

NOVI MEDIJI IN GOVORJENA SLOVENŠČINA: ZAZNAVANJE, KODIFIKACIJA, ANALIZA

^{1,2}Ana Zwitter Vitez, ²Darja Fišer

¹Fakulteta za humanistiko, Koper; ²Filozofska fakulteta, Ljubljana

UDK 811.163.6'322.5'42:004.738.52

Snovalci jezikovnih virov in priročnikov bodo imeli pomembno nalogo, da ugotovijo odnos med zaznavo jezika novih medijev, dejanskim stanjem in smiselnostjo vključevanja opisov njihovih tipičnih prvin v jezikovne priročnike. S korpusno analizo smo raziskali ključne govorne prvine v besedilih novih medijev na besednovrstni, oblikoskladenjski in leksikalni ravni, s čimer želimo izboljšati poznavanje načel delovanja govornjenih in spletnih besedil.

spletna slovenščina, govornjeni diskurz, računalniško posredovana komunikacija, korpusna analiza

The authors of future language resources and handbooks will have the important task of analyzing the characteristics and perception of internet language, as well as the opportunities of including them in language reference books. In this paper we present a corpus analysis of typical spoken elements in Slovene new media texts at the part-of-speech, morphosyntactic and lexical levels, which can help us to understand the key principles of spoken and internet language.

internet Slovene, spoken discourse, computer-mediated communication, corpus analysis

1 Uvod

Zaradi velike količine spletnih besedil, ki jih vsakodnevno ustvarijo in objavijo uporabniki, imajo snovalci jezikovnih virov in priročnikov pomembno nalogo, da temeljito preučijo besedila novih medijev ter ugotovijo odnos med zaznavo lastnosti jezika novih medijev, dejanskim stanjem in smiselnostjo vključevanja njihovih tipičnih prvin v jezikovne priročnike. Ker je eden od najmočnejših stereotipov glede spletne slovenščine povezan ravno z govornimi značilnostmi besedil novih medijev, smo s korpusno analizo raziskali ključne govorne prvine v besedilih novih medijev, s čimer smo želeli izboljšati poznavanje načel delovanja govornjenih in spletnih besedil.

2 Preučevanje elementov govora in besedil novih medijev

Preučevanje spontanega govornih izmenjav je pomembno prispevalo k poznavanju specifik spontanega govora v odnosu do standardnih pisnih besedil: M. Schlamberger Brezar (1998) na avtentičnih posnetkih analizira retorične principe v argumentaciji, M. Smolej (2006) analizira vpliv diskurzivnega žanra na leksikalno in oblikoskladenj-

sko strukturo govora, D. Verdonik analizira diskurzne označevalce in popravljanja v spontanem govoru, A. Zwitter Vitez (2009) pa preučuje realizacijo komunikacijskih strategij v spontanem govoru. Omenjene raziskave predstavljajo teoretski okvir za analizo specifik govornih prvin, ki jih bomo v analitičnem delu poskušali identificirati v besedilih novih medijev.

Specifike besedil novih medijev obravnavajo naslednje raziskave: M. Kalin Golob (2008) analizira kratka sporočila in v njih identificira številne pojave nestandardne rabe ločil, fonetični zapis, opuščanje pomožnega glagola in nedoločnika. M. Michelizza (2008) ugotavlja, da se jezik kratkih sporočil spreminja glede na vpliv 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 med rezultati omenjajo tudi govorne prvine v besedilih novih medijev, zato se zdi smiselno izvesti sistematično primerjavo specifik govorjenih besedil in spletnih pisnih besedil.

3 Metodologija

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

- uravnoteženi korpus pisnih besedil Kres, ki zajema 100 milijonov pojavnic (Logar Berginc idr. 2012),
- korpus govorjene slovenščine Gos v obsegu milijon pojavnic (Verdonik, Zwitter Vitez 2010),
- korpus spletnih besedil Janes s 161 milijoni pojavnic (Fišer idr. 2015).

Analiza je potekala v dveh fazah: najprej smo iz korpusa Gos izluščili tipične prvine govora na podlagi primerjave z uravnoteženim korpusom pisnih besedil korpusa Kres, potem pa smo pojavljanje identificiranih govornih prvin preverili v korpusu spletnih besedil Janes. Kvantitativno in kvalitativno analizo smo izvedli na besednovrstni, oblikoslovni in leksikalni ravni.

