

Govorne prvine v nestandardni spletni slovenščini

Ana Zwitter Vitez, Darja Fišer

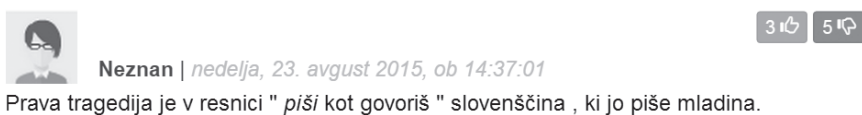
Izvleček

Komunikacija na forumih, družbenih omrežjih in novičarskih portalih je pogosto označena kot hibrid med govorom in standardnim pisnim jezikom. Da bi presegli ta stereotip, smo poskušali raziskati dejanske specifične govornega diskurza in računalniško posredovane komunikacije s pomočjo analize seznamov ključnih besednih oblik v korpusih Gos in Janes glede na korpus Kres. Rezultati kažejo, da so na besednovrstni ravni besedila računalniško posredovane komunikacije bližje pisnim besedilom kot govornemu diskurzu ter da se od standardnega zapisa najbolj odmikajo tviti in pozitivno naravnani komentarji na novičarskih portalih, najmanj pa besedila forumov. Na ravni besedišča so posebej zanimivi za govor specifični elementi interakcije z drugimi udeleženci, ki so najpogosteje prisotni v tvitih in komentarjih. Rezultate raziskave bi lahko uporabili pri razmisleku o teoriji zvrstnosti za slovenščino in jih vključili v snovanje novih jezikovnih priročnikov.

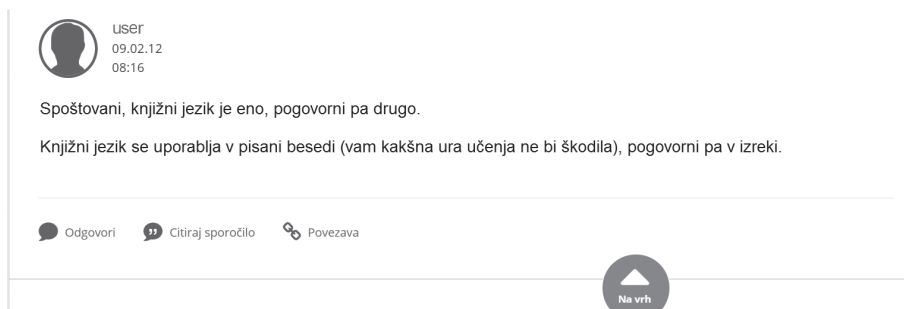
Ključne besede: govorneni diskurz, računalniško posredovana komunikacija, neformalna komunikacija, korpusna analiza, elementi interakcije

1 UVOD

Računalniško posredovana komunikacija je v zadnjih desetletjih dobila pomembno vlogo na različnih raziskovalnih področjih zaradi vedno večje količine javnih, prosto dostopnih neformalnih besedil, s katerimi se uporabniki vsakodnevno odzivajo na aktualne družbene dogodke. Zaradi kombinacije javnosti besedil in dejstva, da jih lahko ustvarja slehernik in ne več poklicni pisec, kot je to veljalo nekoč, so tovrstna besedila pogosto deležna stereotipnih oznak (Sliki 1 in 2). Prvi stereotip sporoča, da je jezik družbenih medijev in računalniško posredovane komunikacije nasploh podoben govoru. Drugi stereotip sporoča, da je govorna komunikacija neformalna, pisna pa formalna.



Slika 1: Stereotip o podobnosti med govorom in jezikom družbenih medijev.¹



Slika 2: Prepričanje, da morajo biti pisna besedila vedno standardna.²

Velik del neformalne komunikacije najverjetneje dejansko poteka v govorni obliki, vendar so njeni elementi vse bolj prisotni tudi v pisni slovenščini. To velja predvsem za besedila novih medijev (forumi, komentarji, blogi in družbena omrežja), ki jih tvorimo v interakciji z drugimi uporabniki. Čeprav se pri marsikateri prvini takih besedil ponuja misel »vpliva govora na pisni jezik« (Slika 1), velja tovrstne razlage vzeti z veliko mero previdnosti in uzavestiti misel D. Tannen (1982: 14), da ravnamo zmotno, »kadar primerjamo besedila različnih žanrov, razlike med njimi pa pripišemo razliki med govorjenim in pisnim jezikom.«

1 <http://www.delo.si/kultura/knjizevni-listi/znanstvena-slovenscina-tragedija-v-dveh-dejanjih.html>

2 <https://med.over.net/forum5/viewtopic.php?t=1146255>

Toporišičev model zvrstnosti (Toporišič 1991) sicer prevede, da sta tako govorjeni kot pisni jezik lahko vpeta v bolj ali manj formalne okoliščine vsakokratnega sporazumevalnega položaja, zato vpeljuje kategorijo socialnih zvrsti, vendar favorizira pisni knjižni jezik, normativizacijo govorjenega jezika pa skuša reševati s splošnim pogovornim jezikom (Gruden 2013). Čeprav lahko govorjeni jezik znotraj Toporišičevega modela nastopa tako znotraj knjižnih in neknjižnih zvrsti, se je v slovensko zavest veliko bolj kot celotna slika zvrstnosti zakoreninila dialektika med pogovornim (tj. neformalnim, govorjenim) jezikom (Slika 1) in knjižnim (tj. formalnim, pisnim) jezikom (Slika 2), ki je Toporišič ni predvidel.

Tovrstne stereotipe je mogoče preseči z empiričnimi raziskavami. Zato smo poskušali raziskati stopnjo prekrivnosti med prvinami, ki so tipične tako za besedila računalniško posredovane komunikacije kot za govorjeni diskurz. S pomočjo frekvenčnih seznamov in seznamov ključnih besed smo identificirali specifične korpusov tvitov, komentarjev, forumov in korpusa Gos glede na referenčni pisni korpus Kres in nato zaznane specifične primerjali med seboj na besednovrstni in leksikalni ravni ter na ravni odstopanja od standardnega zapisa oziroma izgovorjave.

2 RAZISKAVE GOVORJENEGA DISKURZA IN RAČUNALNIŠKO POSREDOVANE KOMUNIKACIJE

Analize računalniško posredovane komunikacije so izjemno uporabne za področje marketinga (Tedeschi et al. 2015) in mnenjskih raziskav (Pang et al. 2008, Cambria et al. 2013, Smailović et al. 2013), saj lahko s tovrstnimi metodami ugotavljamo politično naravnano volilcev, ugotavljamo zadovoljstvo strank s produkti ali zaznavamo centre nestrpnosti v družbi. Številne raziskave so že vzpostavile povezavo med specifikami spletnega in govorjenega diskurza. Jezikoslovno naravnane študije so precej široko zasnovane in pogosto zajemajo ravni zapisa, leksične, skladnje in pragmatike (Chovanec 2009, Nyström 2013), vendar je zaradi nereprezentativnosti vzorcev njihove zaključke težko posploševati. Kvantitativne analize so izvedene na velikih vzorcih besedil in večinoma usmerjene v posamezno jezikovno raven ali strukturo, kot so na primer frekvence besed (Leech in dr. 2001), raznolikost besedišča (Bamman in dr. 2014) ali segmentiranje diskurzivnih enot (Baron 2010). Nekatere študije podobnost med govorjenim in spletnim diskurzom vidijo tako očitno, da specifične jezika novih medijev uporabljajo kot izhodišče za raziskovanje govorjenega diskurza pri avtomatskem razpoznavanju govora (Vaufreydaz et al. 1999).

