

Elementi interakcije v govornih in spletnih besedilih

Ana Zwitter Vitez,*† Darja Fišer†

*Oddelek za uporabno jezikoslovje, Fakulteta za humanistične študije

Titov trg 5, 6000 Koper

ana.zwitter@guest.arnes.si

† Oddelek za prevajalstvo, Filozofska Fakulteta

Aškerčeva 2, 1000 Ljubljana

darja.fiser@ff.uni-lj.si

Povzetek

Na družbenih omrežjih, spletnih forumih in komentarjih avtorji pri interakciji z drugimi udeleženci uporabljajo posebne jezikovne vzorce, ki se zdijo precej podobni govornemu diskurzu. Zato smo izvedli primerjavo korpusov Gos in Janes (s podkorpusi Tviti, Komentarji in Forumi). Kvantitativna analiza je pokazala visoko stopnjo ujemanja med kategorijami interakcije v korpusu Gos in podkorpused Janes, kvalitativna analiza pa je razkrila pomembne razlike, ki so posledica prostorske oddaljenosti udeležencev, sočasnosti načrtovanja in tvorjenja govornih besedil in (ne)identifikacije z dejansko glasovno podobo nekaterih pogostih izrazov.

Interaction structures in spoken discourse and computer-mediated communication

On social networks, web forums and news comments, the authors interact with other participants and use specific language patterns which seem to be similar to spoken discourse. In this study, we present a comparison of the Gos and the Janes corpora (with subcorpora Tweets, Comments and Forums). Quantitative analysis showed a high degree of interactional correlation between the analysed corpora, while qualitative analysis revealed significant differences which seem to be related to spatial and temporal distance between CMC participants, to simultaneous planning and producing of spoken texts and to authors' perception of the acoustic image of frequent linguistic structures.

1 Uvod

V besedilih družbenih omrežij, spletnih forumov in komentarjev na novičarskih portalih lahko zaznamo jezikovne vzorce, precej drugačne od tistih, ki so tipični za tradicionalna pisna besedila. Taki jezikovni vzorci se kažejo skozi neposredno naslavljanje drugih udeležencev, sredstva sprotnega tvorjenja besedil in rabo neformalnih besed, saj avtorji skozi nenehno interakcijo z drugimi udeleženci utrjujejo svojo identiteto, gradijo samopodobo in razkrivajo svoj odnos do zunajjezikovne realnosti. Tovrstni jezikovni vzorci so tradicionalno značilni za govorni diskurz, zato se je besedil računalniško posredovane pisne komunikacije (CMC) prijala stereotipna oznaka hibrida med govornimi in pisnimi besedili. Ugotavljanje dejanskih zakonitosti besedil CMC lahko omogočijo empirične raziskave, ki temeljijo na njihovi primerjavi z govornim diskurzom in tradicionalnimi pisnimi besedili.

Cilj pričujoče analize je ugotoviti, kateri tipični elementi interakcije v govornih izmenjavah so prisotni v besedilih računalniško posredovane komunikacije in kakšna je njihova vloga v vsakodnevni komunikaciji na spletu. To bomo izvedli s pomočjo primerjave seznamov ključnih besednih oblik v korpusih GOS (Verdonik in Zwitter Vitez, 2011) glede na posamezne podkorpuse korpusa spletnih uporabniških vsebin Janes (Fišer et al., 2014). Predvidevamo, da lahko prek identifikacije elementov interakcije in njihovih funkcij sklepamo o komunikacijskem namenu tvorca besedila in posledicah njegovega jezikovnega udejstvovanja.

2 Preučevanje elementov interakcije v govornih in pisnih besedilih

Pomen raziskovanja značilnosti jezika, ki ga na spletu vsakodnevno uporabljajo blogerji, komentatorji ter

uporabniki forumov in družbenih omrežij, prvi sistematično izpostavi Crystal (2001) z ugotovitvijo, da se prvine t. i. spletnega jezika (v izvorniku *netspeak*) širijo na druga področja jezikovne rabe in niso več omejene na računalniško posredovano komunikacijo. Pri tem imajo pomembno vlogo specifični jezikovni vzorci (Wray, 2005), ki nastopajo kot orodje za manipulacijo z naslovniki, vzpostavljajo lastne identitete in zagotavljajo družbene identitete. Herring (2006) kot ključne značilnosti jezikovnih specifik CMC izpostavlja na ravneh strukture (krajšanje), pomena (tematika), interakcije (stopnja strinjanja) in družbene vloge (ekspresivnost, identifikacija).