Za pridobivanje podatkov iz korpusov smo uporabili orodje Sketch Engine¹ (Kilgarriff idr. 2004), ki omogoča vpogled v konkordance in primerjavo različnih korpusov s pomočjo frekvenčnih in ključnih seznamov besednih oblik, lem in oblikoskladenjskih oznak.

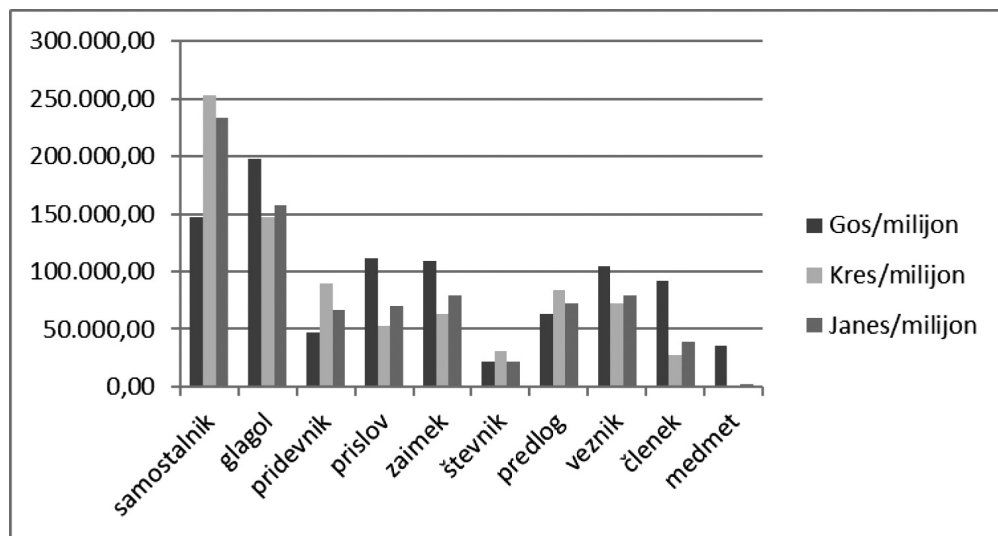
4 Analiza

V nadaljevanju predstavljamo primerjavo korpusov Gos, Kres in Janes na ravni besednih vrst, besednih oblik in besedišča.

¹ <http://www.sketchengine.co.uk>

4.1 Besedne vrste

Najprej smo s pomočjo frekvenčnih seznamov analizirali zastopanost posameznih besednih vrst v korpusih Gos, Kres in Janes. Določanje besednih vrst je bilo opravljeno ročno, in sicer v skladu s Priporočili za oblikoslovno označevanje JOS (Erjavec idr. 2010). V primeru prekrivnih besednih oblik (npr. *niti*, ki je lahko samostalnik, prislov ali veznik) smo izbrali besedno vrsto, v kateri se oblika v korpusu največkrat pojavi.



Slika 1: Besedne vrste v korpusih Gos, Kres in Janes

4.1.1 Specifike govora

Kvantitativna primerjava porazdelitve besednih vrst pokaže, da je v korpusu govorne slovenščine glede na korpus Kres za polovico manj samostalnikov in pridevnikov, števnikov in predlogov pa za tretjino, medtem ko je v korpusu Gos nekoliko več glagolov, veznikov, zaimkov, dvakrat več prislovov in trikrat več členkov. Razmerje med številom medmetov v govornjeni in pisni slovenščini je kar 130 proti 1.

Da se bodo govorne prvine pokazale skozi večjo zastopanost medmetov, členkov in zaimkov, smo pričakovali, nekoliko pa preseneča večja zastopanost glagolov in veznikov. Podrobnejša analiza gradiva kaže, da razliko v veliki meri tvorijo številna ponavljanja in glagoli v vlogi diskurzivnih označevalcev, kot na primer *misliti* in *vedeti* (3):

(1)

// zakaj kako **a veš mislim** eee poznaš eee [ime] od prej? //

V govornjenih besedilih izstopa tudi pogosta raba veznikov, ki bi lahko bila posledica večje skladenjske kompleksnosti besedil. Analiza gradiva pokaže, da je to povezano s pojavom ponavljanja zaradi sprotnega procesa tvorjenja govornjenih besedil:

