring, S. C., in Androutsopoulos, J. (2015). Computer-Mediated Discourse 2.0. V: D. Tannen, H. E. Hamilton in D. Schiffrin (ur.), *The Handbook of Discourse Analysis*, str. 127–151. Hoboken, USA: John Wiley & Sons.
- Herring, S. C., Das, A., in Penumarthy, S. (2005). *CMC Act Taxonomy*.
<http://ella.slis.indiana.edu/~herring/cmc.acts.html>
- Herring, S. C., Stein, D., in Virtanen, T. (2013a). Introduction to the pragmatics of computer-mediated communication. V: S. C. Herring, D. Stein in T. Virtanen (ur.), *Pragmatics of Computer-Mediated Communication*, str. 3–31. Berlin, Boston: De Gruyter Mouton.
- Herring, S. C., Stein, D., in Virtanen, T. (ur.) (2013b). *Pragmatics of computer-mediated communication*. Berlin, Boston: De Gruyter Mouton.
- Hirschberg, J., in Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3), str. 501–530.
- Hobbs, J. (1985). *On the coherence and structure of discourse*. CSLI Technical Report, str. 85–37. Stanford, CA, USA.
- Holozan, P., Krek, S., Pivec, M., Rigač, S., Rozman, S., in Velušček, A. (2008). *Specifikacije za učni korpus - projekt 'Sporazumevanje v slovenskem jeziku' - Kazalnik 2*.
http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik2/SSJ_Kazalnik_2_Specifikacije-ucni-korpus_v1.pdf
- Hopper, P. J., in Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.
- Hutchinson, B. (2003). Mining the Web for Discourse Markers. V: *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, str. 407–410.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. Bloomsbury Publishing.
- Ibn Faiz, S., in Mercer, R. E. (2013). Identifying Explicit Discourse Connectives in Text. *Lecture Notes in Computer Science*, 7884, str. 64–76.
- Iruskieta, M., Aranzabe, M. J. M. J., Ilaraza, A. D. De, de Ilaraza, A. D., Gonzalez, I., Lersundi, M., in de la Calle, O. L. (2013). The RST Basque TreeBank: an online search interface to

- check rhetorical relations. V: *Proceedings of the 4th Workshop »RST and Discourse Studies«*, str. 40–49.
- Jakop, N. (2006). *Pragmatična frazeologija*. Ljubljana: Založba ZRC SAZU.
- Johnston, J. E., Berry, K. J., and Mielke, P. W. (2006). Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills*, 103(2), str. 412–414.
- Jucker, A. H., in Dürscheid, C. (2012). The Linguistics of Keyboard-to-screen Communication. A New Terminological Framework. *Linguistik Online*, 56(6), str. 1–26.
- Jucker, A. H., in Ziv, Y. (ur.). (1998). *Discourse Markers*. Amsterdam: John Benjamins Publishing.
- Kaltenböck, G. (2007). Spoken parenthetical clauses in English: a taxonomy. V: N. Dehé in Y. Kavalova (ur.), *Parentheticals*, str. 25–52. Amsterdam: John Benjamins Publishing.
- Kaltenböck, G. (2009). Initial I think: main or comment clause. *Discourse and Interaction*, 2(1), str. 49–70.
- Kay, P., in Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: the 'What's X doing Y?' construction. *Language*, 75, str. 1–33.
- Kehler, A. (2002). *Coherence, reference, and the theory of grammar*. CSLI Publications.
- Kilgarriff, A. (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, 6(1), str. 97–133.
- Kilgarriff, A. (2012). Getting to know your corpus. V: P. Sojka, A. Horak, I. Kopecek idr. (ur.), *Proceedings of the 15th International Conference on Text, Speech and Dialogue (TSD2012)*, str. 3–15. Brno, Czech Republic: Springer.
- Kilgarriff, A., Baisa, V., Jakubiček, M., in Rychlý, P. (2015). Longest-commonest Match. V: *Electronic lexicography in the 21st century: linking lexical data in the digital age. Proceedings of the eLex 2015 conference*, str. 397–404,
- Kilgarriff, A., Rychly, P., Kovar, V., in Baisa, V. (2012). Finding Multiwords of More Than Two Words. V: *Proceedings of the 15th EURALEX International Congress*, str. 1–7.
- Knott, A. (1996). *A Data-Driven Methodology for Motivating a Set of Coherence Relations*. Doktorska disertacija. University of Edinburgh.
- Knott, A., in Dale, R. (1994). Using linguistic phenomena to motivate a set of coherence relations. *Discourse Processes*, 18, str. 35–62.

- Knott, A., in Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30(2), str. 135–175.
- Knox, J. S. (2009). Punctuating the home page: image as language in an online newspaper. *Discourse & Communication*, 3(2), str. 145–172.
- Koops, C., in Lohmann, A. (2015). A quantitative approach to the grammaticalization of discourse markers: Evidence from their sequencing behavior. *International Journal of Corpus Linguistics*, 20(2), str. 232–259.
- Kosem, I., Gantar, P., in Krek, S. (2013). Avtomatizacija leksikografskih postopkov. *Slovenščina* 2.0, 2, str. 139–164.
- Kosem, I., in Verdonik, D. (2012). Key word analysis of discourses in Slovene speech: differences and similarities. *Linguistica*, 1(52), str. 309–322.
- Kranjc, S. (1999). *Razvoj govora predšolskih otrok*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Krek, S. (2010). *Pridobivanje jezikovnih podatkov iz besedilnih korpusov za namen izdelave enojezičnih slovarjev in slovníc: Doktorska disertacija*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Kuiper, K. (1996). *Smooth Talkers: The Linguistic Performance of Auctioneers and Sportscasters*. Mahwah, NJ: Erlbaum.
- Kunst-Gnamuš, O. (1981). *Pomenska sestava povedi*. Ljubljana: Pedagoški inštitut pri Univerzi Edvarda Kardelja v Ljubljani.
- Labov, W., in Fanshel, D. (1977). *Therapeutic Discourse: Psychotherapy As Conversation*. New York: Academic Press.
- Lapshinova-Koltunski, E., in Kunz, K. (2014a). Conjunctions across Languages, Registers and Modes: semi-automatic extraction and annotation. V: A. Diaz-Negrillo in J. Diaz-Perez Francisco (ur.), *Specialisation and Variation in Language Corpora*, str. 77–104. Peter Lang.
- Lapshinova-Koltunski, E., in Kunz, K. A. (2014b). Annotating cohesion for multilingual analysis. V: *Proceedings 10th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*, str. 57–64.
- Lenk, U. (1998). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, 30(2), str. 245–257.
- Lewis, D. M. (2006). Discourse markers in English: a discourse-pragmatic view. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str 43–59. Amsterdam: Elsevier.

- Lewis, D. M. (2011). A discourse-constructional approach to the emergence of discourse markers in English. *Linguistics*, 49(2), str. 415–443.
- Lin, Y.-L. (2013). Discourse functions of recurrent multi-word sequences in online and spoken intercultural communication. V: J. Romero-Trillo (ur.), *Yearbook of Corpus Linguistics and Pragmatics 2013*, str. 105–129. Springer.
- Ljubešić, N., Dobrovoljc, K., and Fišer, D. (2015). *MWElex – MWE Lexica of Croatian, Slovene and Serbian Extracted from Parsed Corpora. *Informatica*, 39(3), str. 293–300.
- Ljubešić, N., in Erjavec, T. (2016). Corpus vs. Lexicon Supervision in Morphosyntactic Tagging: The Case of Slovene. V: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, str. 1527–1531.
- Ljubešić, N., Erjavec, T., in Fišer, D. (2014). Standardizing tweets with character-level machine translation. V: *Proceedings of the 15th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2014)*, str. 164–175. New York: Springer-Verlag.
- Logar, N., Gantar, P., in Kosem, I. (2014). Collocations and examples of use: a lexical-semantic approach to terminology. *Slovenščina 2.0*, 2(1), str. 41–61.
- 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.
- Louis, A., Joshi, A., in Nenkova, A. (2010). Discourse indicators for content selection in summarization, V: *SIGDIAL '10 Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, str. 147–156.
- Louwerse, M. M., in Mitchell, H. H. (2003). Towards a taxonomy of a set of discourse markers in dialog: a theoretical and computational linguistic account. *Discourse Processes*, 35(1), str. 199–239.
- Lü, X., Zhang, L., in Hu, J. (2005). Statistical substring reduction in linear time. V: *Natural Language Processing – IJCNLP 2004*, str. 320–327. Berlin, Heidelberg: Springer.
- Maat, H. P., in Degand, L. (2002). Scaling causal relations and connectives in terms of speaker involvement. *Cognitive Linguistics*, 12(3), str. 211–245.
- Macaulay, R. (2002). You know, it depends. *Journal of Pragmatics*, 34(6), str. 749–767.
- Mann, W. C., in Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes*, 9(1), str. 57–90.

- Mann, W. C., in Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3), str. 243–281.
- Manning, C., in Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Marcu, D. (1998). A surface-based approach to identifying discourse markers and elementary textual units in unrestricted texts. V: *Proceedings of COLING-ACL Workshop on Discourse Relations and Discourse Markers*, str. 1–7.
- Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, MA: MIT Press.
- Marcu, D., Carlson, L., in Watanabe, M. (2000). The automatic translation of discourse structures. V: *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference (NAACL 2000)*, str. 9–17.
- Martin, J. R. (1992). *English Text: System and structure*. John Publishing.
- Maschler, Y., in Schiffrin, D. (2015). Discourse markers: Language, meaning, and context. V: D. Tannen, H. E. Hamilton in D. Schiffrin (ur.), *The Handbook of Discourse Analysis*, str. 189–221. Hoboken, NJ, USA: John Wiley & Sons.
- Maskey, S., in Hirschberg, J. (2005). Comparing lexical, acoustic/prosodic, structural and discourse features for speech summarization. V: *Proceedings of Eurospeech '05*, str. 621–624.
- Matthiessen, C. M. I. M., in Thompson, S. A. (1988). The structure of discourse and ‘subordination’. V: J. Haiman in S. A. Thompson (ur.), *Clause Combining in Grammar and Discourse*, str. 275–329. Amsterdam, Philadelphia: John Benjamins.
- Maynor, N. (1994). The Language of Electronic Mail: Written Speech?. V: G. D. Little in M. Montgomery (ur.): *Centennial Usage Studies*, str. 48–54. Tuscaloosa, AL: University of Alabama Press.
- McCarthy, M., in Carter, R. (2006). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. V: M. McCarthy (ur.), *Explorations in Corpus Linguistics*, str. 7–26. Cambridge: Cambridge University Press.
- McEnery, T., Xiao, R., in Tono, Y. (2006). *Corpus-based language studies. An advanced resource book*. London: Routledge.

- Meyer, T., Popescu-Belis, A., Zufferey, S., in Cartoni, B. (2011). Multilingual annotation and disambiguation of discourse connectives for machine translation. V: *Proceedings of 12th SIGdial Meeting on Discourse and Dialogue*, str. 194–203.
- Meyer, T., in Popescu-Belis, A. (2012). Using sense-labeled discourse connectives for statistical machine translation. V: *Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra)*, str. 129–138.
- Michelizza, M. (2015). *Spletna besedila in jezik na spletu: primer blogov in wikipedije v slovenščini*. Založba ZRC, ZRC SAZU.
- Misra, A., in Walker, M. (2013). Topic Independent Identification of Agreement and Disagreement in Social Media Dialogue. V: *Proceedings of the SIGDIAL 2013 Conference*, str. 41–50.
- Mittal, N., Agarwal, B., Agarwal, S., Agarwal, S., in Gupta, P. (2013). A Hybrid Approach for Twitter Sentiment Analysis. V: *Proceedings of ICON-2013: 10th International Conference on Natural Language Processing*, str. 116–120.
- Moon, R. (1998). *Fixed expressions and idioms in English : a corpus-based approach*. Oxford, New York: Clarendon Press.
- Murray, D. E. (1990). CmC. *English Today*, 6(3), str. 42–46.
- Nagao, M., in Mori, S. (1994). A new method of N-gram statistics for large number of n and automatic extraction of words and phrases from large text data of Japanese. V: *COLING '94 Proceedings of the 15th conference on Computational linguistics - Volume 1*, str. 611–615.
- Nemo, F. (2006). Discourse particles as morphemes and as constructions. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 375–402. Amsterdam: Elsevier.
- Nesi, H., and Basturkmen, H. (2006). Lexical bundles and discourse signalling in academic lectures. *International Journal of Corpus Linguistics*, 11(3), str. 283–304.
- O'Donnell, M. B. (2010). The adjusted frequency list: A method to produce cluster-sensitive frequency lists. *ICAME Journal*, 35, str. 135–170.
- O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies*, 1, str. 17.
- Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.
- Overstreet, M. (2000). *Whales, candlelight, and stuff like that: General extenders in English discourse*. New York: Oxford University Press.

- Overstreet, M. (2014). The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics*, 24(1), str. 105–129.
- Oza, U., Prasad, R., Kolachina, S., Sharma, D. M., in Joshi, A. (2009). The Hindi Discourse Relation Bank. V: *Proceedings of the Third Linguistic Annotation Workshop (LAW 2009)*, str. 158–161.
- Palmer, F. (2001). *Mood and modality*. Cambridge: Cambridge University Press.
- 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*, str. 191–226. London: Longman
- Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation*, 44(1–2), str. 137–158.
- Pekarek Doehler, S. (2016). Not Just an Epistemic Hedge: French je sais pas ‘I don’t know’ as a Resource for Sequentially Organizing Turns and Actions. *Journal of Pragmatics*, 106, str. 148–162.
- Peldszus, A. (2014). Towards segment-based recognition of argumentation structure in short texts. V: *Proceedings of the First Workshop on Argumentation Mining*, str. 88–97. Baltimore: Association for Computational Linguistics.
- Petrič, T. (1999). *Nemški naklonski členki in njihovi slovenski ustrezniki*. http://splet-stari.fnm.uni-mb.si/pedagoska/programi/nem/Germanistik_files/virtual/petric/pdf/partikeln/LK99.pdf
- Petrović, S., Šnajder, J., in Bašić, B. D. (2010). Extending lexical association measures for collocation extraction. *Computer Speech & Language*, 24(2), str. 383–394.
- Pisanski Peterlin, A. (2005). Text-organising metatext in research articles: an English–Slovene contrastive analysis. *English for Specific Purposes*, 24(3), str. 307–319.
- Pisanski Peterlin, A. (2007). Raziskave metabesedilnosti v uporabnem jezikoslovju. *Jezik in slovstvo*, 52(3/4), str. 7–19.
- Pisanski Peterlin, A. (2010). Hedging devices in Slovene-English translation: a corpus-based study. *Nordic Journal of English Studies*, 9(2), str. 171–193.
- Pisanski Peterlin, A. (2011). *Metabesedilo med dvema kulturama*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Pisanski Peterlin, A. (2014). Nagovor bralca v angleških, slovenskih in prevedenih poljudnoznanstvenih člankih: medkulturna retorična primerjava. V: A. Pisanski Peterlin in

- M. Schlamberger Brezar (ur.), *Prevodoslovno usmerjene kontrastivne študije*, str. 172–187. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Pisanski Peterlin, A. (2015a). Problematika veznika kot besedne vrste v enojezičnem slovarju. V: V. Gorjanc, P. Gantar, I. Kosem in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve*, str. 514–523. Ljubljana: Znanstvena založba Filozofske fakultete.
- Pisanski Peterlin, A. (2015b). Sentence-initial adversative connectives in Slovene-English translation of academic discourse: a corpus study. V: M. Schlamberger Brezar, D. Limon, in A. Gruntar Jermol (ur.), *Contrastive analysis in discourse studies and translation*, str. 68–82. Ljubljana: Znanstvena založba Filozofske fakultete.
- Pisanski Peterlin, A. (2015c). So prevedena poljudnoznanstvena besedila v slovenščini drugačna od izvirnih? Korpusna študija na primeru epistemske naklonskosti. *Slavistična revija*, 63(1), str. 29–42.
- Pitler, E., in Nenkova, A. (2009). Using syntax to disambiguate explicit discourse connectives in text. V: *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, str. 13–16.
- Pojanapunya, P., in Todd, R. W. (2016). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*, v tisku.
- Poláková, L., Mírovský, J., Nedoluzhko, A., Jínová, P., Zikánová, Š., in Hajičová, E. (2013). Introducing the Prague Discourse Treebank 1.0. V: *International Joint Conference on Natural Language Processing*, str. 91–99.
- Pollak, S. (2015a). Identifikacija spletno specifičnih kolokacij pogostega besedišča. V: *Zbornik konference Slovenščina na spletu in v novih medijih*, str. 57–62.
- Pollak, S. (2015b). Luščenje kolokacij iz korpusa uporabniških spletnih vsebin. V: M. Smolej (ur.), *Simpozij OBDOBJA 34: Slovničarstvo in slovar – aktualni jezikovni opis*, str. 601–607.
- Pollak, S., in Božinovski, B. (2014). Luščenje borzne terminologije. V: *Zbornik 9. konference Jezikovne tehnologije*, str. 114–119.
- Popescu-Belis, A., in Zufferey, S. (2011). Automatic identification of discourse markers in dialogues: An in-depth study of like and well. *Computer Speech and Language*, 25(3), str. 499–518.
- Prasad, R., in Bunt, H. (2015). Semantic relations in discourse: The current state of ISO 24617-8. V: *Proceedings of the 11th Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, str. 80–92.

- Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A., in Webber, B. (2008). The Penn Discourse TreeBank 2.0. V: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, str. 1–4.
- Prasad, R., Joshi, A., in Webber, B. (2010). Realization of discourse relations by other means: alternative lexicalizations. V: *COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics*, str. 1023–1031.
- Prasad, R., Miltsakaki, E., Dinesh, N., Lee, A., in Joshi, A. (2007). *The Penn Discourse Treebank 2.0 Annotation Manual*.
- Prasad, R., Webber, B., in Joshi, A. (2014). Reflections on the penn discourse treebank, comparable corpora, and complementary annotation. *Computational Linguistics*, 40(4), str. 921–950.
- Puc, K., in Erjavec, T. (2006). Uporaba korpusa pri urejanju spletnega terminološkega slovarja. V: *Jezikovne tehnologije: Zbornik 9. mednarodne multikonference Informacijska družba IS 2006*, str. 156–161.
- Ramisch, C., Villavicencio, A., in Boitet, C. (2010). mwetoolkit : a Framework for Multiword Expression Identification. V: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, str. 662–669.
- Rashid, R. A. (2016). Topic continuation strategies employed by teachers in managing supportive conversations on Facebook Timeline. *Discourse Studies*, 18(2), str. 188–203.
- Rayson, P. (2008a). From key words to key semantic domains. *International Journal of Corpus Linguistics*, 13(4), str. 519–549.
- Rayson, P. (2008b). *Log-likelihood and effect size calculator*.
<http://ucrel.lancs.ac.uk/llwizard.html>
- Rayson, P. (2013). Corpus Analysis of Key Words. V: *The Encyclopedia of Applied Linguistics*. Oxford: Wiley-Blackwell.
- Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics*, 14(3), str. 367–381.
- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), str. 1139–1172.
- Redeker, G. (2000). Coherence and structure in text and discourse. V: H. Bunt in W. Black (ur.), *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics*, str. 233–264. John Benjamins Publishing.
- Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 339–358. Amsterdam: Elsevier.

- Redeker, G., Berzlánovich, I., Vliet, N. Van Der, in Egg, M. (2012). Multi-Layer Discourse Annotation of a Dutch Text Corpus. V: *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, str. 2820–2825.
- Redeker, G., in Gruber, H. (2014). Introduction: The pragmatics of discourse coherence. V: H. Gruber in G. Redeker (ur.), *The pragmatics of discourse coherence: Theories and applications*, str. 1–20. John Benjamins.
- Reese, B., Hunter, J., Asher, N., Denis, P., in Baldridge, J. (2007). *Reference Manual for the Analysis and Annotation of Rhetorical Structure (Version 1.0)*. Technical report. Austin: University of Texas, Departments of Linguistics and Philosophy.
- Rehbein, I. (2015). Filled pauses in User-generated Content are Words with Extra-propositional Meaning, V: *Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics (ExProM 2015)*, str. 12–21.
- Rehbein, I., in Ruppenhofer, J. (2013). *Investigating orality in speech, writing, and in between*. Corpus Linguistics 2013 conference.
- Roberts, C., in Street, B. (1998). Spoken and Written Language. V: *The Handbook of Sociolinguistics*, str. 187–204. Oxford, UK: Blackwell Publishing.
- Romero-Trillo, J. (2013). Pragmatic Markers. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- Rossari, C. (2006). Formal properties of a subset of discourse markers: connectives. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 299–314. Amsterdam: Elsevier.
- Rouchota, V. (1996). Discourse Markers: What do they Link. *UCL Working Papers in Linguistics*, 8, str. 199–214.
- Roulet, E. (2006). The description of text relation markers in the Geneva model of discourse organization. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 115–132. Amsterdam: Elsevier.
- Roulet, E., Auchlin, A., Moeschler, J., Schelling, M., in Rubattel, J. (1985). *L'articulation du discours en français contemporain*. Bern: Lang.
- Roze, C., Danlos, L., in Muller, P. (2012). LEXCONN: a French lexicon of discourse connectives. *Discours*, 10. <https://discours.revues.org/8645>
- Rychlý, P. (2008). A lexicographer-friendly association score. V: *Proceedings of Recent Advances in Slavonic Natural Language Processing (RASLAN 2008)*, str. 6–9.

- Rysová, M., in Rysová, K. (2014). The Centre and Periphery of Discourse Connectives. V: *Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing (PACLIC28)*, str. 452–459.
- Rysová, M., in Rysová, K. (2015). Secondary Connectives in the Prague Dependency Treebank. V: *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, str. 291–299.
- Sag, I., Baldwin, T., in Bond, F. (2002). Multiword expressions: A pain in the neck for NLP. V: *CICLing '02 Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing*, str. 1–15.
- Saint-Georges, I. De. (2013). Context in the Analysis of Discourse and Interaction. V: *The Encyclopedia of Applied Linguistics*. Oxford, UK: Blackwell Publishing.
- Sanders, T., Demberg, V., Evers-Vermeul, J., Hoek, J., Scholman, M., in Zufferey, S. (2016). Unifying dimensions in discourse relations. How various annotation frameworks are related. *Corpus Linguistics and Linguistic Theory* (v tisku).
- Sanders, T., in Noordman, L. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, 29(1), str. 37–60.
- Sanders, T., in Pander Maat, H. (2006). Cohesion and Coherence: Linguistic Approaches. V: *Encyclopedia of Language & Linguistics*, str. 591–595. Elsevier.
- Sanders, T., in Spooren, W. (2015). Causality and subjectivity in discourse: The meaning and use of causal connectives in spontaneous conversation, chat interactions and written text. *Linguistics*, 53(1), str. 53–92.
- Sanders, T., Spooren, W., in Noordman, L. (1992). Toward a taxonomy of coherence relations. *Discourse Processes*, 15(1), str. 1–35.
- Sanders, T., Spooren, W., in Noordman, L. (1993). Coherence relations in a cognitive theory of discourse representation. *Cognitive Linguistics*, 4(2), str. 93–134.
- Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O., in Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, 9(2), str. 191–217.
- Scheffler, T. (2014). A German Twitter Snapshot. V: *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, str. 2284–2289.
- Scheffler, T., Schülter, R., in Stede, M. (2016). *Discourse Structuring Devices on Twitter*. COST TextLink Second Action Conference, Budimpešta.

- Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis (Volume 1)*. Cambridge: Cambridge University Press.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schiffrin, D. (1994). *Approaches to discourse*. Blackwell Oxford.
- Schiffrin, D. (2001). Discourse markers: language, meaning, and context. V: D. Schiffrin, D. Tannen, in E. H. Hamilton (ur.), *The handbook of discourse analysis*, str. 54–75. Blackwell Oxford.
- Schiffrin, D. (2006). Discourse marker research and theory: revisiting and. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 315–338. Amsterdam: Elsevier.
- Schilder, F. (2002). Robust discourse parsing via discourse markers, topicality and position. *Natural Language Engineering*, 8(2-3), str. 235–255.
- Schlamberger Brezar, M. (2007). Vloga povezovalcev v govorjenem diskurzu. *Jezik in slovstvo*, 52(3-4), str. 21–32.
- Schlamberger Brezar, M. (2009). *Povezovalci v francoščini: od teoretičnih izhodišč do analize v diskurzu*. Ljubljana: Znanstvena založba FF UL.
- Schneider, A., in Stede, M. (2012). Ambiguity in German Connectives: A Corpus Study. V: *Proceedings of KONVENS 2012*, str. 254–258.
- Scholman, M., Evers-Vermeul, J., in Sanders, T. J. M. (2016). A step-wise approach to discourse annotation: Towards a reliable categorization of coherence relations. *Dialogue & Discourse*, 1(1), str. 1–28.
- Schoonjans, S. (2013). Modal particles: Problems in defining a category. V: L. Degand, B. Cornillie in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and description*, str. 133–162. John Benjamins.
- Schourup, L. (1999). Discourse markers. *Lingua*, 107(3), str. 227–265.
- Schwenter, S., in Traugott, E. (2000). Invoking scalarity: The development of in fact. *Journal of Historical Pragmatics*, 1(1), str. 7–25.
- Scott, M. (1997). PC analysis of key words—and key key words. *System*, 25(2), str. 233–245.
- Scott, M. (2010). Problems in investigating keyness, or clearing the undergrowth and marking out trails.... V: M. Bondi in M. Scott (ur.), *Keyness in Texts*, str. 43–57. John Benjamins Publishing.
- Scott, M. (2016). *WordSmith Tools version 7*. Stroud: Lexical Analysis Software.

- Selting, M. (2000). The construction of units in conversational talk. *Language in Society*, 29(4), str. 477–517.
- Shriberg, E. (1996). Disfluencies in SWITCHBOARD. V: *Proceedings International Conference on Spoken Language Processing*, str. 11–14.
- Siepmann, D. (2005). *Discourse Markers Across Languages: A Contrastive Study of Second-level Discourse Markers in Native and Non-native Text with Implications for General and Pedagogic Lexicography*. Psychology Press.
- Siepmann, D. (2006). Collocation, colligation and encoding dictionaries. Part II: Lexicographical aspects. *International Journal of Lexicography*, 19(1), str. 1–39.
- Silva, J. Da, Dias, G., Guilloiré, S., in Lopes, J. (1999). Using localmaxs algorithm for the extraction of contiguous and non-contiguous multiword lexical units. V: *Progress in Artificial Intelligence: 9th Portuguese Conference on Artificial Intelligence (EPIA '99)*, str. 113–132.
- Simmons-Mackie, N., in Damico, J. (1996). The contribution of discourse markers to communicative competence in aphasia. *American Journal of Speech-Language Pathology*, 5(1), str. 37–43.
- Simon-Vandenberg, A. (2000). The functions of I think in political discourse. *International Journal of Applied Linguistics*, 10(1), str. 41–63.
- Simpson-Vlach, R., in Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, 31(4), str. 487–512.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Sindoni, M. (2014). *Spoken and written discourse in online interactions: A multimodal approach*. New York, London: Routledge.
- Skubic, A. (1999). Ogled kohezijske vloge slovenskega členka. *Slavistična Revija*, 47(2), str. 211–238.
- Smolej, M. (2004a). Členki kot besedilni povezovalci. *Jezik in slovstvo*, 49(5), str. 45–57.
- Smolej, M. (2004b). Obvezni in neobvezni členki. *Slavistična revija*, 52(2), str. 141–155.
- Smolej, M. (2012). *Besedilne vrste v spontanem govoru*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Smolej, M. (2015a). Členek v sodobnem slovenskem jezikoslovju. V: M. Smolej (ur.), *Simpozij Obdobja 34: Slovnica in slovar – aktualni jezikovni opis (2. del)*, str. 671–678.

- Smolej, M. (2015b). Particle Clauses. *Slovenski Jezik - Slovene Linguistic Studies*, 10, str. 127–137.
- Somasundaran, S., in Wiebe, J. (2009). Recognizing stances in online debates. V: *Proceedings of the ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1*, str. 226–234.
- Sotillo, S. (2000). Discourse functions and syntactic complexity in synchronous and asynchronous communication. *Language Learning & Technology*, 4(1), str. 82–119.
- Sperber, D., in Wilson, D. (1986). *Relevance: communication and cognition*. Harvard University Press.
- Spooren, W., in Degand, L. (2010). Coding coherence relations: Reliability and validity. *Corpus Linguistics and Linguistic Theory*, 6(2), str. 241–266.
- Squartini, M. (2013). From TAM to discourse. V: L. Degand, B. Cornillie in P. Pietrandrea (ur.), *Discourse Markers and Modal Particles: Categorization and description*, str. 163–190.
- Stede, M. (2002). DiMLex: A Lexical Approach to Discourse Markers. V: A. Lenci in V. Di Tomaso (ur.): *Exploring the Lexicon - Theory and Computation*, str. 1–15. Alessandria: Edizioni dell'Orso, 2002.
- Stede, M. (2012). *Discourse Processing*. Morgan & Claypool Publishers.
- Stede, M. (2014). Resolving connective ambiguity :A prerequisite for discourse parsing. V: G. Redeker in H. Gruber (ur.), *The pragmatics of discourse coherence: Theories and applications*, str. 121–141. John Benjamins Publishing.
- Stede, M., in Neumann, A. (2014). Potsdam Commentary Corpus 2.0: Annotation for Discourse Research. V: *Proceedings of the Ninth Language Resources and Evaluation Conference (LREC 2014)*, str. 925–929.
- Stede, M., in Umbach, C. (1998). DiMLex: A lexicon of discourse markers for text generation and understanding. V: *Proceedings of the 17th international conference on Computational linguistics*, str. 1238–1242.
- Stepanov, E. A., in Riccardi, G. (2015). Sentiment Polarity Classification with Low-level Discourse-based Features. V: *Proceedings of the Second Italian Conference on Computational Linguistics CLiC-it 2015*, str. 269–273.
- Swan, M. (1992). *Practical English Usage*. Oxford University Press.
- Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press.

- Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics*, 30, str. 485–496.
- Škrjanec, I., Sobočan, A. M., in Pollak, S. (2015). Uporabniške spletne vsebine, spol in nasilje. V: *Sociologija med produkcijo vednosti in so-oblikovanjem družbe: zbornik povzetkov*, str. 31.
- Taboada, M. (2004). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*. John Benjamins Publishing.
- Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4), str. 567–592.
- Taboada, M., in Das, D. (2013). Annotation upon Annotation: Adding Signalling Information to a Corpus of Discourse Relations. *Dialogue & Discourse*, 3(2), str. 249–281.
- Taboada, M., in Mann, W. C. (2006a). Applications of Rhetorical Structure Theory. *Discourse Studies*, 8(4), str. 567–588.
- Taboada, M., in Mann, W. C. (2006b). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3), str. 423–459.
- Tadić, M., in Šojat, K. (2003). Finding Multiword Term Candidates in Croatian. V: *Proceedings of IESL2003 Workshop*, str. 102–107.
- Georgakopoulou, A. in Spilioti, T. (2015). *The Routledge Handbook of Language and Digital Communication*. London: Routledge.
- Thurlow, C., in Mroczek, K. (2011). *Digital Discourse: Language in the New Media*. Oxford University Press.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins Publishing.
- Tonelli, S., Riccardi, G., Prasad, R., in Joshi, A. A. (2010). Annotation of Discourse Relations for Conversational Spoken Dialogs. V: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, str. 2084–2090.
- Toporišič, J. (1974). Kratko oblikoslovje slovenskega knjižnega jezika. Slovenski Jezik, Literatura in Kultura.
- Toporišič, J. (1984). *Slovenska slovnica*. Maribor: Obzorja.
- Toporišič, J. (2004). *Slovenska slovnica (Četrta prenovljena in razširjena izdaja)*. Maribor: Založba Obzorja Maribor.