Vendar Crystal, ki je že pred leti preučeval specifične različnih zvrsti govorjenega in pisnega jezika (1995), opozarja (2001), da spletne komunikacije ne smemo

obravnavati kot govora, ki bi ga nekdo zapisal, saj gre za pisno komunikacijo, ki jo pogosto zaznamujejo neformalne komunikacijske okoliščine. Baron pa dodaja (2010), da tudi sodobne raziskave novih medijev svoje izsledke posplošujejo na celoten diskurz računalniško posredovane komunikacije, čeprav primerjajo zelo različne spletne besedilne zvrsti (blogi, spletne klepetalnice, forumi ipd.).

V slovenskem raziskovalnem prostoru govorjeni diskurz in pisna računalniško posredovana komunikacija do zdaj bili sistematično obravnavani skupaj, zato navajamo raziskave, ki so prispevale k poznavanju govorjenega jezika, in tiste, ki se ukvarjajo s pisnim spletnim diskurzom. Schlamberger Brezar (1998) v spontanem govorjenem diskurzu analizira retorične principe v argumentaciji, Smolej (2006) skladnjo in pragmatične prvine, Verdonik (2006) diskurzne označevalce in popravljanja, Zwitter Vitez (2009) realizacijo komunikacijskih strategij, Huber (2013) poudarek in pavzo, Tivadar (2015) razmerje med domačimi in tujimi prvini v govorjenem jeziku.

Omenjene študije so izjemnega pomena za poznavanje produkcijskih razlik med govorjenimi in pisnimi besedili, ki močno vplivajo na zgradbo govornih oziroma pisnih enot diskurza, saj brez dvoma drži, da je »med govorjenim in pisnim jezikom celo vesolje«³ (Morel in Danon Boileau 1998: 7). Ne obravnavajo pa specifik govorjenega diskurza brez vnaprej postavljenih hipotez. Korpusno zasnovane analize spontanega diskurza, ki temeljijo izključno na analizi gradiva, so izvedli Verdonik in Kosem (2012) ter Zwitter Vitez (2016), ki so na korpusu Gos izvedli raziskavo tipičnih prvin posameznih besedilnih zvrsti glede na formalnost sporazumevalnega položaja.

Specifike besedil novih medijev obravnavajo naslednje raziskave: Kalin Golob (2008) analizira kratka sporočila in v njih identificira številne pojave nestandardne rabe ločil, fonetični zapis, opuščanje pomožnega glagola in nedoločnika. Michelizza (2008) ugotavlja, da je jezik kratkih sporočil drugačen od pisnega jezika zaradi vpliva svetovnega spleta, slenga, angleščine in omejitev dolžine. Tudi rezultati raziskav o rabi slovenščine v elektronskih sporočilih (Dobrovoljc 2008) in na spletnih forumih (Jakop 2008) kažejo, da v tovrstni komunikaciji izstopa vpliv narečnih glasovnih posebnosti, prevzete besede iz tujih jezikov, čustveno zaznamovane besede, mašila in krajšave.

Predstavljene raziskave besedil novih medijev so zasnovane z vnaprej postavljenimi hipotezami o prisotnosti določenih jezikovnih struktur v analiziranih besedilnih žanrih, vendar med rezultati omenjajo tudi prisotnost govornih prvin. Zato se zdi smiselno izvesti celovito primerjavo specifik govorjenega diskurza in pisnih besedil računalniško posredovane komunikacije.

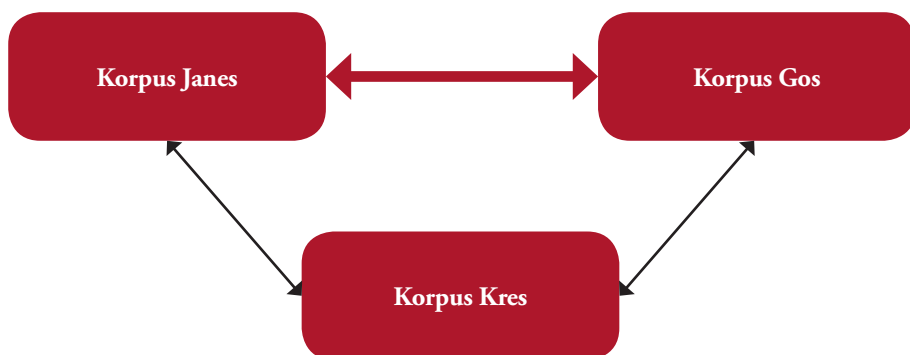
3 V izvirniku: »De l'oral à l'écrit, il y a un monde.«

3 METODOLOGIJA

Prvine govora v besedilih novih medijev smo analizirali s primerjavo treh korpusov slovenskega jezika:

- uravnoreženi korpus pisnih besedil Kres, ki zajema 100 milijonov pojavnic (Logar Berginc et al. 2012),
- korpus govornjene slovenščine Gos v obsegu milijon pojavnic (Verdonik in Zwitter Vitez 2011),
- korpus slovenskih spletnih uporabniških vsebin z 200 milijoni pojavnic (Erjavec et al. 2018).

Analiza je potekala v dveh fazah: najprej smo iz korpusa Gos izluščili tipične prvine govora na podlagi primerjave z uravnoreženim korpusom pisnih besedil Kres, potem pa smo pojavljanje identificiranih govornih prvin preverili v korpusu spletnih besedil Janes (Slika 3). Kvantitativno in kvalitativno analizo smo izvedli na ravni besednih vrst, besedišča in stopnje odmika od standarda .



Slika 3: Primerjava specifik korpusa Janes in korpusa Gos glede na korpus Kres.

Za pridobivanje podatkov iz korpusov smo uporabili orodje Sketch Engine⁴ (Kilgarriff et al. 2004), ki omogoča vpogled v konkordance in primerjavo različnih korpusov s pomočjo relativnih frekvenčnih seznamov in seznamov ključnih oblik, ki se v določenem korpusu pojavijo zelo pogosto in obenem v primerjanem korpusu niso pogoste.