(2)

zarad tega **ker** smo **ker** eee **ker** živimo oziroma **ker** so takrat že živeli v času ko so informacije dejansko z lahkoto krožle

4.1.2 Govorne prvine v spletnih besedilih

Korpus Janes se po zastopanosti analiziranih besednih vrst sicer razlikuje od korpusa Kres, vendar se umešča bliže korpusu Kres kot govorjenim 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, saj imamo pri pisanju spletnih besedil vseeno več časa za tvorjenje koherentnih enot kot v govoru:

(3)

srečkovička **haha** .. ti prvošm

Na ravni števnikov in pridevnikov so besedila novih medijev bliže govorjenim kot standardnim pisnim besedilom. To bi lahko bila posledica tematske usmerjenosti posameznih besedilnih zvrsti korpusa Kres 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. Vendar v spletnih besedilih ne opazimo povečanega števila glagolov in veznikov, ki so posledica ponavljanj v procesu tvorjenja govorjenih besedil.

4.2 Oblikoskladenjske oznake

Analizo oblikoskladenjskih lastnosti smo izvedli v dveh fazah: najprej smo prek seznama ključnih oblikoskladenjskih oznak korpusa Gos glede na korpus Kres izluščili tipične prvine govora, v naslednjem koraku pa smo ugotavljali, katere od identificiranih govornih prvin so prisotne tudi med ključnimi specifikami korpusa Janes.

4.2.1 Specifike govora

Oblikoslovne specifike govora smo identificirali s pomočjo analize seznama prvih 50 ključnih oblikoskladenjskih oznak v korpusu Gos glede na korpus Kres (v Tabeli 1 je prikazanih prvih 10).

Tabela 1: Ključne oblikoskladenjske oznake korpusa Gos

Oblikoskladenjska oznaka	Primer	Frekvenca	Frekv./milijon
osebni zaimek (prva oseba, ednina, imenovalnik)	jaz	6.296	5.668,8
glavni glagol (nedovršni, sedanjik, druga oseba, ednina)	veš	5.485	4.938,6
števnik (glavni, množina, tožilnik)	pet	5.745	5.172,7
glavni glagol (dovršni, velelnik, prva oseba, množina)	poglejmo	2.146	1.932,2
zaimek (osebni, tretja oseba, moški, množina, imenovalnik)	oni	985	886,9
glavni glagol (nedovršnik, sedanjik, druga oseba, ednina, nezanikani)	imaš	1.185	1.066,9
zaimek (kazalni, srednji spol, ednina, imenovalnik)	ta	15.427	13.890,1
zaimek (osebni, druga oseba, ednina, dajalnik, klitika)	ti	2.141	1.927,7
glagol (glavni, dvovidski, sedanjik, druga oseba ednina)	vidiš	1.151	1.036,3

Rezultati analize so naslednji: med prvimi 50 ključnimi oblikoskladenjskimi oznakami so le tri besedne vrste: glagoli, števnik in zaimki. Največ različnih oblik je med zaimki (24), sledijo glagoli (19), nato števnik (7). Daleč najbolj ključna besedna oblika je osebni zaimek v imenovalniku ednine (*jaz*), močno izstopajo tudi drugi osebni zaimki:

(5)

kaj **ti** sploh hočeš? // kaj kaj **jz** hočem kaj **ona** hoče joj //

Tipična glagolska oblika v sedanjiku druge osebe ednine pogosto nastopa v vlogi diskurznega označevalca:

(6)

veš pol spiš **veš** po pet ur cel teden in te zmanjka enkrat pač se moraš naspat ne

Tudi velelniške oblike največkrat nastopajo v vlogi diskurznega označevalca (*recimo, dej no*):

(7)

eni so šli bolj k Alpam a ne to so bli **recimo** naši predniki

Kazalni zaimki se v večini primerov realizirajo kot *ta* in *un* v vlogi besedilnih aktualizatorjev (Smolej 2006):

(8)

mislim dobiš **un** denar ki je komaj da [premor] več ... ja da stroške pokriješ

Analiza je pokazala, da so za govornjena besedila najznačilnejše oblikoskladenjske oznake skrčene na ozek nabor besednih vrst (glagoli, števnik in zaimki). Najbolj tipične oblike so osebni zaimki v imenovalniku, glagoli v sedanjiku druge osebe

ednine (*veš*) in velelniku (*recimo*) v vlogi diskurzivnih označevalcev in kazalni zaimki v vlogi besedilnih aktualizatorjev (*un ta rdeč*).