Tudi v govoru se pojavljajo specifični jezikovni vzorci, ki so posledica sočasnega načrtovanja in tvorjenja besedil in neposredne interakcije z drugimi udeleženci (Morel in Danon Boileau, 1998). Zato je korpusno jezikoslovje govora osredotočeno na rabo diskurzivnih označevalcev (Adolphs in Carter, 2013) in elementov interakcije med udeleženci, znotraj katere Tracey in Robles (2013) kot ključne strukture preučujeta elemente menjavanja vlog, diskurzne označevalce, vprašanja in odgovore, elemente govorne skupnosti (identifikacije).

Omenjene raziskave na besedilih CMC in govornem diskurzu zaznavajo primerljive jezikovne vzorce, ki odstopajo od standardnega pisnega jezika. V raziskavi Zwitter Vitez in Fišer (2015) smo izvedli vzporedno analizo jezikovnih specifik govornega diskurza in besedil računalniško posredovane komunikacije glede na standardna pisna besedila. Ta je pokazala, da so prvine interakcije med udeleženci prisotne tudi v primarno drugače razporejenih kategorijah (npr. pri izrazu *pejd*, pri katerem lahko zaznamo tako odstop od standardnega zapisa kot interakcijo z drugimi udeleženci). Zato se zdi smiselno prvine interakcije v govoru in spletnih besedilih podrobneje preučiti.

3 Metodologija

Cilj analize je bil identificirati točke prekrivnosti in razhajanj nestandardnih in interakcijskih elementov govornih in spletnih prvin, kar smo izvedli s primerjavo treh korpusov slovenskega jezika:

- 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 spletnih besedil Janes s 161 milijoni pojavnic (Fišer et al., 2014).

Najprej smo izluščili tipične besedne oblike (pojavnice) korpusa Gos in korpusa Janes glede na korpus Kres. Sezname ključnih besednih oblik smo izdelali z orodjem SketchEngine (Kilgarriff et al., 2004). Na korpusu Gos smo izvedli eno samo analizo, korpus Janes pa smo razdelili na tri podkorpuse (Tviti, Forumi in Komentarji), ker smo zaradi različnih okoliščin nastajanja omenjenih spletnih žanrov pričakovali različne rezultate. Ker nas je zanimala neformalna komunikacija, smo v korpusu Janes analizirali le nestandardna besedila (tista, ki imajo pripisano 2. ali 3. raven jezikovne nestandardnosti) (Ljubešić et al., 2015).

GOS	Forumi	Twitter	Komentarji
<i>eee</i>	<i>avto</i>	<i>btw</i>	<i>ane</i>
<i>mhm</i>	<i>tud</i>	<i>oz.</i>	<i>nebi</i>
<i>eem</i>	<i>mal</i>	<i>cca</i>	<i>nevem</i>
<i>sej</i>	<i>tko</i>	<i>slo</i>	<i>ala</i>
<i>tud</i>	<i>blo</i>	<i>lol</i>	<i>kriv</i>
<i>zdej</i>	<i>tut</i>	<i>cez</i>	<i>krivi</i>
<i>tko</i>	<i>gor</i>	<i>bos</i>	<i>obsojen</i>
<i>aha</i>	<i>jst</i>	<i>nic</i>	<i>fajn</i>
<i>blo</i>	<i>mam</i>	<i>prevec</i>	<i>cel</i>
<i>tak</i>	<i>gume</i>	<i>mogoce</i>	<i>neprimerno</i>

Tabela 1: 10 najbolj ključnih besednih oblik v analiziranih korpusih.

Na prvih 200 besednih oblikah z vsakega seznama smo identificirali elemente, pri katerih je zaznati interakcijo z drugimi udeleženci. Nato smo elemente interakcije analizirali z metodologijo Tracey in Robles (2013), razširjeno s kategorijami, ki sledijo shemi jezikovnih funkcij (Jakobson, 1966):

deiktika	<i>jst</i>
vprašalnica	<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>

Tabela 2: Kategorizacija elementov interakcije.