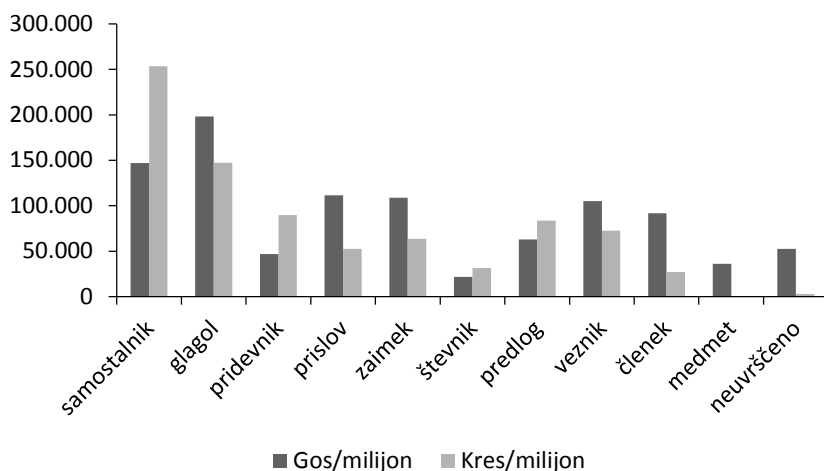
⁴ <http://www.sketchengine.co.uk>

4 ANALIZA

4.1 Besedne vrste

S pomočjo frekvenčnih seznamov smo analizirali zastopanost posameznih besednih vrst v korpusih Gos, Kres in Janes. Besedne vrste so bile določene avtomatsko v skladu s priporočili za oblikoslovno označevanje JOS (Erjavec et al. 2010).

4.1.1 Specifike govora



Graf 1: Zastopanost besednih vrst v korpusu Gos in korpusu Kres.

Kvantitativna primerjava porazdelitve besednih vrst pokaže, da je v korpusu govorjene slovenščine glede na korpus Kres izrazito več glagolov, prislovov, zaimkov, veznikov, členkov in medmetov.

Da se bodo specifike govora pokazale skozi večjo zastopanost medmetov, členkov in zaimkov, smo pričakovali, saj te besedne vrste lajšajo sprotno tvorjenje in razumevanje govorjenih besedil. Nekoliko preseneča večja zastopanost glagolov, vendar podrobnejša analiza gradiva pokaže, da razliko v veliki meri tvorijo glagoli v vlogi diskurzivnih označevalcev, kot na primer *mislim* in *vedeti* (3):

- (1) *zakaj kako a veš mislim eee poznaš eee [ime] od prej?*

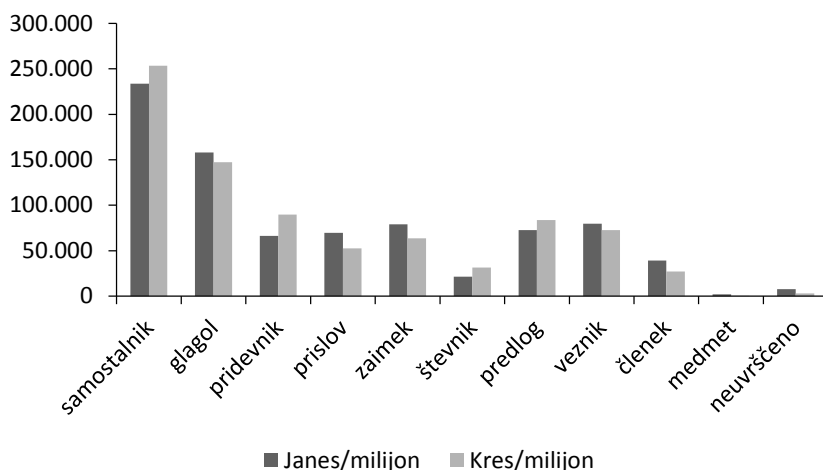
Poleg glagolov v vlogi diskurzivnih označevalcev prihaja v spontanem govoru do pogostih ponavljanj, ki so posledica zanimivega »uresničevanja paradigmatične osi

na sintagmatski osi« (Blanche Benveniste 1991, Smolej 2004), ko se elementi, med katerimi izbiramo znotraj svojega jezikovnega sistema, nakopičijo eden za drugim (*zakaj, kako, a veš, mislim, eee*). S pojavom kopičenja paradigmatičnih elementov lahko razložimo tudi pogosto rabo veznikov (primer 2), ki sicer v pisnih besedilih kažejo na večjo skladenjsko kompleksnost besedil.

- (2) zaradi tega **ker** smo **ker** eee **ker** živimo oziroma **ker** so takrat že živeli v času ko so informacije dejansko z lahkoto krožle

Tako glagoli s pragmatično vlogo diskurzivnih označevalcev kot uresničevanje paradigmatične osi na sintagmatski omogočata sprotno tvorjenje govornih besedil, številni premori in ponovljene strukture pa pogosto olajšata tudi razumevanje govornih besedil.

4.1.2 Govorne prvine v spletnih besedilih



Graf 2: Besedne vrste v korpusu Janes in korpusu Kres.

Korpus Janes se po zastopanosti analiziranih besednih vrst sicer razlikuje od korpusa Kres, vendar se statistično umešča bliže pisnim kot govornim besedilom. V besedilih novih medijev so medmeti sicer sedemkrat bolj pogosti kot v korpusu Kres, vendar večinoma oponašajo neverbalne izraze in smeh (*haha, aja, hm*), ne odsevajo pa sprotnega procesa tvorjenja besedila kot v govoru (*eee, eem, mmm*), saj pisanje spletnih besedil vseeno omogoča več časa za tvorjenje koherentnih enot kot govor:

- (3) srečkovička **haha** .. ti prvošm

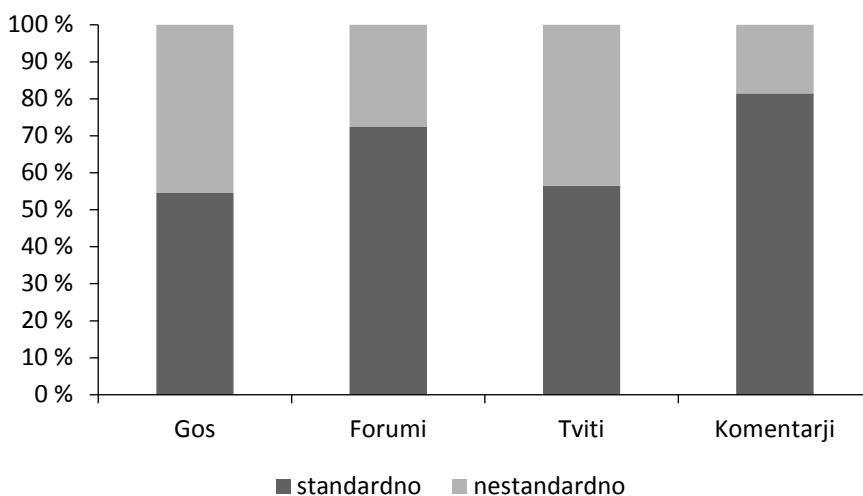
Podobnost med govorjenimi besedili in računalniško posredovano komunikacijo se kaže v manjši stopnji števnikov in pridevnikov glede besedila korpusa Kres. To bi lahko bila posledica tematske usmerjenosti posameznih besedilnih zvrsti korpusa Kres (primer 4) ali bistveno krajšega postopka tvorjenja spletnih in govorjenih besedil.