- Tottie, G. (2011). Uh and Um as sociolinguistic markers in British English. *International Journal of Corpus Linguistics*, 16(2), str. 173–197.
- Tree, J., in Schrock, J. (2002). Basic meanings of you know and I mean. *Journal of Pragmatics*, 34(6), str. 727–747.
- van Dijk, T. A. (1980). The Semantics and Pragmatics of Functional Coherence in Discourse. V: A. Ferrara (ur.), *Speech act theory: Ten years later. Special issue of Versus*, str. 49–65.
- Verdonik, D. (2006). Mhm, ja, no, dobro, glejte, eee ...: diskurzni označevalci v telefonskih pogovorih. *Jezik in slovstvo*, 51(2), str. 19–36.
- Verdonik, D. (2007). *Jezikovni elementi spontanosti v pogovoru: diskurzni označevalci in popravljanja*. Maribor: Slavistično društvo.
- Verdonik, D. (2008). Označevanje vrste diskurznihih označevalcev. V: *Zbornik Šeste Konference Jezikovne tehnologije*, str. 25–28.
- Verdonik, D. (2011). Teorija mentalnih modelov v raziskovanju pogovora: pot k socio-kognitivnemu modelu pogovora? *Jezik in Slovstvo*, 56(3-4), str. 93–110.
- Verdonik, D. (2013). Koncept konteksta v jezikoslovnih in diskurznihih teorijah. *Slavistična revija*, 61(4), str. 631–649.
- Verdonik, D. (2014). Vprašanja zapisovanja govora v govornem korpusu Gos. V: *Zbornik 9. konference Jezikovne tehnologije*, str. 151–156.
- Verdonik, D. (2015a). Govorjeni proti pisnemu ali katera leksika je ‘tipično govorjena’. V: V. Gorjanc, P. Gantar, I. Kosem, in S. Krek (ur.), *Slovar sodobne slovenščine: problemi in rešitve*, str. 392–405. Ljubljana: Znanstvena založba Filozofske fakultete.
- Verdonik, D. (2015b). Internal variety in the use of Slovene general extenders in different spoken discourse settings. *International Journal of Corpus Linguistics*, 20(4), str. 445–468.
- Verdonik, D. (2015c). Slovarski opis pragmatično-diskurzne razsežnosti leksikalne rabe. V: M. Smolej (ur.), *Simpozij OBDOBJA 34: Slovnica in slovar - aktualni jezikovni opis*, str. 793–799.
- Verdonik, D., in Kačič, Z. (2012). Označevalci odprte propozicije v govorjenem diskurzu. *Slavistična revija*, 60(1), str. 79–94.
- Verdonik, D., Kosem, I., Vitez, A. Z., Krek, S., in Stabej, M. (2013). Compilation, transcription and usage of a reference speech corpus: the case of the Slovene corpus GOS. *Language Resources and Evaluation*, 47(4), str. 1031–1048.

- Verdonik, D., in Maučec, M. S. (2016). A speech corpus as a source of lexical information. *International Journal of Lexicography* (v tisku).
- Verdonik, D., Rojc, M., in Stabej, M. (2007a). Annotating discourse markers in spontaneous speech corpora on an example for the Slovenian language. *Language Resources and Evaluation*, 41(2), str. 147–180.
- Verdonik, D., in Zwitter Vitez, A. (2011). *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Verdonik, D., Žgank, A., in Peterlin, A. P. (2007b). Diskurzni označevalci v dveh pogovornih žanrih. *Jezik in slovstvo*, 52(6), str. 19–32.
- Versley, Y., in Gastel, A. (2012). Linguistic Tests for Discourse Relations in the TüBa-D/Z Corpus of Written German. *Dialogue & Discourse*, 1(2), str. 1–24.
- Vičar, B. (2009). Zgodovinskorazvojni pregled pojmovanja vrinjenega stavka (parenteze) v slovenističnem jezikoslovju. *Jezikoslovni zapiski*, 15(1-2), str. 113–126.
- Vičar, B. (2011). *Parenteza v novinarskem in parlamentarnem diskurzu*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta.
- Vidovič-Muha, A. (2000). *Slovensko leksikalno pomenoslovje: govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.
- Vidovič-Muha, A. (2015). Propozicija v funkcijski strukturi stavčne povedi – vprašanje besednih vrst (poudarek na povedkovniku in členku). *Slavistična revija*, 63(4), str. 389–406.
- Vintar, S. (2010). Bilingual term recognition revisited: The bag-of-equivalents term alignment approach and its evaluation. *Terminology*, 16(2), str. 141–158.
- Vogel, J. (2004). Razvijanje razumevanja in vrednotenje govorjenega besedila. *Jezik in slovstvo*, 49(2), str. 37–56.
- Waltereit, R. (2006). The rise of discourse markers in Italian: a specific type of language change. V: K. Fischer (ur.), *Approaches to discourse particles*, str. 61–67. Amsterdam: Elsevier.
- Waltereit, R., in Detges, U. (2007). Different functions, different histories. Modal particles and discourse markers from a diachronic point of view. *Catalan Journal of Linguistics*, 6, str. 61–80.
- Webber, B., Egg, M., in Kordoni, V. (2012). Discourse structure and language technology. *Natural Language Engineering*, 18(4), str. 437–490.
- Webber, B., Stone, M., Joshi, A., in Knott, A. (2003). Anaphora and Discourse Structure. *Computational Linguistics*, 29(4), str. 545–587.

- Wei, N., in Li, J. (2013). A new computing method for extracting contiguous phraseological sequences from academic text corpora. *International Journal of Corpus Linguistics*, 18(4), str. 506–535.
- Wiechmann, D. (2008). On the computation of construction strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory*, 4(2), str. 253–290.
- Wilson, A. (2013). Embracing Bayes Factors for Key Item Analysis in Corpus Linguistics, V: M. Bieswanger in A. Koll-Stobbe (ur.), *New approaches to the study of linguistic variability*, str. 3–11. Frankfurt: Peter Lang.
- Wilson, D., in Sperber, D. (1993). Linguistic form and relevance. *Lingua*, 90(1-2), str. 1–25.
- Wolf, F., in Gibson, E. (2005). Representing Discourse Coherence: A Corpus-Based Study. *Computational Linguistics*, 31(2), str. 249–287.
- Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2013). Formulaic Language. *Language Teaching*, 46(3), str. 316–334.
- Yates, S. J. (1996). Oral and written linguistic aspects of computer conferencing. V: S. C. Herring (ur.), *Computer-Mediated Communication: Linguistic, social, and cross-cultural perspectives*, str. 29–46. John Benjamins Publishing.
- Yimam, S. M., Gurevych, I., Eckart de Castilho, R., in Biemann, C. (2013). WebAnno: A flexible, web-based and visually supported system for distributed annotations. V: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, str. 1–6.
- Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins Publishing.
- Zappavigna, M. (2012). *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web*. A&C Black.
- Zeevat, H. (2006). *A dynamic approach to discourse particles*. V: K. Fischer (ur.), *Approaches to Discourse Particles*, str. 133–148. Amsterdam: Elsevier.
- Zeyrek, D., Demirşahin, I., Sevdik-Çalli, A., Balaban, H. Ö., Yalçinkaya, İ., in Turan, Ü. D. (2010). The annotation scheme of the Turkish Discourse Bank and an evaluation of inconsistent annotations. V: *LAW IV '10 Proceedings of the Fourth Linguistic Annotation Workshop*, str. 282–289.

- Zhou, Y., in Xue, N. (2012). PDTB-style discourse annotation of Chinese text. V: *ACL '12 Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, str. 69–77.
- Zufferey, S., in Degand, L. (2013). Annotating the meaning of discourse connectives in multilingual corpora. *Corpus Linguistics and Linguistic Theory*, 10, str. 1–24.
- Zufferey, S., Degand, L., Popescu-Belis, A., in Sanders, T. (2012). Empirical validations of multilingual annotation schemes for discourse relations. V: *Proceedings of the 8th Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation*, str. 77–84.
- Zuljan Kumar, D. (2007). *Narečni diskurz: diskurzivna analiza briških pogovorov*. Ljubljana: Založba ZRC ZRC SAZU.
- Zwitter Vitez, A., in Fišer, D. (2015a). Elementi interakcije v govorjenih in spletnih besedilih. V: Zbornik konference *Slovenščina na spletu in v novih medijih*, str. 87–90.
- Zwitter Vitez, A., in Fišer, D. (2015b). Novi mediji in govorjena slovenščina: zaznavanje, kodifikacija, analiza. V: M. Smolej (ur.), *Simpozij OBDOBJA 34: Slovnica in slovar – aktualni jezikovni opis*, str. 881–890.
- Žele, A. (2014a). Medpropozicijskost z vidika družljivosti glagolov oz. povedkov. *Slavia Centralis*, 2, str. 5–17.
- Žele, A. (2014b). *Slovar slovenskih členkov*. Ljubljana: Založba ZRC.

Priloge

Priloga 1: Seznam večbesednih diskurznihih označevalcev v vzorčenem korpusu Gos

V tej prilogi navajamo seznam večbesednih diskurznihih označevalcev, prepoznanih v vzorčenem korpusu govorjene slovenščine, s podatkom o številu pojavitev v vzorčenem korpusu, pripisanih diskurznihih funkcijah po shemi MDMA (maksimalistična tipologija) in jedrni pomenski skupini (minimalistična tipologija). Nizi so razvrščeni po pogostosti pojavljanja v vzorčenem korpusu.

	označevalec	št. poj.	maksimalistične funkcije	minimalistične funkcije
1	<i>tako da</i>	49	posledica (P), sklep (R), opis (R), zapolnjevanje (S), zapiranje (S), nadaljevanje (S)	rezultat
2	<i>a ne</i>	33	preverjanje pozornosti (M), zapolnjevanje vrzeli (S)	dvogovor
3	<i>ne vem</i>	33	zapolnjevanje vrzeli (S), zapiranje (S), omilitev (M), nestrinjanje (M)	nedoločnost
4	<i>v bistvu</i>	28	poudarjanje (R), nasprotje (R), sprememba ubeseditve (R), zapolnjevanje vrzeli (S)	nasprotje
5	<i>a veš</i>	18	preverjanje pozornosti (M), zapolnjevanje vrzeli (S)	dvogovor
6	<i>se pravi</i>	17	sklep (R), opis (R), sprememba ubeseditve (R), nadaljevanje (S), sprememba teme (S), zapolnjevanje vrzeli (S)	rezultat
7	<i>tako kot</i>	14	opis (R)	specifikacija
8	<i>pa to</i>	13	elipsa (M), zapiranje (S)	nedoločnost
9	<i>glede na to da</i>	10	vzrok (P), motivacija (R)	razlog
10	<i>zato ker</i>	9	vzrok (P), motivacija (R)	razlog
11	<i>ali pa</i>	8	alternativa (P), dodajanje (S), zapiranje (S)	dodajanje
12	<i>v redu</i>	7	preverjanje pozornosti (M), strinjanje (M), zapiranje (S), odpiranje (S), menjava teme (S)	dvogovor
13	<i>zdaj pa</i>	7	časovno (P), dodajanje (S), menjava teme (S)	razvrščanje
14	<i>v glavnem</i>	6	sklep (R), nadaljevanje (S)	rezultat
15	<i>in sicer</i>	6	opis (R)	specifikacija
16	<i>ali kaj</i>	6	preverjanje pozornosti (M), omilitev (M), elipsa (M), zapiranje (S)	dvogovor
17	<i>ali kako</i>	6	preverjanje pozornosti (M), elipsa (M), omilitev (M)	dvogovor
18	<i>sicer pa</i>	5	nasprotje (R), dodajanje (S), nadaljevanje (S)	nasprotje
19	<i>pol pa</i>	5	čas (P), sklep (R), dodajanje (S)	razvrščanje
20	<i>potem pa</i>	5	čas (P), sklep (R), naštevovanje (S), dodajanje (S)	razvrščanje
21	<i>samo da</i>	4	posledica (P), izjema (P), nasprotje (R)	rezultat, nasprotje
22	<i>veš kaj</i>	4	preverjanje pozornosti (M), odpiranje (S)	dvogovor
23	<i>kot da</i>	4	opis (R)	specifikacija

24	<i>zato da</i>	4	posledica (P)	rezultat
25	<i>bi rekel</i>	4	zapolnjevanje vrzeli (S)	tvorjenje
26	<i>in tako naprej</i>	4	elipsa (M)	nedoločnost
27	<i>ja no</i>	3	nestrinjanje (M), odpiranje (S), sklep (R)	dvogovor
28	<i>no ja</i>	3	nestrinjanje (M), menjava teme (S)	dvogovor
29	<i>na primer</i>	3	opis (R)	specifikacija
30	<i>pa tako</i>	3	elipsa (M), zapolnjevanje vrzeli (S)	nedoločnost
31	<i>s tem</i>	3	sklep (R)	rezultat
32	<i>zaradi tega</i>	3	posledica (P)	rezultat
33	<i>veš kaj je</i>	2	preverjanje pozornosti (M), odpiranje (S)	dvogovor
34	<i>v nadaljevanju</i>	2	naštevane (S)	razvrščanje
35	<i>medtem ko</i>	2	čas (P), nasprotje (R)	razvrščanje, nasprotje
36	<i>saj veš</i>	2	ohranjanje dostojanstva (M), preverjanje pozornosti (M)	dvogovor
37	<i>v resnici</i>	2	poudarjanje (R), nasprotje (R)	nasprotje
38	<i>edino če</i>	2	izjema (P)	pogoj
39	<i>saj pravim</i>	2	opis (R)	tvorjenje
40	<i>bom rekel</i>	2	zapolnjevanje vrzeli (S)	tvorjenje
41	<i>na koncu</i>	2	čas (P), naštevane (S)	razvrščanje
42	<i>recimo temu</i>	2	omilitev (M)	tvorjenje
43	<i>pri tem</i>	2	opis (R), dodajanje (S)	dodajanje
44	<i>za začetek</i>	2	naštevane (S)	razvrščanje
45	<i>s tem da</i>	2	dodajanje (S), opis (R)	dodajanje, specifikacija
46	<i>zaradi tega ker</i>	2	vzrok (P), motivacija (R)	razlog
47	<i>za to da</i>	2	posledica (P)	rezultat
48	<i>tako ka</i>	2	sklep (R)	rezultat
49	<i>a vidite</i>	2	preverjanje pozornosti (M)	dvogovor
50	<i>sploh če</i>	2	relevantnost (R)	pogoj
51	<i>in pa</i>	2	dodajanje (S)	dodajanje
52	<i>kaj jaz vem</i>	2	omilitev (M), zapolnjevanje vrzeli (S)	nedoločnost
53	<i>zato pa</i>	2	sklep (R)	rezultat
54	<i>prav tako</i>	2	dodajanje (S)	dodajanje, nedoločnost
55	<i>predvsem pa</i>	2	dodajanje (S)	dodajanje
56	<i>pa vendarle</i>	2	nasprotje (R), menjava teme (S)	nasprotje
57	<i>ne vem kaj</i>	1	elipsa (R)	nedoločnost
58	<i>v končni fazi</i>	1	poudarjanje (R)	rezultat
59	<i>te pa</i>	1	sklep (R)	razvrščanje
60	<i>pa take fore</i>	1	elipsa (M)	nedoločnost
61	<i>vendar pa</i>	1	nasprotje (R)	nasprotje
62	<i>pa take stvari</i>	1	elipsa (M)	nedoločnost
63	<i>kot smo rekli</i>	1	dodajanje (S)	tvorjenje

64	<i>druga zadeva</i>	1	naštevande (S)	razvrščanje
65	<i>tudi če</i>	1	dopustnost (P)	nasprotje
66	<i>pa tako naprej</i>	1	elipsa (M)	nedoločnost
67	<i>v smislu</i>	1	opis (R)	specifikacija
68	<i>ali karkoli</i>	1	elipsa (M)	nedoločnost
69	<i>ne da</i>	1	opis (R)	specifikacija
70	<i>a la</i>	1	opis (R)	specifikacija
71	<i>znaš kaj</i>	1	preverjanje pozornosti (M), odpiranje (S)	dvogovor
72	<i>po drugi strani</i>	1	nasprotje (R)	nasprotje
73	<i>tako torej</i>	1	zapiranje (S)	dvogovor
74	<i>po svoje</i>	1	omilitev (R)	nedoločnost
75	<i>to pomeni</i>	1	opis (R)	rezultat
76	<i>ali ne</i>	1	preverjanje pozornosti (M), omilitev (M)	dvogovor
77	<i>bom rekla</i>	1	zapolnjevanje vrzeli (S)	tvorjenje
78	<i>poleg tega pa</i>	1	dodajanje (S)	dodajanje
79	<i>moram reči</i>	1	poudarjanje (R)	tvorjenje
80	<i>potem ko</i>	1	čas (P)	razvrščanje
81	<i>v tem primeru</i>	1	posledica (P)	rezultat
82	<i>glj tako</i>	1	dodajanje (S)	dodajanje, rezultat
83	<i>na ta način</i>	1	posledica (P), sklep (R)	rezultat
84	<i>ali nekaj takega</i>	1	elipsa (M)	nedoločnost
85	<i>zaradi tega da</i>	1	posledica (P)	rezultat
86	<i>a veš kaj</i>	1	preverjanje pozornosti (M), odpiranje (S)	dvogovor
87	<i>danes ko</i>	1	čas (P)	razvrščanje
88	<i>prej ko</i>	1	čas (P)	razvrščanje
89	<i>pa take</i>	1	elipsa (M)	nedoločnost
90	<i>in tako</i>	1	elipsa (M)	nedoločnost
91	<i>tako rekoč</i>	1	omilitev (M)	tvorjenje
92	<i>ali pa nekaj</i>	1	omilitev (M)	nedoločnost
93	<i>takrat ko</i>	1	čas (P)	razvrščanje
94	<i>in to</i>	1	opis (R)	nedoločnost, specifikacija
95	<i>tedaj pa</i>	1	sklep (R)	razvrščanje
96	<i>in vse skupaj</i>	1	elipsa (M)	nedoločnost
97	<i>to se pravi</i>	1	nadaljevanje (R)	rezultat
98	<i>saj ne vem</i>	1	omilitev (M)	nedoločnost
99	<i>ma ja</i>	1	strinjanje (M), odpiranje (S)	dvogovor
100	<i>bi človek rekel</i>	1	zapolnjevanje vrzeli (S)	tvorjenje
101	<i>v kolikor</i>	1	pogoj (P)	pogoj
102	<i>ali kaj je</i>	1	elipsa (M), omilitev (M)	dvogovor
103	<i>medtem ko pa</i>	1	nasprotje (R)	nasprotje

Priloge

104	<i>kaj je</i>	1	zapolnjevanje vrzeli (S)	dvogovor
105	<i>če dovolite</i>	1	odpiranje (S)	dvogovor
106	<i>kaj naj rečem</i>	1	odpiranje (S)	tvorjenje
107	<i>v smislu da</i>	1	opis (R)	specifikacija
108	<i>se pravi da</i>	1	sklep (R)	rezultat
109	<i>v vsakem primeru</i>	1	sklep (R)	rezultat
110	<i>sedaj pa</i>	1	čas (R)	razvrščanje
111	<i>če ne</i>	1	nasprotje (R)	rezultat
112	<i>kakorkoli že</i>	1	nadaljevanje (S)	rezultat
113	<i>namesto da</i>	1	nasprotje (R)	nasprotje
114	<i>kljub temu da</i>	1	dopustnost (P)	nasprotje
115	<i>ne pa da</i>	1	protivnost (P)	nasprotje
116	<i>sploh pa</i>	1	dodajanje (S)	dodajanje
117	<i>če ne drugega</i>	1	poudarjanje (R)	dodajanje
118	<i>še posebej če</i>	1	pogoj (P)	pogoj
119	<i>nekaj takega</i>	1	elipsa (M)	nedoločnost
120	<i>tako ali tako</i>	1	poudarjanje (R)	rezultat
121	<i>ob tem da</i>	1	dodajanje (S), opis (R)	dodajanje
122	<i>a ne da</i>	1	preverjanje pozornosti (M)	dvogovor
123	<i>že odkar</i>	1	čas (P)	razvrščanje
124	<i>kot rečeno</i>	1	dodajanje (S)	tvorjenje
125	<i>tako ko</i>	1	opis (R)	specifikacija
	SKUPAJ	444		

Priloga 2: Seznam 100 najpogostejših besednih nizov v korpusu Gos glede na običajno metodo štetja

V tej prilogi navajamo 100 najpogostejših besednih nizov dolžine od 2 do 6 standardiziranih besednih pojavnic z minimalno pogostostjo 5 pojavitev na milijon pojavnic in pojavljanjem pri vsaj dveh različnih govornicah, razvrščenih glede na običajno pogostost. Nizi so razvrščeni po (običajni) pogostosti.

	Niz	Običajna pogostost v Gos	Število besed	Število različnih govorcev
1	<i>to je</i>	3853	2	642
2	<i>ja ja</i>	3420	2	392
3	<i>ne vem</i>	2743	2	504
4	<i>da je</i>	2697	2	640
5	<i>je pa</i>	2014	2	498
6	<i>je to</i>	1867	2	507
7	<i>a ne</i>	1836	2	289
8	<i>tako da</i>	1804	2	463
9	<i>je bilo</i>	1681	2	431
10	<i>da se</i>	1539	2	509
11	<i>pa je</i>	1514	2	454
12	<i>da bi</i>	1438	2	470
13	<i>se je</i>	1359	2	418
14	<i>ali pa</i>	1358	2	422
15	<i>jaz sem</i>	1316	2	432
16	<i>je bil</i>	1176	2	352
17	<i>ne ne</i>	1170	2	354
18	<i>ki je</i>	1069	2	377
19	<i>mislim da</i>	1038	2	379
20	<i>je bila</i>	1031	2	347
21	<i>kaj pa</i>	1016	2	345
22	<i>ja ja ja</i>	999	3	176
23	<i>v bistvu</i>	985	2	293
24	<i>kaj je</i>	938	2	333
25	<i>pa ne</i>	929	2	381
26	<i>pa to</i>	917	2	323
27	<i>a veš</i>	839	2	146
28	<i>v redu</i>	833	2	292
29	<i>se pravi</i>	832	2	214
30	<i>ker je</i>	777	2	345
31	<i>pa še</i>	775	2	320
32	<i>mi je</i>	753	2	270

Priloge

33	<i>je v</i>	744	2	361
34	<i>ja saj</i>	729	2	278
35	<i>ne bi</i>	728	2	359
36	<i>če bi</i>	726	2	338
37	<i>pol pa</i>	721	2	232
38	<i>da ne</i>	709	2	345
39	<i>zdaj pa</i>	693	2	248
40	<i>da so</i>	679	2	331
41	<i>se mi</i>	671	2	289
42	<i>ko je</i>	665	2	302
43	<i>to pa</i>	646	2	288
44	<i>ki so</i>	645	2	287
45	<i>in to</i>	640	2	299
46	<i>v tem</i>	626	2	253
47	<i>ne da</i>	623	2	272
48	<i>ne bo</i>	614	2	296
49	<i>pa tudi</i>	604	2	289
50	<i>s tem</i>	597	2	305
51	<i>je že</i>	587	2	306
52	<i>ja to</i>	584	2	293
53	<i>da bo</i>	580	2	283
54	<i>kako je</i>	571	2	265
55	<i>je tudi</i>	567	2	290
56	<i>če je</i>	565	2	291
57	<i>ja pa</i>	552	2	254
58	<i>in je</i>	551	2	243
59	<i>to da</i>	550	2	282
60	<i>bi bilo</i>	545	2	276
61	<i>to se</i>	520	2	257
62	<i>je tako</i>	513	2	256
63	<i>pa se</i>	511	2	285
64	<i>to so</i>	511	2	238
65	<i>zato ker</i>	508	2	260
66	<i>potem pa</i>	507	2	185
67	<i>ali kaj</i>	507	2	209
68	<i>tako je</i>	500	2	216
69	<i>pa tako</i>	499	2	221
70	<i>da da</i>	493	2	201
71	<i>bi se</i>	493	2	275
72	<i>to ne</i>	490	2	247
73	<i>se pa</i>	489	2	277
74	<i>tako kot</i>	484	2	267
75	<i>in tako</i>	468	2	200

Priloge

76	<i>in pol</i>	459	2	159
77	<i>če se</i>	458	2	267
78	<i>na to</i>	456	2	238
79	<i>se ne</i>	454	2	276
80	<i>sem se</i>	448	2	206
81	<i>pa v</i>	447	2	259
82	<i>ni bilo</i>	442	2	227
83	<i>dobro jutro</i>	440	2	105
84	<i>ja ne</i>	439	2	222
85	<i>ne to</i>	434	2	237
86	<i>in da</i>	432	2	223
87	<i>ne pa</i>	431	2	221
88	<i>pa da</i>	429	2	252
89	<i>ne in</i>	423	2	193
90	<i>da si</i>	418	2	222
91	<i>ko sem</i>	413	2	200
92	<i>veš kaj</i>	409	2	164
93	<i>je ta</i>	402	2	222
94	<i>se bo</i>	399	2	208
95	<i>sem pa</i>	394	2	189
96	<i>pa kaj</i>	392	2	214
97	<i>je še</i>	391	2	208
98	<i>in potem</i>	391	2	164
99	<i>je zdaj</i>	389	2	219
100	<i>a je</i>	384	2	178

Priloga 3: Seznam 100 najpogostejših besednih nizov v korpusu Gos glede na prilagojeno metodo štetja

V tej prilogi navajamo 100 najpogostejših besednih nizov dolžine od 2 do 6 standardiziranih besednih pojavnic z minimalno pogostostjo 5 pojavitev na milijon pojavnic in pojavljanjem pri vsaj dveh različnih govornicah, razvrščenih glede na prilagojeno pogostost in s pripisanim podatkom o spremembi razvrstitve glede na razvrščanje z običajnim načinom štetja (Priloga 2). Nizi so razvrščeni po prilagojeni pogostosti.

	Niz	Običajna pogostost v Gos	Prilagojena pogostost v Gos	Število besed	Število različnih govorcev	Sprememba uvrstitve glede na obič. pog.
1	<i>ja ja</i>	3420	1727	2	392	↑2
2	<i>a ne</i>	1836	1118	2	289	↑7
3	<i>to je</i>	3853	810	2	642	↓1
4	<i>ali pa</i>	1358	659	2	422	↑14
5	<i>tako da</i>	1804	617	2	463	↑8
6	<i>ne vem</i>	2743	599	2	504	↓3
7	<i>ja ja ja</i>	999	547	3	176	↑22
8	<i>ki je</i>	1069	535	2	377	↑18
9	<i>da je</i>	2697	502	2	640	↓4
10	<i>ne ne</i>	1170	497	2	354	↑17
11	<i>a veš</i>	839	430	2	146	↑27
12	<i>je pa</i>	2014	429	2	498	↓5
13	<i>da se</i>	1539	424	2	509	↓10
14	<i>se pravi</i>	832	386	2	214	↑29
15	<i>pa je</i>	1514	375	2	454	↓11
16	<i>je bil</i>	1176	361	2	352	== 16
17	<i>ali kaj</i>	507	359	2	209	↑66
18	<i>se je</i>	1359	347	2	418	↓13
19	<i>dobro jutro</i>	440	347	2	105	↑83
20	<i>kaj pa</i>	1016	341	2	345	↑21
21	<i>zdaj pa</i>	693	327	2	248	↑39
22	<i>pa to</i>	917	326	2	323	↑26
23	<i>da bi</i>	1438	320	2	470	↓12

Priloge

24	<i>ker je</i>	777	316	2	345	↑30
25	<i>je bila</i>	1031	308	2	347	↓20
26	<i>pa še</i>	775	304	2	320	↑31
27	<i>pol pa</i>	721	297	2	232	↑37
28	<i>kaj je</i>	938	295	2	333	↓24
29	<i>mhm mhm</i>	383	295	2	99	↑101
30	<i>ki so</i>	645	292	2	287	↑44
31	<i>v redu</i>	833	291	2	292	↓28
32	<i>ko je</i>	665	287	2	302	↑42
33	<i>jaz sem</i>	1316	283	2	432	↓15
34	<i>tako je</i>	500	283	2	216	↑68
35	<i>je to</i>	1867	282	2	507	↓6
36	<i>pa tudi</i>	604	282	2	289	↑49
37	<i>ja saj</i>	729	281	2	278	↓34
38	<i>in je</i>	551	272	2	243	↑58
39	<i>da bo</i>	580	268	2	283	↑53
40	<i>v bistvu</i>	985	267	2	293	↓23
41	<i>je bilo</i>	1681	259	2	431	↓9
42	<i>pa ne</i>	929	251	2	381	↓25
43	<i>potem pa</i>	507	249	2	185	↑67
44	<i>če je</i>	565	244	2	291	↑56
45	<i>je tudi</i>	567	234	2	290	↑55
46	<i>pa tako</i>	499	233	2	221	↑69
47	<i>da so</i>	679	232	2	331	↓40
48	<i>ja pa</i>	552	231	2	254	↑57
49	<i>tudi v</i>	329	227	2	182	↑127
50	<i>s tem</i>	597	224	2	305	== 50
51	<i>ne pa</i>	431	219	2	221	↑87
52	<i>če bi</i>	726	216	2	338	↓36
53	<i>ne bo</i>	614	216	2	296	↓48
54	<i>zato ker</i>	508	211	2	260	↑65
55	<i>v v</i>	261	210	2	145	↑176
56	<i>pa se</i>	511	209	2	285	↑63

Priloge

57	<i>še enkrat</i>	294	207	2	152	↑150
58	<i>in to</i>	640	204	2	299	↓45
59	<i>to so</i>	511	200	2	238	↑64
60	<i>ne in</i>	423	198	2	193	↑89
61	<i>je v</i>	744	196	2	361	↓33
62	<i>da da</i>	493	196	2	201	↑70
63	<i>in da</i>	432	193	2	223	↑86
64	<i>ja ne</i>	439	192	2	222	↑84
65	<i>ja ja ja ja</i>	340	192	4	80	↑120
66	<i>in tudi</i>	262	192	2	158	↑174
67	<i>in se</i>	336	191	2	203	↑124
68	<i>in pa</i>	247	190	2	99	↑190
69	<i>je že</i>	587	189	2	306	↓51
70	<i>tako kot</i>	484	188	2	267	↑74
71	<i>pa v</i>	447	188	2	259	↑81
72	<i>in potem</i>	391	185	2	164	↑97
73	<i>za to</i>	356	184	2	228	↑111
74	<i>ne da</i>	623	183	2	272	↓47
75	<i>so se</i>	365	183	2	202	↑108
76	<i>ni bilo</i>	442	181	2	227	↑82
77	<i>so pa</i>	353	181	2	186	↑115
78	<i>a res</i>	200	179	2	99	↑257
79	<i>da si</i>	418	178	2	222	↑90
80	<i>in pol</i>	459	177	2	159	↓76
81	<i>mi je</i>	753	176	2	270	↓32
82	<i>v tem</i>	626	176	2	253	↓46
83	<i>ki ga</i>	269	176	2	158	↑171
84	<i>ki se</i>	347	175	2	199	↑118
85	<i>ne ne ne</i>	298	172	3	134	↑146
86	<i>ja no</i>	328	171	2	164	↑128
87	<i>ne samo</i>	273	171	2	189	↑167
88	<i>mislim da</i>	1038	170	2	379	↓19
89	<i>na primer</i>	270	169	2	113	↑170