4.2.2 Govorne prvine v spletnih uporabniških vsebinah

V naslednjem koraku smo želeli ugotoviti, v kolikšni meri se identificirane oblikoskladenjske govorne pojavnosti pojavljajo v spletnih uporabniških vsebinah. Zato smo tipične govorne prvine primerjali s seznamom prvih 50 ključnih oblikoskladenjskih oznak korpusa Janes (in vseh podkorpusov) glede na korpus Kres (Tabela 2 prikazuje prvih 10).

Tabela 2: Ključne oblikoskladenjske oznake korpusa Janes (prekrivnost s korpusom Gos)

Oznaka	Primer	Frekvenca	Frekv./milijon
glavni glagol (dovršni, namenilnik)	pokazat	157.240	974,9
glavni glagol (nedovršni, namenilnik)	hecat	181.617	1.126,0
glavni glagol (dovršni, sedanjik, druga oseba, ednina)	dobiš	257.332	1.595,5
glavni glagol (dvovidski, sedanjik, prva oseba, ednina)	stavim	243.385	1.509,0
osebni zaimek (prva oseba, ednina, imenovalnik)	jaz	274.770	1.703,6
glavni glagol (nedovršni, sedanjik, druga oseba, ednina)	veš	354.300	2.196,7
glavni glagol (nedovršni, sedanjik, prva oseba, ednina)	podpiram	993.952	6.162,5
glavni glagol (nedovršni, sedanjik, prva oseba, ednina, nezanikani)	hočem	129.332	801,9
osebni zaimek (druga oseba, ednina, dajalnik, klitika)	ti	164.260	1.018,4
glavni glagol (sedanjik, dvovidski, druga oseba, ednina)	zgledaš	84.001	520,8

Do ujemanj med govornimi besedili in spletnimi uporabniškimi vsebinami prihaja na ravni glagolov prve in druge osebe ednine:

(9)

Ono, ko **prideš** na šiht pa maš 113 neprebranih mailov

Raba velelnika v korpusu Janes za razliko od prevladujoče rabe v govoru ni izključno pragmatizirana z vlogo diskurzivnih označevalcev:

(10)

Priznaj, da ti gre na živce, ker te @zzTurk ignorira :)

Na ravni osebnih zaimkov prevladuje druga oseba ednine, ki kaže visoko stopnjo interakcije z drugimi avtorji:

(11)

Ampak **ti** si povrhu še ženska, bogtenimarad!

Tudi vprašalni zaimki so posledica aktivne interakcije z drugimi uporabniki:

(12)

Kaj so dacarji pojedli in popili v okviru nadzora gotovinskega poslovanja ?

Kazalni zaimki v vlogi deiktičnih izrazov izpostavljajo prostor in čas, v katerem je nastalo besedilo:

(13)

To ni jamranje. **To** je tvitprotest ! :)

Tipične govorne oznake v spletnih uporabniških vsebinah so torej povezane s prvo in drugo glagolsko obliko, rabo velelnika, poudarjenih osebnih zaimkov v prvi in drugi osebi ednine, deiktični izrazi v vlogi besedilnih aktualizatorjev. Našteti elementi so posledica interakcije z drugimi avtorji besedil.

Govorjena slovenščina ima največ skupnih oblikoslovnih pojavnosti s podkorpusom Forumi (40.000 skupnih oznak), sledi podkorpus Komentarji s 30.000 skupnimi oznakami, najmanj prekrivna z govorjeno slovenščino pa sta korpusa Blogi in Tviti s približno 20.000 enakimi oznakami. Na ravni oblikoskladenjskih oznak smo zaradi dejavne interakcije s sogovorniki pričakovali največ podobnosti med tviti in govorom, zato bi bilo smiselno preveriti ustreznost statistične primerjave frekvence ključnih oznak.

4.3 Besedišče

Tudi analizo besedišča smo izvedli v dveh korakih: najprej smo na podlagi ključnega seznama lem v korpusu Gos glede na korpus Kres identificirali leme, tipične za govorjeni diskurz, potem pa smo analizirali prisotnost identificiranih govorjenih prvin v korpusu Janes.

4.3.1 Specifike govora

Tipično govorjeno besedišče smo izluščili na podlagi analize seznama 50 ključnih lem v korpusu Gos glede na leme korpus Kres (Tabela 3 prikazuje prvih 10).

