

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/330898753>

Raba tipično govorjenih diskurzivnih označevalcev na spletu

Article in *Slavistica Revija* · October 2018

CITATIONS

0

READS

28

1 author:



[Kaja Dobrovoljc](#)

Jožef Stefan Institute

22 PUBLICATIONS 66 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



New Grammar of Modern Slovene [View project](#)



Slovenščina na Dlaní Interactive e-Learning Portal [View project](#)

UDK 811.163.6'322.5:81'373:004

Kaja Dobrovoljc

Inštitut Jožef Stefan, Ljubljana

Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru

Filozofska fakulteta Univerze v Ljubljani

kaja.dobrovoljc@cjvt.si

RABA TIPIČNO GOVORJENIH DISKURZNIH OZNAČEVALCEV NA SPLETU

V članku predstavimo primerjavo rabe treh tipično govorjenih diskurznihih označevalcev v korpusu govorjene slovenščine Gos in korpusu slovenskih uporabniških spletnih vsebin Janes. Rezultati potrjujejo, da so ti izrazi na spletu rabljeni bistveno redkeje kot v spontanem govoru, vendarle pa njihova raba ni zanemarljiva, zlasti v besedilnih tipih s poudarjeno interaktivno oz. dialoško izmenjavo uporabniških sporočil. Pri tem se označevalci na spletu pojavljajo predvsem v semantično motiviranih funkcijah, kot so vzpostavljanje stika z naslovnikom, preverjanje strinjanja ali omiljevanje izrečenega, ki se hkrati prepletajo tudi z besedilnimi funkcijami poudarjanja ter menjave vlog. Prav tako na spletu razvijajo nekatere nove kontekste rabe, kot so nagovarjanje neznanega ali neudeleženega naslovnika, stilizacija in vstopanje v nove stalne besedne zveze.

Ključne besede: diskurzni označevalci, govorjeni jezik, uporabniške spletne vsebine, korpusno jezikoslovje

This article compares the usage of three speech-specific discourse markers in the Gos corpus of spoken Slovenian and the Janes corpus of user-generated content. The results confirm that the online usage of these expressions is much less common than in speech, but nevertheless significant, in particular on platforms encouraging interactive or dialogue-like communication. In online discourse, markers' semantic-based functions prevail, such as those relating to the addressee, checking for agreement, or attenuating an assertion. They intertwine with discourse-structuring functions, such as creating emphasis or turn managing. All three markers also exhibit some new contexts of online usage, such as addressing an unknown or generic addressee, producing a specific stylistic effect, and entering into new types of fixed multi-word expressions.

Keywords: discourse markers, spoken language, user-generated content, corpus linguistics

1 Uvod

Tako kot v jezikoslovju nasploh se tudi na področju analize diskurza in pragmatike računalniško posredovane komunikacije (Yus 2011, Herring idr. 2013) v zadnjih letih odpira povsem novo področje zvrstnostnih raziskav, analize t. i. uporabniških spletnih vsebin, kot so družbena omrežja, forumi, blogi in druge participativne spletne platforme, ki so se vzpostavile s spletom druge generacije (Herring in Androutsopoulos 2015).

Kot posebej aktualen vidik njihovega preučevanja se kaže raba tipično govornih leksikalnih prvin, saj tako kot za računalniško posredovano komunikacijo nasploh tudi za uporabniško spletno komunikacijo velja, da zaradi posebnih okoliščin, v katerih običajno nastaja, selektivno vsebuje tako prvine pisnega kot govornega jezika, ki se odražajo na različnih jezikovnih ravneh (Crystal 2011, Sindoni 2014).

Pri tem so se dosedanje domače raziskave prehajanja tipično govornih leksikalnih prvin v spletno komunikacijo osredotočale predvsem na vprašanja nestandardnega zapisa (Jakop 2008, Fišer idr. 2018), besedišča (Michelizza 2015, Zwitter Vitez in Fišer 2018) in skladnje (Arhar Holdt 2018), presenetljivo malo pa je raziskav spletne rabe diskurzni prvin, čeprav sorodne diskurznoanalitične raziskave za tuje jezike ugotavljajo, da govorniki v spletni komunikaciji razvijajo zelo specifične strategije upravljanja diskurza, od organizacije tem (Condon in Čech 2010, Bou-Franch idr. 2012) do menjave vlog (Garcia in Baker Jacobs 1999, Herring 1999) in uravnavanja medosebnih odnosov (Harrison in Allton 2013). Vprašanje, v kolikšni meri si pri tem pomagajo z rabo tipično govornih diskurznofunkcijskih jezikovnih sredstev, ki jih v nadaljevanju imenujemo diskurzni označevalci,¹ pa z izjemo analiz rabe specifičnih označevalcev v specifičnih tipih uporabniških spletnih vsebin (Fox Tree 2015, Rehbein 2015, Scheffler idr. 2016) ne v tujem ne v slovenskem prostoru doslej še ni bilo podrobneje raziskano.

Namen pričujočega prispevka² je zapolniti to vrzel in predstaviti celovito empirično raziskavo rabe tipično govornih diskurzni označevalcev v slovenskih uporabniških spletnih vsebinah. Po predstavitvi izbrane metodologije v drugem razdelku v jedrnem tretjem razdelku skušamo odgovoriti, (i) kako pogosto se najbolj tipični govorniki diskurzni označevalci rabijo v uporabniških spletnih vsebinah, (ii) v kolikšni meri je pogostost rabe najbolj tipičnih govornih diskurzni označevalcev odvisna od tipa uporabniških spletnih vsebin, in (iii) ali najbolj tipični govorniki diskurzni označevalci v uporabniških spletnih vsebinah opravljajo enako vlogo kot v govoru. Prispevek v četrtem razdelku sklenemo s sintezo najpomembnejših ugotovitev in njihovim ovrednotenjem z vidika možnosti nadaljnjih raziskav.