Priloge

90	<i>na na</i>	238	169	2	128	↑203
91	<i>v glavnem</i>	279	168	2	127	↑164
92	<i>je tako</i>	513	167	2	256	↓62
93	<i>pa na</i>	286	167	2	194	↑157
94	<i>če se</i>	458	165	2	267	↓77
95	<i>in tako naprej</i>	235	165	3	97	↑206
96	<i>o tem</i>	338	164	2	178	↑121
97	<i>ali ne</i>	236	164	2	143	↑205
98	<i>aha aha</i>	220	164	2	60	↑227
99	<i>so bili</i>	379	162	2	188	↑105
100	<i>hvala lepa</i>	225	162	2	90	↑223

Priloga 4: Seznam statistično prepoznanih večbesednih diskurznihih označevalcev v korpusu Gos

V tej prilogi navajamo seznam večbesednih diskurznihih označevalcev, prepoznanih na podlagi pregleda prvih 2.000 statistično najbolj izstopajočih besednih nizov korpusa Gos glede na osem različnih načinov razvrščanja (), s podatkom o posamezni statistični vrednosti in pripadajoči razvrstitvi. Nizi so razvrščeni po abecednem redu.

	niz	funkcijska skupina	OP	PP	MI	MI3	χ^2	LL	Dice	t-score
1	<i>a ne</i>	dvogovor	1836 (7.)	1118 (2.)	3,0 (10143.)	24,7 (12.)	0,012 (970.)	5161,1 (9.)	0,095 (801.)	37,8 (5.)
2	<i>a ne da</i>	dvogovor	86 (739.)	59 (667.)	2,0 (14035.)	14,8 (2489.)	0,000 (6584.)	28,6 (10016.)	0,011 (3779.)	7,0 (599.)
3	<i>a veste</i>	dvogovor	52 (1378.)	47 (968.)	4,4 (6413.)	15,8 (1852.)	0,001 (2982.)	230,6 (1325.)	0,014 (3121.)	6,8 (637.)
4	<i>a veš</i>	dvogovor	839 (27.)	430 (11.)	5,4 (4576.)	24,8 (11.)	0,034 (663.)	5023,9 (10.)	0,172 (541.)	28,2 (18.)
5	<i>a veš kaj</i>	dvogovor	33 (2415.)	22 (2962.)	2,6 (11770.)	12,7 (4774.)	0,000 (7673.)	8,6 (16697.)	0,007 (5265.)	4,8 (1655.)
6	<i>ali kaj</i>	dvogovor	507 (66.)	359 (17.)	3,4 (8906.)	21,4 (231.)	0,004 (1464.)	1596,3 (52.)	0,071 (976.)	20,5 (37.)
7	<i>ali kaj je</i>	dvogovor	40 (1894.)	27 (2241.)	1,7 (15250.)	12,4 (5241.)	7,013 (10228.)	2,4 (21508.)	0,003 (9521.)	4,4 (1968.)
8	<i>ali kako</i>	dvogovor	111 (532.)	81 (395.)	2,8 (10922.)	16,4 (1568.)	0,000 (3986.)	254,1 (1132.)	0,027 (1874.)	9,0 (326.)
9	<i>ali ne</i>	dvogovor	236 (205.)	164 (97.)	0,6 (20449.)	16,4 (1591.)	5,003 (11545.)	43,0 (7568.)	0,012 (3426.)	5,5 (1141.)
10	<i>ali nekaj</i>	nedoločnost	69 (972.)	37 (1357.)	2,5 (12149.)	14,7 (2614.)	0,000 (5824.)	128,6 (2647.)	0,018 (2558.)	6,8 (648.)
11	<i>ali nekaj takega</i>	nedoločnost	32 (2482.)	32 (1766.)	6,9 (2563.)	16,9 (1338.)	0,003 (1572.)	205,9 (1549.)	0,024 (2093.)	5,6 (1114.)
12	<i>ali pa</i>	dodajanje	1358 (14.)	659 (4.)	3,2 (9535.)	24,0 (28.)	0,010 (1033.)	4143,8 (16.)	0,079 (899.)	33,0 (10.)
13	<i>ali pa kaj</i>	nedoločnost	111 (531.)	51 (825.)	3,9 (7694.)	17,5 (1154.)	0,001 (2505.)	255,4 (1126.)	0,027 (1879.)	9,8 (279.)
14	<i>ali pa kaj podobnega</i>	nedoločnost	9 (11039.)	9 (9565.)	8,4 (1582.)	14,7 (2610.)	0,003 (1772.)	77,8 (4496.)	0,008 (4785.)	2,9 (5470.)
15	<i>ali pa kaj takega</i>	nedoločnost	41 (1806.)	41 (1191.)	8,2 (1631.)	18,9 (729.)	0,012 (974.)	323,7 (758.)	0,036 (1545.)	6,3 (784.)
16	<i>alla fine</i>	razvrščanje	6 (17422.)	6 (15681.)	16,4 (47.)	21,6 (201.)	0,545 (75.)	137,7 (2448.)	0,705 (81.)	2,4 (9029.)

Priloge

17	<i>bi človek rekel</i>	tvorjenje	12 (7972.)	12 (6724.)	7,6 (1986.)	14,8 (2521.)	0,002 (1954.)	17,6 (12889.)	0,005 (6907.)	3,4 (3883.)
18	<i>bi rekel</i>	tvorjenje	190 (285.)	109 (223.)	4,1 (7034.)	19,3 (647.)	0,003 (1752.)	774,5 (150.)	0,043 (1376.)	13,0 (143.)
19	<i>bi rekla</i>	tvorjenje	46 (1586.)	20 (3357.)	2,5 (12052.)	13,5 (3686.)	0,000 (6868.)	87,2 (4012.)	0,010 (3975.)	5,6 (1109.)
20	<i>bi rekli</i>	tvorjenje	42 (1759.)	23 (2798.)	3,4 (8883.)	14,2 (3002.)	0,000 (4908.)	129,4 (2624.)	0,010 (4032.)	5,9 (975.)
21	<i>bolj tako</i>	nedoločnost	46 (1594.)	28 (2115.)	1,4 (16848.)	12,4 (5153.)	4,954 (11587.)	33,8 (9032.)	0,007 (5217.)	4,2 (2251.)
22	<i>bom rekel</i>	tvorjenje	97 (637.)	68 (540.)	5,2 (4813.)	18,4 (863.)	0,003 (1649.)	532,1 (292.)	0,060 (1061.)	9,5 (291.)
23	<i>bom rekla</i>	tvorjenje	31 (2597.)	31 (1865.)	4,0 (7410.)	13,9 (3346.)	0,000 (4622.)	115,6 (2974.)	0,021 (2270.)	5,2 (1319.)
24	<i>bomo rekli</i>	tvorjenje	36 (2152.)	36 (1449.)	5,1 (5037.)	15,5 (2052.)	0,001 (2772.)	190,3 (1713.)	0,028 (1831.)	5,8 (1000.)
25	<i>čakaj malo</i>	dvogovor	46 (1579.)	46 (1004.)	5,2 (4915.)	16,2 (1651.)	0,001 (2390.)	248,2 (1173.)	0,029 (1789.)	6,6 (717.)
26	<i>če ne</i>	rezultat	279 (166.)	104 (244.)	0,5 (21053.)	16,7 (1438.)	3,650 (12849.)	32,4 (9264.)	0,014 (3075.)	4,9 (1511.)
27	<i>če že</i>	dodajanje, pogoj	52 (1397.)	42 (1134.)	0,9 (19213.)	12,3 (5445.)	2,168 (14988.)	17,1 (13047.)	0,009 (4286.)	3,3 (4065.)
28	<i>čisto tako</i>	nedoločnost	41 (1832.)	41 (1195.)	1,9 (14329.)	12,6 (4809.)	9,015 (9292.)	51,8 (6516.)	0,007 (5448.)	4,7 (1700.)
29	<i>druga zadeva</i>	razvrščanje	6 (17820.)	6 (15970.)	7,2 (2354.)	12,4 (5284.)	0,000 (3274.)	47,6 (6989.)	0,028 (1825.)	2,4 (9506.)
30	<i>drugače pa</i>	nasprotje	167 (328.)	134 (143.)	3,0 (10283.)	17,8 (1049.)	0,001 (2921.)	448,7 (400.)	0,011 (3884.)	11,3 (189.)
31	<i>glede na to</i>	rezultat	178 (304.)	20 (3346.)	5,9 (3781.)	20,8 (329.)	0,010 (1052.)	1139,6 (83.)	0,036 (1556.)	13,1 (141.)
32	<i>glede na to da</i>	razlog	122 (483.)	37 (1353.)	6,5 (3044.)	20,4 (418.)	0,011 (1019.)	792,9 (143.)	0,032 (1664.)	10,9 (212.)
33	<i>glede tega</i>	dodajanje	25 (3404.)	20 (3414.)	4,8 (5719.)	14,0 (3183.)	0,000 (3865.)	119,2 (2868.)	0,020 (2360.)	4,8 (1648.)
34	<i>glih tako</i>	dodajanje, rezultat	59 (1189.)	48 (931.)	3,3 (9180.)	15,1 (2298.)	0,000 (4335.)	176,4 (1877.)	0,010 (3990.)	6,9 (617.)
35	<i>in pa</i>	dodajanje	247 (190.)	190 (68.)	-0,9 (25707.)	14,9 (2446.)	0,000 (8474.)	136,6 (2477.)	0,010 (3969.)	-14,4 (27130.)
36	<i>in podobno</i>	nedoločnost	26 (3257.)	26 (2421.)	3,8 (7863.)	13,2 (4047.)	0,000 (5311.)	95,9 (3636.)	0,003 (10657.)	4,7 (1725.)
37	<i>in sicer</i>	specifikacija	191 (283.)	149 (122.)	4,0 (7365.)	19,2 (682.)	0,002 (1834.)	765,7 (153.)	0,022 (2200.)	12,9 (145.)
38	<i>in tako</i>	nedoločnost	468 (75.)	120 (181.)	1,4 (16621.)	19,2 (680.)	0,000 (4195.)	373,4 (601.)	0,035 (1593.)	13,8 (120.)

Priloge

39	<i>in tako dalje</i>	nedoločnost	30 (2675.)	30 (1933.)	6,4 (3122.)	16,3 (1640.)	0,002 (1885.)	184,9 (1778.)	0,007 (5489.)	5,4 (1218.)
40	<i>in tako naprej</i>	nedoločnost	235 (206.)	165 (95.)	6,5 (3042.)	22,3 (120.)	0,021 (775.)	1457,9 (62.)	0,052 (1191.)	15,1 (89.)
41	<i>in to</i>	nedoločnost, specifikacija	640 (45.)	204 (58.)	1,1 (18327.)	19,7 (554.)	0,000 (4826.)	309,3 (809.)	0,036 (1532.)	13,5 (133.)
42	<i>in vse</i>	nedoločnost	72 (931.)	57 (699.)	0,6 (20316.)	13,0 (4372.)	1,638 (16105.)	13,8 (14230.)	0,007 (5241.)	3,1 (4713.)
43	<i>ja no</i>	dvogovor	328 (128.)	171 (86.)	1,4 (16694.)	18,1 (948.)	0,000 (4995.)	258,1 (1110.)	0,021 (2260.)	11,5 (184.)
44	<i>jaz ne vem</i>	nedoločnost	116 (513.)	54 (748.)	3,6 (8471.)	17,3 (1217.)	0,001 (2780.)	349,7 (653.)	0,035 (1585.)	9,8 (274.)
45	<i>kaj jaz vem</i>	nedoločnost	62 (1106.)	44 (1059.)	6,0 (3601.)	17,9 (1003.)	0,004 (1563.)	136,3 (2484.)	0,018 (2584.)	7,7 (472.)
46	<i>kaj je</i>	dvogovor	938 (24.)	295 (29.)	1,4 (16927.)	21,1 (286.)	0,001 (3052.)	713,3 (166.)	0,040 (1453.)	19,1 (46.)
47	<i>kaj je že</i>	dvogovor	54 (1326.)	47 (967.)	3,4 (8930.)	14,9 (2416.)	0,000 (4387.)	72,1 (4865.)	0,013 (3224.)	6,6 (694.)
48	<i>kaj pa jaz vem</i>	nedoločnost	22 (3950.)	16 (4597.)	7,1 (2455.)	16,0 (1782.)	0,002 (1783.)	35,9 (8660.)	0,009 (4486.)	4,6 (1804.)
49	<i>kaj sem hotela reči</i>	tvorjenje	5 (21815.)	5 (19829.)	7,9 (1804.)	12,5 (4955.)	0,001 (2765.)	15,6 (13589.)	0,003 (11152.)	2,2 (11971.)
50	<i>kaj takega</i>	nedoločnost	119 (502.)	53 (772.)	5,2 (4812.)	19,0 (716.)	0,004 (1498.)	688,2 (180.)	0,024 (2089.)	10,6 (230.)
51	<i>kaj že</i>	dvogovor	42 (1801.)	32 (1734.)	-0,0 (23151.)	10,7 (8379.)	3,121 (27457.)	0,0 (27411.)	0,006 (6294.)	-0,0 (23154.)
52	<i>kako bi rekel</i>	tvorjenje	25 (3392.)	25 (2545.)	5,9 (3733.)	15,2 (2237.)	0,001 (2480.)	100,2 (3469.)	0,020 (2387.)	4,9 (1559.)
53	<i>kako naj rečem</i>	tvorjenje	9 (11054.)	9 (9575.)	7,6 (1977.)	14,0 (3243.)	0,001 (2227.)	53,0 (6404.)	0,010 (4073.)	2,9 (5505.)
54	<i>kako se reče</i>	tvorjenje	23 (3766.)	23 (2845.)	7,2 (2353.)	16,2 (1647.)	0,003 (1679.)	155,7 (2142.)	0,023 (2113.)	4,7 (1712.)
55	<i>kako se že reče</i>	tvorjenje	6 (17653.)	6 (15835.)	9,1 (1282.)	14,3 (2958.)	0,003 (1684.)	35,0 (8797.)	0,009 (4407.)	2,4 (9276.)
56	<i>kar se tega tiče</i>	dodajanje	10 (9725.)	10 (8332.)	9,4 (1186.)	16,1 (1722.)	0,007 (1240.)	78,0 (4486.)	0,014 (3148.)	3,1 (4802.)
57	<i>kljub temu</i>	nasprotje	58 (1209.)	12 (6597.)	10,7 (865.)	22,4 (107.)	0,097 (401.)	798,6 (141.)	0,259 (400.)	7,6 (499.)
58	<i>kljub temu da</i>	nasprotje	33 (2398.)	19 (3632.)	5,7 (3998.)	15,8 (1848.)	0,001 (2271.)	195,6 (1650.)	0,006 (6012.)	5,6 (1097.)
59	<i>kljub vsemu</i>	nasprotje	16 (5670.)	16 (4641.)	13,1 (458.)	21,1 (289.)	0,143 (328.)	283,1 (962.)	0,273 (381.)	3,9 (2661.)
60	<i>konec koncev</i>	rezultat	44 (1664.)	38 (1326.)	11,5 (704.)	22,4 (110.)	0,127 (355.)	670,3 (188.)	0,268 (388.)	6,6 (711.)
61	<i>kot da</i>	specifikacija	147 (392.)	60 (644.)	1,0 (18412.)	15,4 (2062.)	8,875 (9353.)	67,6 (5159.)	0,012 (3603.)	6,4 (773.)

Priloge

62	<i>kot rečeno</i>	tvorjenje	32 (2483.)	32 (1767.)	6,9 (2584.)	16,9 (1351.)	0,003 (1585.)	260,3 (1091.)	0,018 (2600.)	5,6 (1115.)
63	<i>lahko rečem</i>	tvorjenje	26 (3247.)	14 (5477.)	5,0 (5242.)	14,4 (2857.)	0,000 (3429.)	134,5 (2524.)	0,012 (3605.)	4,9 (1540.)
64	<i>ma ja</i>	dvogovor	91 (689.)	69 (524.)	1,7 (15271.)	14,7 (2572.)	0,000 (7313.)	99,1 (3514.)	0,006 (5641.)	6,7 (674.)
65	<i>med drugim</i>	specifikacija	30 (2671.)	20 (3380.)	8,6 (1474.)	18,4 (870.)	0,011 (997.)	310,5 (806.)	0,065 (1017.)	5,4 (1191.)
66	<i>medtem ko</i>	razvrščanje, nasprotje	70 (953.)	47 (953.)	7,5 (2070.)	19,7 (542.)	0,012 (964.)	674,4 (185.)	0,031 (1699.)	8,3 (400.)
67	<i>moram priznati</i>	tvorjenje	20 (4376.)	11 (7374.)	9,9 (1052.)	18,5 (829.)	0,019 (815.)	248,0 (1174.)	0,069 (990.)	4,4 (1988.)
68	<i>moram reči</i>	tvorjenje	108 (553.)	23 (2783.)	9,1 (1291.)	22,6 (96.)	0,059 (518.)	1207,7 (78.)	0,24 (238.)	
69	<i>na koncu</i>	razvrščanje	227 (218.)	139 (133.)	6,0 (3622.)	21,7 (196.)	0,014 (915.)	1721,9 (48.)	0,036 (1538.)	14,8 (102.)
70	<i>na neki način</i>	nedoločnost	126 (482.)	107 (232.)	7,3 (2275.)	21,1 (281.)	0,019 (820.)	859,0 (127.)	0,038 (1500.)	10,9 (207.)
71	<i>na primer</i>	specifikacija	270 (170.)	169 (90.)	5,6 (4193.)	21,8 (184.)	0,013 (956.)	1796,2 (44.)	0,043 (1371.)	16,1 (77.)
72	<i>na ta način</i>	rezultat	83 (769.)	69 (526.)	7,2 (2361.)	19,9 (502.)	0,012 (984.)	518,2 (303.)	0,026 (1978.)	9,0 (331.)
73	<i>na tak način</i>	rezultat	34 (2308.)	22 (2958.)	7,2 (2320.)	17,4 (1176.)	0,005 (1414.)	153,3 (2174.)	0,010 (3980.)	5,7 (1031.)
74	<i>na vem</i>	nedoločnost	71 (946.)	53 (783.)	0,7 (20096.)	13,0 (4348.)	1,876 (15561.)	15,6 (13581.)	0,009 (4548.)	3,3 (4175.)
75	<i>na začetku</i>	razvrščanje	119 (501.)	100 (269.)	5,9 (3800.)	19,6 (569.)	0,006 (1258.)	856,3 (128.)	0,019 (2488.)	10,7 (219.)
76	<i>ne da</i>	specifikacija	623 (47.)	183 (74.)	-0,0 (23338.)	18,5 (848.)	8,270 (24493.)	0,8 (23918.)	0,023 (2109.)	-0,9 (24329.)
77	<i>ne glede na to</i>	rezultat	38 (2011.)	18 (3915.)	4,0 (7283.)	14,5 (2727.)	0,000 (4072.)	15,0 (13809.)	0,004 (8013.)	5,8 (1028.)
78	<i>ne pa da</i>	nasprotje	32 (2529.)	32 (1787.)	1,5 (16406.)	11,5 (6884.)	0,000 (12494.)	167,5 (1982.)	0,002 (13109.)	3,6 (3285.)
79	<i>ne vem</i>	nedoločnost	2743 (3.)	599 (6.)	4,6 (6124.)	27,4 (1.)	0,063 (496.)	15398,5 (1.)	0,156 (582.)	50,2 (2.)
80	<i>ne vem kaj</i>	nedoločnost	242 (198.)	75 (441.)	3,7 (8053.)	19,6 (582.)	0,002 (1797.)	204,7 (1561.)	0,021 (2244.)	14,4 (108.)
81	<i>ne znam</i>	nedoločnost	140 (417.)	90 (330.)	4,3 (6609.)	18,6 (816.)	0,002 (1859.)	699,4 (173.)	0,008 (4662.)	11,2 (197.)
82	<i>nekaj takega</i>	nedoločnost	101 (605.)	50 (853.)	6,9 (2597.)	20,2 (446.)	0,012 (983.)	811,2 (137.)	0,070 (977.)	9,9 (270.)
83	<i>no ja</i>	dvogovor	98 (635.)	73 (465.)	-0,2 (24078.)	12,9 (4466.)	3,814 (21243.)	4,1 (19826.)	0,006 (5900.)	-2,1 (25242.)
84	<i>ob koncu</i>	razvrščanje	16 (5715.)	16 (4663.)	5,9 (3771.)	13,9 (3360.)	0,000 (3174.)	100,2 (3466.)	0,026 (1964.)	3,9 (2781.)

Priloge

85	<i>ob tem</i>	dodajanje	53 (1350.)	37 (1370.)	4,7 (5860.)	16,1 (1691.)	0,001 (2676.)	249,2 (1169.)	0,034 (1623.)	7,0 (604.)
86	<i>od takrat naprej</i>	razvrščanje	8 (12597.)	8 (11045.)	8,7 (1455.)	14,7 (2640.)	0,003 (1702.)	26,0 (10589.)	0,008 (4581.)	2,8 (6338.)
87	<i>pa ne vem kaj</i>	nedoločnost	38 (2014.)	21 (3147.)	3,5 (8669.)	14,0 (3223.)	0,000 (4985.)	0,4 (24757.)	0,005 (6685.)	5,6 (1096.)
88	<i>pa take</i>	nedoločnost	49 (1486.)	30 (1895.)	1,9 (14506.)	13,1 (4160.)	0,000 (8697.)	62,0 (5612.)	0,003 (10389.)	5,1 (1364.)
89	<i>pa tako</i>	nedoločnost	499 (69.)	233 (46.)	0,7 (20142.)	18,6 (818.)	0,000 (7987.)	110,5 (3129.)	0,025 (2031.)	8,7 (354.)
90	<i>pa tako naprej</i>	nedoločnost	28 (2959.)	20 (3391.)	2,6 (11527.)	12,3 (5462.)	0,000 (7995.)	26,2 (10533.)	0,003 (9651.)	4,4 (1984.)
91	<i>pa to</i>	nedoločnost	917 (26.)	326 (22.)	0,7 (19908.)	20,4 (414.)	0,000 (5705.)	234,3 (1287.)	0,038 (1493.)	12,5 (158.)
92	<i>pa vse</i>	nedoločnost	197 (266.)	135 (140.)	1,2 (17549.)	16,5 (1544.)	0,000 (7120.)	122,6 (2789.)	0,012 (3604.)	8,2 (410.)
93	<i>pa vse to</i>	nedoločnost	10 (10409.)	10 (8898.)	1,6 (15745.)	8,3 (15400.)	1,534 (16381.)	202,9 (1579.)	0,000 (22105.)	2,1 (13166.)
94	<i>pač pa</i>	nasprotje	35 (2278.)	35 (1534.)	-1,2 (26206.)	9,0 (13207.)	2,862 (13866.)	36,7 (8497.)	0,002 (14070.)	-8,1 (26763.)
95	<i>po drugi strani</i>	nasprotje	49 (1467.)	27 (2232.)	8,1 (1690.)	19,4 (634.)	0,013 (935.)	339,2 (697.)	0,045 (1315.)	6,9 (612.)
96	<i>po pa</i>	razvrščanje	47 (1561.)	47 (971.)	-1,1 (26033.)	9,9 (10335.)	3,085 (13538.)	38,8 (8173.)	0,002 (11563.)	-8,0 (26753.)
97	<i>po svoje</i>	nedoločnost	32 (2490.)	32 (1772.)	4,5 (6181.)	14,5 (2728.)	0,000 (3686.)	144,1 (2315.)	0,016 (2843.)	5,4 (1213.)
98	<i>po tem</i>	razvrščanje	64 (1068.)	47 (957.)	3,0 (10204.)	15,0 (2351.)	0,000 (4746.)	162,4 (2035.)	0,022 (2198.)	7,0 (593.)
99	<i>pol ko</i>	razvrščanje	66 (1037.)	37 (1360.)	1,8 (14777.)	13,9 (3321.)	0,000 (8031.)	76,5 (4555.)	0,015 (2966.)	5,9 (965.)
100	<i>pol pa</i>	razvrščanje	721 (37.)	297 (27.)	2,5 (12009.)	21,5 (221.)	0,003 (1774.)	1463,9 (60.)	0,043 (1375.)	22,2 (30.)
101	<i>poleg tega</i>	dodajanje	47 (1539.)	24 (2644.)	7,6 (2022.)	18,7 (800.)	0,009 (1121.)	423,5 (455.)	0,044 (1337.)	6,8 (655.)
102	<i>potem ko</i>	razvrščanje	65 (1054.)	34 (1549.)	2,2 (13295.)	14,2 (3005.)	0,000 (6799.)	99,7 (3491.)	0,017 (2754.)	6,3 (792.)
103	<i>potem pa</i>	razvrščanje	507 (67.)	249 (43.)	2,4 (12501.)	20,3 (421.)	0,001 (2217.)	940,4 (115.)	0,031 (1727.)	18,3 (56.)
104	<i>prav tako</i>	dodajanje, nedoločnost	86 (736.)	68 (543.)	2,7 (11490.)	15,5 (2022.)	0,000 (4850.)	180,9 (1824.)	0,014 (3080.)	7,8 (460.)
105	<i>pred tem</i>	razvrščanje	34 (2322.)	28 (2149.)	4,3 (6775.)	14,4 (2817.)	0,000 (3974.)	139,9 (2410.)	0,022 (2162.)	5,5 (1156.)
106	<i>predvsem pa</i>	dodajanje	40 (1890.)	40 (1240.)	2,0 (14115.)	12,6 (4847.)	9,496 (9102.)	54,5 (6248.)	0,002 (12085.)	4,7 (1705.)
107	<i>pri čemer</i>	dodajanje	22 (3946.)	22 (3012.)	7,8 (1904.)	16,7 (1451.)	0,004 (1447.)	207,9 (1523.)	0,017 (2660.)	4,6 (1790.)

Priloge

108	<i>pri tem</i>	dodajanje	101 (606.)	91 (326.)	4,2 (6844.)	17,6 (1119.)	0,001 (2274.)	415,7 (476.)	0,044 (1350.)	9,5 (296.)
109	<i>prvo kot prvo</i>	razvrščanje	5 (21524.)	5 (19587.)	11,0 (782.)	15,7 (1929.)	0,010 (1041.)	44,6 (7344.)	0,035 (1582.)	2,2 (11622.)
110	<i>razen če</i>	pogoj	23 (3786.)	23 (2855.)	4,9 (5397.)	14,0 (3259.)	0,000 (3751.)	117,9 (2915.)	0,007 (5409.)	4,6 (1819.)
111	<i>recimo če</i>	pogoj	55 (1304.)	41 (1172.)	2,3 (12855.)	13,8 (3389.)	0,000 (6962.)	90,8 (3856.)	0,013 (3280.)	5,9 (951.)
112	<i>recimo da</i>	pogoj	65 (1061.)	46 (986.)	0,8 (19505.)	12,8 (4530.)	2,378 (14632.)	19,2 (12393.)	0,005 (6483.)	3,5 (3488.)
113	<i>s tega stališča</i>	dodajanje	6 (17701.)	6 (15872.)	8,2 (1642.)	13,4 (3864.)	0,001 (2243.)	40,0 (7987.)	0,005 (7356.)	2,4 (9340.)
114	<i>s tega vidika</i>	dodajanje	8 (12617.)	8 (11057.)	8,2 (1649.)	14,2 (3036.)	0,002 (1967.)	63,4 (5491.)	0,006 (5775.)	2,8 (6372.)
115	<i>s tem</i>	rezultat	597 (50.)	224 (50.)	5,8 (3828.)	24,3 (18.)	0,034 (665.)	3952,7 (17.)	0,174 (535.)	24,0 (23.)
116	<i>s tem da</i>	dodajanje, specifikacija	84 (757.)	60 (650.)	3,6 (8391.)	16,4 (1571.)	0,000 (3229.)	76,8 (4538.)	0,012 (3449.)	8,4 (382.)
117	<i>saj ne vem</i>	nedoločnost	58 (1219.)	40 (1219.)	3,3 (9384.)	15,0 (2374.)	0,000 (4497.)	151,3 (2211.)	0,022 (2232.)	6,8 (647.)
118	<i>saj pravim</i>	tvorjenje	129 (456.)	81 (393.)	7,1 (2460.)	21,1 (292.)	0,017 (852.)	1105,3 (88.)	0,060 (1067.)	11,2 (195.)
119	<i>saj veste</i>	dvogovor	25 (3411.)	25 (2554.)	4,1 (7071.)	13,4 (3823.)	0,000 (4847.)	98,7 (3528.)	0,011 (3810.)	4,7 (1741.)
120	<i>saj veš</i>	dvogovor	106 (566.)	58 (674.)	3,1 (9843.)	16,6 (1480.)	0,000 (3538.)	284,9 (946.)	0,030 (1748.)	9,1 (314.)
121	<i>samo če</i>	pogoj	45 (1650.)	32 (1732.)	0,8 (19742.)	11,7 (6362.)	1,448 (16635.)	11,7 (15087.)	0,008 (4723.)	2,8 (6075.)
122	<i>samo da</i>	rezultat, nasprotje	152 (374.)	76 (430.)	0,8 (19550.)	15,3 (2162.)	5,417 (11261.)	44,1 (7401.)	0,012 (3594.)	5,4 (1201.)
123	<i>samo pa</i>	nasprotje	11 (9028.)	11 (7659.)	-3,4 (27381.)	3,4 (26959.)	0,000 (8783.)	171,2 (1939.)	0,000 (23789.)	-33,1 (27379.)
124	<i>se prav</i>	rezultat	38 (2051.)	38 (1347.)	0,9 (19224.)	11,4 (7085.)	1,589 (16215.)	12,5 (14752.)	0,004 (8116.)	2,8 (6005.)
125	<i>se pravi</i>	rezultat	832 (29.)	386 (14.)	5,3 (4761.)	24,7 (15.)	0,031 (681.)	5206,5 (8.)	0,096 (795.)	28,1 (19.)
126	<i>se pravi da</i>	rezultat	92 (678.)	55 (728.)	3,2 (9659.)	16,2 (1645.)	0,000 (3676.)	0,3 (25036.)	0,009 (4250.)	8,5 (368.)
127	<i>sedaj pa</i>	razvrščanje	36 (2177.)	36 (1454.)	1,9 (14436.)	12,2 (5495.)	7,812 (9822.)	45,9 (7213.)	0,002 (13058.)	4,4 (2017.)
128	<i>sicer pa</i>	nasprotje	81 (796.)	63 (599.)	1,9 (14380.)	14,6 (2690.)	0,000 (6954.)	105,1 (3300.)	0,005 (6926.)	6,6 (693.)
129	<i>še enkrat poudarjam</i>	tvorjenje	7 (14760.)	7 (13089.)	8,0 (1736.)	13,6 (3610.)	0,001 (2208.)	76,1 (4575.)	0,003 (9383.)	2,6 (7618.)