- (4) Tudi **dvakratni olimpijski** zmagovalec Gianni Romme in »**srebrni**«
Bob de Jong napovedala svoje slovo iz reprezentance

Analiza besednih vrst je pokazala, da je v govorjenih in spletnih uporabniških vsebinah bistveno več zaimkov, členkov in medmetov, manj pa pridevnikov in števnikov. Vendar kvalitativna analiza gradiva pokaže, da so v govoru prisotni drugi medmeti kot v računalniško posredovani komunikaciji, kjer prav tako ne opazimo povečanega števila glagolov in veznikov, ki so posledica sprotnega tvorjenja govorjenih besedil.

4.2 Odmik od standardne izgovorjave in zapisa

Med tipičnimi besednimi oblikami korpusov Gos in podkorpusov Janes smo želeli ugotoviti, v kolikšni meri besedišče, značilno za govor in računalniško posredovano komunikacijo, odstopa od priporočil izgovorjave (brez upoštevanja mesta naglase in drugih prozodičnih lastnosti) oz. zapisa v SSKJ. Ključnim besednim oblikam v korpusih Gos in Janes smo pripisali oznako standardno/nestandardno (Graf 3).



Graf 3: Odmik od standarda v govorjenem korpusu in korpusih Janes.

Delež tipičnih besed v govoru in spletnih besedilih, ki imajo drugačno standardno obliko po priporočilih SSKJ, se giblje med 19 % in 46 %: največ takšnih besed najdemo v govoru (46 %) in v korpusu tvitov (43 %). Ko podrobneje analiziramo govorno gradivo, lahko zasledimo trend padanja standardne izgovarjave oz. naraščanja nestandardne izgovarjave od javnega informativnega diskurza proti nejavnemu zasebnemu diskurzu (Zwitter Vitez 2016). Pa vendar v prispevku Zwitter Vitez (2016) ugotavljamo, da je nestandardna izgovorjava precej prisotna tudi v formalnih besedilnih zvrsteh (32 %), kot kaže primer 5.

- (5) smo policisti srečujemo se b reku z vsemi oblikami eee b reku bolečin stisk ampak men bi tuki zdaj največ pomenlo če bi recimo mmm ta punca ostala živa ne

Tudi računalniško posredovana komunikacija pogosto poteka v obliki neformalnih diskusij. Zato ne preseneča visoka stopnja nestandardnega zapisa na Twitterju (43 %), kar odseva prevladujočo sproščenost vzdušja med Twitteraši (primer 6).

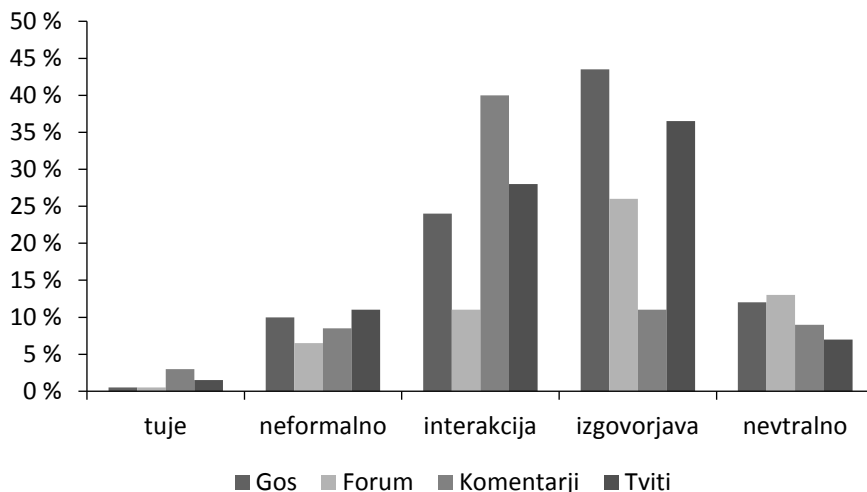
- (6) - ma čak, a ne bi mogu bit model že v čuzi? koga se čaka?
- sodne počitnice so. al neki.
- a to tud za Furs vela?

Komunikacija na forumih je drugačne narave, saj avtorji besedila po navadi objavlajo, da rešijo konkreten problem z določenega področja (zdravstvo, gospodinjstvo, avtomobilizem ipd.), zato so tudi odgovori strukturirani bolj formalno. Komentarji na spletne novice so specifičen žanr, pri katerem smo v raziskavi Zwitter Vitez in Fišer (2016) zaznali zanimiv pojav, da naravnost komentarja vpliva na stopnjo standardnosti sporočila. V pozitivnih komentarjih smo namreč v primerjavi z negativnimi zaznali veliko primerov zapisa, ki odstopa od standarda (*dejmo*), ter pogosto rabo velikih tiskanih črk (BO), nestandardno rabljenih ločil (*Dajmo klobasica!!!*;) in emotikonov (:)))). Pri negativnih komentarjih pa je opaziti manj nestandardnega zapisa (*Če zaupaš našim medijem, so še skoraj vsako leto bile kritike glede naših pesmi pozitivne, ampak rezultata pa nobenega in isto bo letos*). Možno razlago za zaznano razliko na ravni standardnosti komentarjev vidimo v dejstvu, da se avtorji pozitivnih komentarjev lažje identificirajo s pripadnostjo družbeno-demografskim skupinam s tipičnimi jezikovnimi specifikami (*Dejmo, klobasica!!!*), pozicioniranje negativnega mnenja pa je bolj odvisno od sprejemanja skupnosti, zaradi česar avtorji uporabljajo manj zaznamovane jezikovne strukture.

4.3 Besedišče

Seznami ključnih besed (angl. *keywords*; Scott 1997) omogočajo preučevanje tipičnega konteksta, v katerem se te besede uporabljajo, na podlagi prevladujočega

konteksta pa naprej analiziramo njihovo vlogo v jezikovni rabi. Tipične rabe besed s seznamov ključnih besed smo kategorizirali v štiri skupine: nestandardna izgovorjava ali zapis (*lobk*), elementi interakcije (*hej*), neformalno besedišče (*spedenan*) in tuje besede (*jes*).



Graf 4: Kategorizacija besedišča v korpusu Gos in podkorpusi korpusa Janes.

Kot je pokazal že odmik od standardne izgovorjave oz. zapisa, v korpusu tvitov najbolj izstopa izrazita podobnost z govorno slovenščino glede na posebnosti izgovorjave oz. zapisa (43 % Gos in 36 % Twitter).

Kategorija interakcije je v vseh analiziranih korpusih dobro zastopana, posebej pa izstopa v korpusu komentarjev, ki mu sledi korpus Gos. Kvalitativna analiza pa pokaže, da so komentarji v opazovanih besedilnih zvrsteh precej različni. V komentarjih najdemo elemente, ki izražajo naravnost oz. sodbo avtorja besedila (*bravo, verjetno, škoda, sramota*), medtem ko se v tvitih in v govoru pojavljajo drugačni elementi interakcije, ki vzdržujejo stik z naslovnikom (*hej, mhm, a veš*).