Na koncu smo primerjali rezultate analize govornega korpusa in podkorpusev novih medijev ter poskušali določiti stopnjo in mesta prekrivanja elementov

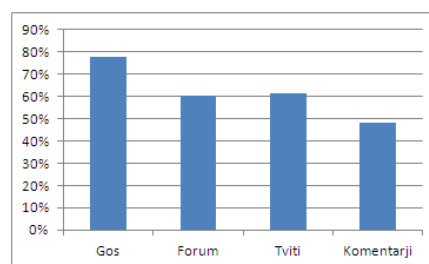
interakcije v govornem in spletnem neformalnem diskurzu.

4 Analiza

Elemente interakcije v govoru in uporabniških vsebinah smo analizirali na dveh ravneh. S kvantitativno raziskavo smo želeli proučiti tipe prvin, ki so za interakcijo v proučevanih besedilnih žanrih najznačilnejši, ter preveriti, kakšne podobnosti in razlike veljajo med njimi. Kvantitativno analizo smo nato nadgradili še s kvalitativno, pri kateri smo za vsako od proučevanih kategorij preverili, v kakšni meri so posamezni interakcijski elementi univerzalni oz. žanrsko specifični.

4.1 Kvantitativna analiza

Nabor interaktivnih besednih oblik se razteza čez vse kategorije nestandardnosti, identificirane v raziskavi Zwitter Vitez in Fišer (2015), najpogosteje pa je interakcija poleg primarne kategorije prisotna v kategoriji specifik izgovorjave oz. zapisa (43 % Gos, 25 % Forum, 37 % Tviti). Le pri podkorpusu komentarji je največ elementov interakcije vezanih na tematiko besedila (17 %). To dokazuje, da podrobnejša analiza zgolj kategorije primarno interaktivnih besed ne bi podala celotne slike, temveč je interaktivne elemente v komunikaciji treba opazovati na celotnem besedišču.



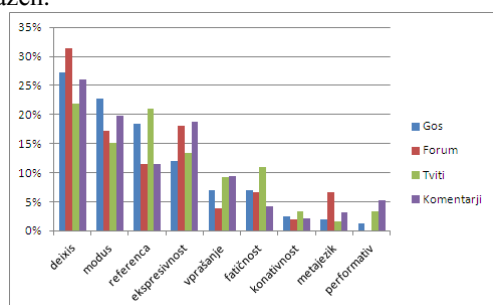
Graf 1: Prisotnost elementov interakcije v korpus Gos in podkorpuseh Janes.

Končna slika analize korpusov Gos, Forum, Tviti in Komentarji kaže, da pri treh od štirih analiziranih medijev več kot 60 % analiziranih ključnih besed vsebuje elemente interaktivnosti.

Najvišjo stopnjo interakcije najdemo v govoru, razlike pa se kažejo tudi med različnimi družbenimi mediji. Največ elementov interakcije vsebuje podkorpus 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 komentarji na spletne novice, saj je primarni motiv komentatorjev sporočanje osebnega mnenja o članku/dogodkih/osebah v članku, ne toliko interakcija z drugimi.

Prevladujoča kategorija v vseh podkorpuseh 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.



Graf 2: Elementi interakcije v korpusu Gos in podkorpusih Janes.

V korpusu Gos so zelo pogosti elementi modalnosti (*dejansko, itak, pač*), ki pogosto “definirajo stopnjo gotovosti, s katero govorec izraža svoje mnenje” (Morel in Danon Boileau, 1998). Največ elementov modalnosti najdemo v komentarjih, kar je skladno z namenom komentiranja na novičarskih portalih. Podobno vlogo imajo ekspresivni izrazi, le da so bolj osredotočeni na avtorja samega (*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 sporočanske položaje, v Janesu pa pogosto predstavljajo ventil za izražanje mnenj, deljenje izkušenj in sproščanje frustracij uporabnikov.

Posebnost tvitov je izstopanje referencialnih elementov, ki se nanašajo na udeležence v komunikaciji (*poznam, videl, imajo*), ki je najvišja med vsemi analiziranimi podkorpusi. To kaže na visoko stopnjo interaktivnosti tega medija, v katerem udeleženci sebe in svoje sogovornike eksplicitno vključujejo v sporočila. Tviti izstopajo tudi po deležu fatičnih elementov (*aja, btw, čao*), saj gre za priljubljen kanal efemerne komunikacije, medtem ko komentarji niso namenjeni vzpostavljanju in vzdrževanju stika med sogovorniki, temveč za izražanje lastnega mnenja.