130	<i>še nekaj</i>	razvrščanje	147 (390.)	89 (334.)	3,0 (10372.)	17,4 (1184.)	0,000 (3206.)	367,3 (612.)	0,030 (1763.)	10,6 (227.)
131	<i>tako ali tako</i>	rezultat	50 (1442.)	35 (1488.)	6,2 (3393.)	17,5 (1144.)	0,003 (1622.)	15,2 (13713.)	0,009 (4359.)	6,9 (611.)
132	<i>tako da</i>	rezultat	1804 (8.)	617 (5.)	3,0 (10190.)	24,7 (14.)	0,012 (986.)	4925,3 (11.)	0,116 (701.)	37,4 (6.)
133	<i>tako ka</i>	rezultat	52 (1382.)	47 (969.)	2,6 (11766.)	14,0 (3252.)	0,000 (6326.)	104,3 (3329.)	0,009 (4472.)	6,0 (908.)
134	<i>tako kakor</i>	specifikacija	45 (1625.)	39 (1281.)	3,2 (9769.)	14,1 (3087.)	0,000 (5278.)	123,4 (2757.)	0,008 (4889.)	5,9 (936.)
135	<i>tako ko</i>	specifikacija	49 (1506.)	39 (1272.)	0,1 (22607.)	11,3 (7167.)	4,667 (25214.)	0,4 (24938.)	0,006 (5798.)	0,6 (21811.)
136	<i>tako kot</i>	specifikacija	484 (74.)	188 (70.)	3,7 (8074.)	21,6 (208.)	0,005 (1370.)	1724,8 (47.)	0,070 (982.)	20,4 (38.)
137	<i>tako rekoč</i>	tvorjenje	30 (2674.)	30 (1932.)	6,5 (2995.)	16,4 (1597.)	0,002 (1826.)	272,1 (1020.)	0,005 (6560.)	5,4 (1214.)
138	<i>takrat ko</i>	razvrščanje	126 (468.)	33 (1627.)	5,1 (5115.)	19,0 (712.)	0,004 (1542.)	672,7 (187.)	0,049 (1256.)	10,9 (214.)
139	<i>te pa</i>	razvrščanje	67 (1023.)	51 (835.)	-0,4 (24450.)	11,7 (6491.)	5,464 (20225.)	6,0 (18277.)	0,004 (8601.)	-2,6 (25552.)
140	<i>ti povem</i>	dvogovor	30 (2681.)	24 (2672.)	5,4 (4503.)	15,2 (2204.)	0,001 (2719.)	173,8 (1920.)	0,013 (3366.)	5,3 (1249.)
141	<i>to pomeni</i>	rezultat	223 (224.)	67 (550.)	4,4 (6532.)	20,0 (495.)	0,004 (1501.)	1038,2 (96.)	0,023 (2138.)	14,2 (111.)
142	<i>to se pravi</i>	rezultat	152 (372.)	107 (229.)	4,2 (6910.)	18,7 (792.)	0,002 (1896.)	559,7 (265.)	0,028 (1834.)	11,6 (179.)
143	<i>toliko da</i>	rezultat	56 (1276.)	30 (1891.)	1,7 (15206.)	13,4 (3899.)	9,928 (8940.)	61,1 (5687.)	0,005 (7094.)	5,3 (1275.)
144	<i>tudi če</i>	nasprotje	140 (419.)	50 (850.)	1,5 (16479.)	15,7 (1905.)	0,000 (7137.)	113,1 (3045.)	0,019 (2457.)	7,6 (491.)
145	<i>v bistvu</i>	nasprotje	985 (23.)	267 (40.)	5,8 (3965.)	25,7 (5.)	0,054 (540.)	7963,6 (2.)	0,105 (743.)	30,8 (13.)
146	<i>v glavnem</i>	rezultat	279 (164.)	168 (91.)	5,8 (3969.)	22,0 (152.)	0,015 (900.)	2231,1 (34.)	0,030 (1742.)	16,4 (73.)
147	<i>v nadaljevanju</i>	razvrščanje	77 (854.)	60 (653.)	5,7 (4025.)	18,3 (916.)	0,004 (1557.)	594,9 (244.)	0,008 (4735.)	8,6 (365.)
148	<i>v nadaljevanju pa</i>	razvrščanje	6 (19391.)	6 (17403.)	2,3 (12791.)	7,5 (18156.)	0,000 (15275.)	214,5 (1461.)	0,000 (25504.)	1,9 (15501.)
149	<i>v redu</i>	dvogovor	833 (28.)	291 (31.)	5,8 (3980.)	25,2 (9.)	0,045 (588.)	6662,5 (5.)	0,089 (827.)	28,3 (17.)
150	<i>v resnici</i>	nasprotje	109 (546.)	92 (321.)	5,5 (4331.)	19,1 (703.)	0,005 (1433.)	765,2 (154.)	0,012 (3626.)	10,2 (251.)
151	<i>v smislu</i>	specifikacija	30 (2694.)	25 (2528.)	4,0 (7350.)	13,8 (3411.)	0,000 (4612.)	120,5 (2830.)	0,003 (10216.)	5,1 (1376.)
152	<i>v tem času</i>	razvrščanje	41 (1809.)	36 (1435.)	6,4 (3125.)	17,2 (1261.)	0,003 (1642.)	222,4 (1385.)	0,008 (4672.)	6,3 (794.)
153	<i>v tem primeru</i>	rezultat	83 (770.)	58 (676.)	6,6 (2938.)	19,3 (637.)	0,008 (1182.)	554,6 (269.)	0,017 (2693.)	9,0 (335.)

Priloge

154	<i>v tem smislu</i>	rezultat	57 (1243.)	28 (2107.)	6,7 (2860.)	18,3 (895.)	0,005 (1353.)	438,8 (417.)	0,012 (3600.)	7,4 (526.)
155	<i>v vsakem primeru</i>	rezultat	40 (1871.)	40 (1239.)	6,7 (2802.)	17,4 (1189.)	0,004 (1522.)	211,4 (1485.)	0,008 (4641.)	6,2 (819.)
156	<i>vendar pa</i>	nasprotje	38 (2030.)	38 (1339.)	2,1 (13408.)	12,6 (4803.)	0,000 (8574.)	59,3 (5812.)	0,002 (12519.)	4,8 (1647.)
157	<i>veste kaj</i>	dvogovor	54 (1323.)	41 (1177.)	4,0 (7358.)	15,5 (2012.)	0,000 (3456.)	209,7 (1501.)	0,010 (3927.)	6,9 (631.)
158	<i>veš kaj</i>	dvogovor	409 (92.)	142 (129.)	3,9 (7726.)	21,2 (264.)	0,005 (1399.)	1519,8 (57.)	0,066 (1013.)	18,8 (49.)
159	<i>veš kaj je</i>	dvogovor	87 (729.)	39 (1248.)	3,2 (9531.)	16,1 (1700.)	0,000 (3712.)	67,9 (5135.)	0,008 (4848.)	8,3 (392.)
160	<i>veš kaj je fora</i>	dvogovor	6 (17655.)	6 (15837.)	9,1 (1288.)	14,2 (2982.)	0,003 (1701.)	50,9 (6620.)	0,010 (4097.)	2,4 (9278.)
161	<i>za konec</i>	razvrščanje	38 (2010.)	38 (1332.)	4,1 (7171.)	14,6 (2698.)	0,000 (4015.)	150,9 (2215.)	0,009 (4493.)	5,8 (1019.)
162	<i>za to</i>	rezultat	356 (113.)	184 (73.)	1,2 (17508.)	18,2 (933.)	0,000 (5462.)	222,6 (1383.)	0,026 (1930.)	11,1 (202.)
163	<i>za to da</i>	rezultat	101 (607.)	41 (1163.)	4,1 (7216.)	17,4 (1177.)	0,001 (2420.)	55,6 (6134.)	0,013 (3287.)	9,4 (300.)
164	<i>za začetek</i>	razvrščanje	32 (2489.)	32 (1771.)	5,6 (4211.)	15,6 (1973.)	0,001 (2444.)	200,9 (1598.)	0,007 (5062.)	5,5 (1152.)
165	<i>zaradi tega</i>	rezultat	371 (107.)	76 (426.)	7,9 (1831.)	24,9 (10.)	0,088 (420.)	3616,5 (20.)	0,269 (387.)	19,1 (45.)
166	<i>zaradi tega ker</i>	razlog	187 (288.)	87 (339.)	7,6 (2026.)	22,7 (90.)	0,036 (649.)	1533,1 (56.)	0,122 (679.)	13,6 (128.)
167	<i>zato da</i>	rezultat	154 (363.)	71 (493.)	2,4 (12321.)	17,0 (1329.)	0,000 (4049.)	288,5 (920.)	0,014 (3197.)	10,1 (257.)
168	<i>zato ker</i>	razlog	508 (65.)	211 (54.)	6,3 (3319.)	24,2 (19.)	0,039 (625.)	3702,9 (19.)	0,165 (553.)	22,2 (31.)
169	<i>zato pa</i>	rezultat	39 (1990.)	33 (1664.)	-0,0 (23230.)	10,5 (8891.)	7,439 (27169.)	0,0 (27064.)	0,002 (12592.)	-0,0 (23299.)
170	<i>zdaj ko</i>	razvrščanje	105 (579.)	47 (947.)	1,9 (14429.)	15,3 (2133.)	0,000 (6360.)	130,3 (2609.)	0,019 (2461.)	7,5 (500.)
171	<i>zdaj ne vem</i>	nedoločnost	62 (1112.)	33 (1634.)	2,7 (11193.)	14,6 (2650.)	0,000 (5418.)	121,0 (2824.)	0,019 (2506.)	6,7 (682.)
172	<i>zdaj pa</i>	razvrščanje	693 (39.)	327 (21.)	1,9 (14656.)	20,7 (341.)	0,001 (2545.)	864,3 (125.)	0,038 (1479.)	19,2 (44.)
173	<i>znaš kaj</i>	dvogovor	41 (1819.)	25 (2501.)	4,1 (7230.)	14,8 (2539.)	0,000 (3904.)	162,4 (2036.)	0,008 (4830.)	6,0 (912.)

Priloga 5: Seznam večbesednih diskurznihih označevalcev govorne slovenščine

V tej prilogi navajamo končni seznam prepoznanih večbesednih diskurznihih označevalcev govorne slovenščine, ki so bili prepoznani v procesu ročnega označevanja vzorčenega korpusa govorne slovenščine in/ali ročnega pregleda več različnih seznamov statistično najbolj izstopajočih besednih nizov korpus Gos, skupaj s podatki o: krovni metodi prepoznavi (R: ročno oz. kvalitativno, S: statistično oz. kvantitativno), minimalistični funkcijski skupini, pogostosti rabe diskurznihih označevalca v korpusih Gos in Kres s pripisom deleža diskurznihih rab glede na vse pojavitve oblike ter izračunu ključnosti za govorni jezik z metodama LL in OR. Nizi so razvrščeni glede na ključnost, izračunano po metodi LL.

	Niz	Št. bes.	Funkcijska skupina	Metoda	Pogostost Gos	Pogostost Kres	Ključnost (LL)	Ključnost (OR)
1	<i>a ne</i>	2	dvogovor	R in S	1726 (94 %)	1038 (23 %)	12267,9	165,1
2	<i>ne vem</i>	2	nedoločnost	R in S	1399 (51 %)	310 (2 %)	11280,7	448,1
3	<i>a veš</i>	2	dvogovor	R in S	705 (84 %)	111 (30 %)	5848,4	630,3
4	<i>pol pa</i>	2	razvrščanje	R in S	714 (99 %)	901 (90 %)	4379,7	78,7
5	<i>v bistvu</i>	2	nasprotje	R in S	985 (100 %)	4920 (99 %)	3852,6	19,9
6	<i>ja no</i>	2	dvogovor	R in S	328 (100 %)	142 (100 %)	2449,8	229,3
7	<i>se pravi</i>	2	rezultat	R in S	666 (80 %)	3856 (72 %)	2433,1	17,1
8	<i>tako da</i>	2	rezultat	R in S	1804 (100 %)	42181 (100 %)	2414,9	4,2
9	<i>pa to</i>	2	nedoločnost	R in S	339 (37 %)	240 (5 %)	2345,1	140,1
10	<i>ali kaj</i>	2	dvogovor	R in S	345 (68 %)	509 (20 %)	2035,4	67,2
11	<i>zdaj pa</i>	2	razvrščanje	R in S	693 (100 %)	8081 (100 %)	1700,6	8,5
12	<i>v glavnem</i>	2	rezultat	R in S	193 (69 %)	379 (9 %)	1051,6	50,4
13	<i>pa tako</i>	2	nedoločnost	R in S	110 (22 %)	0 (0 %)	1000,4	10899,8
14	<i>v redu</i>	2	dvogovor	R in S	183 (22 %)	536 (7 %)	883,2	33,9
15	<i>ali pa</i>	2	dodajanje	R in S	815 (60 %)	23052 (65 %)	866,2	3,5
16	<i>veš kaj</i>	2	dvogovor	R in S	176 (43 %)	563 (32 %)	821,4	31,0
17	<i>saj pravim</i>	2	tvorjenje	R in S	119 (92 %)	89 (84 %)	811,8	132,3
18	<i>zato ker</i>	2	razlog	R in S	498 (98 %)	9923 (58 %)	787,4	5,0
19	<i>in tako naprej</i>	3	nedoločnost	R in S	235 (100 %)	1820 (100 %)	741,1	12,8
20	<i>tako kot</i>	2	specifikacija	R in S	315 (65 %)	5039 (22 %)	606,8	6,2
21	<i>in tako</i>	2	nedoločnost	R in S	66 (14 %)	0 (0 %)	593,5	6505,3
22	<i>ma ja</i>	2	dvogovor	R in S	91 (100 %)	90 (100 %)	589,6	100,4
23	<i>to se pravi</i>	3	rezultat	R in S	126 (83 %)	454 (82 %)	564,0	27,6
24	<i>zaradi tega ker</i>	3	razlog	R in S	170 (91 %)	1220 (73 %)	559,4	13,9
25	<i>drugače pa</i>	2	nasprotje	samo S	164 (98 %)	1533 (91 %)	462,5	10,6

26	<i>tako ka</i>	2	rezultat	R in S	52 (100 %)	4 (100 %)	450,5	1290,8
27	<i>pol ko</i>	2	razvrščanje	samo S	65 (98 %)	61 (58 %)	425,0	105,0
28	<i>bi rekel</i>	2	tvorjenje	R in S	89 (47 %)	291 (14 %)	414,1	30,4
29	<i>potem pa</i>	2	razvrščanje	R in S	507 (100 %)	17678 (100 %)	397,2	2,8
30	<i>ali pa kaj takega</i>	4	nedoločnost	samo S	41 (100 %)	1 (100 %)	368,4	4071,0
31	<i>po pa</i>	2	razvrščanje	samo S	47 (100 %)	17 (94 %)	359,4	274,5
32	<i>ali kako</i>	2	dvogovor	R in S	60 (54 %)	87 (7 %)	355,0	68,1
33	<i>recimo če</i>	2	pogoj	samo S	55 (100 %)	63 (93 %)	345,1	86,7
34	<i>to pomeni</i>	2	rezultat	R in S	54 (24 %)	74 (1 %)	321,7	72,2
35	<i>če ne</i>	2	rezultat	R in S	92 (33 %)	635 (5 %)	308,8	14,4
36	<i>bom rekel</i>	2	tvorjenje	R in S	61 (63 %)	157 (33 %)	306,6	38,5
37	<i>a veste</i>	2	dvogovor	samo S	43 (83 %)	27 (21 %)	303,3	157,6
38	<i>ali pa kaj</i>	3	nedoločnost	samo S	36 (32 %)	4 (5 %)	301,5	881,7
39	<i>kaj jaz vem</i>	3	nedoločnost	R in S	46 (74 %)	50 (51 %)	292,0	91,3
40	<i>glih tako</i>	2	dodajanje, rezultat	R in S	41 (69 %)	26 (51 %)	288,9	156,6
41	<i>na neki način</i>	3	nedoločnost	samo S	126 (100 %)	1785 (100 %)	268,3	7,0
42	<i>in pa</i>	2	dodajanje	R in S	69 (28 %)	364 (16 %)	264,4	18,9
43	<i>čisto tako</i>	2	nedoločnost	samo S	35 (85 %)	18 (8 %)	255,3	194,8
44	<i>jaz ne vem</i>	3	nedoločnost	samo S	31 (27 %)	8 (4 %)	250,4	409,2
45	<i>na vem</i>	2	nedoločnost	samo S	26 (37 %)	0 (0 %)	231,1	2581,7
46	<i>ne znam</i>	2	nedoločnost	samo S	24 (17 %)	0 (0 %)	211,0	2363,0
47	<i>saj ne vem</i>	3	nedoločnost	R in S	29 (50 %)	18 (13 %)	205,4	161,7
48	<i>kaj je že</i>	3	dvogovor	samo S	25 (46 %)	5 (3 %)	205,0	540,8
49	<i>samo da</i>	2	rezultat, nasprotje	R in S	123 (81 %)	2413 (63 %)	198,0	5,1
50	<i>pa tako naprej</i>	3	nedoločnost	R in S	28 (100 %)	19 (100 %)	195,0	146,3
51	<i>no ja</i>	2	dvogovor	R in S	98 (100 %)	1677 (100 %)	178,6	5,8
52	<i>ali nekaj</i>	2	nedoločnost	samo S	20 (29 %)	0 (0 %)	176,3	1985,9
53	<i>in vse</i>	2	nedoločnost	samo S	20 (28 %)	0 (0 %)	176,3	1985,8
54	<i>saj veš</i>	2	dvogovor	R in S	67 (63 %)	718 (49 %)	173,0	9,2
55	<i>pa vse</i>	2	nedoločnost	samo S	26 (13 %)	22 (1 %)	171,0	116,5
56	<i>zdaj ne vem</i>	3	nedoločnost	samo S	20 (32 %)	4 (4 %)	162,8	496,5
57	<i>bomo rekli</i>	2	tvorjenje	samo S	21 (58 %)	9 (10 %)	157,1	231,7
58	<i>nekaj takega</i>	2	nedoločnost	R in S	20 (20 %)	7 (1 %)	154,3	272,1
59	<i>ne vem kaj</i>	3	nedoločnost	R in S	17 (7 %)	0 (0 %)	148,4	1682,0
60	<i>tako ko</i>	2	specifikacija	R in S	21 (43 %)	13 (13 %)	147,8	155,7
61	<i>pri tem</i>	2	dodajanje	R in S	55 (54 %)	19653 (76 %)	145,1	0,3
62	<i>znaš kaj</i>	2	dvogovor	R in S	16 (39 %)	0 (0 %)	139,9	1588,7
63	<i>zdaj ko</i>	2	razvrščanje	samo S	105 (100 %)	2576 (87 %)	133,1	4,0
64	<i>se pravi da</i>	3	rezultat	R in S	70 (76 %)	1148 (88 %)	132,3	6,1

65	<i>veš kaj je</i>	3	dvogovor	R in S	15 (17 %)	0 (0 %)	130,8	1489,3
66	<i>ali nekaj takega</i>	3	nedoločnost	R in S	32 (100 %)	149 (100 %)	129,0	21,3
67	<i>glede na to</i>	3	rezultat	samo S	34 (19 %)	182 (2 %)	127,9	18,4
68	<i>kako bi rekel</i>	3	tvorjenje	samo S	25 (100 %)	65 (78 %)	125,4	38,2
69	<i>kaj pa jaz vem</i>	4	nedoločnost	samo S	19 (86 %)	22 (79 %)	118,9	85,8
70	<i>bolj tako</i>	2	nedoločnost	samo S	17 (37 %)	14 (22 %)	114,3	120,6
71	<i>a veš kaj</i>	3	dvogovor	R in S	15 (45 %)	6 (14 %)	113,2	248,2
72	<i>se prav</i>	2	rezultat	samo S	17 (45 %)	17 (1 %)	110,4	101,5
73	<i>ne da</i>	2	specifikacija	R in S	25 (4 %)	11838 (77 %)	109,9	0,2
74	<i>zato da</i>	2	rezultat	R in S	139 (90 %)	4799 (67 %)	109,8	2,9
75	<i>medtem ko</i>	2	razvrščanje, nasprotje	R in S	65 (93 %)	18779 (100 %)	108,6	0,3
76	<i>te pa</i>	2	razvrščanje	R in S	22 (33 %)	63 (6 %)	106,9	34,7
77	<i>čakaj malo</i>	2	dvogovor	samo S	24 (52 %)	91 (99 %)	105,2	26,2
78	<i>veste kaj</i>	2	dvogovor	samo S	27 (50 %)	137 (13 %)	104,9	19,6
79	<i>glede na to da</i>	4	razlog	R in S	106 (87 %)	3399 (81 %)	94,9	3,1
80	<i>bom rekla</i>	2	tvorjenje	R in S	19 (61 %)	53 (23 %)	93,3	35,8
81	<i>edino če</i>	2	pogoj	samo R	16 (100 %)	28 (100 %)	90,3	56,7
82	<i>pa ne vem kaj</i>	4	nedoločnost	samo S	13 (34 %)	9 (9 %)	90,2	143,4
83	<i>na začetku</i>	2	razvrščanje	samo S	107 (90 %)	3665 (43 %)	86,5	2,9
84	<i>bi človek rekel</i>	3	tvorjenje	R in S	12 (100 %)	7 (54 %)	85,7	170,2
85	<i>ali kaj je</i>	3	dvogovor	R in S	10 (25 %)	1 (1 %)	85,5	992,9
86	<i>ali pa nekaj</i>	3	nedoločnost	samo R	10 (45 %)	0 (0 %)	85,5	993,0
87	<i>kaj takega</i>	2	nedoločnost	samo S	10 (8 %)	0 (0 %)	81,1	945,1
88	<i>prav tako</i>	2	dodajanje, nedoločnost	R in S	73 (85 %)	17800 (80 %)	80,7	0,4
89	<i>kaj je</i>	2	dvogovor	R in S	9 (1 %)	0 (0 %)	79,9	931,3
90	<i>samo pa</i>	2	nasprotje	samo S	10 (91 %)	3 (8 %)	78,2	331,0
91	<i>s tem</i>	2	rezultat	R in S	113 (19 %)	23402 (51 %)	78,0	0,5
92	<i>med drugim</i>	2	specifikacija	samo S	29 (97 %)	10495 (99 %)	77,8	0,3
93	<i>za to</i>	2	rezultat	samo S	36 (10 %)	495 (2 %)	76,9	7,1
94	<i>takrat ko</i>	2	razvrščanje	R in S	111 (88 %)	4160 (56 %)	76,7	2,6
95	<i>a ne da</i>	3	dvogovor	R in S	14 (16 %)	31 (34 %)	73,8	44,8
96	<i>pa vse to</i>	3	nedoločnost	samo S	10 (100 %)	5 (6 %)	73,2	198,6
97	<i>zaradi tega</i>	2	rezultat	R in S	145 (39 %)	6413 (67 %)	72,2	2,2
98	<i>moram reči</i>	2	tvorjenje	R in S	18 (17 %)	92 (9 %)	71,9	19,9
99	<i>kaj že</i>	2	dvogovor	samo S	12 (29 %)	17 (13 %)	71,6	70,0
100	<i>tako kakor</i>	2	specifikacija	samo S	33 (73 %)	512 (44 %)	65,3	6,4
101	<i>recimo temu</i>	2	tvorjenje	samo R	17 (81 %)	91 (92 %)	64,5	18,5
102	<i>pa take fore</i>	3	nedoločnost	samo R	10 (100 %)	12 (100 %)	62,1	82,8
103	<i>poleg tega</i>	2	dodajanje	samo S	33 (70 %)	9840 (67 %)	59,2	0,3

104	<i>zaradi tega da</i>	3	rezultat	samo R	17 (100 %)	110 (54 %)	59,0	15,4
105	<i>pa take stvari</i>	3	nedoločnost	samo R	9 (100 %)	9 (60 %)	58,2	99,3
106	<i>tudi če</i>	2	nasprotje	R in S	125 (89 %)	5713 (89 %)	57,6	2,2
107	<i>in vse skupaj</i>	3	nedoločnost	samo R	9 (90 %)	9 (2 %)	57,6	95,1
108	<i>kako se reče</i>	3	tvorjenje	samo S	8 (35 %)	5 (9 %)	56,5	158,9
109	<i>kljub temu</i>	2	nasprotje	samo S	20 (34 %)	7404 (57 %)	56,1	0,3
110	<i>na koncu</i>	2	razvrščanje	R in S	186 (82 %)	10156 (69 %)	54,5	1,8
111	<i>pri čemer</i>	2	dodajanje	samo S	22 (100 %)	7384 (100 %)	50,8	0,3
112	<i>ali pa kaj podobnega</i>	4	nedoločnost	samo S	9 (100 %)	17 (100 %)	49,7	52,6
113	<i>a vidite</i>	2	dvogovor	samo R	6 (32 %)	0 (0 %)	49,6	595,7
114	<i>alla fine</i>	2	razvrščanje	samo S	6 (100 %)	0 (0 %)	49,6	595,8
115	<i>veš kaj je fora</i>	4	dvogovor	samo S	6 (100 %)	0	49,6	595,7
116	<i>ali karkoli</i>	2	nedoločnost	samo R	11 (79 %)	39 (13 %)	49,4	27,9
117	<i>in tako dalje</i>	3	nedoločnost	samo S	30 (100 %)	583 (100 %)	48,6	5,1
118	<i>ob tem</i>	2	dodajanje	samo S	32 (60 %)	8698 (82 %)	46,4	0,4
119	<i>kot smo rekli</i>	3	tvorjenje	samo R	8 (89 %)	14 (54 %)	45,2	56,7
120	<i>prej ko</i>	2	razvrščanje	samo R	24 (92 %)	424 (74 %)	42,5	5,6
121	<i>v končni fazi</i>	3	rezultat	samo R	20 (100 %)	288 (100 %)	42,0	6,9
122	<i>kaj sem hotela reči</i>	4	tvorjenje	samo S	5 (100 %)	0 (0 %)	40,7	496,5
123	<i>kot rečeno</i>	2	tvorjenje	R in S	32 (100 %)	786 (100 %)	40,5	4,0
124	<i>vendar pa</i>	2	nasprotje	R in S	38 (100 %)	8880 (100 %)	37,5	0,4
125	<i>pa take</i>	2	nedoločnost	R in S	6 (12 %)	8 (4 %)	36,6	76,4
126	<i>če ne drugega</i>	3	dodajanje	samo R	15 (100 %)	178 (100 %)	36,4	8,4
127	<i>na ta način</i>	3	rezultat	R in S	83 (100 %)	3923 (100 %)	35,8	2,1
128	<i>v smislu da</i>	3	specifikacija	samo R	5 (100 %)	3 (60 %)	35,6	165,5
129	<i>tedaj pa</i>	2	razvrščanje	samo R	20 (95 %)	366 (77 %)	34,3	5,4
130	<i>za konec</i>	2	razvrščanje	samo S	34 (89 %)	1004 (61 %)	34,1	3,4
131	<i>moram priznati</i>	2	tvorjenje	samo S	9 (45 %)	48 (8 %)	34,1	18,5
132	<i>v vsakem primeru</i>	3	rezultat	R in S	40 (100 %)	1325 (100 %)	34,0	3,0
133	<i>bi rekla</i>	2	tvorjenje	samo S	7 (15 %)	25 (3 %)	31,6	28,4
134	<i>konec koncev</i>	2	rezultat	samo S	44 (100 %)	1635 (100 %)	31,0	2,7
135	<i>poleg tega pa</i>	3	dodajanje	samo R	9 (100 %)	3545 (100 %)	28,4	0,3
136	<i>kaj naj rečem</i>	3	tvorjenje	samo R	11 (85 %)	121 (37 %)	28,0	9,0
137	<i>s tem da</i>	3	dodajanje, specifikacija	R in S	66 (79 %)	3146 (69 %)	27,9	2,1
138	<i>kar se tega tiče</i>	4	dodajanje	samo S	10 (100 %)	99 (100 %)	27,3	10,0
139	<i>druga zadeva</i>	2	razvrščanje	R in S	5 (83 %)	11 (19 %)	26,4	45,1
140	<i>v tem smislu</i>	3	rezultat	samo S	31 (54 %)	1066 (79 %)	24,8	2,9

141	<i>če že</i>	2	dodajanje, pogoj	samo S	52 (100 %)	2414 (100 %)	23,4	2,1
142	<i>pa vendarle</i>	2	nasprotje	samo R	25 (100 %)	780 (100 %)	23,2	3,2
143	<i>recimo da</i>	2	pogoj	samo S	23 (35 %)	684 (97 %)	22,9	3,3
144	<i>če dovolite</i>	2	dvogovor	samo R	9 (75 %)	103 (88 %)	22,4	8,7
145	<i>sploh če</i>	2	pogoj	samo R	17 (94 %)	423 (100 %)	21,2	4,0
146	<i>saj veste</i>	2	dvogovor	samo S	18 (72 %)	485 (42 %)	20,4	3,7
147	<i>kako se že reče</i>	4	tvorjenje	samo S	5 (83 %)	24 (51 %)	19,9	20,7
148	<i>toliko da</i>	2	rezultat	samo S	32 (57 %)	1300 (67 %)	19,1	2,4
149	<i>in sicer</i>	2	specifikacija	R in S	191 (100 %)	13612 (100 %)	18,7	1,4
150	<i>na tak način</i>	3	rezultat	samo S	27 (79 %)	1017 (72 %)	18,6	2,6
151	<i>lahko rečem</i>	2	tvorjenje	samo S	6 (23 %)	60 (6 %)	16,3	9,9
152	<i>bi rekli</i>	2	tvorjenje	samo S	6 (14 %)	66 (6 %)	15,4	9,1
153	<i>ti povem</i>	2	dvogovor	samo S	6 (20 %)	66 (12 %)	15,3	9,0
154	<i>v smislu</i>	2	specifikacija	R in S	10 (33 %)	209 (7 %)	15,1	4,7
155	<i>kako naj rečem</i>	3	tvorjenje	samo S	6 (67 %)	68 (78 %)	15,0	8,8
156	<i>medtem ko pa</i>	3	nasprotje	samo R	5 (100 %)	45 (100 %)	14,5	11,0
157	<i>potem ko</i>	2	razvrščanje	R in S	61 (94 %)	9344 (92 %)	13,2	0,6
158	<i>prvo kot prvo</i>	3	razvrščanje	samo S	5 (100 %)	53 (100 %)	13,1	9,4
159	<i>pač pa</i>	2	nasprotje	samo S	29 (83 %)	5257 (100 %)	12,9	0,5
160	<i>tako ali tako</i>	3	rezultat	R in S	50 (100 %)	2851 (100 %)	12,7	1,7
161	<i>za začetek</i>	2	razvrščanje	R in S	26 (81 %)	1170 (46 %)	12,6	2,2
162	<i>še enkrat poudarjam</i>	3	tvorjenje	samo S	5 (71 %)	65 (73 %)	11,4	7,6
163	<i>po drugi strani</i>	3	nasprotje	R in S	32 (65 %)	1699 (35 %)	10,2	1,9
164	<i>v kolikor</i>	2	pogoj	samo R	7 (41 %)	1889 (100 %)	10,0	0,4
165	<i>sedaj pa</i>	2	razvrščanje	R in S	36 (100 %)	2062 (100 %)	9,0	1,7
166	<i>a la</i>	2	specifikacija	samo R	5 (63 %)	91 (53 %)	8,7	5,5
167	<i>že odkar</i>	2	razvrščanje	samo R	6 (100 %)	139 (100 %)	8,1	4,3
168	<i>in podobno</i>	2	nedoločnost	samo S	20 (77 %)	3533 (94 %)	8,1	0,6
169	<i>in to</i>	2	nedoločnost, specifikacija	R in S	58 (9 %)	8085 (43 %)	7,7	0,7
170	<i>na primer</i>	2	specifikacija	R in S	270 (100 %)	22568 (100 %)	7,5	1,2
171	<i>predvsem pa</i>	2	dodajanje	R in S	40 (100 %)	5925 (100 %)	7,3	0,7
172	<i>v nadaljevanju pa</i>	3	razvrščanje	samo S	6 (100 %)	158 (100 %)	7,0	3,8
173	<i>od takrat naprej</i>	3	razvrščanje	samo S	8 (100 %)	262 (100 %)	6,9	3,0
174	<i>s tega stališča</i>	3	dodajanje	samo S	5 (83 %)	114 (100 %)	6,9	4,4
175	<i>ne pa da</i>	3	nasprotje	R in S	22 (69 %)	1192 (100 %)	6,6	1,8
176	<i>pred tem</i>	2	razvrščanje	samo S	19 (56 %)	3229 (68 %)	6,6	0,6
177	<i>še nekaj</i>	2	razvrščanje	samo S	9 (6 %)	330 (4 %)	6,1	2,7
178	<i>tako torej</i>	2	dvogovor	samo R	4 (27 %)	85 (49 %)	5,9	4,7