(7) **A veš, v bistvu** blokira sama sebe. (Twitter)

A veš in v bistvu na papirju izzveni kot da je to neka tuja firma **ne**. (Gos)

Neformalnih besed ni prav veliko (med 6 % in 11 %), so pa v različnih podkorpusih precej podobne (*ful, glih, kao*). Kot kaže primer 8, se v tvitih neformalne besede (*razirat, našajfat*) prepletajo z elementi interakcije (*jah, sej veš*), in emotikoni (☺), kar ustvarja sproščeno vzdušje na Twitterju.

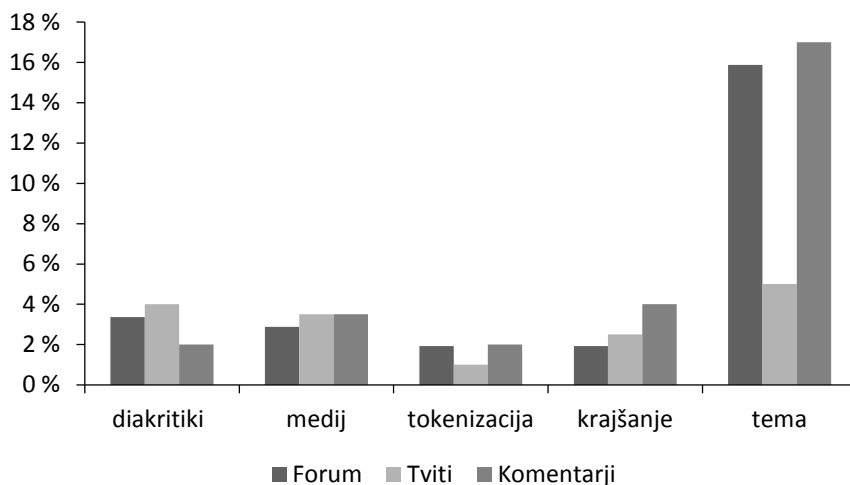
(8) jah sej veš.. za razirat se, morš bit našajfan ☺

Tujih besed je zanemarljivo malo, vendar so v tem primeru rezultati analize ključnih besed vprašljivi, saj je zajem besedil zaradi zaznavanja jezika potekal s pomočjo izločanja gradiva z neslovenskimi besedami.

Na ravni besedišča je govornemu jeziku najmanj podoben podkorpus forumskih sporočil, kjer smo zaznali manj nestandardnega zapisa (25 %), neformalnih izrazov (6 %) in manj elementov interakcije kot v korpusu Gos. To verjetno korelira s formalnostjo forumskih tematik (višja stopnja pri zdravstvenih vprašanjih, nižja pri avtomobilističnih temah). Medtem ko uporabniki Twitterja uživajo v besednih igrah, so forumski uporabniki pragmatično usmerjeni v pridobivanje in izmenjavo informacij.

4.3.1 Elementi računalniško posredovane komunikacije

Tipične elemente računalniško posredovane komunikacije, ki zaradi narave komunikacijskega kanala ne morejo biti prisotni v korpusu Gos, smo analizirali posebej. To so kategorije izrazov, tipičnih za tematiko diskusije (*volitve*), odsotnosti diakritičnih znamenj (*a ves*), besede, vezane na medij (*lajkaš*), specifične v tokenizaciji (*neboš*) in postopki krajšanja besed (*slo*).



Graf 5: Tipični elementi računalniško posredovane komunikacije.

Izrazi, povezani s tematiko diskusije, so največkrat samostalniki in so najpogosteje prisotni v forumih (e.g. *avto*, *problem*) in v komentarjih (e.g. *tekma*, *volitve*). To ne preseneča, saj so bili forumi korpusa Janes zajeti glede na tematiko diskusije,

komentarji na novičarskih portalih pa so že po definiciji povezani z aktualnimi tematikami (npr. volitve). Besedila korpusa Gos in tviti pa so bili zajeti neodvisno od teme diskusije.

Vsi trije podkorpusi Janes kažejo veliko število ključnih besed, ki odsevajo jezikovno rabo družbenih omrežij (*com*, *všeč*, *videoposnetek*). Ti izrazi so bili prvotno specifični za eno platformo komunikacije (Youtube, Facebook), kasneje pa so se razširili na celotno računalniško posredovano komunikacijo (*všečkati*, *lajkati*).

Izpuščanje diakritičnih znamenj, krajšanje besed in nestandardna tokenizacija so se pri analizi sicer pojavile, vendar niso relevantne za kvantitativni del analize. Gre namreč za prvine, ki jih določeni uporabniki uporabljajo na različnih besedah, na seznamu ključnih 200 besed pa opazujemo le najpogostejše besede, na katerih smo te prvine zaznali (*veš/ves*). Tako ne moremo izvedeti, v kolikšnem deležu računalniško posredovane komunikacije je ta pojav dejansko prisoten. Podobno težavo opazimo pri pojavu nestandardne tokenizacije, ki ga prav tako lahko opazujemo le na primeru najpogostejših besednih oblik (*ne bi/nebi*). Prav tako opazimo, da v preučeni primerih nestandardne tokenizacije, ki je sicer bolj prisotna v forumskih sporočilih in komentarjih na spletne novice, ni opaziti znakov intencionalne igre z jezikom, kar morda odseva pomanjkanje jezikovne kompetence in ne jezikovne kreativnosti.

4.4 Elementi interakcije

Posebej zanimiva kategorija, ki smo jo zaznali tako v govorjenih kot v spletnih besedilih, so elementi interakcije z ostalimi udeleženci, prek katerih avtorji besedil utrjujejo svojo identiteto, gradijo samopodobo in razkrivajo svoj odnos do zunajjezikovne realnosti. Z dodatno analizo te kategorije smo želeli ugotoviti, kateri tipični elementi govorne interakcije so prisotni v besedilih računalniško posredovane komunikacije in kakšna je njihova vloga v vsakodnevni komunikaciji na spletu. Predvidevali smo, da lahko prek identifikacije elementov interakcije in njihovih funkcij sklepamo o komunikacijskem namenu tvorca besedila in posledicah njegovega jezikovnega udejstvovanja.