Po številu vprašalnih izrazov izstopajo tviti in komentarji, kjer jih je celo več kot v korpusu Gos, ki ga zaradi majhnosti nismo razdelili na podkorpuse, tako da smo vanj zajeli tudi manj interaktivne govorne zvrsti, kot so radijske oddaje, predavanje ipd. V kategorijo metajezik smo zajeli izraze, ki se nanašajo na izrekanje samo, kar lahko 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 in performativi, ki jih v vsakem korpusu najdemo le po nekaj primerov.

4.2 Kvalitativna analiza

4.2.1 Osrednje kategorije

Kvalitativno analizo pričenjamo s kategorijami, ki zavzemajo največji delež analiziranega ključnega besedišča, obenem pa pri njih na kvantitativni ravni prihaja do največjih razlik med posameznimi proučevanimi korpusi, in sicer elementi modalnosti, referencialni elementi, deiktika in vprašalnice. Podrobnejša analiza elementov modalnosti pokaže, da lahko večino razlik med rezultati pripišemo različnim

izgovornim in pisnim variantam besed, ki so precej bogatejša v korpusu Gos, kar nakazuje na večjo regionalno razpršenost zajetega gradiva, kot ga vsebuje korpus Janes.

To lahko ponazorimo z deiktičnimi elementi, od katerih so v korpusu Gos najznačilnejše različne izgovorne variante osebnih in kazalnih zaimkov, kot so *jaz, tisti* (npr. *jst, jest, jez, jaz*), ter časovnih in prostorskih prislovov, kot so *zdaj/prej/potem* ter *tukaj/tam* (npr. *zdej, zaj, zej, zdele*). Čeprav se jih večina pojavi tudi v vseh podkorpusih Janes, jih tam najdemo le v eni ali dveh variantah. Za spletne žanre pa so specifični svojilni in kazalni zaimki (npr. *moj, tvoj, tale, tole*), ki za razliko od govora v računalniško posredovani komunikaciji eksplicitno izraženi, saj je v tem načinu komunikacije onemogočena gestikulacija, sogovorniki pa si tudi ne delijo fizičnega in/ali časovnega konteksta. Podobno je s kategorijo vprašalnic, ki imajo v korpusu Gos precej več variant (npr. *kok, kuk, kolko*) kot v tvitih (npr. *kolk, kok*), še manj pa jih je v forumih in komentarjih (npr. *koliko*). Sicer se vprašanja *kdo/kje/kdaj/koliko* pojavijo v vseh analiziranih žanrih, povsod razen v forumih, ki v primerjavi s preostalimi korpusi močno izstopajo po majhnem naboru identificiranih vprašalnic, najdemo *zakaj*, v tvitih, ki jih od spletnih žanrov vsebujejo največ, pa še *koga* in *kdaj*. Tudi analiza referencialnih elementov, ki večinoma sestoji iz različnih oblik glagolov *biti, imeti, videti, iti* pokaže podobne rezultate, vendar tokrat za razliko od ostalih kategorij več variantnosti zanje najdemo v spletnih žanrih (npr. v tvitih *imaš, maš, mas*), najverjetneje zato, ker v govoru varianta brez začetnega *i* tako močno prevladuje, poleg tega pa je za Twitter tipično izpuščanje diakritik *čšž*.

Sekundarni razlog za razlike smo našli v mediju, preko katerega poteka komunikacija, ki diktira težavnost in posledično ekonomičnost izražanja. V govoru in na forumih, kjer oteževalnih okoliščin praktično ni, je tako elementov modalnosti nekoliko več in so hkrati tudi bolj raznoliki kot v tvitih in komentarjih, kjer uporabnika že tehnične okoliščine (omejena dovoljena dolžina, uporaba mobilnih naprav) in komunikacijske konvencije (komuniciranje v realnem času) spodbujajo k čim krajšim besedilom in posledično omejeni rabi elementov modalnosti. Tako v govoru kot vseh tipih spletnih uporabniških vsebin se pojavljajo tipični prislovi za neformalne govorne okoliščine, kot so *ful, itak, kao, pač*. Za spletne žanre so specifični omilitveni izrazi, kot so *baje, verjetno, zgleda*, pa tudi izrazi, *nebi, nevem, pomoje* (zaradi drugačnih načel transkripcije govora, ki ju obravnava kot niza dveh besed), medtem ko v domeni govora ostajajo regionalno zaznamovani izrazi *glih, provzaprov, zlo*, pa tudi *okej* (zaradi drugačnih načel transkripcije govora, saj je isti izraz v korpusu Janes zapisan kot *ok*).