2 Gradivo in metoda

2.1 Izbrani tipično govorniki diskurzni označevalci

Dosedanje raziskave diskurzni označevalcev v govornem slovenščini (npr. Verdonik 2007, Smolej 2012) o *tipičnosti* za govorniki jezik sklepajo predvsem na podlagi njihove izstopajoče pogostosti v analiziranih korpusih govornega jezika. Z metodološkega vidika je tovrstno sklepanje o tipičnosti za govor težko upravičljivo, saj ni nujno, da

¹ V skladu s trenutnimi smernicami na področju aplikativne analize diskurza (Fischer 2014, Crible 2017) v tej raziskavi med diskurzne označevalce prištevamo vse propozicijsko neobvezne izraze s prepoznavno funkcijo diskurzne povezovanja, strukturiranja ali upravljanja. Za podrobnejšo opredelitev, merila in opis (nadpomenskega) razmerja do drugih sorodnih 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, glej disertacijo K. Dobrovoljc (2018: 75–100).

² Prispevek je nastal na podlagi analiz in ugotovitev doktorske disertacije K. Dobrovoljc (2018, 5. poglavje).

so vsi pogosti izrazi v govoru rabljeni predvsem v govoru, kakor tudi ni nujno, da so izrazi, rabljeni predvsem v govoru, v govoru tudi izstopajoče pogosti. To stanje je pred kratkim skušala preseči raziskava K. Dobrovoljc (2018), ki je s statistično primerjavo pogostosti pojavljanja diskurznihi označevalcev v referenčnih korpusih govornje in pisne slovenščine izluščila nabor tipično govorenih večbesednih diskurznihi označevalcev, med katerimi so se kot trije najbolj tipični izkazali označevalci *a ne*, *ne vem* in *a veš* (Dobrovoljc 2018: 189–205). Čeprav se omenjena raziskava osredotoča zgolj na večbesedne izraze, menimo, da ti izrazi predstavljajo ustrezno izhodišče za raziskave tipično govorenih označevalcev nasploh, saj se omenjajo v večini sorodnih raziskav označevalcev v govoru, s svojo kratkostjo in visoko stopnjo gramatikaliziranosti pa relativizirajo tudi sam kriterij večbesednosti (Dobrovoljc 2017).

2.2 Referenčna korpusa govornje in spletne slovenščine

Korpus govornje slovenščine Gos je največja prostodostopna zbirka posnetkov in transkripcij spontanega govora v slovenskem jeziku (Verdonik in Zwitter Vitez 2011). Vsebuje transkripcije približno 120 ur posnetkov (1 milijon pojavnic) spontanega oz. nepripravljenega govora v različnih vsakodnevnih sporazumevalnih situacijah, ki so uravnotežene glede na demografske lastnosti govorcev (spol, starost, regija, izobrazba), prenosnik (TV, radio, telefon, osebni stik) in vrsto govornega dogodka (javni informativni in izobraževalni, javni razvedrilni, nejavni nezasebni in nejavni zasebni govor). V tej raziskavi smo uporabljali različico Gos 1.4, kakršna je dostopna preko prostodostopnega konkordančnika noSketchEngine³ (Erjavec 2013), do izvornih posnetkov govora pa smo dostopali preko specializiranega konkordančnika na uradni spletni strani.⁴

Korpus slovenskih uporabniških spletnih vsebin Janes je nastal v okviru istoimenskega projekta, posvečenega jezikoslovni analizi nestandardne slovenščine na spletu.⁵ Čeprav zaradi omejitev, povezanih s tehničnimi in pravnimi vidiki pridobivanja in shranjevanja uporabniških spletnih vsebin, korpus Janes vsebuje zgolj en segment uporabniških spletnih vsebin, predstavlja doslej največjo zbirko računalniško posredovane komunikacije v slovenščini. V tej raziskavi smo uporabili različico Janes 0.4 (Fišer idr. 2016), ki vsebuje nekaj več kot 200 milijonov pojavnic iz šestih različnih besedilnih tipov: tvtov (51,4 % pojavnic celotnega korpusa), forumov s treh specializiranih forumskih portalov (22,6 %), blogov in komentarjev nanje na dveh bloggerskih portalih (16,6 %), komentarjev na novice treh novičarskih portalov (7,1 %) ter pogovorne strani na slovenski Wikipediji (2,3 %). Tudi do tega korpusa smo dostopali preko konkordančnika noSketchEngine.

2.3 Vzorčenje in izračun pogostosti rabe diskurznihi označevalcev

V prvem koraku smo za izbrane označevalce pridobili podatek o pogostosti rabe v obeh korpusih. Ker se vsi trije izrazi v jeziku pojavljajo tudi v drugih vlogah, denimo

³ <https://www.clarin.si/noske/>

⁴ <http://www.korpus-gos.net/>

⁵ <http://nl.ijs.si/janes/>

kot del propozicijske vsebine (npr. *tega ne vem*) ali kot del daljšega diskurznega označevalca (*saj ne vem*), izračun pogostosti rabe teh označevalcev v korpusu zahteva predhodno razdvoumljanje. Za ta namen so bile na področju korpusne analize diskurza v literaturi predlagane različne metode (Stede 2014), ki se opirajo na površinsko obliko besedila, različne tipe jezikoslovnih oznak in/ali lastnosti sobesedilnega konteksta in s tem predpostavljajo obstoj korpusa, ki je na razmeroma zanesljiv način (najpogosteje ročno) označen na več različnih jezikoslovnih ravneh. Ker korpusa Janes in še zlasti Gos tovrstnih informacij ne vsebujeta oz. so bile te pripisane z omejeno strojno natančnostjo, smo za izračun pogostosti v obeh korpusih uporabili preprosto metodo sklepanja po vzorcu (Dobrovoljc 2018: 191–194), s katero smo o številu pojavitev diskurznega označevalca v korpusu sklepali na podlagi deleža diskurznih rab v ročno pregledanem naključnem vzorcu 100 konkordanc celotnega korpusa (Gos)⁶ oz. posameznih besedilnih tipov (Janes). Ti vzorci so bili nato v drugem koraku uporabljeni tudi za kvalitativno primerjavo funkcij označevalcev.