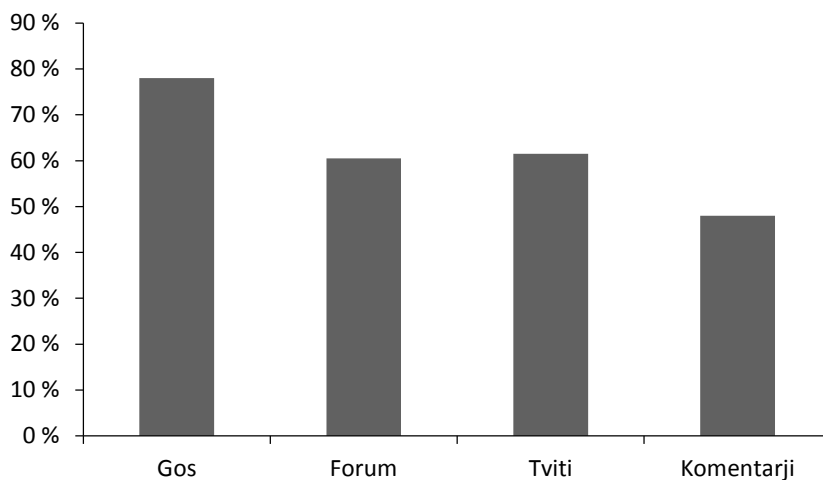
Ker je primarna raziskava besedišča pokazala, da so prvine interakcije med udeleženci prisotne tudi v sicer drugače razporejenih kategorijah (npr. pri izrazu *pejd*, pri katerem lahko zaznamo tako odstop od standardnega zapisa kot interakcijo z drugimi udeleženci), smo prvine interakcije v govoru in spletnih besedilih na novo kategorizirali in podrobneje preučili. Izluščili smo ključne besedne oblike (pojavnice) korpusa Gos in korpusa Janes. Nato smo označili vse elemente, ki označujejo interakcijo s sogovorci, in te elemente analizirali z metodologijo

Tracey in Robles (2013), razširjeno s kategorijami, ki sledijo shemi jezikovnih funkcij (Jakobson 1989). Dobili smo kategorije, ki jih ponazarjajo primeri v desnem stolpcu Tabele 1, pri katerih smo v primeru dvoma izbrali tisto jezikovno funkcijo, ki je za dani element dominantna.

Tabela 1: Kategorizacija elementov interakcije.

deiktika	<i>Jst</i>
vprašalnice	<i>Kaj</i>
modus	<i>Sigurno</i>
konativnost	<i>Gremo</i>
ekspresivnost	<i>uau, vidim</i>
fatičnost	<i>dobro jutro, evo</i>
performativ	<i>Obljubim</i>
metajezik	<i>Tvitam</i>
referenca	<i>glede, rekel</i>

Na koncu smo primerjali rezultate analize govornega korpusa in podkorpusov novih medijev ter poskušali določiti stopnjo in mesta prekrivanja elementov interakcije v govorjenem in spletnem neformalnem diskurzu.

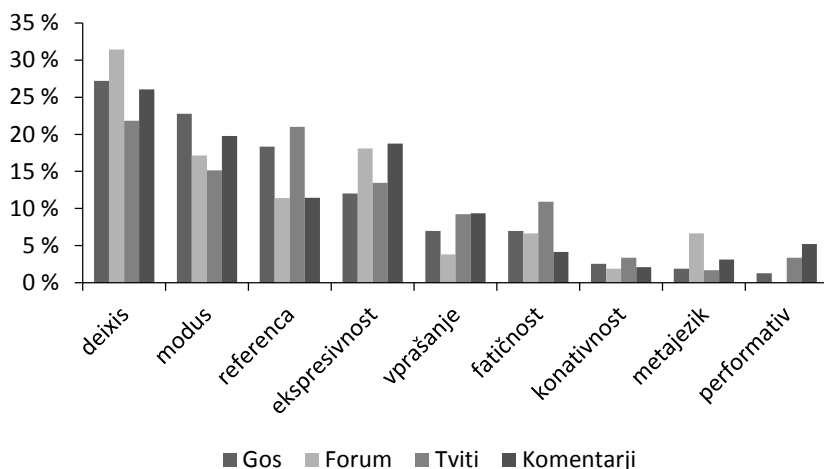


Graf 6: Elementi interakcije v korpusu Gos in podkorpusih Janes.

Analiza korpusov Gos, Janes-Forum, Janes-Tviti in Janes-Komentarji pokaže, da pri treh od štirih analiziranih besedilnih žanrov več kot 60 % analiziranih ključnih besednih oblik vsebuje elemente interaktivnosti. Najvišjo stopnjo interakcije

najdemo v govoru, razlike pa se kažejo tudi med različnimi družbenimi mediji. Največ elementov interakcije vsebuje podkorporus Tviti, ki ga zaznamujejo hitro izmenjana kratka sporočila (pogosto prek pametnih telefonov). Pri spletnih forumih so besedila daljša, a v njih kljub temu opazimo izjemno veliko interakcije med uporabniki (svetovanje, izmenjevanje mnenj in izkušenj). Najmanj interaktivni so komentariji na spletne novice, saj je primarni motiv komentatorjev sporočanje osebnega mnenja o članku/dogodkih/osebah v članku, ne toliko interakcija z drugimi.

V naslednjem koraku smo zaznane elemente interakcije glede na prevladujočo pragmatično vlogo razvrstili v predstavljenih devet kategorij (Graf 7).



Graf 7: Analiza elementov interakcije v korpusu Gos in podkorporisih Janes.

Prevladujoča kategorija v vseh podkorporisih so deiktični izrazi, največ jih je v forumih. Gre večinoma za osebne zaimke (*jst, tale, tist*) in časovne in prostorske deiktične izraze (*zdaj, tam, ven*), ki se nanašajo na zunanji kontekst konverzacije in udeležence v njej. Te izraze verjetno pogojujejo tehnične okoliščine konverzacije na spletu, saj udeleženci ne delujejo v istem fizičnem in časovnem kontekstu, ki mora biti zato bolj eksplicitno izražen. Kvalitativna analiza gradi-va pokaže, da so v korpusu Gos pogoste različne izgovorne variante osebnih in kazalnih zaimkov, kot so *jaz* (npr. *jst, jest, jez, jaz*), ter časovnih in prostorskih prislovov, (npr. *zdej, zaj, zej, zdele*). V podkorporisih Janes pa deiktične izraze najdemo le v eni ali dveh variantah.

V korpusu Gos so zelo pogosti elementi modalnosti (*dejansko, itak, pač*), ki »izražajo stopnjo gotovosti, s katero govorec izraža svoje mnenje« (Morel in Danon Boileau 1998). Največ elementov modalnosti (20 %) najdemo v komentarijih (*očitno, sigurno, zgleda*), kar je skladno z namenom komentiranja na novinarskih

portalih. V tvitih je modalnih elementov manj, saj uporabnika tehnične okoliščine (omejena dovoljena dolžina, uporaba mobilnih naprav) in komuniciranje v realnem času spodbujajo k čim krajšim besedilom.

Podobno vlogo imajo ekspresivni izrazi, le da so bolj osredotočeni na naravnost in razpoloženje samega avtorja (*haha, kul, super*) in jih je na forumih in komentarjih celo več kot v govoru. To je verjetno posledica dejstva, da korpus Gos zajema formalne in neformalne sporazumevalne položaje, v Janesu pa pogosto predstavljajo ventil za izražanje mnenj, deljenje izkušenj in sproščanje frustracij uporabnikov. Poleg tega v analizo nismo zajeli neverbalnih prvin (npr. smeha), ki označujejo razpoloženje govorca, vendar so v korpusu Gos označeni kot dogodek in niso transkribirani kot besedilo.

Posebnost tvitov je izstopanje referencialnih elementov (*poznam, glede, tiče.*). To kaže na visoko stopnjo interaktivnosti tega medija, kjer udeleženci v pogovore vključujejo svoje izkušnje z zunanjim, realnim svetom. Tviti izstopajo tudi po deležu fatičnih elementov (*aja, btw, čao*), medtem ko komentarji niso namenjeni vzpostavljanju in vzdrževanju stika med sogovorci, temveč izražanju lastnega mnenja.