179	<i>kot da</i>	2	specifikacija	R in S	137 (93 %)	16339 (95 %)	5,0	0,8
180	<i>namesto da</i>	2	nasprotje	samo R	23 (100 %)	3447 (100 %)	4,5	0,7
181	<i>v tem primeru</i>	3	rezultat	R in S	78 (94 %)	6180 (96 %)	3,7	1,3
182	<i>kljub vsemu</i>	2	nasprotje	samo S	16 (100 %)	2464 (95 %)	3,6	0,6
183	<i>sicer pa</i>	2	nasprotje	R in S	81 (100 %)	9780 (100 %)	3,3	0,8
184	<i>ob tem da</i>	3	dodajanje	samo R	6 (100 %)	269 (100 %)	2,9	2,2
185	<i>v nadaljevanju</i>	2	razvrščanje	R in S	49 (64 %)	6075 (89 %)	2,6	0,8
186	<i>tako rekoč</i>	2	tvorjenje	R in S	30 (100 %)	3764 (100 %)	1,8	0,8
187	<i>za to da</i>	3	rezultat	R in S	17 (17 %)	1242 (20 %)	1,5	1,4
188	<i>razen če</i>	2	pogoj	samo S	23 (100 %)	2909 (100 %)	1,4	0,8
189	<i>sploh pa</i>	2	dodajanje	samo R	22 (96 %)	1675 (100 %)	1,4	1,3
190	<i>ne glede na to</i>	4	rezultat	samo S	5 (13 %)	295 (7 %)	1,1	1,7
191	<i>glede tega</i>	2	dodajanje	samo S	13 (52 %)	1661 (100 %)	0,9	0,8
192	<i>danes ko</i>	2	razvrščanje	samo R	8 (100 %)	560 (100 %)	0,9	1,4
193	<i>zato pa</i>	2	rezultat	R in S	39 (100 %)	3369 (100 %)	0,7	1,1
194	<i>s tega vidika</i>	3	dodajanje	samo S	8 (100 %)	1050 (324 %)	0,7	0,8
195	<i>še posebej če</i>	3	pogoj	samo R	7 (100 %)	932 (100 %)	0,7	0,7
196	<i>samo če</i>	2	pogoj	samo S	16 (36 %)	1311 (100 %)	0,5	1,2
197	<i>kakorkoli že</i>	2	rezultat	samo R	15 (94 %)	1235 (85 %)	0,5	1,2
198	<i>ali ne</i>	2	dvogovor	R in S	7 (3 %)	542 (7 %)	0,4	1,3
199	<i>v tem času</i>	3	razvrščanje	samo S	41 (100 %)	4360 (98 %)	0,2	0,9
200	<i>ob koncu</i>	2	razvrščanje	samo S	8 (50 %)	699 (12 %)	0,1	1,1
201	<i>po tem</i>	2	razvrščanje	samo S	9 (14 %)	806 (9 %)	0,1	1,1
202	<i>v resnici</i>	2	nasprotje	R in S	109 (100 %)	11066 (100 %)	0,1	1,0
203	<i>kljub temu da</i>	3	nasprotje	R in S	33 (100 %)	3221 (100 %)	0,0	1,0
204	<i>po svoje</i>	2	nedoločnost	R in S	13 (41 %)	1300 (40 %)	0,0	1,0

Priloga 6: Vzorčni primeri rabe označevalca *a ne* v korpusu Gos

V tej prilogi navajamo 94 primerov rabe diskurznega označevalca *a ne*, ki so bili identificirani v analizi 100 naključnih konkordanc skupno 1.836 pojavitev te oblike v celotnem korpusu Gos.

	tip ¹³²	id izjave	Konkordanca ¹³³
1	JII	001-0166	zdej jasno ključno temeljno vprašanje v kakšnem jeziku je to potekalo a ne se na žalost ne ve zaradi tega ker ni ohranjenih zapisov
2	JII	004-0006	ker vsi vemo a ne v kakšnem svetu živimo v bistvu v svetu ki je kr zahtevn za preživetje a ne
3	JII	004-0036	smo ravno meli ravno včeraj smo meli skupino za starše in smo neko podobno temo meli a ne in smo že mi odrasli v bistvu govorili čisto [gap] a ne
4	JII	004-0301	verjetno doskrat sploh [:voice] mislim ženske mamo po navadi kr težave z mamami a ne se kr težko nekaj //
5	JII	004-0392	tu [gap] ko greš v šolo pa si prvič ocenjen recimo a ne sicer je [gap] ampak prideš v eno okolje kje pa se kr neki od tebe začne pričakvat
6	JII	004-0421	ja če bi pa naš otrok šel v stran bi pa bi pa šli ne vem v Južno Ameriko al pa v Afriko po njih a ne sam da bi prišli nazaj
7	JII	004-0431	pa je to čist normalna stvar k jo bom nardila in vse mame tk delajo a ne ampak moja hčerka pa tam doživi da da sn jaz šla a ne
8	JII	004-0475	rečt ja mam službo pa to pa ono a ne pa ne ne uspem ne morem a ne niam časa je v bistu res tk en izgovor na nek način a ne okej se pravi [gap] voljo
9	JII	004-0603	odnosa tud z otrokom a ne in tud otrokove samopodobe
10	JII	005-0147	eee zdaj načeloma seveda če govorimo o univerzalnem temeljn dohodku vsi zagovorniki morajo priznat da eee bi bilo edino pravično da bi čisto vsi ljudje v tem univerzumu prejemali temeljn dohodek a ne nimaš dobene [gap] dobene zgovora nimaš //
11	JII	010-0329	tele tri a ne dvojka
12	JII	015-0192	nekeje Štajerske Dolenjsko pa nekeje Notranjske a ne ja
13	JII	015-0277	zapelja tak je naglas a ne mhm //
14	JII	015-0453	ja drugačne besede so bile a ne čemu je temu tako da so ble druge besede rabljene? //
15	JII	015-0455	ja [gap] narečje drug pogovorni jezik a ne se smo rekli da se ko se prenašajo iz roda v rod da eee se jih jezik spreminjal //
16	JII	017-0468	ja men tud ni sla [gap] mislm pač karkoli se ti bo zgodil vedno ti bo stal ob strani tud če bo pač ne vem v njegovo škodo vedno bo naredu to zate k pač tud vse najboljš a ne in je v bistvu kokr ena tvoja tvoje drugo srce
17	JII	020-0061	torej nacisti so eee nekatere skupine ljudi dali na tisti seznam a ne ki in jih je bilo treba [gap] na nek način odstraniti
18	JII	020-0200	ravno o kristalni noči tu vidimo kako so torej zažigali uničevali židovske trgovine židovske cerkve a ne po na [gap] eee po ukazu Gebelsa to je bil pa Hitlerjev zlo tesn sodelavec a ne ki je bil tudi odgovorn za marsikatero dejanje in dogajanje med drugo svetovno vojno s strani Nemcev //
19	JII	020-0310	navajeni a ne // pogledite še dva otroka zopet starejši mlajšega drži a ne v naročju in pač vidite te starejše žide torej niso mogli it ven iz geta in prav tako tud v geto nihče ni mogu prit pa jim kaj prineset tako da so bli prebivalci tukaj not dejansko na kaj obsojeni //
20	JII	020-0371	še kaj? kaj je še zaskrbljujoče v tem njegovem razmišljanju? ker to je njegovo razmišljanje ki ga v dnevnik piše a ne ja da bi se vrnli //
21	JII	020-0390	taborišča smo so že poznali a ne smo jih že omenjali in sicer so bila to koncentracijska taborišča ki so jih v tem obdobju [gap] širiti in graditi si kr napišimo novo točko //
22	JII	026-0091	ampak eee lej jz niti eno minuto na leto ne gledam te televizije a ne se prav da jz njim doniram vsako leto sto dvaintrideset evrov do zdej a ne // okej //

¹³² JII = javni informativno/izobraževalni, JR = javni/razvedrilni, NN = nejavni/nezasebni, NZ = nejavni/zasebni.

¹³³ Zaradi tehničnih omejitev izvoza konkordanc v orodju noSketchEngine so izjave (z izjemo standardizirane oblike označevalca) zapisane v pogovornem zapisu. Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/zxh5lex>.

23	JII	033-0025	mism da si vsi niso sami krivi nekateri si sicer to želijo nekateri so si tud krivi za nekatere je pač tko da jih je da je življenje jim tko nanese a ne da so se pač eee najdl v taki situaciji pa mogoče nimajo moči da bi se s tega pobrali a ne ///
24	JII	033-0241	jez eee tud k sem bil eee sem mov recimo v zavetišču ponoči čez dan sem bil pa uzunaj ne recimo sem delal v društvu eee vmes k je bil vm [gap] vmesni čs recimo sem pa šel na kosilo in sem si kupu recimo že toplo narjeno eee recimo mineštro karkoli a ne in se nekam usedu in pojedu a ne recimo ///
25	JII	036-0158	ja če si igralc potem tukej ne sme bit problem a ne če pa tega nimaš potem pa bolj da se takga projekta ne ločš lotiš ne
26	JII	051-0215	no tako je tud Puškin verjetno doživljal a ne a veste kako pa drugače ne kot da dve leti ne bosta nikoli minili
27	JII	051-0233	no in zde kaj se je zej dogajalo a ne eee zatekal se je on potem v razmišljanje branje a ne kot b s [gap] bi tudi vi napravili in pa seveda ustvarjanje
28	JII	051-0249	dobro lepo si to prebrala no zde imamo tole lejte da je prevedel to pesem še Oton Župančič to se prav sto let nazaj a ne pa nam ne deluje ta jezik arhaično al nam eem a ste vse razumel? ///
29	JII	051-0552	točno tako a ne točno tako
30	JII	051-0636	da je potem vzbudila v njem ta ta spomin a ne da je moralo bit no ampak dejstvo pa je da je vmes nanjo ///
31	JII	052-0336	skoraj zaščitni znak a ne Mehike ///
32	JII	053-0100	tri elektrone ima a ne no [name:surname] kar sedi bo še kdo pršu h tabli [name:surname] ///
33	JII	053-0247	manjša a ne ta razlika tukaj a ne delta a ne je manjša kot tale razlika ///
34	JII	053-0306	prve ionizacijske energije elementov glavnih skupin periodnega sistema a ne /// ja ///
35	JII	053-0366	[:voice] in njihovo ionizacijsko energijo a ne ki nam lepo kaže da recimo v drugi periodi od litija do neona narašča potem od natrija do a ne spet od natrija do argona spet narašča in tako naprej a ne
36	JII	053-0597	kloridni ion je večji ga takole narišemo a ne /// [pause] ///
37	JII	055-0391	izvrši se dejanje krsta a ne ko krsti
38	JII	056-0965	ne ... eee desna je bolj eee čustvena a ne tko [gap]
39	JII	056-0974	nisi motivirana za učenje a ne [gap] ///
40	JII	056-1077	pravo stvar je vprašal a ne /// [gap] zakaj me nč o fuzbalu ne vprašate ///
41	JII	060-0625	tk da [gap] postopku smo to včrtali v kocko a ne pa je osnovna ploskev kvadrat ///
42	JII	069-0086	pa najbolj nevaren je to d je zjutri ful teme a ne pa tko eee ful avtov je v bistvu ///
43	JII	074-0098	zmaga za en procent v proporcionalnem volilnem sistemu ni resna zmaga a ne pač eee to ne morte govort da je to da je to [:voice] mandat ljudi za neke resne spremembe ///
44	JII	081-0011	znajo delat to kar znajo to se pravi u [gap] usnje a ne in to je neka specifika ki je posebne vrste in se težko prilagajat nekim drugim delo [gap]
45	JII	129-0127	mi smo dolgove eem ti [gap] o tist o čemer se je tolko govorilo a ne že poravnali ///
46	JR	157-0708	to iz tega izhaja a ne to je najin cilj
47	JR	161-0304	[speaker:laughter] sej roke lahko uporabm a ne drugo roko
48	JR	162-0385	sej to sprašujem kerih več nardim dolgih a ne ? ///
49	JR	163-0210	teta volna nam je čist nam je zbila moralo včeraj smo še malo česali eee pol smo probal na različne načine jo jo plect v v nit a ne tud z pomočjo eem s pomočjo sukanca
50	JR	164-2221	je blo prejšn Sladko vince bol a ne ? ... ja ja ///
51	NN	175-0086	kaj se splača kaj se splača je grozn težko rečt a ne /// no to je sigurno ne [:voice] ///
52	NN	175-0090	trden pa stojim za tem de so zavarovalnice vse strašno podobne med sabo da mam cene prbližn enake vem po tem k en drugmu v police kukamo a ne eee da eee mmm funkcioniramo zlo po istih pravilih zlo po istih principih ///
53	NN	175-0105	takrat sem provzaprov pršu k [name:organisation] a ne /// aha v glavnem ... mhm ... mhm ... mhm mhm ///
54	NN	175-0278	mhm /// a ne /// je ta polica dražja eee tko mislm ni dražja sto osemnajst ampak eee ker je znesek premije nižji so nekteri popusti nekej nižji se to sem pa clo nekaj uspela nekaj uspela nekaj uspela to razložit ///
55	NN	175-0314	ampak to bi rad v m [gap] v bistvu v miru naredu a ne /// tako tako ///
56	NN	175-0367	re [gap] zavaruješ zato de se zavaruješ od tam naprej kjer ne morš več sam a ne in eno eee mislm popravilo enga kolesa je nekej kar čist v celoti čist v celoti zmorte a ne de se vam pa en mulc prekucne z z balkona dol ///
57	NN	175-0396	al pa pol ure a ne ja ... ja ja ///
58	NN	175-0411	se pravm so so so take stvari torej ko s [gap] ko spoznaš stranko začneš pol o takmu primeru razmišljat a ne /// ja a [gap] da v glavnem se se se da tudi v takih primerih pomagat no ampak pustimo to za nazaj ///

59	NN	177-0244	ampak je dolgi i a ne nasproti kratkemu i se prav ta razlika je ///
60	NN	177-0391	ja mislm to bom prebrala natančno sej pravm pregledala sem že kr eee mmm pa pa se p [gap] mislm da se mi bo dejansko odpiral a ne ... ja ///
61	NN	179-0385	Dzs k naj ne bi mel nobene zveze z mediji a ne /// a veš [:voice] ///
62	NN	181-0381	mhm tud a ne ?
63	NN	185-0023	res je ja res ja samoza [gap] tko čist sproščen je tko jz sem pač tak in tazga me sprejmite in konc a ne res tist k da nama on sporoča to na ta način a ne /// to se prav da je podoben k jz sem misla da mogoče ... ja ja s tem s tem mammo dejansko kr probleme ne ker ///
64	NN	185-0041	midve že lahko sam problem je v tem da v družbi potem ko se on igra ne on se odloči da bo mel tale svinčnik a ne tud če ga ma nekdo drug recimo a ne ///
65	NN	185-0042	in po so pač tle konflikti s tem ne ... zdej se je začel počasi ne vem kok časa zdele neki za [gap] zadnih štrnast ja n [gap] neki tazga zadnih štrnajst dni a ne aha a se je to že zgodil da da ... [gap] ... aja aha
66	NN	185-0205	jz sem mela en neki časa tak občutek ja tak občutek da eee ja lej sej jst vse zmorem ampak dej mi ti a ne /// da mu ni izziv k da mu ni izziv no na tak način ne ... ja ///
67	NN	185-0287	pa odvisno zej kako pač vidva reagirata na na te situacije a ne na na n [gap] ta njegov jok a ne dej ///
68	NN	185-0349	[gap] ni problema a ne pa tud ko tečemo tlele ko gremo ko gremo na sprehod k tečemo od lučke do lučke a ne [gap] o to je super
69	NN	185-0353	prej s [gap] prej je bil zmer pr men zdej se pa spusti in pol k se njemu pa zdi a ne da je pa treba pa poišče eee
70	NN	185-0359	pa ga morm pol počakat a ne ... po pa [gap] ///
71	NN	189-0347	z veseljem zdaj pa glejte če je karkoli ne pišete sporočite [gap] tako boste lahko a ne ta čas dokler še traja ///
72	NN	203-0048	eem [gap] študentski račun je eee mislm da tud na internetu da piše a ne da so pač stroški vodenja te mesečni a ne so brezplačni ///
73	NN	203-0052	eem ta s [gap] eee študentski paket je samo do eem dokler imate status a ne se pravi vi za odprtje tega študentkega računa potrebujete eee osebni dokument davčno številko pa potrdilo o vpisu a ne in to so te ugodnosti ///
74	NN	203-0071	eee tam kjer pa ni eee valuta evro se pravi na primer na Hrvaškem a ne kjer so kune a ne pa lahko pč na eee bankomatih eee banke ki je v naši skupini [name:organisation] na primer na Hrvaškem je to [name:organisation] ///
75	NN	203-0086	ja eem posojilo a ne podjetjem je eem ponavadi pač na podlagi eee predhodnega a ne poslovanja podjetja ///
76	NN	203-0101	in boste dobival prihodke a ne /// ja ///
77	NZ	229-0098	de tole dol potrgal bomo šle pa na Bled to se prav pozim a ne [audience:laughter] januarja ///
78	NZ	229-0201	je rekva a ne sej vaša ta mlada je pa z Bohina ///
79	NZ	229-0211	eee z [gap] in sem se ze le spomnva a ne de je ... de j poj za to pot tud uhka vedu a ne [speaker:laughter] ///
80	NZ	230-0091	in po je šla mami pucat kupalncu pa je rekla m jz to probala a ne pa s tismu je neki ne vem kva je treba v bistvu k da res pometaš tko da prot seb a veš pa se nekek kupček dela al kva ampak to nej bi zavgal
81	NZ	231-0099	ona četrtega februarja še dela pa j pr ns a ne po pa neki frej
82	NZ	231-0102	sevede sej sej pravm se zdej sam sam ena [gap] ja sej to je važn a ne /// a ne ... da se na razmnožujejo zastopeš ... jz sem ji poj povedala sn reka sva jih pa p [gap] eee ///
83	NZ	231-0376	ne vem v glavnem se ... koker so jo hotel men je to vseen a ne /// po nej pa se ja sej točno ja ///
84	NZ	232-0295	ja no sej tko k tale [name:personal] a ne k je zdej z nami plesala k sem pršla jz zanč na treningu pa no sej tebe pa že poznam [gap] ///
85	NZ	235-0165	in eee takrat je [name:personal] eee v bistvu prvič povedov misim jez sem prvič takat ga vprašala kako je blo p [gap] težko na maturi pa če je že k seb pršu od mature a ne /// k matura je bla ju [gap] že junija ... ja men se je zdel to precej težku ne ///
86	NZ	235-0264	ja mmm ti boš loh k povedov eee o Mensi a ne /// aja spet sem vablen ampak ne smem ///
87	NZ	235-0375	ja ja takrat ja sej s reku de boš dolg živu a ne /// kaj? sto šest let sem reku ... ja
88	NZ	241-0071	eem za diplomsko jez upam da bom dobil eee to k mammo mi mmm praktično izobraževanje a ne pol leta ///
89	NZ	253-0149	ja ni lega ne sej [gap] lega je a ne /// ni ne uspeva ///
90	NZ	266-0184	vi pa bi a ne hrvaško ///
91	NZ	283-0014	ja ne vem samo meni [gap] meni pa je kul v Maribori ne jz ne ne bi šel domu pa če je ne vem kaj a ne /// pa dobro veš te pri tebi je mal drugačik ne ti lej še ti si reden študent jaz pa nekšno delo obveznosti duma ampak nej to pogrešam doma svoje svoje svoje okolje svoje kolege eee ///

Priloge

92	NZ	286-0329	in da ma tud sedež na prevajalstvu in tko naprej a ne k pol lahk dejansko od sam [gap] oddelka za prevajalstvo tud financira sofinancira pa take fore a ne //
93	NZ	286-0466	ja ful mislm pr nas se eee ta prvo [gap] [name:personal] ma pol devetih a ne // ja //
94	NZ	286-0652	ja to je res pa tud če niso neki višji letniki je dost težko dobit tko da delaš na domu mislm k vsej kar se pr agencijah prjavlaš je več al mn eee delo pr njih a ne // pa to mi ni jasn zakaj //

Priloga 7: Vzorčni primeri rabe označevalca *a veš* v korpusu Gos

V tej prilogi navajamo 84 primerov rabe diskurznega označevalca *a veš*, ki so bili identificirani v analizi 100 naključnih konkordanc skupno 839 pojavitev te oblike v celotnem korpusu Gos.¹³⁴

	tip ¹³⁵	id izjave	Konkordanca ¹³⁶
1	JR	157-0493	še hočeš? rada dobr žuraš kaj kaj kaj bi rada kaj bi rada? a veš si ambiciozna punca kokr poznam take punce kt si ti blazno ambiciozne kaj je tvoja naslednja ambicija? //
2	JR	157-0802	ne ne eee Šp [gap] kako [name:personal] gleda na to na to da na ta na ta eee a veš misim //
3	JR	157-0958	[gap] material a veš to k se posta [gap] //
4	JR	157-1016	ja kok deleč ga pol spustiš? // a veš a si upaš it do njega kako to zgleda kaj rečeš preden se ... ja ja ja //
5	JR	159-1528	jah no sej o tem pa govorim a veš pa cigare kadi hohštapler //
6	JR	159-1557	tu si to je tu si rejver ampak to je taka ime [gap] ume [gap] umetniška bolj in ta [gap] tako da je mal ja a veš ja tko da se ne ve a veš //
7	NN	176-0503	da [gap] a veš okej se strinjam //
8	NN	176-0585	ampak to je naji [gap] ne to je najin nekaj dogovor z [name:personal] zato ker potem raje pr kekšnih urah ali pa eee tko a veš kekšen prost dan pa to //
9	NN	176-0602	ja mislm jz vse kar ho [gap] hočem reč je tko ne sej najbrž to delate tud zarad logike kda [gap] kdaj uni ljudje pridejo najlaž [gap] najlažje ne em ka sam mislim po drugi strani se mi zdi pa da blo to bl pač bl smiselno met ne vem v četrtek dopoldne ka mene briga a veš // ne m [gap] ne mormo ne m [gap] ja ja //
10	NN	176-1030	ne se so taki domovi so še v Šentvidu [gap] a veš [gap]
11	NN	177-0051	[gap] ... mhm mhm mhm // a veš ja [speaker:laughter] ti si v strahotno visoki kategoriji ampak še vseen je to super vem zakaj ne //
12	NN	177-0064	jst upam da je to dovolj eem za en doktorski predmet a veš če bo pa treba bomo pa napisale eno stran literature ki si jo ti mogla eem prebret //
13	NN	177-0068	mhm mhm // a veš // ja ja ja
14	NN	177-0213	ja // a veš j [gap] jz jih ne slišm a ti slišiš razliko če rečeš nit pa nit pa rečeš eno je takani al kako so že te finte a veš //
15	NN	177-0279	dobr pol je pa že označevanje itak ne // a veš to boš itak mogla prej al slej prebret tko da loh si še ští [gap] eee ... dva do štir označiš ne //
16	NN	177-0367	nnn nimaš pa s tem zvočnih posnetkov a veš nimaš li [gap] polinkan na zvočne posnetke ne //
17	NN	178-0159	ja sam v tem smislu je to še zmeri s [gap] spontani govor a veš tud če ni spontana situacija //
18	NN	178-0265	ja // a veš k ti te eee mmm če se ti kej prtožuješ ti povejo da se vaš pogovor snema zaradi bol [gap] zagotavljanja boljše vatever //
19	NN	178-0419	ah sevede ne mislm lej a veš zde je blo tk [gap] een tko je blo no mi smo na začetku vedl da ne mormo plač [gap] plačvet tega tko kot princip a ne
20	NN	178-0690	čez trinajst dni pet ur in dve [gap] [speaker:laughter] ... pa vseen nekje oš mogla sedet tiste dneve a veš [gap] aja pa č [gap] pa čist paralizirana //
21	NN	179-0605	ampak še vseen ne a veš tukej podobne zadeve recimo maš pri maš [name:surname] ne
22	NN	179-0652	ne ker ponekod potrebuješ peeset a veš tko kot recimo v parlamentu //

¹³⁴ JII = javni informativno/izobraževalni, JR = javni/razvedrilni, NN = nejavni/nezasebni, NZ = nejavni/zasebni. Zaradi tehničnih omejitev izvoza konkordanc v orodju noSketchEngine v primerih manjkajo oznake za konec izjave in/ali vloge.

¹³⁵ JII = javni informativno/izobraževalni, JR = javni/razvedrilni, NN = nejavni/nezasebni, NZ = nejavni/zasebni.

¹³⁶ Zaradi tehničnih omejitev izvoza konkordanc v orodju noSketchEngine so izjave (z izjemo standardiziranega oblike označevalca) zapisane v pogovornem zapisu. Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/j9zp7m7>.

23	NN	179-0670	odločata mimo tega a veš to so te pol pomembne tud stvari ne
24	NN	181-0097	sej se prilagajaš ampak tko a veš mislm s [gap]
25	NN	181-0682	to ne vem a veš ker e [gap] dva razreda a ne
26	NN	182-0121	lepe sprehajalne sam a veš namest da bi bla trole naravnost ma take mini stopničke tak tak tak tak tak tak
27	NN	182-0225	ne več najbrž je ostov tak kt je bil a veš takrat k sem bla deset stara in po so me mal postaral [all:laughter]
28	NN	182-0230	ja a veš puberteta je huda zadeva
29	NN	182-0492	sem rekla ja pa sej je men tud sej men je to tud ful dobr jz bi tud če b ga mela tam vsak dan ga naskočila še dvakrat če bi mela cajt a veš ne
30	NN	182-0495	sn rekla pa še mal se pretelovadeš pa to a veš to je pa ja super in pol se začne po tej steni dž dž dž a veš ful našigat ne
31	NN	182-0562	odzuni sem odzuni mam sam še fasado za prebarvat a veš pa tud sem tlakovce lan na nov po celem dvorišču dala kurac pa so se tam nek pogreznal pa tam nek je bula pa t [gap] zdej a veš
32	NN	182-0587	sn rekla pa kaj je to men kamorkoli pridem voda nagaja a veš edin doma mi ne no
33	NN	201-0126	aha teste a veš [speaker:laughter]
34	NN	201-0146	[audience:laughter] okej ti borm a veš eee ene liste ti morm pol dt d izpolneš
35	NZ	225-0471	[audience:laughter] in tud tko a veš ne mal več pove kaj je kej familija uno tretjo včasih ni nč kej velik [:voice] no ne vem
36	NZ	225-0477	e če zdej sem hotu it z njim je bla njegova cura pa ne vem mama pa še ena njegova sorodna pa [gap] hotu it zato ker je bla mama a veš čuj [name:personal]
37	NZ	226-0671	ampak ja važno je to ne da maš zdej prej smo ble s ta malo skupi v sobi ta ma [gap] čist ta mala je bla z njima v sobi ne pa ni blo nobenga takga p [gap] placa kamr bi se lah udmaknu a veš ne da b šov nekam ne vem ko rabiš mir pa greš ne vem v sobo se zapreš pa to je to ne ... aha prej prej je blo to [gap] sedeminpetdeset
38	NZ	229-0214	ja sam mmm tko je a veš Bled je ful turističen ful bl turističen je
39	NZ	229-0229	pa tm al mi k smo bli a veš dvanaj [gap] ja dvanajzga aprila pol
40	NZ	230-0060	ampak ne vem nekek je tko blo naštiman da je da je ne vem da je neki prtisnla na knof al kva pa se je kr tko dež a veš kokr da je to neki tropskega k so ble kr tko ploščice po tleh da je poj to odtekal a veš pa da se kr use zalil pa tko
41	NZ	230-0228	ponvad čokolešnik al pa une zej majo tud u šparu sem kupla ene k nisem šla v Merkator po tiste čokoladne kuglce ampak sm šla v špar pa so eni tisti krivi čolnički a veš so tud taki čokoladni
42	NZ	231-0223	po sem še tisto žirafa a veš k je mela nalepleno za eee da smo jo merl kok je velka je bla vsa putrgana že
43	NZ	231-0395	po moje de ji je [name:personal] tud reku a veš d nej ne opleta tok pa kva vse narobe
44	NZ	232-0517	ja sem gledala najprej sam tle a veš tm k so f [gap] eem te
45	NZ	233-0005	ja a veš vsak dan ista zgodba
46	NZ	239-0207	po pa človk lih eem z b [gap] torej v živlejn ne premišluješ takih sej veš tist k ja un na slone mislu a veš tist?
47	NZ	239-0313	eee skor ne eee veš zakva je eee se mi zdi eee ker jz nisem znov francosk a veš in sem se mogu [gap] ... čez mejo sva šla ja ja ja ja ja ja ... ja ja po sva šla najbrž res skupi
48	NZ	239-0476	no po sva šla pa a veš po sva bla še skupi po sva se nekaj alternativno eee alternativno eee dobivala eee nisva čist skoz štopala sama
49	NZ	240-0450	sej je ziher [name:personal] [name:personal] povedov sej je un je bl taku a veš mmm
50	NZ	241-0018	po pa tko nak [gap] naenkat pokaže sam tko une une tko čevle a veš od vokalista une bele bosanske a veš [all:laughter] tko z visoko konico pa tko počas tko prot njega gor videt tko vsi v črnem a veš un pa en tak bel b [gap] bel Balkanc tam pa začne gor tult a veš [all:laughter]
51	NZ	241-0018	po pa tko nak [gap] naenkat pokaže sam tko une une tko čevle a veš od vokalista une bele bosanske a veš [all:laughter] tko z visoko konico pa tko počas tko prot njega gor videt tko vsi v črnem a veš un pa en tak bel b [gap] bel Balkanc tam pa začne gor tult a veš [all:laughter]
52	NZ	243-0252	je reku dej bi za promocijski koncert cedeja pa b radi naem ka de de de vse špila pa de [gap] jez sem reku kva ni problema se bomo zmenil za en pir pa za en pa za benz mivste dal pa tu je tu a veš [:voice]
53	NZ	243-0277	in tko sem ob zaveso zapel k je bla napeta če se a veš da so ozadje naredl ne
54	NZ	243-0359	pa pol po Zizijih na na žurko takoj ... na na žurko k so mel pa pr Grosupljem Vrhovci žurko a veš na zize [all:laughter] ... po Zizijih še na zize
55	NZ	243-0398	sem reku jez bom lepo počas spizdu pa to a veš sem se v avto usedu sem zakurblov k me je zebl
56	NZ	245-0517	mislm fak dej no a veš ja k te zbudi ne k je tok ži [gap] živčna ... ja ja