V primerjavi s korpusom Gos, v katerem so bile standardizirane ustreznice pogovornim oblikam (npr. *ne em, a vejš, ane*) pripisane ročno, smo glede na omejeno natančnost strojne normalizacije oblik v korpusu Janes (Ljubešić idr. 2014) pred tem identificirali še vse različice izvirnega zapisa označevalcev na podlagi tipologije nestandardnih zapisovalnih posebnosti v slovenščini (Čibej 2018). S hevrstičnimi poizvedbami smo tako identificirali več kot 120 različnih zapisov izbranih označevalcev, kot relevantne pa smo v nadaljnji analizi upoštevali samo tiste z vsaj pet pojavitvami v korpusu, 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 veeem, ne um, neuem, ne vemm*), 14 različnih zapisov niza *a ne* (*a ne, ane, anne, a n, aneee, a neee, aaa ne, anee, a neeee, aneeee, aaaa ne, aneeeeee, 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 veeeeeš, aveeeeš, a wes*).⁷

3 Rezultati

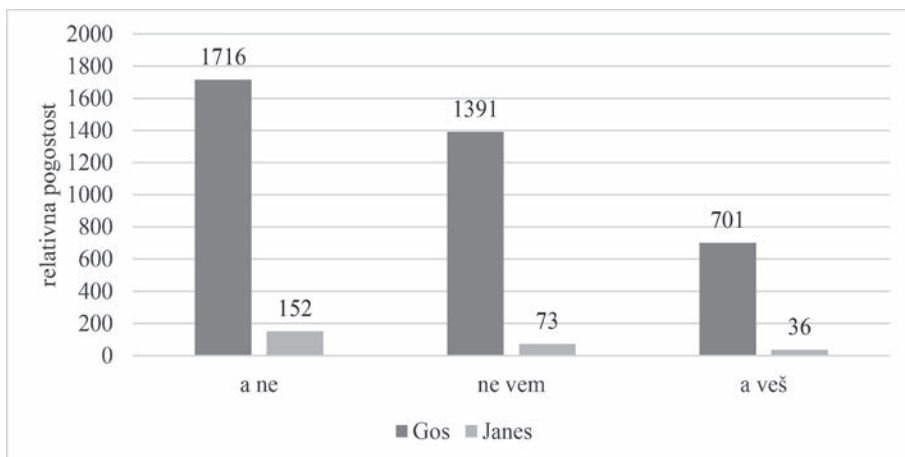
3.1 Pogostost rabe v uporabniških spletnih vsebinah

Kot prikazujejo izračunani podatki o pogostosti rabe tipično govorjenih označevalcev v korpusih Gos in Janes na Sliki 1, ki jih obenem sopostavljamo še s pogostostjo rabe v referenčnem korpusu pisne slovenščine Kres (Logar Berginc idr. 2012), se tako kot v govorjeni rabi v uporabniških spletnih vsebinah najpogosteje pojavlja označevalec

⁶ Čeprav funkcija izdelave naključnega vzorca v konkordančniku noSketchEngine ne upošteva distribucije besedilnih metapodatkov, analiza vzorčenih pojavitev (za sezname glej Dobrovoljc 2018) kaže, da vzorci izkazujejo podobno razmerje med javnim in nejavnim diskurzom kot celota vseh pojavitev v korpusu, tj. 51 % proti 49 % za *a ne* (celota: 56 % proti 44 %); 22 % proti 79 % za *ne vem* (celota: 29 % proti 71 %); 13 % proti 87 % za *a veš* (celota: 18 % proti 82 %).

⁷ Pogosta raba oblik v zapisu skupaj (npr. *nevem, aveš, ane*) potrjuje potrebo po relativizaciji koncepta večbesednosti pri tovrstnem tipu stalnih besednih zvez, obenem pa tudi razkriva izjemen potencial korpusa Janes za nadaljnje raziskave tega vprašanja, saj o govorčevi percipiji večbesednosti izrazov doslej ni bilo mogoče sklepati ne v (lektoriranih) korpusih pisnega jezika ne v (konsenzualno transkribiranih) korpusih govora.

a ne (152 pojavitev na milijon pojavnic), približno enkrat manj pogosto se uporablja označevalec *ne vem* (83 pojavitev na milijon), najredkeje pa se uporablja označevalec *a veš* (36 pojavitev na milijon). To dokazuje, da se tipično govoreni diskurzni označevalci z nezanemarljivo pogostostjo pojavljajo tudi v spletni rabi, na podlagi česar lahko ovržemo predpostavke nekaterih starejših razprav (prim. npr. Crystal 2011: 22), ki so domnevale, da se tovrstni izrazi zaradi svoje specifične vloge v govoru v računalniško posredovani komunikaciji ne rabijo.



Slika 1: Primerjava relativne pogostosti najbolj tipično govorenih diskurznih označevalcev v korpusu govorne slovenščine Gos, korpusu uporabniških spletnih vsebin Janes in korpusu pisne slovenščine Kres.

Vendarle pa nadaljnja primerjava pogostosti rabe v vseh treh tipih diskurza kaže, da so tipično govoreni diskurzni označevalci v uporabniških spletnih vsebinah 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 diskurznih označevalcev v govoru predvsem posledica tistih okoliščin tvorjenja govornega 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 diskurznih 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.⁸ Da pa njihov potencialni vpliv na rabo označevalcev vendarle

⁸ Dejstvo, da E. Fox Tree (2015) v svoji raziskavi rabe nekaterih tipično govorenih diskurznih označevalcev v spletnih klepetalnicah ugotavlja, da sta označevalca *well* in *oh* tam rabljena približno enako pogosto kot v spontanem pogovoru (mi pa tako kot I. Rehbein (2015) v analizi rabe zapolnjenih premorov na Twitterju za vse tri označevalce ugotavljamo bistveno večji prepad), obenem sugerira, da je delež prehajanja leksikalnih prvin govornega jezika v računalniško posredovano komunikacijo močno odvisen od tipa tovrstne komunikacije (npr. spletne klepetalnice na eni in uporabniške spletne vsebine na drugi strani).