Tabela 3: Ključne leme korpusa Gos

Lema	Frekv./milijon	Besedna vrsta
eee	20.917,5	medmet
ja	22.797,5	členek
mhm	4.033,7	medmet
eem	2.682,2	medmet
aha	1.855,7	medmet
mmm	1.373,1	medmet
no	4.229,1	členek
pol	3.771,7	prislov
aja	1.186,7	medmet
tule	850,0	prislov

Med ključnimi lemmi korpusa Gos močno prevladujejo prislovi (40 %), medmeti (14 %) in členki (14 %). Daleč najbolj ključni so medmeti v vlogi označevalcev procesa tvorjenja:

(14)

eee mism velka **eee** mism vse je blo zoran tko da se nisem mogu kam ulež

Govorno besedišče zaznamuje zelo malo glagolov (12 %) in samostalnikov (6 %), pa še ti pogosto nastopajo v vlogi diskurznihi označevalcev (*veš, mislim, v bistvu*).

4.3.2 Govorne prvine v spletnih besedilih

Identificirano govorno besedišče smo primerjali s seznamom prvih 50 ključnih lem korpusa Janes (Tabela 4 prikazuje prvih 10).

Tabela 4: Ključne leme korpusa Janes (prekrivnost s korpusom Gos)

Janes	Blog	Comment	Forum	Tvit
slovenija	slovenija	slovenija	lp	slovenija
lp	blog	janša	avto	videoposnetek
avto	pač	sds	guma	via
tale	edini	slovenec	pozdravljen	tale
ljubljana	mogoče	zda	hvala	ljubljana
mogoče	ljubljana	edini	forum	slo
edini	itak	evropa	pozdrav	všeč
hvala	ampak	bravo	mogoče	rabiti
janša	komentar	rusija	tale	jutri
itak	tale	maribor	tih	lej

Leme, ki so tipične tako za govor kot za spletna besedila, pripadajo besednim vrstam prislovov, medmetov in členkov. Največ analiziranega govornega besedišča se pojavi v podkorpusu Forum (7.000 pojavitev/milijon), sledi podkorpus Blog (5.700 pojavitev/milijon), nato podkorpus Komentar (4.000 lem/milijon) in nazadnje korpus Tvit (3.700 lem/milijon).

Identificirane govorne prvine nastopajo predvsem v vlogi diskurznihi označevalcev, zato smo jih po metodologiji D. Verdonik (2008) kategorizirali glede na njihovo pragmatično vlogo:

- ideacijski (priedni vezniki in prislovi):

(15)

Verjetno si mislijo, da nisem čisto gladka. **Ampak** mi je prav vseeno.

- interakcijski (izražanje strinjanja in dogovarjanje o poteku diskurza):

(16)

Pravzaprav ni imel česa zgrešiti. Vsaka njegova pesem je **itak** hit.

– označevalci procesa tvorjenja besedila:

(17)

Recimo, da reče neka uslužbenka par let pred penzijo, da se **pač** ne bo učila na računalnik 5 let pred penzijo.

Naslednja tipična govorna kategorija v spletnih besedilih so količinski prislovi (*ful, malo, dosti*):

(18)

Luštn se imajo. **Mal** državo rešujejo, pa **mal** dolžnike države.

Zadnja zaznana kategorija so kazalni zaimki v vlogi besedilnih aktualizatorjev:

(19)

Se ti malo vleče **tale** teden, ha ?

Analiza besedišča je pokazala, da statistično največ govornega besedišča vsebuje podkorpus Forumi, najmanj pa podkorpus Tviti, kar je morda povezano z omejitvijo dolžine besedil na Twitterju. Ključno govorno besedišče je zgoščeno na prislove, medmete in členke v vlogi diskurznihih označevalcev, ki kažejo močno prisotnost interakcije z drugimi avtorji besedil.

5 Zaključek

V prispevku smo identificirali govorne prvine v besedilih novih medijev na ravni besednih vrst, besednih oblik in besedišča. Analiza besednih vrst v korpusih Gos, Janes in Kres je pokazala, da je v govoru in v besedilih novih medijev prisotnih več zaimkov, členkov in medmetov kot v klasičnih pisnih besedilih. Vendar v spletnih besedilih ne opazimo povečanega števila glagolov in veznikov, ki so posledica ponavljanj v procesu tvorjenja govornih besedil.