Po številu vprašalnic izstopajo tviti in komentarji, ki jih je celo več kot v korpusu Gos. V kategorijo metajezik smo zajeli izraze, ki se nanašajo na izrekanje samo, kar je pogosto povezano z glagoli izrekanja (*povedal, govorim, reku*), v novih medijih pa tudi z glagoli *tvitnil, napisal* ipd. Najmanj izraziti kategoriji predstavljajo konativni elementi (*gremo, mors*) in performativi (*obljubim, se strinjam*), ki jih v vsakem korpusu najdemo le po nekaj primerov.

Analiza elementov interakcije je pokazala, da najmočnejšo podobnost med govorjenim diskurzom in računalniško posredovano komunikacijo predstavljajo deiktični izrazi in referencialni elementi v tvitih, medtem ko so modalni elementi prisotni predvsem v komentarjih. V računalniško posredovani komunikaciji je variantnost zapisa manjša, opaziti pa je tudi določene omejitve zaradi medija samega, ki diktira ekonomičnost izražanja.

5 DISKUSIJA

Zdi se, da analiza na leksikalni in besednovrstni ravni ter na ravni odmika od standarda pove marsikaj o specifikah posameznih spletnih zvrsti v primerjavi z govorjenim diskurzom. Vendar predstavljena raziskava ne zajema skladišne ravni analize, ki bi bistveno več povedala o tem, katere besedne zveze, tipi povedi in značilnosti upovedovanja so tipični za preučevana tipa diskurza in opazovane

besedilne zvrsti. Predvidevamo, da bi se s skladenjsko analizo dokončno pokazalo, kako specifičen je spontani govor zaradi sočasnega načrtovanja in izrekanja. Obenem pa verjamemo, da so programi za avtomatsko skladenjsko razčlenjevanje spontanega govora še v razvoju prav zato, ker je enota segmentiranja spontanega govora tako temeljno drugačna od tistih, ki smo jih vajeni analizirati v pisnih besedilih (Morel in Danon Boileau 1998).

Z analizo seznamov ključnih besednih oblik oziroma besednih vrst prav tako nismo zajeli prvin, ki jih lahko zaznamo s kvalitativnim vpogledom v gradivo spletnega in govorjenega diskurza. Hiter pogled na dogajanje na Twitterju namreč pokaže močno prisotno ironijo, ki je kvantitativne analize ne zaznajo (*Drugi tir? Ni važno. Važno, da bo referendum*). Podobne primere lahko zaznamo tudi v korpusu Gos (*eee kaj ti praviš bo dobr? ha? al ni to dobr? no dobr pol važno da gremo pol na čaj [smeh]*). Zato bi bilo v prihodnosti smiselno sistematično raziskati jezikovne prvine, s katerimi uporabniki izražajo svojo čustveno naravnost (Martinc et al. 2017), nato pa s pomočjo identificiranih jezikovnih prvin raziskati različne tipe naravnosti uporabnikov v govoru in v spletnih žanrih.

6 SKLEP

V raziskavi smo primerjali specifikke govorjenega diskurza in računalniško posredovane komunikacije glede na standardna pisna besedila. Ugotovili smo, da je ortografska variantnost tvitov primerljiva z variantnostjo izgovorjave, kakršno najdemo v govorjenih besedilih, ker avtorji besedil v sproščenem vzdušju radi pokažejo svojo pripadnost določenim jezikovno prepoznavnim skupinam. Komentarje zaznamuje predvsem visoka stopnja za govor značilnih modalnih elementov, ki kaže avtorjevo naravnost v zvezi z obravnavano tematiko. Vendar kvalitativna analiza gradiva pokaže, da so modalni elementi v govoru povezani z avtorjevo stopnjo gotovosti (*itak*), v komentarjih pa z njegovimi sodbami (*škoda*). Pri forumih je formalnost diskurza povezana s stopnjo formalnosti forumske tematike, zaznana odstopanja od pravopisnih pravil pa ne kažejo jezikovne ustvarjalnosti, ampak delujejo kot pomanjkanje jezikovne kompetence (*nebi*).

Vsak od analiziranih žanrov računalniško posredovane komunikacije ima seveda številne prvine, ki jih lahko zaznamo tudi v neformalnem govorjenem diskurzu, vendar je vloga prenosnika še vedno bistvenega pomena za razumevanje delovanja obeh analiziranih zvrsti. Sočasnega tvorjenja in načrtovanja besedil, ki temeljno zaznamuje spontani govorjeni diskurz, ne moremo opazovati v pisnih besedilih, kar se pokaže tudi prek osnovne primerjave zastopanosti besednih vrst. Verjetno pa razlog za pogosto izpostavljeno podobnost med računalniško posredovano komunikacijo in neformalnimi zvrstmi govorjenega diskurza predstavlja dejstvo, da

so specifične interakcije in neformalnih elementov na prvi pogled bolj vpadljive kot nekoliko ponotranjene, vendar vseprisotne prvine sprotnega tvorjenja govorenega diskurza.

Smo z rezultati analize uspeli preseči stereotip o računalniško posredovani komunikaciji, ki je podobna govorni slovenščini, ta pa predstavlja sinonim za ne-standardno slovenščino? Vsekakor lahko rečemo, da o računalniško posredovani komunikaciji, pa tudi o govornem diskurzu, vemo več. Vemo, katere specifične govora so povezane s procesom sprotnega tvorjenja besedil, katere specifične računalniško posredovane komunikacije s tehničnimi okoliščinami medija, katere z vzdrževanjem stika z ostalimi udeleženci in katere z neformalnostjo okoliščin, do katerih lahko pride tako v govornem kot v pisnem jeziku. Zasnova raziskave lahko prispeva k razmisleku o teoriji zvrstnosti za slovenščino in zaključi debato o pogovornem (govornem) in knjižnem (pisnem) jeziku, rezultate raziskave pa bi bilo smiselno vključiti tudi v snovanje novih jezikovnih priročnikov.

Literatura

- Bamman, David, Jacob Eisenstein in Tyler Schnoebelen, 2014: Gender identity and lexical variation in social media. *Journal of Sociolinguistics* 1/2. 135–160.
- Blanche-Benveniste, Claire, 1991: Les études sur l'oral et le travail d'écriture de certains poètes contemporains. *Langue française* 89. 52–71.
- Cambria, Erik, Björn Schuller, Yunqing Xia, Catherine Havasi, 2013: New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems* 28/2. 15–21. <http://sentic.net/new-avenues-in-opinion-mining-and-sentiment-analysis.pdf>
- Chovanec, Jan, 2009: Simulation of Spoken Interaction in Written Online Media Texts. *Brno Studies in English* 35/2. 109–129.
- Crystal, David, 1995: *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press.
- Crystal, David, 2001: *Language and the Internet*. Cambridge: Cambridge University Press.
- Dobrovoljc, Helena, 2008: Jezik v e-poštnih sporočilih in vprašanja sodobne normativistike. Košuta, Miran (ur.): *Slovenščina med kulturami*, Ljubljana: Slavistično društvo Slovenije. 295–314.
- Erjavec, Tomaž, Simon Krek, Špela Arhar, Darja Fišer, Nina Ledinek, Amanda Saksida, Breda Sivec in Blaž Trebar, 2010: *Priporočila za oblikoslovno označevanje JOS*. <http://nl.ijs.si/jos/msd/html-sl/msd.index.msds.html>
- Erjavec, Tomaž, Nikola Ljubešić in Darja Fišer, 2018: Korpus slovenskih spletnih uporabniških vsebin Janes. Fišer, Darja (ur.): *Viri, orodja in metode za analizo spletne slovenščine*. Ljubljana: Znanstvena založba Filozofske fakultete v Ljubljani. 16–43.