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.

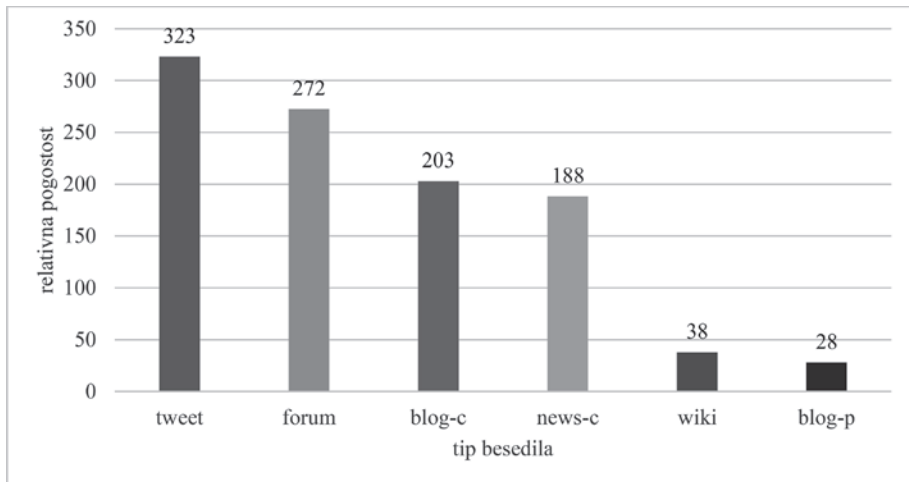
3.2 Raba glede na tip uporabniških spletnih vsebin

Kot smo izpostavili že v uvodu, uporabniških spletnih vsebin kljub nekaterim skupnim značilnostim ni mogoče obravnavati kot enotne jezikovne zvrsti, saj na jezikovno rabo znotraj teh platform vplivajo številni tehnološki in družbeni dejavniki (Herring 2007). To potrjujejo tudi rezultati primerjave rabe tipično govornih diskurzni označevalcev, saj se pogostost rabe med posameznimi besedilnimi tipi precej razlikuje (Slika 2). Ti so najpogosteje rabljeni na družbenem omrežju Twitter, 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 zapisah.

Kot dejavnik, ki najbolj vpliva na splošno rabo diskurzni označevalcev v spletnih besedilih, se glede na temeljne razlike v namenu in tehnoloških možnostih kaže predvsem stopnja interaktivnosti oz. dialoški posameznih besedilnih tipov oz. platform, na katerih se ti pojavljajo. Najbolj dinamična je namreč izmenjava besedil na družbenem omrežju Twitter, kjer uporabniki z namenom jedrnat 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 s strani administratorjev običajno sankcionirane. Podobno velja tudi za komentarje statičnih vsebin, 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 izhodiščno statično vsebino.

Če bi torej blogovske zapise v primerjavi z dialoški (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 diskurzni označevalcev, pa to predpostavko, ki bi jo veljalo v prihodnje preveriti tudi eksperimentalno, pod vprašaj postavlja presenetljivo nizka raba diskurzni označevalcev na pogovornih straneh Wikipedije. Te strani so namreč prav tako namenjene izmenjavi mnenj med uporabniki, in sicer bodisi izražanju predlogov glede izboljšave vsebin posameznih geselskih člankov (t. i. geselske pogovorne strani) bodisi pošiljanju sporočil posameznemu uporabniku Wikipedije (t. i. uporabniške pogovorne strani), najpogosteje glede uporabnikovih geselskih člankov ali (ne)spoštovanja pravil urejanja. V obeh podtipih uporabniških strani Wikipedije se namreč pogosto razvije diskusija v obliki več replik dveh ali več uporabnikov, podobna tistim na spletnih forumih. Pri pojasnjevanju izrazito redkejša raba diskurzni označevalcev bi zato poleg stopnje interaktivnosti (ki je vsekakor bližje dialoški kot statični) veljalo upoštevati tudi

druge specifike komunikacije na pogovornih straneh slovenske Wikipedije (Michelizza 2015), kot so struktura udeležencev (majhna skupnost), odnosi med njimi (spoštljivost), namen platforme (izobraževanje), tehnološke omejitve (platforma ne spodbuja diskusij), obstojnost sporočil (dolgotrajno viden arhiv) itd. Vsi ti dejavniki lahko namreč pomembno vplivajo na razvoj ustaljenih konvencij komuniciranja v določeni skupini in posledično sooblikujejo jezikovne izbire uporabnikov, ki želijo tej skupini pripadati.

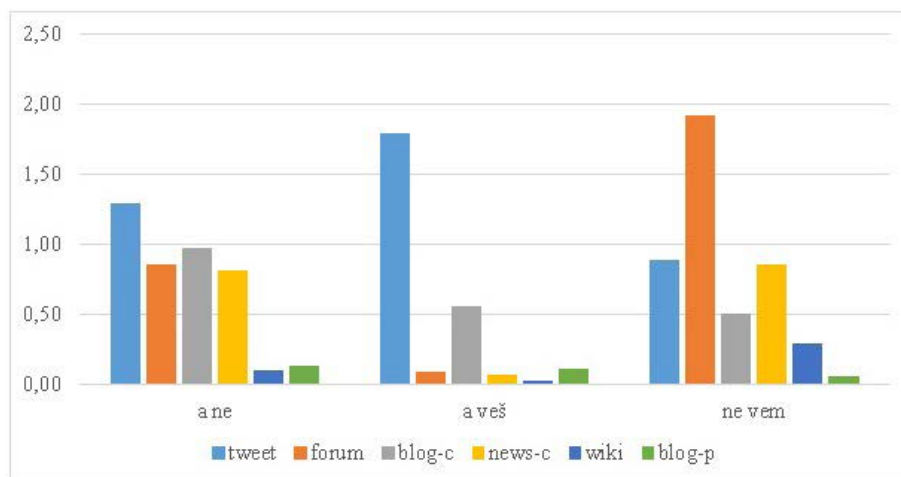


Slika 2: Primerjava relativne pogostosti tipično govorjenih diskurzivnih označevalcev v šestih besedilnih tipih korpusa Janes.