Priloge

57	NZ	245-0680	to j ta to j ta scena za Nemce a veš // mhm //
58	NZ	245-0892	to j to // a veš tko pač //
59	NZ	258-0022	ampak stari tu je blu taku unu jez sem mene je blo kar strah razumeš vmes unu ubi ubi Srbina a veš uno tisto ful taki jz tist lajk jo unu //
60	NZ	258-0305	ja ne k tiste vaje k smo mel je blu ful unu [gap] a veš // ja //
61	NZ	259-0064	ne eem a veš en mesec nazaj sem bil v Nemčiji //
62	NZ	260-0018	sao baje ni zdravo baje sploh ne bi smeli mleka pit a veš // mmm ja ja ja //
63	NZ	264-0243	ja pa uno z morskimi sadeži a eš [gap] jz ful ne jem morskih sadežov in so tk nadevali take konkretne a veš // mhm //
64	NZ	265-0259	a a veš sej prajim jz ti ne morm poveet //
65	NZ	266-0220	se tak dela eni se znajo tak na [gap] na [gap] [gap] dela [speaker:laughter] // a veš [gap] sej te mam rada vseen [speaker:laughter] //
66	NZ	268-0027	ja sej glih tk je tiska ka [gap] tist eee spis mu je rekla ga da ni naredu ne po mu je pa tk rekla ja lahk nekok lah se prieš z dreka vn potegnati oziroma ne na tak grd ampak tak podobn a veš // mhm //
67	NZ	268-0046	ja to je ne sam ka pa še karkoli drugiga a veš // mhm //
68	NZ	268-0052	takrat sn se js rodila ne ni takrat Cankar Hlapcov pisu a veš // [pause] //
69	NZ	277-0437	ja sej jz tud ne vem ampak a veš zdej bo vedla da je njen etbra derpe //
70	NZ	278-0205	ja s [gap] [speaker:laughter] sam pomaga a veš v takih mrzlih dnevih tko tko eem drži te pr življenju no [audience:laughter] k že misleš da boš zmrznu [speaker:laughter] //
71	NZ	280-0030	eee ti in jest čist v dobri vjeri dejta mi mav za pit a veš pa me zebe sem šla se obelč in kostanj lepu lupmo tam k eee [name:personal] ga pa peče //
72	NZ	280-0226	pol v skrino sem pa dala a veš // ja pa a tu stoji tu //
73	NZ	280-0233	rastejo rastejo sam treba jih je nabirat a veš // aja ... ja ja ja //
74	NZ	280-0345	glih dns sem študirala sem šla po stopnicah kuk cajt je še mama provzaprov živela a veš // ja vejž de //
75	NZ	280-0422	v dnevni sob oziroma v hiš k prav ma nč ... sama bla ona je bla v ta star hiš a veš uni pa v ta novi ... prav //
76	NZ	280-0470	tu jest ne morem pozabet sej pravm eee zmer sem rekla pr nabenmu ne b bla rada pr smrti ti pa sem bla pr očet pr mami in pr teti [name:personal] ne morem pozabit tega a veš // ja no vidiš //
77	NZ	280-0599	ma sej je tok trapast narjen tu a veš sej si misleš
78	NZ	280-0677	je re [gap] po mi je pa rekla jaj je rekla veš tuk me sekirajo kdaj bom eee z orehi to ona prodaja // a veš ... ja tu majo vse naročen recimo eee tale eee [name:surname] pa vse tu za eee čaki za kuga žje //
79	NZ	280-0833	[gap] o ja pride zdej so mel osensetletnco a veš // ja ja //
80	NZ	281-0034	ja dobr sem a veš zdej spet da se mečeš v stroške boš dal petsto evrov pismo za tist ruter klinčev pol je //
81	NZ	281-0355	pa s na istem a veš // ja ja //
82	NZ	281-0499	[pause] // a veš ni mus da bo ni mus da je kej pametnga blo povedano smo pač //
83	NZ	282-0291	ne tes [gap] ja onga je klical a veš ja testo od [name:personal] prfoksa //
84	NZ	287-0379	mmm bom ja // a veš // mhm //

Priloga 8: Vzorčni primeri rabe označevalca *ne vem* v korpusu Gos

V tej prilogi navajamo 51 primerov rabe diskurznega označevalca *ne vem*, ki so bili identificirani v analizi 100 naključnih konkordanc skupno 2.743 pojavitev te oblike v celotnem korpusu Gos.

	tip ¹³⁷	id izjave	Konkordanca ¹³⁸
1	JII	007-0216	recimo za primerjavo vseh zvezd v prejšnjem stoletju na ta način pomerjenih je blo tam velikost ne vem štrdeset tisoč zdele smo trenutno eee ja na tem računalniku jih je eee neki čez štiristo tisoč ne
2	JII	030-0075	zakaj štirideset metrov zakaj ne ne vem dvesto al pa šestdeset? ///
3	JII	033-0187	eee ste mi pa ne vem [name:personal] ste mi vi rekli al e [gap] en od vaju je reku da je pa sicer na eee božični večer ljubljana zoprna ker se trgovine zaprejo ker ljudi ni ker eee nekaj se sprazni ne ker vsi grejo domov a ne? ///
4	JII	056-0973	nisem motivirana glih [gap] ... ne mislm ne vem /// nisi motivirana za učenje a ne [gap] ///
5	JR	137-0058	a? ... ne vem so in bodo ne ... kot vedno ///
6	JR	142-0263	ne glih neki na v [gap] za Val dvesto dva [name:personal] ne vem dej boš mogu pogledat [speaker:laughter] ///
7	JR	143-0199	pa eee ne vem v neki mogoče da obstaja kje en kraj ki je lepši od tega ne
8	JR	163-0013	eee za hlapca sem pa izbral osebo ki pač ma eem veliko življenjskih izkušenj eee pa mogoč tud eem ne vem mogoče na nek način tud sam sluti da pač bo da pač bo postal hlapec ta tedn ane
9	JR	164-1838	ja ne vem /// najbolj ///
10	NN	172-0148	eni so rekli ja jz nimam nobenga problema pa kej čem pisat tku /// ne vem jz bi šla eee za prvi nivo je važnu sej ostalu bos [gap] bojo pol ///
11	NN	174-0051	ne zdej edino ta [gap] takrat ko ste se vidve nekej dogovarjale ne vem al potem je o [gap] mmm ta verzija obdržala s temi eee različnimi eee večjezičnimi? ///
12	NN	174-0071	aha ma pol [gap] čim krajše tem boljše ja ... ker mene je že to motlo da ni recimo vvvv ne pika ne vem neki ne da je že ta http me je moto ampak tko je ///
13	NN	176-0747	ker tko neko servisiranje nečesa pa tud [:voice] res nima smisla ne vem pol t [gap] eee to pomen da se pač [gap] z [name:personal] to še pomenmo seveda kaj zdej to je pa da se pač aja ona ve bl kaj bomo zej mi to misim z [gap] loh tud iz naklonjenosti neki delamo ampak res ne vem naklonjenosti do česa ///
14	NN	176-1304	jah sej vemo kako je čez ne vem junija nam bodo začel težiti z ministrstva verjetno ne vem /// jz nisem nič foruma brala ///
15	NN	181-0601	k majo tisto igrišče k so mel otvoritu se mi je zdev dost mirno no ne vem /// sej je ///
16	NN	185-0302	zej če vidva ne vem da je enmu res ne vem kaj se je zgodil ne potem ja drgač pa ponavad ne vem al ignorirava al pa ///
17	NN	189-0354	absolutno ne pričakujem zdaj tukaj povsem čistega jezika ne tako nimamo vsi ampak samo to da ko stojiš pred š [gap] [gap] tudenti da vam ni vseeno da kr rečemo sevea nea ne vem te besede ampak da skušaš glede na socialno okolje malo dvigniti svoj jezik se potruditi ///
18	NN	192-0132	je kul neka ne vem nekaj dogaja ne ///
19	NN	192-0543	okej ne hecam se misim pač ne vem meni se je že kdaj sprožila želja po očetovstvu in to a ne ///
20	NN	201-0083	ja ne vem men eee mism [name:personal] je men reku da enih par v bistvu je blo kr predavatelj k ne morjo [gap] ///
21	NN	205-0181	eee bi tam pomalcali nekej in pol pol ne [:voice] ja k drgač če se začne ne vem ob dveh šele je pol [:voice] ///
22	NN	216-0211	in s [gap] tud smo pazli da recimo pozimi že s temi klasičnimi kolektorji na sončn dan se je voda zagrela na ne vem dvajset stopinj al pa trenutno je se je zagrela celo na ne vem trideset petintrideset ///
23	NN	216-0233	ja [gap] ... ja razumem ja /// ne vem z biomaso je tako biomaso imate vi ze tri vrste bi reku tri vrste al mate klasična polena ///
24	NZ	221-0222	kjere tu? /// ne vem lani sem gledla k jih je blu pač une so igrale tiste [name:personal] in teve ku zmjerm nje in puo ene par jih je menavla ///

¹³⁷ JII = javni informativno/izobraževalni, JR = javni/razvedrilni, NN = nejavni/nezasebni, NZ = nejavni/zasebni.

¹³⁸ Zaradi tehničnih omejitev izvoza konkordanc v orodju noSketchEngine so izjave (z izjemo standardizirane oblike označevalca) zapisane v pogovornem zapisu. Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/jrdvrlly>.

25	NZ	222-0197	tko da bi bli ne vem po dva v sobi ne //
26	NZ	224-0245	ja sej pravm de [gap] // ne vem se mi zdi de sta de je [name:personal] dol pr [name:surname] //
27	NZ	225-0089	zdejle sem mel tud srečo [gap] pet nas je blo pa sem mel ne vem menda čip lider sn bil pa dvignem [gap] z [gap] na čuka osem po pa me en k me je mel čuka majn od mene
28	NZ	226-0671	ampak ja važno je to ne da maš zdej prej smo ble s ta malo skupi v sobi ta ma [gap] čist ta mala je bla z njima v sobi ne pa ni blo nobenga takga p [gap] placa kamr bi se lah udmaknu a veš ne da b šov nekam ne vem ko rabiš mir pa greš ne vem v sobo se zapreš pa to je to ne ... aha prej prej je blo to [gap] sedeminpetdeset
29	NZ	236-0161	je rekla [name:personal] da je bla da so mel ne vem pred kratkim en žur pr enmu tipu tm iz Slovenj Gradca al v okolici //
30	NZ	241-0182	eem tisti senzori pa dajo na izhodu ne dajo linearen signal ampak je ne vem pač nelinearen a ne
31	NZ	241-0328	potem da preračuna okolco da ne bo ne vem ob ograjo udaru in tko naprej ne
32	NZ	243-0161	mu pa reč [:voice] pol da da enostavno ne vem ne vem kva kva če maš pol reč ne mism //
33	NZ	245-0020	a veš če mam ne vem trinajst predavanj morm jst okej pač šest pa pol al pač šest al sedem iz [gap] verjeten sedem izdelkov narest ne //
34	NZ	245-0069	pač itak una [incident] ga j psihirala na polno in eee ne vem pač jst sem tko sklenu kompromis //
35	NZ	247-0084	kr to je zdaj tak majo točno stvari ja zato da ti u [gap] umestijo hišo na eee v prostor ne vem je šeststo evrov ne pol pa maš ne vem petdeset evrov to soglasje pa petdeset evrov ono pa tretjo odvisno ne kaj na tvojem področju potrebuješ
36	NZ	247-0602	pol lahko interakcije ne ne vem // ja
37	NZ	249-0094	Industrijska kovinarska šola poklicna šola tu san se ja našo tu je avtomehanik nam je povedo en da je to na da je to na eee kod nije bila borza takrat takrat se je reklo drugač ne vem eee tam je bio en moški in oni su poslali razpis in da pridejo
38	NZ	255-0428	se lahko poročajo pa to otroke pa to ne vem // jah ja to nega kaj se to samo eee katoliška vera nema //
39	NZ	258-0489	in ma svojo sobco isto pač tko ful pocen ne vem sto pa še neki evrov [gap] da //
40	NZ	259-0147	ful izgrajen ne vem zadnjič ko sem šel pred enim letom eee ne vem ni bilo vsega tega ne [speaker:laughter] [gap] //
41	NZ	259-0291	a kaj misliš eee mmm kak nate gledajo punce kot eee južnjaka ali a mmm ne vem ali majo e [gap] etikete pa ne vem [gap] ... še //
42	NZ	262-0250	oja [speaker:laughter] ne no saj kaj ma saj ta diploma saj po mojem bo ena sošolka moja ne vem pri [gap] pač sedem nas je
43	NZ	262-0314	ja ne vem v glav ne saj ne saj po sem [name:personal] napisala ne ker pač je napisu da je s Pivke pol sem misla okej [name:personal] [gap] dobr rečmo ne //
44	NZ	266-0281	pa če pride zdej ne vem po krstu v Svetem Duhu ona še na seminar
45	NZ	268-0010	pa angleščino s štiri pa ne vem [gap] fajn //
46	NZ	271-0053	ja ko m [gap] ne vem več možnosti mam še pol odprtih če se ne prepišem no //
47	NZ	281-0094	tko da mogoče ni tok sleb ne sm ne vem // ja jst bom tkule reku jst mam dvejstga sestank za animatorij u Nov mest
48	NZ	284-0031	nej pa ne vem mogoče kakšna reakcija ka je ceplen je bil prejšnji keden če je kaj tou krivo ali //
49	NZ	285-0461	[pause] // ne vem neki je aja
50	NZ	286-0334	ja pač ne vem sej sam oddelk ma nek svoj fo [gap] a ma tud svoj fond a ne in bi pol [gap] financiral al bi cel [gap] drekt drekt to šlo sej d [gap] mislm s filozofske fakultete nism zihr kk to deluje zdej k eee se nisem poglablal še v to a ne //
51	NZ	286-0452	to je ful smešn ne vem mislim ko to pa je res en izmed teh izpitov ko ga greš na prvi rok že //

Priloga 10: Vzorčeni primeri rabe označevalca *a ne* v korpusu Janes

V tej prilogi navajamo primere rabe diskurznega označevalca *a ne*, ki so bili identificirani v analizi 100 naključnih konkordanc izmed skupno 2.894 pojavitev te oblike v podkorpusu komentarjev na bloge (68 primerov), 1.190 pojavitev te oblike v podkorpusu blogovskih zapisov (25 primerov), 7.435 pojavitev te oblike v podkorpusu forumov (69 primerov), 24.747 pojavitev te oblike v podkorpusu tvitov (71 primerov), 206 pojavitev te oblike v podkorpusu pogovornih strani wikipedije (29 primerov) in 2.336 pojavitev te oblike v podkorpusu komentarjev na novice (66 primerov).

Komentarji na bloge ¹³⁹		
	ID dokumenta	Konkordanca
1	blog.publishwall.1991.15	Naj dajo pa zastoj, a ne ?
2	blog.publishwall.2733.15	morda malo čudna primerjava, drži pa kot pribito, a ne ?
3	blog.publishwall.4596.6	potem pa mora imeti prav, ane ?
4	blog.publishwall.7215.13	Težava je v tem, da se vedemo tako, kot da je nimamo! A ne ??
5	blog.publishwall.13342.2	(z bolečinami, a v bolnici brez njih ne gre, a ne ?
6	blog.publishwall.17659.12	Ja jaz, a ne .
7	blog.rtvsl.560.1	janković je čist drgač, on je nedolžen, dokler se mu krivda ne dokaže. ane .
8	blog.rtvsl.819.11	ricardo_kaka22, saj se hecaš, ane :)
9	blog.rtvsl.988.1	Kajti le to legalizira samoljubje, a ne ?
10	blog.rtvsl.1098.1	kaj vse se lahko spiše, ko se včasih pozabi temeljno , da je človek pač po naravi tak, da teži in je usmerjen k preseganju in presegajočem... a ne ..
11	blog.rtvsl.1130.12	igrala odprto proti barci... a ne ??..
12	blog.rtvsl.1166.3	Saj si še mlada. A ne ?
13	blog.rtvsl.1875.47	ali pa kake analize konsekvenc kakih izhodišč ali pojmovnih izrekanj.. a ne ..
14	blog.rtvsl.2611.15	sej nisi ljubosumen, a ne ?
15	blog.rtvsl.2906.29	Tako kot vsi ostali, ane .
16	blog.rtvsl.3046.12	", se dere lopov, a ne Konsidere!!!
17	blog.rtvsl.3088.25	Vredno je včasih osvežiti kaj lepega, a ne .
18	blog.rtvsl.3886.21	Ukrepali bomo pa že vsak po svojih močeh, a ne ?
19	blog.rtvsl.4066.43	Pa saj nismo otroci ali neumni ane .
20	blog.rtvsl.4491.15	Imamo možnost izbire, a ne ?
21	blog.rtvsl.4775.20	ponavadi to dvojje mešate, pač kakor vam odgovarja... a ne ??
22	blog.rtvsl.6078.126	Pa Avstrici tudi, a ne, ki so krampasti kot Švabi, pa Italjani a ne , ki v vsaki dobri mrhi iščejo svojo mamico.
23	blog.rtvsl.6418.16	Vidiš, pri tej zelenjavi se vsaj vidi, da je domača, da ni gola :-))) Uživaš, ko nabiraš, a ne ?
24	blog.rtvsl.6913.21	Se priporočam :) nish, čudno, a ne , da nam ni vedno do smeha...

¹³⁹ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/jy58cwu>.

25	blog.rtvsllo.7597.28	Tudi, če bi pisala do konca leta, bi lahko šel mimo, a ne ?
26	blog.rtvsllo.7617.28	res je že čas! ¶ A ne Kouvrn - takole bi morali skupaj držati, da bi pri čem uspeli.
27	blog.rtvsllo.8450.8	Saj je bila, a ne ?
28	blog.rtvsllo.8480.3	1. a ne no?
29	blog.rtvsllo.8502.49	net revolucionarji, borci, vstajniki.. a ne ?
30	blog.rtvsllo.9085.14	Ne pozabimo, da so atomsko orožje prvi in zadnji uporabili prav Američani, ki jim eni požete hvalice, a ne Tonček?
31	blog.rtvsllo.9398.5	@jakec taprvi in @catelyn , fajn se je kdaj malo vrniti po že prehojenih poteh, a ne ?
32	blog.rtvsllo.9466.15	naj raja uživa prej poletje, ane .
33	blog.rtvsllo.9515.4	Na splošno se na babe dobro spozna, ane andromaha.
34	blog.rtvsllo.10404.33	Iz "sosedovih" se pa tako lepše sliši a ne ?
35	blog.rtvsllo.11423.25	Čebela, ti bi peteršilj, ane ?
36	blog.rtvsllo.11453.19	hi hi hi no vsi imamo takšne al mal drgačne spomine ane ;) ;) nice zapisan ;)
37	blog.rtvsllo.12742.11	ker je resničen, ane ...
38	blog.rtvsllo.13781.60	Prej ali slej jo bomo pili kjerkoli se nam zahoče. A ne Komsomolec?
39	blog.rtvsllo.13793.8	Fajn ane :)
40	blog.rtvsllo.13935.45	eni se zgleda že pripravljajo na novo vojno.. a ne ??..
41	blog.rtvsllo.13953.19	ti gre ane ?
42	blog.rtvsllo.14313.16	Saj se vsaka nauči, a ne ?
43	blog.rtvsllo.14484.14	Najhuje, ane ;) se je ponoči obleči in iti po cigarete...
44	blog.rtvsllo.14788.2	Ampak - še smo in še bomo kaj doživeli, a ne ?
45	blog.rtvsllo.15358.27	Pomembno je, da je strokovnjakinji v laganju SMEJKI všeč, a ne ?
46	blog.rtvsllo.15432.20	včas sam seboj ni tko slabo bit , glasba pa je to za dušco ane ;)
47	blog.rtvsllo.15772.7	spodbudno... a ne ?...
48	blog.rtvsllo.16152.13	Da je podmazan, ane Kobilček :))) Javaša, isto, le v manjših količinah.
49	blog.rtvsllo.16425.4	Alenčica ane !
50	blog.rtvsllo.16805.3	ko ga daš ven sam..... a ne
51	blog.rtvsllo.16895.17	Ojdip je konstrukt ne fakt, to ve vsak otrok, želja ni retrogardna, marveč je parcialna, zaradi sistema nenehnih restrikcij in dominacij - saj si bral Deleuza, a ne ?
52	blog.rtvsllo.17318.5	V tem je razlika med mano in tabo, a ne ?
53	blog.rtvsllo.17706.2	pa je dobr, ane .
54	blog.rtvsllo.18246.22	@Kobilica, v bistvu to ni najslabše, a ne ?
55	blog.rtvsllo.18453.3	Andromaha, je pa lepa. Ane ?
56	blog.rtvsllo.18621.14	Ja, in vedno bo tako , a ne ?!
57	blog.rtvsllo.19565.25	saj tudi moški in ženska sta povsem dva različna pojma, a ne ?
58	blog.rtvsllo.19908.156	Boli a ne ?
59	blog.rtvsllo.20436.10	OK - je prsitranskost, a ne ?
60	blog.rtvsllo.20552.24	Kouvrn - a ne , kako hitro vse mine ?
61	blog.rtvsllo.20884.56	ja, mesič je ukazal napad na hrvatijo.. a ne ?
62	blog.rtvsllo.21575.7	Vsaj lepi spomini nas pokonci lahko še držijo, ane :)
63	blog.rtvsllo.22297.7	:)) ¶ Ane , Joži, kako je svetla, velika tale večer...
64	blog.rtvsllo.22334.117	boš verjel tudi to, ane .
65	blog.rtvsllo.22789.16	Važno je, da prtljago pustiš doma in greš bosa dalje :) ¶ Ane , Zzzzzz-jevka?!

Priloge

66	blog.rtvsllo.22859.2	ja, sej' potem smo pa na istem, kot v komunizmu, ane ...
67	blog.rtvsllo.22884.46	lastni slavi ne gre gledati v zobe, a ne !!
68	blog.rtvsllo.22910.15	@Borut, zanimivo, a ne !

Blogi ¹⁴⁰		
	ID dokumenta	Konkordanca
1	blog.publishwall.4395.0	Ja , mame prenesejo vse , ane ?
2	blog.publishwall.4553.0	Kapirate , ane ?
3	blog.publishwall.7275.0	Saj , konec koncev , kdaj lahko človek pa tudi malo popije , a ne !?
4	blog.publishwall.8910.0	noro a ne ?
5	blog.publishwall.12510.0	» Sama sreča , a ne « ?
6	blog.publishwall.12513.0	Sem že govorila o BRCA , a ne ?
7	blog.publishwall.16412.0	Naj si vsak misli pač svoje , a ne .
8	blog.publishwall.16412.0	Jebi ga , a ne .
9	blog.rtvsllo.264.0	ampak naslednjič bo pa sin , a ne .
10	blog.rtvsllo.5627.0	Pospraviti karkoli čudno se vam zdi , morda tudi komu " krivica " se godi , ane " tastari " spet teži .
11	blog.rtvsllo.6501.0	) ki bo ustrezala vsem (vsakemu delček , ane ?
12	blog.rtvsllo.8161.0	Ona , ki si želi stanovitnega partnerja , ki bi jo ljubil , bi z malo truda dobila sanjskega moškega . A ne , njo zanimajo kratkoročni , hitro dosegljivi cilji .
13	blog.rtvsllo.8198.0	Moj prijatelj g. Peterle , a ne ?
14	blog.rtvsllo.12300.0	Če te ne bi ljubil , bi že odšel , a ne ?
15	blog.rtvsllo.12527.0	Prelisčili smo paznika s poznim nočnim poljubom (Monsanto ne more pregnati ljubezni , a ne !
16	blog.rtvsllo.16067.0	Pa kozlička za potenco , pa jagnje za izpričevanje nedolžnosti , pa odojka , da ne bo zrasel v svinjo , ki bi delala z nami kot z » namiljenim mehurjem « (karkoli to že je , a ne , g. Štuhec ?
17	blog.rtvsllo.16386.0	Sem nora kot vedno , a ne ?
18	blog.rtvsllo.19843.0	" Vi , gospod boste prihajali po sina ob petkih ob šestnajsti uri in ga pripeljali nazaj v nedeljo ob šestnajsti uri , gospa pa ima skrbništvo , a ne ?
19	blog.rtvsllo.20126.0	To , da si » gazda « , ki lahko odloča , baje človeka notranje bogati , a ne ?
20	blog.rtvsllo.20486.0	Kobilica , ti si pa malo poreden ta vikend a ne .
21	blog.rtvsllo.20610.0	Ja sej ga mamo vsi a ne , samo eni ga ne matrajo pa obremenjujejo z ljubeznijo , pa srečo , pa veseljem pa žalostjo .
22	blog.rtvsllo.22345.0	Ti je grozno , a ne .
23	blog.rtvsllo.22687.0	Najbrž pluralizem , pa kvote pa demokracija pa te stvari a ne .
24	blog.rtvsllo.22914.0	Doba bo že vedel , za katero vrsto gre . A ne , Doba .
25	blog.rtvsllo.22948.0	ja babah , a ne !

Forumi ¹⁴¹		
	ID dokumenta	Konkordanca
1	forum.medovernet.9686738	potem je bolje , da pred tekom spijem kvečjemu kak z glukozo bogat gel ali napitek in ne burnerjev , saj ti ravno z glikogenom varčujejo a ne ?
2	forum.medovernet.7305059	registraj se pa boš lahko pošiljala a ne .

¹⁴⁰ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/hq2l8xy>.

¹⁴¹ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/z6v6v87>.

Priloge

3	forum.medovernet.6570089	Yo je morda za v vrtec , ampak težave z živalmi imate pa vi , a ne ?
4	forum.medovernet.6206480	Mislim , se pa ne pusti motit :) Pa tudi ful hitro napreduje , a ne ?
5	forum.medovernet.6528260	Ve ste pa že skoraj vse v službah , ane ??
6	forum.medovernet.9847414	sem stara pet let več , kot leta 2009 , ane .
7	forum.medovernet.9924225	Neko ravnotežje je treba loviti , a ne .
8	forum.medovernet.9937650	Ti si gotovo " precednik " kakšne velike mednarodne korporacije , a ne ?
9	forum.medovernet.6831148	Jst mu pač nisem hotela neki ' skrivat ' , da mi je všeč , čeprav je bilo več kot očitno , ampak tako je , da mogoče njemu pa ni bilo ane .
10	forum.medovernet.5802508	Ženske smo si različne , a ne :))
11	forum.avtomobilizem.2.38134.842060	Sej se nas zbere določeno število in naj da tip ponudbo a ne ?
12	forum.avtomobilizem.4.44073.794342	Ja bomo pobe skup nabral pa bomo pršli ane Zmeden :P
13	forum.avtomobilizem.25.3358.350591	ej folk neki pa mam s čimer se lahko pohvalm ane
14	forum.avtomobilizem.37.55073.1588922	je pa tko ane : če ne znaš nič s telefonom počet , kupi iOS (iphone)
15	forum.avtomobilizem.4.82245.1407068	Sej zato pa se tako dobro pozna a ne ¶ lp
16	forum.avtomobilizem.33.83896.1417869	Še je veliko , sam komentiramo pa lahko a ne .
17	forum.avtomobilizem.7.62474.1177726	ja pa pr mašini se še da velik nardit ane ¶ jst pa mislm , da že cel vikend razmišlaš o mk5 gti-ju ...
18	forum.avtomobilizem.7.68729.1397889	ker se mi zdi , da ima tisti zgornji rob dol zaviti a ne ?
19	forum.avtomobilizem.7.29072.470137	sej ane o kromu bo treba se globoko premislit
20	forum.avtomobilizem.7.29072.533203	jah nc pišuka tole bo kavica ane ¶ ja un uvink mi je se posebj pr srcu ...
21	forum.avtomobilizem.6.62993.1030937	Hej , neke je treba prišparat , če so tako poceni , a ne .
22	forum.avtomobilizem.11.26542.869473	Šit ane .
23	forum.avtomobilizem.30.100522.1638761	Itak boš pa naslednji dan opazil puščanje in takoj urgiral ane ?
24	forum.avtomobilizem.30.7136.105889	Sej plačat tko morm sej konc koncev sem jaz kriv za prekršek in bom tudi nastradu jaz a ne .
25	forum.avtomobilizem.26.84397.1425541	hvala za tole in točno to sm hotla (očitno se da , a ne)
26	forum.avtomobilizem.25.67571.1121183	Aja pa fajn bi blo ane , da bi končno nardil simulacijo in ne arkado , ker tega je pa preveč na trgu .
27	forum.avtomobilizem.3.13212.224366	Tut nebi blo vec zaminivo ane ?
28	forum.avtomobilizem.33.33716.494697	p. s. ja svašta ane ...
29	forum.avtomobilizem.7.63017.1154939	pa še type-r značko na prtlačnik ane ?
30	forum.avtomobilizem.7.86861.1461898	Si bil na izletu čez Belo krajino a ne ?
31	forum.avtomobilizem.7.6143.74688	Tista srebrna mrežica , ki se dobi pri A3M se da brez problema rezat in oblikovat , a ne ...
32	forum.avtomobilizem.11.46574.1038253	Je pa res da imajo BMWjaši druge prioritete kot Audi fanboyi (ker drugače se ne bi prodala niti ena A4 B6 (popravek) a ne Sašo , če bi bila vozna dinamika kakorkoli odločujoča pri nakupu))
33	forum.avtomobilizem.17.17365.261211	Kermogli bi ustimat a ne .
34	forum.avtomobilizem.18.100795.1642344	Ker turbo motorji so pa avtomatsko šrot ane ?
35	forum.avtomobilizem.18.53859.949265	Klical sem te že sva govorila pa ni bil nič drugač , sej usi vemo da pač pr tem tut ti neki zasluž ane , ne bom te zej js pa prosu da mi ga prpeles ...
36	forum.avtomobilizem.30.82295.1406125	Res ane , ko tole berem mi gre na bruhanje .
37	forum.avtomobilizem.30.15638.233776	Sej razumete ane ..
38	forum.avtomobilizem.7.66838.1121646	to si jo kupil samo za honde dol pokat a ne ?
39	forum.avtomobilizem.26.8997.125506	pa na spojlerju mam štop lampo , ti maš na zadnji šipi a ne ?
40	forum.avtomobilizem.20.25127.375346	Huh ja bo letos treba tudi neki nardit , ce bo folk za oz bo udelezba kr ane , pol bi lahko naredili mogoce mal vec kot lansko leto .
41	forum.avtomobilizem.2.59509.971454	če si stopil na suha tla in imel še pri tem mogoče čevlje s suhim gumi podplatom , pol se glih nisi razelektril , a ne ?