Najbolj tipične oblikoskladenjske oznake v korpusih Gos in Janes so povezane s prvo in drugo glagolsko obliko, rabo veleznika, poudarjenimi osebnimi zaimki in deiktičnimi izrazi v vlogi besedilnih aktualizatorjev. Na leksikalni ravni tako v govornih besedilih kot v besedilih novih medijev izstopajo diskurzni označevalci, ki kažejo močno prisotnost interakcije z drugimi avtorji besedil.

V nadaljnjih raziskavah bomo izvedli analizo elementov interakcije z naslovniki, odstopanj od standardnega besedišča in variantnosti izgovora v korpusu govornih slovenščine in zapisa v besedilih novih medijev. Pričakujemo, da bo celovita analiza govornih prvin v besedilih novih medijev tvorno prispevala k snovanju novih priročnikov za slovenščino in dopolnila postopke kvalifikacije obravnavanih prvin.

Zahvala

Raziskava, opisana v prispevku, je bila opravljena v okviru nacionalnega temeljnega projekta Viri, orodja in metode za raziskovanje nestandardne spletne slovenščine (J6-6842, 2014–2017), ki ga financira ARRS.

Literatura

- BERGINC LOGAR, Nataša, GRČAR, Miha, BRAKUS, Marko, ERJAVEC, Tomaž, ARHAR HOLDT, Špela, KREK, Simon, 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- 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, KREK, Simon, ARHAR, Špela, FIŠER, Darja, LEDINEK, Nina, SAKSIDA, Amanda, SIVEC, Breda, TREBAR, Blaž, 2010: *Priporočila za oblikoslovno označevanje JOS. 3.4 Seznam oblikoslovnih oznak*. <http://nl.ijs.si/jos/msd/html-sl/msd.index.msds.html>
- FIŠER, Darja, ERJAVEC, Tomaž, ČIBEJ, Jaka, LJUBEŠIČ, Nikola, 2015: Gradnja in analiza korpusa spletne slovenščine JANES. Mojca Smolej (ur.) *Slovnica in slovar – aktualni jezikovni opis. Obdobja 34*. Ljubljana: Znanstvena založba Filozofske fakultete. 225–231.
- JAKOP, Nataša, 2008: Pravopis in spletni forumi – kva dogaja? Miran Košuta, Miran Hladnik, Zoltan Jan, Pšena Kovačič (ur.): *Slovenščina med kulturami. 19. slovenski slavistični kongres*. Ljubljana: Slavistično društvo Slovenije. 315–327.
- KALIN GOLOB, Monika, 2008: SMS-sporočila treh generacij. Miran Košuta, Miran Hladnik, Zoltan Jan, Pšena Kovačič (ur.): *Slovenščina med kulturami. 19. slovenski slavistični kongres*. Ljubljana: Slavistično društvo Slovenije. 283–294.
- KILGARRIFF, Adam, RYCHLY, Pavel, SMRZ, Pavel, TUGWELL, David, 2004: The Sketch Engine. *Zbornik EURALEX 2004*. Lorient. 105–116.
- KRANJC, Simona, 1996: Govorjeni diskurz. *Jezik in slovstvo* 42/7. 307–319.
- LOGAR, Nataša, 2003: Kratice in tvorjenke iz njih – aktualna poimenovalna možnost. *Wspo łczesna polska i słowen śka sytuacja językowa*. 131–149.
- SMOLEJ, Mojca, 2006: *Vpliv besedilne vrste na uresničitev skladenjskih struktur (primer narativnih besedil v vsakdanjem spontanem govoru)*. *Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- VERDONIK, Darinka, 2008: Označevanje vrste diskurzivnih označevalcev. Tomaž Erjavec, Jerneja Žganec Gros (ur.): *Zbornik Šeste konference Jezikovne tehnologije*. Ljubljana: Institut »Jožef Stefan«. 25–28.
- VERDONIK, Darinka, ZWITTER VITEZ, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- ZEMLJARIČ MIKLAVČIČ, Jana, 2008: Iskanje odgovorov na vprašanja govorenega jezika. *Jezik in slovstvo* 53/1. 89–106.
- ZWITTER VITEZ, Ana, 2009: *Strategije in strukturiranje spontanega govora v francoščini in slovenščini*. *Doktorska disertacija*. Ljubljana: Filozofska fakulteta.