Glede na izrazno in funkcijsko raznolikost vseh treh obravnavanih označevalcev nas je v drugem koraku zanimala še primerjava njihove posamične 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.⁹ Prikaz distribucije po posameznih besedilnih tipih za vsakega izmed označevalcev na Sliki 3 potrjuje, da kljub enotni 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.

Najbolj enakomerno rabo glede na različne tipe uporabniških spletnih vsebin izkazuje označevalec dvogovora *a ne*, ki se najpogosteje 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, kar lahko delno pojasnimo s tem, da tvorci pri pisanju blogov redko nagovarjajo (konkretnega) naslovnika.

⁹ <https://www.sketchengine.eu/documentation/statistics-used-in-sketch-engine/>



Slika 3: Primerjava relativne besedilnotipske pogostosti posameznih označevalcev.

Precej bolj raznolika je besedilnotipska distribucija označevalca *a veš*, ki 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, na katerem uporabniki sami zamejujejo svoj krog sledilcev oz. naslovnikov, medtem ko tvorca besedil na drugih tipih platform običajno naslavljajo širši krog običajno neznanih naslovnikov. Izjema so komentarji na blogovske zapise, v katerih se avtorji komentarjev (vključno s tvorcem izvornega zapisa) v diskusiji pogosto tikajo.

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 *ne vem* najpogosteje uporablja v besedilih spletnih forumov. Glede na to, da ta označevalec 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 mnenj, ne pa drugim tipom dialoških izmenjav ali odzivov.

3.3 Značilnosti rabe v uporabniških spletnih vsebinah

V zadnjem koraku raziskave tipično govorjenih diskurzni označevalcev smo po analizi kvantitativnih kazalnikov njihove rabe na spletu s kvalitativnim pregledom vzorčenih primerov rabe v kontekstu (za seznam glej Dobrovoljc 2018: 310–329) želeli izvedeti še, ali obravnavani diskurzni označevalci v spletni komunikaciji opravljajo podobno vlogo kot v spontanem govoru. Pri tem funkcije označevalcev v govoru, ki

so bile podrobneje predstavljene že v raziskavah K. Dobrovoljc (2018), M. Smolej (2012) in D. Verdonik (2007), zaradi omejenega prostora zgolj na kratko povzemamo.

3.3.1 Raba označevalca *a ne*

3.3.1.1 Raba označevalca *a ne* v korpusu Gos

Najbolj tipični diskurzni označevalec govorjenega jezika *a ne* se v govoru pojavlja v dveh tipičnih vlogah: kot zapolnjevalec vrzeli, s katerim govorec poudarja oz. zapolnjuje prehode med deli besedila in eksplicitno vzdržuje stik z naslovnikom, preverja njegovo razumevanje ali strinjanje (primer [1]), in kot usmerjevalec poteka interakcije, s katerim tvorec zaključuje svojo vlogo in s pozivom k izmenjavi mnenj vlogo predaja naslovniku (primer [2]).

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

3.3.1.2 Raba označevalca *a ne* v korpusu Janes

V enaki krovni skupini lahko v splošnem razdelimo tudi rabo označevalca *a ne* v korpusu Janes – povedne rabe na eni (primer [3]) in vprašalne rabe (primer [4]) na drugi strani – vendar vsaka izkazuje nekaj pomembnih razlik glede na rabo v govoru.

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

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

Najbolj opazna razlika med obema modalnostma je zagotovo ta, da je povedna raba 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, težko govorimo o kognitivni vlogi zapolnjevanja vrzeli med načrtovanjem nadaljnjega besedila, kot bi jo na primer lahko pripisali v primeru [1].

Namesto tega označevalci v spletnih vsebinah opravljajo druge funkcije, ki pa niso enostavno določljive. Kot skupno lastnost lahko sicer prepoznamo tvorčevo namero po vzpostavljanju stika z naslovnikom oz. naslovniki, vendar v nekaterih primerih označevalec deluje predvsem kot poudarjalec splošne veljavnosti soledne trditve, ki

bi jo lahko parafrazirali z »saj se vsi strinjamo« ali »saj vsi vemo« (primer [5]), tudi v pokroviteljskih oz. sarkastičnih rabah (prim. Rehbein 2015), spet v drugih pa se zdi, da je označevalec rabljen predvsem z namenom dodatne omilitve oz. relativizacije izrečenega (primer [6]), ki že prehaja iz trdilne v vprašalno rabo.

[5] *je pa tko **ane**: če ne znaš nič s telefonom počet, kupi iOS (iphone)*

[6] *Sej **ane** o kromu bo treba se globoko premislit.*

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 govornjo rabo, kjer tvorec z rabo označevalca naslavlja konkretnega naslovnika oz. več njih, ki se lahko na vprašanje odzove v realnem času (glej primera [1] in [2] zgoraj), v uporabniških spletnih vsebinah tvorca naslavlja več tipov naslovnikov. Poleg konkretnih posameznikov (primer [4]) se namreč označevalec uporablja tudi pri naslavljanju splošnega naslovnika, tj. vseh uporabnikov z dostopom do tvorčevih vsebin, bodisi za poziv k izmenjavi mnenj bodisi za poudarjanje samoumevnosti izrečenega (primer [7]), prav tako pa uporabniki z označevalcem *a ne* pogosto naslavlja konkretno posameznike, ki v razpravi sploh niso udeleženi, kot so javne osebnosti. To ironično naslavljanje neudeleženega naslovnika je značilno zlasti v komentarjih novic, v katerih so ti posamezniki omenjeni, ali ob omembi tem, ki jih javnost s to osebo običajno povezuje (primer [8]).