42	forum.avtomobilizem.4.8333.135474	No mi smo se včer mal pogovarjal o tem oz. se je bol mat tko da baje gremo a ne mat ?
43	forum.avtomobilizem.7.68185.1160686	bi se loh ze enkat dobli da ksko recemo a ne :P
44	forum.avtomobilizem.8.62147.1093117	to je normalno , da se pri menjavi izpuha uporablja kladivo ane ...
45	forum.avtomobilizem.8.73383.1252386	Ne gre , a ne ?
46	forum.avtomobilizem.7.31391.772637	ampak si verjetno vzel brezisvne a ne
47	forum.avtomobilizem.7.76074.1301164	Res je lep ane .
48	forum.avtomobilizem.7.70786.1248972	drugače pa stvar okusa a ne ¶ ponoči s pozicijo pomojem kr lepo zgledajo
49	forum.avtomobilizem.7.77419.1343247	Jup samo odbijač je malo čudn , ampak to bi že dali kej druga gor a ne only me ?
50	forum.avtomobilizem.11.89934.1505401	Pa sej je sam testna mula ane ...
51	forum.avtomobilizem.11.46492.716266	hudičevo blizu , ane ¶ lp
52	forum.avtomobilizem.18.70193.1178762	Tisti BMW , ki ga je nekdo predlagal bo imal skoraj pol milijona prevoženih po petih letih in ne bo ravo kot nov a ne , razen če zavrtiš kilometre nazaj
53	forum.avtomobilizem.30.32281.506710	Ah no , saj vemo , da se v stiski okleneš vsake bilke , a ne ?
54	forum.avtomobilizem.26.56595.905035	ja naj nekdo proba ane ??
55	forum.avtomobilizem.9.7682.101913	Lepa tale stvarca ane .
56	forum.avtomobilizem.9.7682.102555	Ja res so lepe tele a4-ke sam _spirit_ potolazi se da bos dobil ibizo kot si napisu da te ze caka sej je lbiza tut lep avto pac za kaksen razred nizje od A4 sam vseen lep avto ane Viper 123
57	forum.kvarkadabra.21.4919.79957	Tudi one se borijo za vire , a ne ?
58	forum.kvarkadabra.21.4919.81005	Šolska kreda se hitro zlomi , a ne ?
59	forum.kvarkadabra.21.4919.81203	Vse kar je , je dejansko energija in v energiji ni sebičnosti , zato bi morala nekako nastati , a ne ?
60	forum.kvarkadabra.21.4919.94467	Skupna točka , bi bila UM , a ne ?
61	forum.kvarkadabra.21.4919.95508	Imamo vendar skupne gradnike , mostovi so v nas , a ne ?
62	forum.kvarkadabra.21.4919.96555	Vidiš sedaj , števila in njihov sestav je potrebno jemati resno , še posebej , če niso od boga dana , a ne ?
63	forum.kvarkadabra.21.4973.81517	Dobro no , lahko rečemo , če je znanstveni sistem napačen , potem je prekleto dobro napačen , a ne ?
64	forum.kvarkadabra.21.4973.82827	Kako pa zvezda " umre " veš ; vse je odvisno od količine / mase predhodno izgubljenih , a ne ?
65	forum.kvarkadabra.21.4973.83118	Norost pa je lahko pomanjkanje empatije , a ne ?
66	forum.kvarkadabra.21.4973.91138	Iz te perspektive mora imeti celoten ekosistem na Zemlji neke vrste kolektivno zavest , a ne ?
67	forum.kvarkadabra.7.4767.76454	Torej ni važno , kaj je vsebina flaške , važno je kako flaško prezentiraš , a ne ?
68	forum.kvarkadabra.27.5258.77900	Torej , rečemo lahko , da lahko vemo kaj ni naključnost , nikakor in nikoli ne moremo naključnosti potrditi , a ne ?
69	forum.kvarkadabra.27.5180.75011	Pa še videti je , da ne potrebuje neke posebne koncentracije , a ne ?

Tviti ¹⁴²		
	ID dokumenta	Konkordanca
1	tid.347746416220856320	@GeorgLesev Tale je tvoja , a ne ?
2	tid.343392177721860096	vreme pa je ... ane !
3	tid.649532995678945280	@mihavete sam me boš pelu en krog poj , ane ?
4	tid.340876073967419392	Brez prostovoljcev ni @EuroBasket ! Ane @Lipko2013 ?
5	tid.517760095946813441	@Chateebox A ne .
6	tid.519787092101328896	@karmennovak @tyschew research scientis =/ = data scientist , ane ?

¹⁴² Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/gpw7k43>.

Priloge

7	tid.547129525004668928	pa ne gre ane ...
8	tid.438294565716770816	A bo sam 1 sezona a ne ?
9	tid.591213758753742848	@miskasmetiska tole z germanwings se hecas ane ;)
10	tid.398590692294811648	Pivo je pomembno , a ne .
11	tid.519787010048139264	@MyBlueDragoness @EffeV @Jure_Bajic @miskasmetiska A ne :)) Nam je mala totalno huda , glas in vse ...
12	tid.610323632494088192	@nmusar Če bi zločine nacistov dali v tabelico in jih prikazali v odstotkih , bi bilo tudi fejest , a ne ?
13	tid.433958124676866049	@BorutPahor @TeamSlovenia @lovehokey jebela cesta gospod predsednik , da se ne boste utrudili Risi so v formi , a ne ?
14	tid.601647687293886464	@Jure_Bajic midva pa verjetn prej grema na #pir u Šentjurju :D . A ne ?
15	tid.362864819679936514	@tbesther to pa že veš , da je ljubezen obojestranska , a ne :)
16	tid.407787089460461568	@hruske a ne :))) @zzTurk
17	tid.428986335580262400	Pa ta recept za ananasove zrezke bi tudi lahko zaupala , ane ?
18	tid.510075240396107776	so mogli pohitet a ne ;)
19	tid.403995022775046144	" Se se nkol ni , a ne " << sarkazem
20	tid.649500385913253888	VE se - GRE se : v kino ! ane @iztokgartner ?
21	tid.593471565431439360	@anja_ps a ne ??
22	tid.624227937936654336	@maticej @vampirc4 ane ?
23	tid.431527504793571330	@miharejc Cudno , a ne ;)
24	tid.549860345498832896	@WinstonMiniluv Plemstvo je blo zmeri fer , a ne ?
25	tid.629157166184833024	@lazgor a zdej si 2 x dvignu rit pred peto zjutraj pa se mal greš , a ne ?
26	tid.500276400663068672	@PlanetTV saj se hecate , a ne ?
27	tid.499663354097848320	@KlemenHumerca ane :-)
28	tid.576022978628075520	@teyaena @loudk taki smo , ane ?
29	tid.617402593829105664	@lajnar_EU Silba , to je tam zraven Krete , a ne ?
30	tid.476617570674810880	Za moje pojme mal pretirava . A ne ?
31	tid.499615841743548416	Se mamu česa veselit , a ne @lisjakm @tmoven @ErikRenko @n_e_j_k_o @jure_mastnak @sZlatko
32	tid.365056143061549058	porabi več rastlin kot ljudje , ane ?
33	tid.549617255081074688	Saj je v poljščini , ane ?
34	tid.418359186242961409	Pa dejmo , ane .
35	tid.483716945817391104	pravzaprav za vsako rešitev imaš problem , ane ...
36	tid.654401637830758400	Bruh bruh , včasih si smrdel če si kadil al a ne :S #CrknuTV
37	tid.365757575117029376	@nkran Ali pa če so tista razgaljena kolena ne desni bistvo , a ne ?
38	tid.491534802752114688	saj res @strankaSDS je krasen zgled , a ne ;) V možganov ;)
39	tid.606833781937446912	Jah cool a ne .
40	tid.664118951207682048	pripada nam . a ne .
41	tid.386073004339245056	ko te nekdo vsak dan pokliče , ker mu velikopomeni tvoja dobra volja , čeprav ve , da si čist ' v rit - izjemno , a ne :) http://t.co/WA8DFJnj9s
42	tid.603847704268173313	@KarmelinaHu Hehehe kok mal je včasih treba , ane ☺
43	tid.386219579778662400	To nalašč zajebavajo , a ne ?
44	tid.364397179550449667	@BoPa1 a ne !
45	tid.536469275159326721	Drugace pa bi , a ne ?
46	tid.442687565284769792	Saj smo vsi pricakovali invazijo grupnih selfijev ane ?
47	tid.553238316834103296	Pariz je pa malo bližje kot Sirija in Irak , a ne ?

Priloge

48	tid.380755540785127424	v Slo se #vine nekako še ni prav prijel a ne ?
49	tid.617665390253092864	@agathung Važno , da berejo , a ne ;)
50	tid.677195667492757504	@tretjeoko ja mojeee ane .
51	tid.513636200318763008	Ne , ane .
52	tid.525350150034956288	Racun pa bo ane ¶ @MIBPivac @GligoraCheese biti ce jako ukusno .
53	tid.381846048349491200	@PiaJagodic Smej je pol zdravja , ane .
54	tid.401779912270303233	@sloveniafooty o Rajšlu se je že enkrat neki pisalo a ne ?
55	tid.380427687493656576	Vsem igralcem damo za mrzlo @Lasko_Pivo sam zmagajte :) a ne @MatejRezek #junaki
56	tid.453507417226350592	@KlemenRobnik - a se , ane ?
57	tid.619624218553155584	☹ Ti se nikol nisi nic na puf kupil , ane ?
58	tid.450687343213944832	Funkcionarjem pa vse na pladnju prineso , pa še vprašajo , če kaj rab . A ne Štefanec ?
59	tid.426452982574501892	@FranciKek čudno ane pa glih on kje od tam doma , da za domač kraj ni nč bil pripravljen naredit
60	tid.421362421014724608	@UrškaSoncek ti tut ne , ane ???!!?
61	tid.406067683395125249	@UlaVovk aj , ti stalkas book depository , ane ?
62	tid.623794350619820032	@freeeky ane ? Ane ??
63	tid.525202107088252928	@MorskaMania pa na zalost nc ne pomaga , ane ...
64	tid.419854816874360832	@merineseri a n , a n ?
65	tid.359281881155506177	@UrškaSoncek to mas pa prov , ane @JakaUrankar
66	tid.365837815826747392	@Verjanka naju čez šipo futrajo , a ne ?
67	tid.445586068214394880	@krejcl hehe a ne , softič cel :) Hvalaaa , bomo nazdravli ob prilki :)
68	tid.364692609173692417	@r4vver Če bi boti znal mal sodelovat , bi blo tut ql a ne .
69	tid.434343130222321666	Najbolj fajn pa je , ko ti sredi izpolnjevanja ful podrobne akreditacije crashne comp , a ne ?
70	tid.478459464845389824	" Modrecu " pa ja namo " grozil " ane ?
71	tid.476281682287673344	Predvidevam da samo na fri skor vsako izpitno obdobje rata spletna učilnica nedosegljiva ane ?

Pogovorne strani wikipedije ¹⁴³		
	ID dokumenta	Konkordanca
1	wikipedia.pagetalk.11.1.1	Krogla iz puške že bo , a ne ?
2	wikipedia.pagetalk.60.7.20	' Kul ' , a ne .
3	wikipedia.pagetalk.521.1.9	Glede spornih gradov , mogoče so na seznamu tudi gradovi , ki so bili porušeni in so na njihovih temeljih zgradili druge zgradbe , a ne ?
4	wikipedia.pagetalk.634.1.3	Glede na dejstvo , da je naslov članka , Sir (naziv) , bi potem lady bila pravilna izbira , a ne ;) .
5	wikipedia.pagetalk.920.1.2	Naboj je relativno nova zadeva (prvi se pojavi šele 1812) , tako da arkebuze in muškete niso uporabljale nabojev , a ne ?
6	wikipedia.pagetalk.960.1.5	Vsak kulinarik to ve , a ne ?
7	wikipedia.pagetalk.1182.1.4	Slovenci , a ne ?
8	wikipedia.pagetalk.1416.2.3	Ne , saj , kot sem rekel zgoraj , ti načeloma ni potrebno popravljati članka , če je » Center « res » center « , a ne .
9	wikipedia.pagetalk.1500.1.9	Vprašati nazaj pa tudi ne škodi , a ne .
10	wikipedia.pagetalk.1637.1.5	Vem , da je povprečnemu slovenskem uporabniku popolnoma vseeno , če se napiše trenutnega japonskega predsednika " Juničiro " ali " Džuničiro " ali " Junichiro " ali magari " Džunichiro " , in da

¹⁴³ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/z7b5ayn>.

		malokoga moti , da se ga v medijih vsakič izgovori drugače , ampak mislim da je wikipedia pa le en resen projekt , a ne ?
11	wikipedia.pagetalk.2235.1.4	Zgled dobrega oblikovanja , a ne ?
12	wikipedia.pagetalk.2495.1.15	Zanimivo , a ne ?
13	wikipedia.pagetalk.3482.1.2	Dopiši , a ne ?
14	wikipedia.pagetalk.5609.1.1	Reči , da poznamo samo dve vrsti načinov , je malo naivno , ane ?
15	wikipedia.pagetalk.7053.1.3	Barvna je o. k. Različni okusi , a ne ?
16	wikipedia.usertalk.29.30.5	Poglej si malo posebne strani (saj veš kaj je to , a ne) , tam sta dve strani , ki se imenujeta Želeni / e in sicer članki ter kategorije ; in to so te rdeče ktgr .
17	wikipedia.usertalk.164.58.1	Če jaz obro razumem , mora biti za vse tole avtor mrtev že 70 let , če jih želiš naložiti na Zbirko a ne ?
18	wikipedia.usertalk.435.30.7	Mogoče te ni " nobeden upošteval " , ker ni noben domač na tem področju , a ne ?
19	wikipedia.usertalk.529.28.1	Zakaj bi jaz bil izjema in se matral zastoj - a ne .
20	wikipedia.usertalk.632.27.1	Tam ne bi trdili , da je bil to , če ne bi bil , a ne ?
21	wikipedia.usertalk.1519.17.14	Tu je preveč dela , da bi bili ljudje užaljeni , a ne .
22	wikipedia.usertalk.2282.25.1	Raje povem , ker drugače moram popravljati - boljša pa je preventiva kot kurativa , a ne .
23	wikipedia.usertalk.2458.3.4	Oprosti , toda sem še malo zelen ane .
24	wikipedia.usertalk.4190.16.2	A to kar si ti dodal na strani - rezervacija - zame ne velja (s tistim uporabnikom si mislil name ane ?
25	wikipedia.usertalk.4412.5.1	Unteroffizier je vsekakor podčastnik , junkerji so pa šolarji , a ne .
26	wikipedia.usertalk.5370.6.9	Predlagati pa ni greh , a ne .
27	wikipedia.usertalk.11247.2.5	Naenkrat več ni tako lahko , a ne ?
28	wikipedia.usertalk.11969.26.5	Ah , na pol pa ne bomo delal ' ane ?
29	wikipedia.usertalk.12104.14.1	Torej kar že deluje za nekoga bo tudi za nas , saj to je vendar wikipedia a ne .

Komentarji na novice ¹⁴⁴		
	ID dokumenta	Konkordanca
1	news.rtvsl.345462.75	Ko bo cela Dolenjska na zavodu , država pa v buli , kako bomo šele takrat kričali - Živela Amerika in EU , a ne ?
2	news.rtvsl.345647.235	pod mizo ane ...
3	news.rtvsl.346245.68	be a ne !!?
4	news.rtvsl.291340.28	To ni res , ane , ker ni bil davkoplačevalski denar , ampak finski .
5	news.rtvsl.341511.91	Ko mački stopiš na rep zacvili ane Ferlinc ?
6	news.rtvsl.250187.6	@senca - no ja , si mislu mal fotošopa pa mal fotošopa pa mal solarija pa mal fotošopa pa še mal fotošopa , ane .
7	news.rtvsl.335117.14	Seveda , saj ga je zaslužil s svojim delom , a ne ?
8	news.rtvsl.326981.28	, kot lahko prebiramo na forumu , a ne ?
9	news.rtvsl.323138.50	kako je preprosto a ne .
10	news.rtvsl.332232.84	yoda , ti res ne razumeš ane ?
11	news.rtvsl.347208.7	Čisto naključje , a ne ?
12	news.rtvsl.346404.6	Geniji ane !
13	news.rtvsl.346646.48	Kako se vse obrne ane .
14	news.rtvsl.343631.191	Če zgrešiš prvi penal ne greš streljat še drugega ane .

¹⁴⁴ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/zpg6hyo>.

15	news.rtvsl.123431.136	Aja , žalitve sm pozabu oment , a ne MilanSu ?
16	news.rtvsl.320171.449	to je pa tko k je kitarist za selektorja ane Čeferin .
17	news.rtvsl.345571.519	drugi bi bili ker smo premagali Avstralijo , se pravi bi mi morali kalkilirati in izgubiti prvo tekmo potem pa bi bilo vse OK ane = ?
18	news.rtvsl.237209.57	Ampak treba je kritizirat Klobučarja ane Garni5 .
19	news.rtvsl.330291.443	Gate na glavo , dva svinčnika v nos ane piše in poje , res izvirno !
20	news.rtvsl.332635.72	lahk ce jim zveza priskoc na pomoc , ane
21	news.rtvsl.331411.328	Brez cincanja v drugi vožnji a ne !
22	news.rtvsl.329689.599	Hehe , kaj delamo babe , a ne ?
23	news.rtvsl.346693.37	Kakor vidim še ni ane :) Torej je tvoja logika spet padla na izpitu .
24	news.rtvsl.346147.18	Ti bi najbrz kar sam konkuriral Kitajski a ne ?
25	news.rtvsl.346085.163	To , da so iz maščevanja poskusili zanetis resno vojno s tem , da so sestrelili potniško letalo nima veze a ne ?
26	news.rtvsl.291648.6	druge države bi pa lahko tut prepovedale uvoz alkohola iz češke preden so ble smrtne žrtve ane ...
27	news.rtvsl.265967.34	ne normalno bila je raketa , ki je prvo premaknila generator pred pentagonom potem pa še zadela pentagon ane ?
28	news.rtvsl.265967.159	) pa lahko razsvetliš ti mene in ostale dvomljivce , a ne ?
29	news.rtvsl.345341.195	aja srbi so drugo , a ne ?
30	news.rtvsl.336588.505	se boš javil , k boš v sloveniji .. a ne ??..
31	news.rtvsl.95961.5	tole a ne ¶ saj boma kmalu tam
32	news.rtvsl.344452.2	Glede na to , da se bolezen najlažje omeji z omejitvijo stikov , verjetno podnevi lažje nadzoruješ gibanje ljudi kot ponoči ob neosvetljenih ulicah , a ne ?
33	news.rtvsl.342198.117	In ubogi naši mediji , ki si ne upajo zavzeti nobenega stališča (bojmo se Rusov , a ne) .
34	news.rtvsl.342198.205	Če bi bilo res drugo vojaško letalo sestreljeno na tem območju malo prej , jim nebi bilo treba umikat in brisat twite , a n ?
35	news.rtvsl.292254.90	Pa ce je tako na vseh faksih , se nekomu kr nabere , a ne ?
36	news.rtvsl.296661.401	A za nekatere je kriv JJ a ne ?
37	news.rtvsl.333989.95	je pa nataša precej boljša od rupla , ane ?
38	news.rtvsl.316650.81	Predvsem ga ne znamo pravilno uporabljat , ane ?
39	news.rtvsl.343080.542	rajc je pa podjetniški biser , a ne ?
40	news.rtvsl.339950.625	Ga ni , ane ???
41	news.rtvsl.339907.886	Saj veš , da so le-te same po sebi greh , ane ?
42	news.rtvsl.339907.1058	Mimogrede , je potrebno najprej poznat zgodovinska dejstva , če že pravite da so partizani (niso bili vsi partizani komunisti) , škodili slovenski zemlji ane ?
43	news.rtvsl.346211.107	Kakšen šold je verjetno tudi prispel iz tujine , a ne ?
44	news.rtvsl.80236.14	Ti pa nisi vaju neke kulturne debate , a ne ?
45	news.rtvsl.345833.20	Oooo bog , še malo pa bo to desna " Janševa " vlada ane ...
46	news.rtvsl.339615.40	Ne , ne , to je za " kravatarje " , a ne ?
47	news.rtvsl.343841.49	A ob to se pa vi moralistični dvoličneži nikoli ne obregnete ane ?
48	news.rtvsl.344802.342	Tisto ni problem , saj so " naši " vzeli ane .
49	news.rtvsl.323530.664	Preprosto , a ne .
50	news.rtvsl.327151.336	Malo na pamet , ane ?
51	news.rtvsl.327151.424	Zgolj izpričuješ svojo forumsko opredelitev , a ne ?
52	news.rtvsl.346838.16	- a ne , veliki človekoljubi ...
53	news.rtvsl.336578.72	In ti bi se odločil kar za amputacijo ... a ne ?
54	news.rtvsl.346229.61	" 3000 € za en tak logotip je postena cena , ampak vertjetno bi vas vecina " strokovnjakov " boljsi logo nakracala v 10 ih minutah in to za 10 krat nizjo ceno . A ne ?!

Priloge

55	news.mladina.160942.5	Nemčiji slabo kaže navaden delavec pa si pri njih lahko vilo zgradi a ne ?!?
56	news.mladina.146695.102	Razumel bi , da jo opusti v 1 ali 2 letniku a ne , to odgovarja tudi šolam in vse skupaj postane boj za glavarino .
57	news.mladina.158957.2	Brothers in arms , pa to ... ane ?
58	news.mladina.155413.5	spet smo krivi ljudje a ne ?
59	news.mladina.153842.32	Saj je itak vse vsaj vsem nam , ki tule pisarimo , popolnoma jasno , a ne : 1. država je tu že od nekdaj zgolj zato , da " zabetonira " družbo na tiste zgoraj in tiste spodaj .
60	news.mladina.142887.11	Nekdo ki je vpleten v toliko afer da ga še tujina preiskuje in da segajo lovke vseh nepravilnosti (milo rečeno) vse do severne Finske je zate očitno poštenjak ki samo dobro hoče ane ?
61	news.mladina.116714.11	saj " intelektualcev " vaše baže zunaj nihče ne bi plačal-tu pa se še najdejo molzne krave , a ne ?
62	news.mladina.114993.11	Seveda ne , ampak takšen pac si , a ne .
63	news.mladina.163830.2	Saj je nedolžen kot Janša , a ne ?
64	news.mladina.153763.4	jazorit , si ti mal preprost ane , kultura tebi ni prioriteta , a ?
65	news.reporter.48534.12	Pa spotoma si bil še mal na zdravljenju na kliniki za male živali s prjatlj a ne ?
66	news.reporter.49635.48	V ta kontekst spadajo tudi njegove » priležnice « , od katerih so bile in so še vedno najbolj zaželenne poročene in vezane starejše in mlajše gospe , verjetno zato ker običajno ne delajo težav a ne Stanko ?

Priloga 11: Vzorčeni primeri rabe označevalca *a veš* korpusu Janes

V tej prilogi navajamo primere rabe diskurznega označevalca *a veš*, ki so bili identificirani v analizi 100 naključnih konkordanc izmed skupno 737 pojavitev te oblike v podkorpusu komentarjev na bloge (36 primerov), 163 pojavitev te oblike v podkorpusu blogovskih zapisov (38 primerov), 465 pojavitev te oblike v podkorpusu forumov (28 primerov), 8.483 pojavitev te oblike v podkorpusu tvitov (68 primerov), 10 pojavitev te oblike v podkorpusu pogovornih strani wikipedije (3 primeri) in 170 pojavitev te oblike v podkorpusu komentarjev na novice (19 primerov).

Komentarji na bloge ¹⁴⁵		
	ID dokumenta	Konkordanca
1	blog.publishwall.3077.11	Vanjo ... a veš , nekega dne sem zapustila eDnevnik , zato , ker je en " mesar " , malo bolj primitiven od tebe zahteval , da grem stran od tam .
2	blog.publishwall.3350.2	J'z te pomnim iz tiste zgodbe a veš , ko ti je " preverjen " " prijatelj " poslal perverzno fotografijo ...
3	blog.publishwall.3350.9	Če razmislim , o vsakem atomu in njegovih možganih ... A veš , tukaj se dejansko ločnica med živim in neživim konkretno zabriše .
4	blog.publishwall.4797.4	In a veš ...
5	blog.publishwall.7362.2	:) A veš , človek vse prenese , samo lučko na koncu tunela mora videt ...
6	blog.rtvsl.649.9	Tone Pavček A veš , sta dve kuri laufali pred petelinom pa pravi ena .
7	blog.rtvsl.2010.76	75. A veš , ti Tito , če bi bil kdo takrat proti , ga bi koljač takoj poslal na oni svet ali pa na prelep otok
8	blog.rtvsl.2483.25	sej a veš se vidi ...
9	blog.rtvsl.2720.11	Krščen duš , najprej moja Barbara , zdaj pa še andromaha :) A veš , Kouvran , da tisto , da imajo vsake oči svojega malarja res drži ?
10	blog.rtvsl.2810.14	Če lastnik nosa poda prijavo - a veš ?
11	blog.rtvsl.3320.2	ja a veš v pol ure je težko kaj pametnega povedat
12	blog.rtvsl.4348.16	no a veš ...
13	blog.rtvsl.8004.3	Preden bo kosovarjem kapnilo , v kaj so namočeni , bo prepozno . A veš , tist free na papirju ni vreden nič , če država ne funkcionira samostojno .
14	blog.rtvsl.8371.16	ta je dobra a viš ...
15	blog.rtvsl.8447.26	Ali kdo si ti , da boš presojal kdo je potencialni morilec in kdo ne - jaz sem jih samo poimenoval idioti , a veš ;))
16	blog.rtvsl.8577.11	ojroti a veš
17	blog.rtvsl.10195.24	powim tjem k snuova ... a wiš
18	blog.rtvsl.10222.14	Prvo je zdravje . A veš , takoj zamenjam lopatanje na vrtu za grabljenje sena , če misliš , da je težje .
19	blog.rtvsl.10256.3	Naj se pobijejo med sabo , meni do pobijanja niti ni . A veš , niti :) Ker giljotina je le neke vrste odgovor ...
20	blog.rtvsl.10968.18	Stanislav S a wiš ...
21	blog.rtvsl.13160.32	Absolutno , se strinjam ... A veš , meni je tole naše bloganje všeč iz enega preprostega razloga - včasih sem o sebi mislila , da veliko vem , poznam , kje kaj je , kako stoji ...
22	blog.rtvsl.13198.5	@BINGA - BONGA - možak je prepričan , kajne ? A ves , glej : Vceraj na plazi ogromno zapade dez !

¹⁴⁵ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/j4tvpkb>.

23	blog.rtvsl.15734.7	ČE JE KDO NAPADEL ZARES SO BILI NOVINARJI IN KOALICIJA A veš blažek , eni tu bodo do onemoglosti branili čelavega krvosesa .
24	blog.rtvsl.16805.3	no tist a veš ...
25	blog.rtvsl.16995.4	Yiruma je car A veš tisti film Somrak ?
26	blog.rtvsl.19056.30	de sn jest kej a veš
27	blog.rtvsl.19238.20	še jaz se bom prijala tega konja za rep , a veš .
28	blog.rtvsl.20075.11	Malo treniram prste za slučaj , da jutri dobim delo .. a veš ?
29	blog.rtvsl.20498.87	ID .. a veš ..
30	blog.rtvsl.20671.16	v primeru , če prileti kakšen iz ssskajota , a veš :-)) .
31	blog.rtvsl.20701.10	jst andromaho čisto razumem . a veš .
32	blog.rtvsl.21525.10	mi puojemo kr tk , na puhrehbih al pa na fantušnih a wiš ...
33	blog.rtvsl.21658.24	Neue Ekspres , okej , še ananas ti pretlačim in polijem z grozdnim sokom , dodam led in pisano slamico . a veš ...
34	blog.rtvsl.22282.12	sam a veš mi je pa tut ena medicinska setrra rekla , da bi mogu bolj hodit v
35	blog.rtvsl.22328.19	In zaradi tega me nima kaj peči vest - a veš - Branka
36	blog.rtvsl.23423.4	moru sn se znjim mal pomenit a veš

blogi ¹⁴⁶		
	ID dokumenta	Konkordanca
1	blog.publishwall.2774.0	No , in zdaj , ko sem rekla , da Kljukca s strehe ne bomo več brali , da se mi zdi , da bolj spodbuja k neumnostim kot kaj drugega (no ja , na nek način so vse otroške neumnosti lahko razumljene) , je rekla sedaj že moja velika hči : " A veš , mama , tista tvoja bela jopca , tista , ki sem ji izstrigla kroge iz hrbta ...?
2	blog.publishwall.2812.0	" A veš , Mali , to je kretnja .
3	blog.publishwall.3164.0	Da so mesta vedno " rezervirana " za tiste izbrane , VIP , oni ki imajo veze , zrihtano aveš .
4	blog.publishwall.4303.0	Moja mama gre k svojim sestram in bratu na obisk in reče : " A veš , da nismo niti enkrat prižgali televizijo ?
5	blog.publishwall.4798.0	In , a veš ...
6	blog.publishwall.7325.0	» A veš , ko smo včasih govorili o ljudeh , ki se obračajo po vetru , o takih , ki se vedno znajdejo nekako tako , za svoj prav .
7	blog.publishwall.7325.0	Predvsem pa vedno pravim , da so bili tisto neki drugi časi . A veš , mislim tisto , kar smo kdaj šimfali , da so bili eni rdeči , zdaj pa pojejo na koru in podobne reči ?!
8	blog.publishwall.11955.0	Taki časi in taka doživetja so danes neverjetna ? A veš , ko še sama sebi včasih ne bi verjela .
9	blog.publishwall.12314.0	" Luna naju gleda ... A veš ?
10	blog.rtvsl.1160.0	» A veš , ko smo včasih govorili o ljudeh , ki se obračajo po vetru , o takih , ki se vedno znajdejo nekako tako , za svoj prav .
11	blog.rtvsl.1160.0	» Delavnico so nam pripravili v fabriki . A veš tisto reč , malo predavanja , malo slikc in potem so vsi pametni .
12	blog.rtvsl.5279.0	Jebaj ga zdaj , saj bi bilo tudi malo , (maaaaalo , a veš) pretenciozno pričakovat , da se prvič udeležiš ene take reči in kar zmagaš , jel ?
13	blog.rtvsl.6742.0	Ah saj veš , vse to s tem rožljamnem z orožjem , tako prikrito , a veš .
14	blog.rtvsl.8774.0	 a veš
15	blog.rtvsl.13022.0	12.6.2009 , o poznanstvih A veš ghun ?
16	blog.rtvsl.13703.0	" " No a veš un film .
17	blog.rtvsl.14285.0	Med potjo nazaj me držiš za roko in praviš A veš , tista kočja Ja , vem , tista Zaprem oči in že prehodim prvi travnik Grem skozi gozd nabiram gobe , kostanje pretečem drugi travnik da sedem na razjedeno klop ob prazni koči Ja , to je tista kočja Hvala ker si me spomnil nanjo Mislim da bom ob njej postavila svoje jaslice

¹⁴⁶ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/z3q67js>.