- Fišer, Darja, Tomaž Erjavec in Nikola Ljubešić, 2016: JANES v0. 4: Korpus slovenskih spletnih uporabniških vsebin. *Slovenščina 2.0* 4/2. 67–99. <http://slovenscina2.0.trojina.si/arhiv/2016-2/2016-2-04/>
- Huber, Damjan, 2013: *Poudarek in pavza v standardnem slovenskem govoru*. Doktorska disertacija. Ljubljana: Filozofska fakulteta.
- Gruden, Ana, 2013: *Pojmovanje socialne zvrstnosti v slovenskem prostoru s poudarkom na šolstvu*. Diplomsko delo. Ljubljana: Filozofska fakulteta.
- Jakobson, Roman, 1989: *Lingvistični in drugi spisi*. Ljubljana: Studia Humanitatis.
- Jakop, Nataša, 2008: Pravopis in spletni forumi - kva dogaja? Košuta, Miran (ur.): *Slovenščina med kulturami* – Slovenski slavistični kongres. Ljubljana: Slavistično društvo Slovenije. 315–327.
- Kalin Golob, Monika, 2008: SMS-sporočila treh generacij. Košuta, Miran (ur.): *Slovenščina med kulturami*. Ljubljana: Slavistično društvo Slovenije. 283–294.
- Kilgarriff, Adam, Pavel Richly, Pavel Smrz in David Tugwell, 2004: The Sketch Engine. *Zbornik EURALEX 2004*. Lorient. 105–116.
- Michelizza, Mija, 2008: Jezik SMS-jev in SMS-komunikacija. *Jezikoslovni zapiski* 14. 151–166.
- Morel, Mary Annick in Laurent Danon-Boileau, 1998: *Grammaire de l'intonation*. Pariz: Ophrys.
- Nyström, Stefan, 2003: *Spoken Language Features in Internet Discussion Groups*. Diplomsko delo. Lund University.
- Pang, Bo in Lillian Lee, 2008: Opinion Mining and Sentiment Analysis, *Foundations and Trends in Information Retrieval* 2, 1–135.
- Scott, Mike, 1997: Pc Analysis of Key words – and Key Key words. *System* 25/1. 1–13.
- Smailović, Jasmina, Miha Grčar, Nada Lavrač in Martin Žnidaršič, 2014: Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences* 285, 181–203.
- Smolej, Mojca, 2004: Načini tvorjenja govornega diskurza – paradigmska in sintagmatska os. Kržišnik, Erika (ur.): *Aktualizacija jezikovnozvrstne teorije na Slovenskem: členitev jezikovne resničnosti (Obdobja 22)*. Ljubljana: Center za slovenščino kot drugi/tuji jezik.
- Smolej, Mojca, 2006: *Vpliv besedilne vrste na uresničitev skladijskih struktur (primer narativnih besedil v vsakdanjem spontanem govoru)*. Doktorska disertacija. Ljubljana: Filozofska fakulteta.
- SSKJ – *Slovar slovenskega knjižnega jezika*. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. <http://www.fran.si/130/sskj-slovar-slovenskega-knjiznega-jezika>
- Tannen, Deborah, 1982: *Spoken and written language: exploring orality and literacy*. Norwood: Ablex.

- Tedeschi, Antonio, Francesco Benedetto, 2015: A cloud-based big data sentiment analysis application for enterprises' brand monitoring in social media streams. *Research and Technologies for Society and Industry Leveraging a better tomorrow* Torino, Italija. 186–191. DOI: 10.1109/RTSI.2015.7325096
- Tivadar, Hotimir, 2015: Razmerje med domačim in tujim v govornem knjižnem jeziku 3. tisočletja. Jesenšek, Marko (ur.): *Leopold Volkmer: prvi posvetni pesnik na slovenskem Štajerskem*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti. 139–149.
- Toporišič, Jože, 1991: *Slovenska slovnica*. Maribor: Založba Obzorja.
- Tracey, Karen in Jessica S. Robles, 2013: *Everyday talk. Second Edition. Building and Reflecting Identities*. New York: Guilford Press.
- Vaufreydaz, Dominique, Mohamad Akbar, José Rouillard, 1999: Internet Documents: A Rich Source for Spoken Language Modeling. *Keystone – Colorado*. 277–281
- Verdonik, Darinka, 2006: Popravljanja v spontano tvorjenih izjavah. *Slavistična revija* 54/2. 188–203.
- Verdonik, Darinka in Iztok Kosem, 2012: Key word analysis of discourses in Slovene speech: differences and similarities. *Linguistica* 52/1. 309–321.
- Verdonik, Darinka in Ana Zwitter Vitez, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Zwitter Vitez, Ana, 2009: O spontanem govoru. Vitez, Primož (ur.): *Spisi o govoru*. Ljubljana: Znanstvena založba Filozofske fakultete. 241–254.
- Zwitter Vitez, Ana in Darja Fišer, 2015: Novi mediji in govornjena slovenščina: zaznavanje, kodifikacija, analiza. Smolej, Mojca (ur.): *Slovnica in slovar – aktualni jezikovni opis (Obdobja 34)*. Ljubljana: Znanstvena založba Filozofske fakultete. 881–890. http://centerslo.si/wp-content/uploads/2015/11/34_2-Zwitter.pdf
- Zwitter Vitez, Ana, 2016: Specifike govorne slovenščine glede na formalnost sporazumevalnega položaja. Kržišnik, Erika in Miran Hladnik (ur.). *Toporišičeva obdobja*, Ljubljana: Znanstvena založba Filozofske fakultete. 351–359. <http://centerslo.si/wp-content/uploads/2016/11/Zwitter.pdf>
- Zwitter Vitez, Ana in Darja Fišer, 2016: Linguistic Analysis of Emotions in Online News Comments-an Example of the Eurovision Song Contest. Fišer, Darja in Michael Beißwenger (ur.): *Proceedings of CMC and Social Media Corpora for the Humanities*, Ljubljana: Znanstvena založba Filozofske fakultete. 74–76. http://nl.ijs.si/janes/wp-content/uploads/2016/09/CMC-2016_Zwitter_Fisher_Linguistic-Analysis-of-Emotions.pdf