[7] *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**?*

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

3.3.2 Raba označevalca *a veš*

3.3.2.1 Raba označevalca *a veš* v korpusu Gos

Tudi z označevalcem *a veš* govorniki na semantični ravni vzpostavljajo in vzdržujejo stik z naslovnikom, z besedilnega vidika pa bi njegove raznolike mikrofunkcije v spontanem govoru lahko razvrstili v dve krovni skupini: na rabe, v katerih označevalec *a veš* uvaja novo besedilno enoto v obliki dodatnega pojasnila, argumentacije ali spremembe ubeseditve in torej kaže v besedilu naprej (primer [9]), in rabe, v katerih označevalec zaključuje in poudarja predhodno trditev brez implikacije njenega nadaljevanja (kaže v besedilu nazaj), kot v primeru [10].

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

[10] *rastejo rastejo sam treba jih je nabirat a veš | aja ... ja ja ja*

Poleg teh dveh načinov členjenja besedila, med katerima meja ni vedno nedvoumno določljiva, pa lahko tudi v primeru označevalca *a veš* govorimo o hkratni vlogi zapolnjevalca vrzeli (vzpostavjalca tekočnosti besedila), ki tvorcu omogoča daljši načrtovalni čas, zlasti kadar se označevalec pojavlja med več zaporednimi segmenti tvorčevega diskurza (primer [11]).

[11] *ja sam mmm tko je a veš Bled je ful turističen ful bl turističen je*

3.3.2.2 Raba označevalca *a veš* v korpusu Janes

Tudi v korpusu Janes se v analiziranih primerih rabe označevalca *a veš* uporablja tako v vlogi uvajalca nove vsebine oz. pojasnila (primer [12]) kot v vlogi poudarjalca predhodne vsebine (primer [13]). Raba obeh je približno enako pogosta.

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

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

Podobno kot smo izpostavili že pri analizi označevalca *a ne*, uporabniki z označevalcem *a veš* na spletu nagovarjajo raznolik nabor naslovnikov. Čeprav se ta oblika v spontanem govoru običajno uporablja zgolj v dialoški interakciji z naslovnikom, s katerim je tvorec vzpostavil prijateljski odnos, poleg tovrstnega nagovarjanja znancev ali prijateljev na spletu (primer [14]) uporabniki na spletu s tem označevalcem pogosto naslavljajo tudi uporabnike, ki jih osebno ali poimensko ne poznajo, a jih kljub temu tikajo (primer [15]). 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.¹⁰

[14] *@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>*

[15] *Sam a ves SF010, v Olimpiji se da tko fino dnar oprat pa po raznoraznih zepih poskrit ..*

Zanimivo se poleg nagovarjanja konkretnih znanih ali neznanih uporabnikov kljub svoji izrazni zaznamovanosti v uporabniški spletni komunikaciji uporablja tudi pri

¹⁰ V slovenskem prostoru sistematične 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 skladijskih in diskurzivnih vlogah, kaže, da oblika *a veš* v drugi osebi ednine močno prevladuje (9.339 proti 873 pojavitev v korpusu Janes).

naslavljanju neopredeljenega splošnega naslovnika, denimo v blogovskih zapisih, v katerih se avtorji povezujejo z neidentificiranim neznanim bralcem (primer [16]). V tem besedilnem tipu se sicer označevalec *a veš* uporablja predvsem v poročanem govoru (primer [17]), kar kaže, da so tipično govorniki diskurzni označevalci pomembno slogovno sredstvo pri ustvarjanju učinka pristnosti poročanega govornega dogodka.

[16] *Jebaj ga zdaj, saj bi bilo tudi malo, (maaaalo, 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 [...]*

[17] *„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.*

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* plus samostalniška zveza v imenovalniku (primer [18]), najpogosteje z določilom ali jedrom v obliki kazalnega zaimka, bistveno redkeje pa tudi s časovnim odvisnikom (primer [14] zgoraj).¹¹

[18] *A veš tisto, k si na obisku in se ti neki strese, pa upaš da te noben ni vidu. No to.*

3.3.3 Raba označevalca *ne vem*

3.3.3.1 Raba označevalca *ne vem* v korpusu Gos

Z označevalcem *ne vem* govorniki na semantični ravni relativizirajo svojo stopnjo zavezanosti glede natančnosti posredovanih informacij ali neomajnosti izraženega mnenja, z besedilnega vidika pa lahko njegovo vlogo v diskurzu členimo še na tri tipične skupine, glede na pozicijo v izjavi oz. vlogi. V rabah sredi izjave ali vloge, v katerih je označevalec rabljen pred popravkom ali spremembo načrtovane ubeseditve, pred navedbo primera ali specifikacijo podrobnosti (primer [19]) se epistemska funkcija nedoločnosti najbolj izrazito prepleta z besedilno funkcijo vzdrževanja tekočnosti govora (zapolnjevanjem vrzeli). Z rabo na začetku vloge ali tematske sekvence govorec nakaže

¹¹ 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(li) 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.

prevzem vloge oziroma nadaljevanja diskurza (primer [20]), z rabo ob zaključku vloge pa vlogo z namenom izmenjave mnenj preda sogovorniku (primer [21]).

[19] *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*

[20] *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*

[21] *se lahko poročajo pa to otroke pa to **ne vem** | jah ja to nega kaj se to samo eee katoliška vera nema*

3.3.3.2 Raba označevalca *ne vem* v korpusu Janes

Vse tri položaje oz. vloge označevalca *ne vem* smo zaznali tudi v analiziranih primerih rabe v korpusu Janes, vendar s precej drugačno distribucijo. Če v spontanem govoru prevladuje raba označevalca kot napovedovalca priključka ustrezne vsebine ali ubeseditve, kakršna se običajno povezuje s položajem na sredini izjave (primer [19]), so tovrstni primeri v spletni komunikaciji zelo redki, bolj kot zapolnjevanje vrzeli pri priključku pa je poudarjena njihova semantična vloga omiljevanja vsebine, ki sledi (primer [22]). Kot posebnost uporabniških spletnih vsebin lahko izpostavimo tudi dejstvo, da je ta vloga označevalca pogosto rabljena ironično, kot v primeru [23], v katerih uporabnik z rabo zveze *ne vem* poudarja samoumevnost informacije, ki sledi.