18	blog.rtvsl.15088.0	Politikih . A veš , ko me včasih prideš poklicat , da je Dnevnik ?
19	blog.rtvsl.15143.0	:) Lp A veš , včeraj , ko si poljubil mi dlan ...
20	blog.rtvsl.16011.0	Vpraša ga čemu joče v temini , pa pravi ji , da ga mučijo spomini . A veš , ko sva se za hišo mečkala , še kot fant in punca sva bila ?
21	blog.rtvsl.17197.0	V mojih občutkih se skriva nek čudn občutek , a veš tak , ki ti nekako da mislit , da
22	blog.rtvsl.17197.0	ma to je en tak zlo čudn občutek , a veš , nevsakdanji , ki se pojavi po nekih čudnih dogodkih
23	blog.rtvsl.17840.0	" A veš , tamladi bodo pa dojenčka dobili , drug mesec se mi zdi , da bo , " mi je zaupno povedala .
24	blog.rtvsl.18096.0	unga a veš Francozarja
25	blog.rtvsl.21072.0	L. p. A VEŠ ONO ?
26	blog.rtvsl.21072.0	A veš ono , ko pride tak moment v dnevu , ko se nasmejiš do srca ? A veš ono , ko dobiš filing , da si preprosto srečen , pa četudi se to sliši klišejsko in pocukrano na kubik ?
27	blog.rtvsl.21072.0	A veš ono , ko opustiš strah za bodoče bivanje v študentskem domu zaradi bojzani pred nemirom ? A veš ono , ko kličeš svojo staro mamo , spremeniš ton glasu , pokriješ slušalko , v fori rečeš » O , Cirila !
28	blog.rtvsl.21072.0	A veš ono , ko si poln ciljev in planov za prihodnost , pa četudi so malce preveč idealizirani ? A veš ono , ko spoznaš , da je sreča stvar odločitve ?
29	blog.rtvsl.21072.0	A veš ono , ko spoznaš , da je sreča stvar odločitve ? A veš ono , ko ugotoviš , da če se trudiš iskati v vsaki stvari pozitivno , da je svet veliko bolj rožnat in ne toliko zagrenjen , kot zna biti marsikdaj ?
30	blog.rtvsl.21072.0	A veš ono , ko ugotoviš , da če se trudiš iskati v vsaki stvari pozitivno , da je svet veliko bolj rožnat in ne toliko zagrenjen , kot zna biti marsikdaj ? A veš ono , ko si vesel oblakov , pa četudi smo sredi julija , in ko ignoriraš letni čas ?
31	blog.rtvsl.21072.0	A veš ono , ko si vesel oblakov , pa četudi smo sredi julija , in ko ignoriraš letni čas ? A veš ono , ko pogruntaš svoj nenavaden smeh ; tak , da se smejiš smehu , a tisti ta drugi pa je pravi in iskren ?
32	blog.rtvsl.21072.0	A veš ono , ko med rekreativnim pohajkovanjem po bližnji deželi z nasmehom zreš ljudem v oči , pa dobiš nazaj smeh z njihovih zagrenjenih lic ? A veš ono , ko se zaveš , da tak moment pride res premalokrat ?
33	blog.rtvsl.21072.0	A veš ono , ko si želiš , da bi bil vsak moment tak ? A veš ono , ko si jezen sam nase , ko tako ne more biti vedno , saj veš , da tudi žalost in jeza pašeta ter sta potrebni ?
34	blog.rtvsl.21072.0	A veš ono , ko veš , da boš jutri spet tečen , nergav in ne tako dobre volje kot danes , pa čeprav si želiš biti tak ? A veš ono , ko pač samo paše biti srečen , pa četudi to traja eno uro , potem pa spolzi kot dežna kaplja po avtomobilskem steklu ?
35	blog.rtvsl.21072.0	In a veš ono , ko nimaš pojma , pa četudi preiščeš največje sahare , oaze , monsune ali amazonske gozdove v svojem miselnem pajčolanu , od kje se ta nenadna sreča prikrade ?
36	blog.rtvsl.21072.0	srečen . A veš , mogoče pa je pomagalo , da me je stara mama transformirala v Marinko ?
37	blog.rtvsl.21281.0	Ko je nekega leta razsajal v naših krajih koloradski hrošč , tista progasta golazen a veš , v krompirju , takrat so tudi stuhtali eno pametno .
38	blog.rtvsl.22621.0	Ej stari a veš tko je posluš , ne morš tko govorit da se ve da si z Ruš .

Forumj ¹⁴⁷		
	ID dokumenta	Konkordanca
1	forum.medovernet.6940109	Kako to , da se nisi bila pripravljena vsaj malo potruditi za tistega , ki ti je bil privlačen ? A veš , če hočeš ti imet neke kriterije , se moraš zavedat , da si tudi sama nekim merilom in kriterijem tudi izpostavljena .
2	forum.medovernet.9539015	ampak se mi je pa res obrnil svet , ko ugotovis , da na drugi strani je res nekdo cisto drugacen , drugace razmislja , druge stvari so mu pomembne , avtomastko gleda na svoj lagodje , jaz pa neka dobricina in bi rad ustregel ... A ves , to je po svoje tudi edinstvena priloznost zame , da zrastem , ker lahko brez prblema pritegnem se enega MOMOVca potem ...
3	forum.medovernet.6090002	" In pa : " A veš , kolegica od bratranca od prijateljice od moje sošolke iz gimnazije je že čisto obupala in sta posvojila otroka , potem pa je čisto odmisllila biološko potomstvo in je ČEZ DVA MESECA SPONTANO ZANOSILA Z DVOJČKI .
4	forum.medovernet.9937966	ja , tudi takšni morate obstajati :) . A veš , potem več ostane za nas , ki nam je mar :) .
5	forum.medovernet.6868041	A veš , kar naenkrat se je pojavu korenje , pa zele , pa vse tko lepo . A veš , tut zima je bila lepa in krasna , sej ne morem verjet da to pišem .

¹⁴⁷ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/huny6ff>.

6	forum.medovernet.6879640	Ma a veš Jani , sej tole naše je en mali mešanček , mama je pekinezerka , očka pa čivava !
7	forum.medovernet.9934876	Z drugimi ni nič narobe , če imaš ti probleme , a veš .
8	forum.medovernet.8126740	Ker je on bolj svobodnega duha , sem posledično tudi jaz zraven njega popolnoma svobodna NIKOLI ampak zares nikoli ni bil problem , če sem hotela sama kam it , nikoli me ne omejuje in vedno je za otroke in dom poskrbljeno . A veš na dolgi rok pa to veliko pomeni .
9	forum.medovernet.1987074	vsak strah je odveč , a veš .
10	forum.avtomobilizem.20.7211.94438	Oldi a ves mogoče ce je v Sloveniji kak Opel GT ?
11	forum.avtomobilizem.25.36416.815382	tam pridejo take ritke se bolj do izraza , ko se nad mizo sklanjajo , a ves se opravičujem za zamudo , sem prvič na vrsti za postavljanje uganke .
12	forum.avtomobilizem.25.5295.91991	<iztokizi> ja ne poznaš me preveč osebno ,, zato sm ti reku da bi se osebno bolj zblížala a veš , samo jest in ti , skupi
13	forum.avtomobilizem.25.5295.160661	(_BeBa_) aveš moje gače se lahko zložijo , aveš , tko da so manše
14	forum.avtomobilizem.7.59378.1474736	To je ta VAG refleks , a veš , se ne obračamo za Opli in Renaulti .
15	forum.avtomobilizem.6.77084.1318953	tudi sam vem , da je lahko 30 litrov ali več (odvisno od avta) normalno v klanec , samo ti pravim , da je na istem klancu , prej ko je delal pil nekje do 15 , 16 , 17 , če si bil fajn na gasu ... a veš , imam referenco , ker sem vse to opazil pri vožnji po istih cestah v podobnih razmerah ...
16	forum.avtomobilizem.2.97357.1609293	jaz skozi servis pridem s 100 eur , ti najbrž ne ... a veš ena " bedna " primerjava : to je tako , kot če bi primerjal bodybilderja (tiste , ki tekmujejo na " lepotnih " tekmovanjih) in enega kmečkega gozdarja ...
17	forum.avtomobilizem.9.401.4041	Kar se pa tiče DELANJA pa sem se jest zdej ubistvu modele spravu delat a veš (model da se pol naredi kalup iz tega) , Tunner-Uroš jih pa obleče v poliester .
18	forum.avtomobilizem.3.15214.235733	Je doma iz Kamnika a veš .
19	forum.avtomobilizem.11.93864.675237	ti si en mali hecko a veš ?
20	forum.avtomobilizem.37.14460.213282	Ej kle je še velik manj podkovanega folka , kot si ti aveš ...
21	forum.avtomobilizem.7.75970.1354647	hehe , sej pol to hitr leti a veš ko jo 10 x podrešč .
22	forum.avtomobilizem.11.52826.987326	Recimo novi oblikovalec pri KIA , ki je naredil audija TT , Hyundai i30 , je narisal možakar , ki je prej narisal BMW serije 1. A veš težko je imeti razpoznaven tako široki spekter vozil od Atosa do Santa-fe , Coupeja in nove limuzine .
23	forum.avtomobilizem.26.54409.872922	Ampak takrat sem povedal , nekaj , ki je prepričalo enega moderatorja da mi je odstranil opozorilo . A veš moreš biti res dober da mša dva opozorila , kok jih imaš a ti 0 ?
24	forum.avtomobilizem.25.50390.789044	Cisteasy prodal ga pa nemorem k je se cist nov a ves stari mogoc cez leto dve
25	forum.avtomobilizem.8.62786.1020631	js mislim kupit dtm-a , sam bv bi mi blo dat 110 eur , da pol nebi blo nobenga zvoka a veš ja js mam tud bravota 1.6 16 v , pa ma več moči ko tvoj (105 hp) , zato nevem zakaj bi tvoj meu bolši zvok kt bravo ?
26	forum.kvarkadabra.22.804.5627	Po moje sta vidva predpostavila , da je napetost na obema skupaj enaka , kot če bi bil eden . A veš , na vsakem U / 2. Pol mata prov .

Tvti ¹⁴⁸		
	ID dokumenta	Konkordanca
1	tid.372965932601470976	, ko spominja tviter modra črta na pozitiven test nosečnosti. A veš uno....
2	tid.371932491638378496	, ko pride serviser na dom in aparat dela, ko pa odide pa ponovno neha delati. A veš uno....
3	tid.371360151812317184	, ko odpreš kuharico in moraš polovico receptov izpustiti, ker nimaš vseh gospodinjskih aparatov, potrebnih v recepturi. A veš uno....
4	tid.379238987032170497	Pa vedno samo en zmaga, drugi pa nekako ne. A veš uno....
5	tid.385150280670908417	@IrenaSirena a veš uno....
6	tid.385486738967502848	, ko naklada in naklada in se ne zaveda, da nihce vec ne posluša. A veš uno....
7	tid.385991469242540033	, ko naredi napako in krivi drugega. A veš uno....

¹⁴⁸ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/jqwwogm>.

Priloge

8	tid.376817916744044544	, ko ti je natakak tako simpatičen, da bom še malo pocakala in narocila še eno kavo. A veš uno....
9	tid.475271812893667328	" A veš , k da je važno.
10	tid.576340373850701824	- A veš tist ...
11	tid.494207697894776833	#karantanjski A veš uno ko maš jurja sledilcev na tviterju?
12	tid.649588040067907584	Da mal štumfe posuši.. A veš un kisel...
13	tid.388362372077092864	#RockNRoll A ves tist, k na konc v salci caja ostane samo se med:) mmm #sick #gettingbetter
14	tid.430012247847735296	:) A veš tist filling, k maš še milijon snovi za predelat in se morš skor u temi učit, ker ni elektrike :/ No, to!
15	tid.497440771164094464	@RichieKis ja pol pa tko naslavljej tvite, a veeeš ☺☺☺ @SladkoKotLimona
16	tid.447734864331419648	:) A veš tist, k maš dva dežnika?
17	tid.409554075383246848	A si vstal ali boš šele sedaj legel? A veš tisto, ko te zjutra dete zbudi, ko se naslanje čez te, in ji slinca kaplja po tvojem obrazu, no to!
18	tid.351332105155907586	@Trdosrcnez neuradnih. a veš , k greš laufat, pa neki švistne mimo tebe?
19	tid.616642524820672513	@MyBlueDragoness to je v naših cajtih bil primer, če si dobil eno redko LEGO kocko a veš ...
20	tid.380569956741435393	:) Ni samo tetris, je tudi space invaders, pacman ... A veš tisto, ko te najdražja prosi, da ji pomagaš pri hujšanju ...
21	tid.462263881336094720	Slišim čričke. A veš tisto, k si na obisku in se ti neki strese, pa upaš da te noben ni vidu.
22	tid.560542754095505408	Ja, pol pa se bomo še naprej šalili.. A veš tisto, ko zadeneš 1,2 mio evrov na lotu?
23	tid.678860633597718528	Call it "experience". A veš tist, k te posiljujejo z reklamami, pa greš, pa odplačaš telefon, vezavo in poveš, da boš ob prvi ugodnejši ponudbi konkurence šel?
24	tid.503256379441020928	O ti ti pacek ti. A veš un @Turinek?
25	tid.669626789858291712	So he train with first team. A veš uno..
26	tid.476304831775510528	@MiranStajerc se strinjam. A veš , ko zdravstvo špara?
27	tid.478431636443652096	@Jure_Bajic ker radi šeramo, a veeesh .
28	tid.506489839219650560	:)) A veš unu, ko tok naliva, da greš tri brzine v krivine (stavba) in te odfuka direkt na trtico, nad glava se ti pa zarola cel tviteraj?
29	tid.461224644662800385	Poslušaj folk, ki kliče @toplovodar na @Val202 in ... A veš , ko kdaj pri frizerju pogledaš Lady in začasno zgubiš 20-30 IQ pik?
30	tid.603981843675897857	@alesh33 saj sva se razumela, ane :) A veš tist, da imaš po 14 urah še 60% baterije na 5S?
31	tid.430649166478725120	3000 and it goes to shout out @MRGUMBALL3000 and his @gumball3000 to which I hope (cont) http://t.co/8nvdozAhzy A ves tist k si prevec sestankov stisnes v en dan pa ne ves kako bos to speljal.
32	tid.526443436413829120	@MisaVugrinec @tretjeoko ja kdo bo pa za briketke zaslužil? a veš , mami mora delat :)
33	tid.551070717962641408	http://t.co/qSV7nBn6O0 Toliko o pravopisu :/ A veš un,k je sin od unga,k je brat od sosede,k so tam živel, kjer je bla moja sodelovka,k je sin z uno poročen #mati #Idontknowldontcare
34	tid.567433680373964800	" @peterprevc kmalu po rekordu za @PS_Sportal // http://t.co/nFS1IdNL3n a veš tisto, ko ne moreš nehat razmišljat o kakšni osebi, ups, stvari?
35	tid.413601556249202688	Medejcii :) am a ves , a ti ziher znaš italjansko aneda da znaš?
36	tid.508891671304867840	' a veš un kojn k mu je ime kojn?
37	tid.522775423437242369	@GoldenBlackHawk cuz you are in our hands bitch. A veš tisto, ko se sprehodiš mimo hiše pa vidiš @Svet_KanalA pa @NusaLesar?
38	tid.635517721095106560	@DJoannes lahko na dm? Aves tist k se ti snemanje zacne jutri ob 6.30 pa moras zdaj se na crnivo po vse cunje.
39	tid.506778970076237824	Pol si ful hepi, a veš .
40	tid.631170185307795457	#nasceni A veš tist, ko spletna stran misli, da bolje od tebe ve, kaj hočeš?
41	tid.627922174159011842	@MihaCirman a veš .
42	tid.652098767638908929	Ampak 2013 aveš , to je še dodiplomc, ne moreš verjet, kako se kvaliteta potem spremeni.
43	tid.572308039165988864	Novi podvigi v @Ziferblat_LJ #cedablat http://t.co/vppvxDJiwt A veš , ko ti veka utripa.
44	tid.464852974343553024	" ~ Elisabeth Kubler Ross A veš tisto, ko bereš komentarje in so ponavadi žaljivi.

45	tid.529305462668546048	@MartinKnapic some grammar lessons would be a good buy as well :) #jokingbutnotjoking A veš tisto, ko skušaš zgooglat "Sperbank", v resnici pa misliš "Sberbank"?
46	tid.395160684230475776	tisto ... a veš , no, nedelja, pa ...
47	tid.561240973704171520	"@barjanski: A veš tist, ...
48	tid.560531359123640323	Woop wooop wooop, a veš tist', ko končaš z delom, za danes.
49	tid.658253612263591937	@MiranStajerc @petrovictomi @XtinaTinka razumem tvojo tegobo (se ne smejem), ampak mojstri to znajo popraviti, a veš ;)
50	tid.647133523632111616	#narnija A veš tisti prizor iz LOTR, ko se na kupu dobijo Ents in imajo sestanek?
51	tid.526624828636561408	@Hribar13 @bosstjan V razredu sem bil vedno blagajnik, a veš ?
52	tid.552025545320697856	@illegall_blonde Kdo bo pa rože zalival, a veš !?
53	tid.479070790248591360	@E_nk_a A veš ,ko si imela letos r.d. in sem ti poslala neprimerno pesem?
54	tid.588722490089811968	@vitaminC_si za avto visjega razreda bi pa lahko imel petrolovo kartico A ves uno k se ti nc ne da.
55	tid.678232476372893696	To je moj p'rjatu, a veš :D https://t.co/N83GwwTWnT
56	tid.349929124883734531	zdaj pa modro ugotavljajo, da "so se strasti umirile" a veš un film Sestre Magdalenke?
57	tid.607854228766396416	Pa si na kavi. A veš ?
58	tid.427172744544129025	:) #LionelMessi #FCBarcelona #Barcelona #Spain @LenartRED http://t.co/Sdj69WXogm A ves uno, k se konec januarja obdarujes.....
59	tid.424628756381454336	@barjanski A ves tisto, ko jo obtozujejo, da je udbovski medij?
60	tid.652588875127332868	@crnkovic kino, a veš tist k piratizirajo in lahko istočasno gledajo drugje, tko kot blogje.
61	tid.657619321464750082	Rad grem pogledat arhitekturo, umetnost v cerkev, vera me ne zanima A veš tist, k čist slučajno napišeš neki, kar je slučajno naslov dobrega štikla od @mi2band , pa to favorizirajo, kt da bi bli na FB?
62	tid.347133191661948928	@superwolke kolo valjda A ves k se pot domov podaljša ko vozis v oblikis crke M?
63	tid.373869152001216512	žesneke padajo kot az satvo RT @MrVates: A veš tist, k nardiš par serij sklec preden greš ven?
64	tid.395909373866045440	My #hometown @ Stari grad Celje http://t.co/shsNRBPqAh A veš tisto, ko pol ure izpolnjuješ obrazec v tujem jeziku ob pomoči GTranslate.
65	tid.441154995652931585	Verjetno ni najbolj primerno, če sredi pisarne padeš v jok? A veš tisto "slabše ne more biti"?
66	tid.524562810941218816	V poslovnih prostorih pač optike NE rabimo, a veš ...
67	tid.381369117217136640	Saj na koncu bova še frenda z gospodom :) A veš tisto, ko ti PR ne odgovori na mail niti na SMS, potem pa se odzove na prvi klic in ves prijazen pove vse.
68	tid.519935523482529793	@MrDee25 Saj za Olimpijo bi isto komentiral, a veš ;)

Pogovorne strani Wikipedije ¹⁴⁹		
	ID dokumenta	Konkordanca
1	wikipedia.pagetalk.1377.1.28	Lahko ' naštepaš ' 8.036.0221.383 in eno škrbino , če želiš in če zmoreš - meni je načeloma vseeno , saj niso moji kalifornijski strežniki . A veš .
2	wikipedia.usertalk.529.45.2	P. S. Boš pogovor kaj arhiviral ? A veš , mal ' mi je dost ' takih ljudi , ki se samo zaj
3	wikipedia.usertalk.2458.20.3	Da Klemen ; kje sem se pa zmotil ... A veš , ker sam preverim , kaj trdim , vem , kaj si naredil narobe .

Komentarji na na novice ¹⁵⁰		
	ID dokumenta	Konkordanca
1	news.rtvsl.344089.9	Sketuskii in Zagloba v Sinijevičevi knjigi Z ognjem in mečem sta si že pred stoletji celila rane z mešanico kruhove sredice in pajčevine . A veš tisti občutek , ko sredi ničesar dobiš pajkovo mrežo na obraz ?

¹⁴⁹ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/j5sf86q>.

¹⁵⁰ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/jrpqo2g>.

Priloge

2	news.rtvsl.337150.11	* lapsus , album je seveda My Beautiful Dark Twisted Fantasy . a veš uno k maš ful keša in misliš da lahko vse kar ti prebliskne skozi plac , kjer naj bi bili možgani ?
3	news.rtvsl.330508.51	Fajn je it kdaj malo po svetu , da začneš malo bolj široko razmišljati , a veš ?
4	news.rtvsl.336798.2	Imaš prav , a veš , zgodovino pišejo vedno tisti , za katerimi ostanejo različni pisni dokumenti v takšni ali drugačni obliki (pa naj bo to na papirju , leseni tabli , glini , kamnu ...
5	news.rtvsl.347146.27	Ni dovolj imeti surovin , če jih pol ne moreš plasirati na trgu a veš .
6	news.rtvsl.337465.783	Razumem da je matematika težak predmet . A veš razmerje levica desnica je nekje na pol , malce v levo , verjetno volivcev SLS ne uvrščaš med kaviar volivce .
7	news.rtvsl.127017.222	sam pač KARABATIC je ime , tko k je biv ZIZU pr fuzbalu aveš !!!
8	news.rtvsl.127017.222	ju bol aveš , nkol se ne ve !!!
9	news.rtvsl.291642.11	Sam a ves SF010 , v Olimpiji se da tko fino dnar oprat pa po raznoraznih zepih poskrit ...
10	news.rtvsl.347145.74	analize in svetovaje bi pa itak delalo tisoče ljudi , teh ponudb je nenormalno veliko . a veš , dobil sem v svoj poštni predal vsaj nekaj sto povpraševalnih mejlov strani dužboslovcev , vrhunsko izobraženih , z vsebinskim in ličnim CV-jem .
11	news.rtvsl.344143.209	(a veš tisto : če me nisi prepričal kot zavarovalniški agent , to gotovo pomeni , da me je prepričal tisti od konukrence ;o) - niti ne pomislijo ne , da nas večina ne naseda nobenemu ;)
12	news.rtvsl.301639.135	Ja , šparat bi bilo bolje , ampak , a veš , ogromno ljudi tu že špara na mejah svojega obstoja .
13	news.rtvsl.80236.32	Hja , na zelo nizkem nivoju je , a veš ...
14	news.rtvsl.344213.7	Tudi v Nemčiji imajo take stranke , a veš (Social demokratska stranka Nemčije (Socialdemokratische Partei Deutschland - SPD) , Levica (Die Linke) ...
15	news.rtvsl.340977.78	Demokracija a veš , kjer lahko vsakdo izrazi svoje mnenje , tudi vrhovni sodnik .
16	news.rtvsl.340034.124	tako je to a ves .
17	news.rtvsl.345033.271	the law a ves kuk je raznij javnih uslužbencev poslancev smc ...
18	news.rtvsl.339515.107	cebelica2 A veš , nič nam ne pomaga anonimnost tu gor .
19	news.mladina.108659.35	ma pravico vsak do mnenja a ne Mladina ? aveš kao ...

Priloga 12: Vzorčeni primeri rabe označevalca *ne vem* v korpusu Janes

V tej prilogi navajamo primere rabe diskurznega označevalca *ne vem*, ki so bili identificirani v analizi 100 naključnih konkordanc izmed skupno 8.279 pojavitev te oblike v podkorpusu komentarjev na bloge (6 primerov), 3.477 pojavitev te oblike v podkorpusu blogovskih zapisov (2 primera), 42.920 pojavitev te oblike v podkorpusu forumov (13 primerov), 44.854 pojavitev te oblike v podkorpusu tvitov (13 primerov), 2.761 pojavitev te oblike v podkorpusu pogovornih strani wikipedije (3 primeri) in 8.711 pojavitev te oblike v podkorpusu komentarjev na novice (9 primerov).

Komentarji na bloge ¹⁵¹		
	ID dokumenta	Konkordanca
1	blog.publishwall.2788.12	Ni to uno , da bi ona bila Francozinja in bi jaz kar naenkrat znala francosko ... Ne vem , možno , kaj jaz vem , ampak v svojem primeru ne verjamem .
2	blog.publishwall.16877.1	Namaste , Deja Ne vem , ne vem , ampak vse je ob pravem času , tako , da si tudi to videla ob pravem času ;)
3	blog.rtvsl.7080.3	mah Vse kar je wild mi nekako bolj sede , je bolj prvinško , brez maskare ne vem ...
4	blog.rtvsl.10173.29	Vendar , od tega momenta je začela prvo piti po tem še po malem pasti travo in se je v dveh ali treh mesecih povrnila v normalno stanje . Ne vem .
5	blog.rtvsl.17871.7	a res ? ne vem ...
6	blog.rtvsl.22218.69	Ne vem , ne vem ...

Blogs ¹⁵²		
	ID dokumenta	Konkordanca
1	blog.rtvsl.1806.0	Če pa bo prišel, bo to fenomen. Ne vem , nekaj bo treba ukreniti.
2	blog.rtvsl.19877.0	A, današnja, ne vem , upam, da se motim, a se mi dozdeva, da le redki ohranjajo upornišvo v svojem srcu.

Forumi ¹⁵³		
	ID dokumenta	Konkordanca
1	forum.medovernet.8744617	pa dan mine . ne vem no , v blokovskih naseljih je en kup takih otrok .
2	forum.medovernet.9285161	sledi oglasnik tip 1 ne vem .
3	forum.medovernet.6923021	Znano je da jo lahko zatremo z KT , nosečnost , dojenje ne vem no
4	forum.avtomobilizem.7.92879.1579636	nimam kaj reči , prejšnji lastnik je redno servisiral avto pa vse napake imam zabeležene Ne vem no , prvi avto pri katerem imam res črno na belem in vidim zgodovino avta ter da so kilometri resnično realni .
5	forum.avtomobilizem.11.30874.771816	Joj nevem , težko .

¹⁵¹ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/hp3p9az>.

¹⁵² Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/zgfa6se>.

¹⁵³ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/zjw57m4>.

6	forum.avtomobilizem.30.76923.1319902	Stoji že 2 leti in nič ne kaže da bi blo kaj narobe , streha je gor pločevinasta Hosekra katere teža je okoli 7 kg / m2 , tak da ne vem no , meni se vse skupaj zdi še predimenzionirano in ne poddimenzionirano .
7	forum.avtomobilizem.56.66828.1402946	nastavek je za tipota tko , da ne vem ...
8	forum.avtomobilizem.52.75287.1287887	Za ta denar mogoče Ducato / jumper / boxer , Turbo Daily . Ne vem ko sem delal pri dostavi prek študenta , sem imel kombi brez turbine mal manj kot 90 konjev , brez težav je šel čez 100.
9	forum.avtomobilizem.2.55005.874649	ce se ne motim 5 x108 , lp ne vem , kar se pa tice renaulta pa majo eni 4 x100 eni pa
10	forum.avtomobilizem.6.84065.1420455	nisem lih mehanik Ne vem , bi bli kaki predhodni znaki , če bi cinšpula šla al to cknv v trenutku ?
11	forum.avtomobilizem.6.2079.23996	Ma jst mam ze od nekda nekaksen odpor do teh favoritk , ker nikamor ne grejo pa se ful prevec zrejo , tko da bi si raj za ta dnar kej druga kupu , ne vem , to je samo moje mnenje !
12	forum.avtomobilizem.11.35201.561997	Pa vecji gril , širši , višji , pa spredi bi mal agresivnejša mogla zgledat , vec ravnih ostrih kotov Ne vem men ni vsec
13	forum.avtomobilizem.18.99711.1628297	Ampak se peljem še marsikam drugam kot samo v službo , ne vem a eni sploh nikamor ne greste od bajte če ne greste na šiht ...

Tvit ¹⁵⁴		
	ID dokumenta	Konkordanca
1	tid.58879365278223361	@karolina83 Jaz se jih tudi , ampak kaj , ko prežijo izza vogala . Ne vem , se še učim .
2	tid.650406618576879616	Ampak ne vem , čudna sem .
3	tid.644947452009824257	A to še obstaja ? Ne vem , se mi zdi , da znajo biti alergični na vojake .
4	tid.407587908024303616	@BostjanJerko ne vem , to je riskantno opravilo , kaj če se razbije ?
5	tid.616688739235012608	Tale tekma v Domžalah pa vse bolj smrdi , ker tako felerji , kartoni in penali ... Ne vem .
6	tid.666964294257483778	Kaj pa recimo ... ne vem ...
7	tid.434033463323533312	@pengovsky Se spomniš suma korupcije pri zaposlitvi izraelskega kirurga v korist ... ne vem , pacientov ?
8	tid.410158301624954881	@darkob Ne vem no .
9	tid.435412026844659716	@leaathenatabako @Fin1Mark Bojda je moški , kar po zmedi na svetu verjamem . Ne vem .
10	tid.490942989435666433	@bosstjan ne vem .
11	tid.422742489192480768	@tamejhna v 5 pisem kolokvij . Ne vem , a so na fdvju štenge čez noč povečal ?
12	tid.654630018543419392	" Ne vem , pizda , če ni novga , bi kahko spet @JJansaSDS .
13	tid.373787928188956672	Begunje pa to je ja :) Sam ne vem ...

Pogovorne strani Wikipedije ¹⁵⁵		
	ID dokumenta	Konkordanca
1	wikipedia.pagetalk.1112.1.20	Ampak Janez se gotovo ne jezi tjavdan . Ne vem , sam o osebnih imenih nimam virov in se nikoli pravzaprav nisem ukvarjal z njimi .
2	wikipedia.pagetalk.4355.1.2	Hmm , ne vem - na navedeni spletni strani je Cerkev sv. Simona in Juda Tadeja , ampak pravilno naj bi bilo cerkev sv. Simona in Jude Tadeja .
3	wikipedia.pagetalk.6168.1.1	Članek je preveden iz angleške Wikipedije . ne vem , upam da mi ne bodo določeni osebki pobrisali tega komentarja

Komentarji na novice ¹⁵⁶		
	ID dokumenta	Konkordanca

¹⁵⁴ Povezava do nonkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/jyqgddj>.

¹⁵⁵ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/hcflfrx>.

¹⁵⁶ Povezava do konkordančnega niza v orodju noSketchEngine: <http://tinyurl.com/z8c84j6>.

Priloge

1	news.rtvsl.340058.88	Če so Slaku vzeli oddajo ?! Ne vem , Rosviti pustijo CELO samostojno oddajo večkrat tedensko s precej večjo minutažo , ne vem čemu Možini ne bi .
2	news.rtvsl.336533.1754	no nevem .
3	news.rtvsl.334163.13	Spomnite se nesreče tankerjev , eksplozije naftne ploščadi v Mehiškem zalivu , jezera kemikalij in odpadne surove nafte v Nigeriji in srednji Ameriki , nalivanje in izpuščanje morske vode v tankerje zaradi balasta , ko v pristanišču razložijo surovo nafto Ne vem meni se zdi , da v kolikor je možna enostavna reciklaža baterij tu po manj umazani tehnologiji tehnična preavaga v prid baterijam .
4	news.rtvsl.345504.251	k_ris , ne vem , si ti res prepričan , da smo se res odbili od dna ?
5	news.rtvsl.318057.304	) , ampak iz njih vre samo neka jeza , ljubosumje ... ne vem , delo za psihiatra , samo za obrazložitev zakaj ?
6	news.rtvsl.344207.63	Že do zdaj so bile poškodbe ... ne vem , no .
7	news.rtvsl.345427.49	Spet sami proruski komentarji tu gor . Ne vem al res noben ne zna bit vec objektivni in mu sovrastvo do zahoda resnicno tako zamegli um , da potem ne vidi ruskih barabij .
8	news.rtvsl.296596.5	" Druga " Švica Ne vem no .
9	news.mladina.163830.12	Pa zakaj ga silijo v zapor , če jih pa potem spuščajo ? Ne vem , ali sodišča obsojajo samo za to , da ljudje pridejo v zapor in vidijo kako je tam , potem jih pa spustijo domov .

IZJAVA O AVTORSTVU DOKTORSKE DISERTACIJE

Podpisana **Kaja Dobrovoljc**, z vpisno številko **18139017**, rojena **17. 2. 1986** v kraju **Ljubljana**, sem avtorica doktorske disertacije z naslovom ***Leksikalne prvine govorjenega jezika v uporabniških spletnih vsebinah: Primer večbesednih diskurznih označevalcev.***

S svojim podpisom zagotavljam, da:

- je predložena doktorska disertacija izključno rezultat mojega lastnega raziskovalnega dela;
- sem poskrbela, da so dela in mnenja drugih avtorjev oz. avtoric, ki jih uporabljam v predloženem delu, navedena v seznamu virov in so v delu citirana v skladu z mednarodnimi standardi in veljavno zakonodajo v RS na področju avtorskih in sorodnih pravic;
- je elektronska oblika identična s tiskano obliko doktorske disertacije;
- soglašam z objavo doktorske disertacije na spletnih straneh Filozofske fakultete Univerze v Ljubljani.

V Ljubljani, dne **9. marca 2018**

Podpis avtorice: _____