[22] *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.*

[23] *@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 govornika (glej primera [20] zgoraj in [24] spodaj), v uporabniških spletnih vsebinah prevladuje predvsem raba na začetku sklepnih trditev, ki se običajno pojavljajo proti koncu govorčevega besedila oz. vloge (primer [25]). Ko uporabnik predstavi svoje mnenje, v izogib konfliktu z drugimi uporabniki končni sklep omili z rabo označevalca, s čimer obenem tudi spodbuja nadaljnjo izmenjavo mnenj. 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 (primer [26]).

[24] *_ris, ne vem, si ti res prepričan, da smo se res odbili od dna?*

[25] *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 .*

[26] *Tale tekma v Domžalah pa vse bolj smrdi, ker tako felerji, kartoni in penali ... Ne vem.*

4 Sklep

V prispevku smo s kombinacijo kvantitativnih in kvalitativnih korpusnih metod raziskali rabo treh tipično govornih diskurznihih označevalcev *a ne*, *a veš* in *ne vem* v referenčnih korpusih govornih in spletnih slovenščine. Pokazali smo, da se tudi v pisni komunikaciji uporabnikov spleta druge generacije najbolj tipično govornih diskurznihih označevalci dejansko pojavljajo bistveno redkeje kot v govoru, kar potrjuje, da je pogosta raba analiziranihih diskurznihih označevalcev v govoru predvsem posledica specifičnih prostorskih in kognitivnih okoliščin tvorjenja govornega diskurza. Toda po drugi strani je raba teh označevalcev na spletu vendarle prisotna, podrobnejša primerjava njihovih funkcij v govornih in spletnih komunikacijah pa kaže, da v spletni rabi poleg besedilnih funkcij poudarjanja ter prevzemanja ali predajanja vloge izstopajo predvsem epistemske in dialoške 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 govornih ne v pisni komunikaciji. Raba tipično govornih diskurznihih označevalcev v uporabniških spletnih vsebinah tako dokazuje, da to niso zgolj redundantna, pomensko izpraznjena »mašila«, s katerimi tvorci kompenzirajo kognitivni pritisk tvorjenja v realnem času, temveč jih spletni uporabniki prepoznavajo kot pomembne gradnike posredovanihih sporočil, tudi v kontekstih, kjer je 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 govornih jezikovnih sredstev. Če izhajamo iz predpostavke, da obstaja le en jezikovni kontinuum, ki ga ni mogoče ločevati na ločena sistema pisnega ali govornega jezika, lahko dihotomijo med značilnostmi govornih in pisnih rabe pripišemo predvsem dejstvu, da so konvencije pisnih rabe, ki delno odražajo različnost tudi samih namenov in okoliščin med obema tipoma sporazumevanja (Verdonik 2015 str. 395), jezikovnim uporabnikom običajno priučene skozi proces izobraževanja oz. socializacije. Preklapljanje iz govornih 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 standardiziranihih konvencij preklapljanja ali njihovega sistematičnega usvajanja. Jezikovne izbire spletnih uporabnikov so torej zelo raznolike in individualizirane ter tudi redko družbeno sankcionirane. Čeprav naša raziskava kaže, da raba vseh treh

tipično govorenih diskurzivnih označevalcev govorenega 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 (diskuzijskih) pogovornih straneh Wikipedije, odpirajo širok prostor za nadaljnje raziskave dejavnikov pojavljanja tipično govorenih prvin na spletu. Poleg vpliva specifičnih lastnosti posameznih tipov uporabniške spletne komunikacije – tudi tistih, ki v korpus Janes niso bili zajeti – bi namreč veljalo raziskati tudi vpliv drugih okoliščin, zlasti demografskih značilnosti samih uporabnikov.

VIRI IN LITERATURA

- Špela ARHAR HOLDT, 2018: Korpusni pristop k skladnji računalniško posredovane slovenščine. *Viri, orodja in metode za analizo spletne slovenščine*. Ur. D. Fišer. Ljubljana: FF. 228–53.
- Patricia BOU-FRANCH, Nuria LORENZO-DUS in Pilar GARCES-CONÉJOS BLITVICH, 2012: Social Interaction in YouTube Text-Based Polylogues: A Study of Coherence. *Journal of Computer-Mediated Communication* 17/4. 501–21.
- Sherri L. CONDON in Claude G. ČECH, 2010: Discourse Management in Three Modalities. *Language@Internet* 7. 1–32.
- Ludivine CRIBLE, 2017: Towards an operational category of discourse markers: A definition and its model. *Discourse markers, Pragmatics Markers and Modal Particles: New Perspectives*. Ur. Andrea Sanso in Chiara Fedriani John Benjamins. 101–26.
- David CRYSTAL, 2011: *Internet Linguistics: A Student Guide*. London: Routledge.
- Jaka ČIBEJ, 2018: Regionalne jezikovne različice v slovenski računalniško posredovani komunikaciji: korpusni pristop z ročno označenim korpusom Janes-Geo. *Viri, orodja in metode za analizo spletne slovenščine*. Ur. D. Fišer. Ljubljana: FF. 160–97.
- Kaja DOBROVOLJC, 2017: Multi-word discourse markers and their corpus-driven identification: The case of MWDM extraction from the reference corpus of spoken Slovene. *International Journal of Corpus Linguistics* 22/4. 551–82.
- Kaja DOBROVOLJC, 2018: *Leksikalne prvine govorenega jezika v uporabniških spletnih vsebinah: primer večbesednih diskurzivnih označevalcev: Doktorska disertacija*. Ljubljana.
- Tomaž ERJAVEC, 2013: Korpusi in konkordančniki na strežniku nl.ijs.si. *Slovenščina* 2.0 1/1. 24–49.
- Kerstin FISCHER, 2014: Discourse Markers. *Pragmatics of Discourse*. Ur. K. P. Schneider, A. Barron. Berlin: De Gruyter Mouton. 271–94.
- Darja FIŠER, Tomaž ERJAVEC in Nikola LJUBEŠIĆ, 2016: JANES v0.4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina* 2.0 4/2. 67–100.
- Darja FIŠER, Maja MILČEVIĆ PETROVIĆ in Nikola LJUBEŠIĆ, 2018: Zapisovalne prakse v spletni slovenščini. *Viri, orodja in metode za analizo spletne slovenščine*. Ur. D. Fišer. Ljubljana: FF. 124–39.
- Jean E. FOX TREE, 2015: Discourse markers in writing. *Discourse Studies* 17/1. 64–82.

- Angela Cora GARCIA in Jennifer BAKER JAKOBS, 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. 337–67.
- Sandra HARRISON in Diane ALLTON, 2013: Apologies in email discussions. S. C. Herring, Dieter Stein in Tuija Virtanen: *Pragmatics of Computer-Mediated Communication*. Boston: De Gruyter Mouton.
- Susan C. HERRING, 1999: Interactional Coherence in CMC. *Journal of Computer-Mediated Communication* 4/4.
- Susan C. HERRING, 2007: A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet* 4.
- Susan C. HERRING in Jannis ANDROUTSOPOULOS, 2015: Computer-Mediated Discourse 2.0. *The Handbook of Discourse Analysis*. Ur. D. Tannen, H. E. Hamilton in D. Schiffrin. Hoboken: John Wiley & Sons. 127–51.
- Susan C. HERRING, Dieter STEIN in Tuija VIRTANEN, 2013: *Pragmatics of computer-mediated communication*. Boston: De Gruyter Mouton.
- Nataša JAKOP, 2008: Pravopis in spletni forumi - kva dogaja? *Slovenščina med kulturami* - Slovenski slavistični kongres. Ur. M. Košuta. Ljubljana: Slavistično društvo Slovenije. 283–94.
- Nataša LOGAR BERGINC, Miha GRČAR, Marko BRAKUS, Tomaž ERJAVEC, Špela ARHAR HOLDT in Simon KREK, 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: Gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Založba FDV.
- Nikola LJUBEŠIĆ, Tomaž ERJAVEC in Darja FIŠER, 2014: Standardizing tweets with character-level machine translation. *Lecture Notes in Computer Science*. Ur. A. Gelbukh. Berlin, Heidelberg: Springer. 164–75.
- Mija MICHELIZZA, 2015: *Spletna besedila in jezik na spletu*. Ljubljana: Založba ZRC, ZRC SAZU.
- Ines REHBEIN, 2015: Filled pauses in user-generated content are words with extra-propositional meaning. *Proceedings of the Second Workshop on Extra-Propositional Aspects of Meaning in Computational Semantics*. Ur. E. Blanco. Denver: ACL. 12–21.
- Tatjana SCHEFFLER, Rike SCHÜLTER, Manfred STEDE, 2016: Discourse Structuring Devices on Twitter. *COST TextLink Second Action Conference*. Na spletu.
- Maria Grazia SINDONI, 2014: *Spoken and written discourse in online interactions: A multimodal approach*. London: Routledge.
- Mojca SMOLEJ, 2012: *Besedilne vrste v spontanem govoru*. Ljubljana: ZIFF.
- Manfred STEDE, 2014: Resolving connective ambiguity: A prerequisite for discourse parsing. *The pragmatics of discourse coherence: Theories and applications*. Ur. H. Gruber in G. Redeker. 121–41.
- Darinka VERDONIK, 2007: *Jezikovni elementi spontanosti v pogovoru: Diskurzni označevalci in popravljanja*. Maribor: Slavistično društvo.
- Darinka VERDONIK, 2015: Govorjeni proti pisnemu ali katera leksika je »tipično govorenjenak«. *Slovar sodobne slovenščine: problemi in rešitve*. Ur. V. Gorjanc, P. Gantar, I. Kosem in S. Krek. Ljubljana: ZIFF. 392–405.
- Darinka VERDONIK in Ana ZWITTER VITEZ, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Ana ZWITTER VITEZ in Darja FIŠER, 2018: Govorne prvine v nestandardni spletni slovenščini. *Viri, orodja in metode za analizo spletne slovenščine*. Ur. D. Fišer. Ljubljana: FF. 254–73.

Francisco YUS, 2011: *Cyberpragmatics*. Amsterdam: John Benjamins Publishing Company.

SUMMARY

There is a dearth of research on the usage of speech-specific discourse-signaling devices in user-generated content. This article presents a corpus-based comparison of three typically-spoken discourse markers (*a ne* 'right', *a veš* 'you know', and *ne vem* 'I don't know') in spoken and user-generated Slovenian to determine (i) how often typically spoken discourse markers occur in user-generated content, (ii) to what extent the usage of typically spoken discourse markers depends on the type of user-generated content, and (iii) the similarities and differences in the usage of typically spoken discourse markers in spoken and online communication.

Based on an extensive qualitative and quantitative comparison of the Gos reference corpus of spoken Slovenian and the Janes reference corpus of user-generated content, our findings confirm that the usage of speech-specific discourse-signaling devices in user-generated content is much less common than in speech, but nevertheless significant. The usage of speech-specific discourse markers also depends on the type of user-generated content, such as tweets, forum exchanges, blog posts, or comments. In particular, the degree of interactivity or dialogue-like communication can be seen as the most influential factor in discourse marker usage (i.e., the more a platform encourages dynamic exchange between users, the more frequent the usage of typically spoken discourse markers is). In addition, a comparison of text-type distribution analysis for each of the three shows that their usage also depends on their semantic characteristics and the specific social features of each platform.

A subsequent in-depth analysis of the discourse marker usage in both corpora confirms that, in speech, speech-specific 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 attenuate an assertion), intertwined with discourse-structuring functions, such as emphasising or turn taking/giving. All three markers under investigation also exhibit some new types of usage in user-generated content, such as addressing an unknown or generic addressee, producing a humorous or narrative stylistic effect, and entering into new types of fixed multi-word expressions.