4.2.2 Obrobne kategorije

V drugi del kvalitativne raziskave smo zajeli kategorije, ki predstavljajo manjši delež analiziranega ključnega besedišča, obenem pa pri njih na kvantitativni ravni nismo zaznali velikih razlik med korpusi. Med te kategorije prištevamo ekspresivne, fatične in konativne izraze.

Čprav je delež zastopanosti ekspresivnih elementov v analiziranih korpusih praktično identičen, se ti med seboj močno razlikujejo glede na vrsto ekspresivnih elementov, ki se v njih pojavljajo. V korpusu Gos prevladujejo sredstva sočasnega načrtovanja in tvorjenja besedil (npr. *eee, eem, mmm*), ki jih v pisnih žanrih zaradi drugačnega načina tvorjenja in organizacije besedil ne srečamo. Po drugi strani pa so v spletnih žanrih pogosto eksplicitno verbalizirani elementi komunikacije, ki tipično govor spremljajo kot neverbalni elementi komunikacije (npr. *hehe, hahaha, lol*). Zanimive so tudi razlike med posameznimi žanri uporabniško posredovane komunikacije. Medtem ko se ekspresivni element *super* pojavlja tako v korpusu Gos kot v vseh podkorpusih Janes, tviti vsebujejo še številne druge pozitivne ekspresivne elemente (npr. *kul, top, fajn, všeč*), medtem ko v komentarjih najdemo predvsem negativne (npr. *sramota, škoda, žal, brez veze*). Zelo malo prekrivanja med analiziranimi korpusi smo zaznali tudi pri fatičnih elementih. Medtem ko jih v Gosu večina služi za začetek komunikacije (npr. *dobro jutro*), so v spletnih žanrih uporabljeni predvsem za njeno zaključevanje (npr. *lep pozdrav, lep dan, lep vikend*), za tvite pa so značilna še kratka, pomensko izpraznjena sporočila, katerih primarna funkcija je vzdrževanje komunikacijskega kanala (npr. *sej, aje, aja*).

Povsem drugače je s konativnimi elementi, pri katerih ni praktično nobenih razlik med žanri (npr. *dej, gremo, morš*), z eno samo izjemo, in sicer spodbujanje, naj sogovornik *prijavi* (najverjetneje neprimerno vsebino), kar smo zasledili samo v komentarjih. Razen besedišča, specifičnega za posamezni medij (npr. *odgovor, komentar, forum*), so povsem prekrivni tudi elementi metajezika, kar je zanimivo, saj so tudi v negovorjenih žanrih povsem enakomerno kot v govoru zastopani izrazi, kot so *reku, povem, govorim*, le v forumu poleg naštetih zasledimo še *napisal*. Podobno velja za performative, od katerih se v vseh žanrih pojavi *hvala*, v jeziku vseh analiziranih družbenih omrežij za razliko od govora obstaja potreba po eksplicitnem izražanju strinjanja s sogovornici (*strinjam*), za tvite in komentarje je značilno še voščilo *bravo*, medtem ko smo vljudnostni *prosim* poleg govora identificirali samo v komentarjih.

5 Zaključek in nadaljnje delo

Z raziskavo smo analizirali elemente interakcije v gorjenem diskurzu in besedilih računalniško posredovane komunikacije glede na standardna pisna besedila.

Kvantitativna analiza je pokazala visoko stopnjo ujemanja med kategorijami interakcije v korpusu Gos in pokorpusih Janes, saj se tovrstni elementi raztezajo čez večino analiziranih ključnih besed. Najbolj značilna kategorija interaktivnih elementov so deiktični izrazi, ki so posebej pogosti v forumih, elementi modalnosti, ki jih je največ v komentarjih, ter referencialni in ekspresivni izrazi, ki prevladujejo v tvitih.

Kvalitativna analiza pa je pokazala, da se dejanske oblike interakcije v govorjenih in spletnih besedilih realizirajo z drugačnimi sredstvi in imajo posledično tudi drugačno vlogo v diskurzu. Pri vseh kategorijah se je

namreč izkazalo, da je variantnost izgovora v korpusu Gos precej večja kot variantnost zapisa v korpusu Janes, kar je lahko posledica regionalne razpršenosti diskurza v korpusu Gos ali pa nezmožnost identifikacije prek dejanske glasovne podobe nekaterih pogostih izrazov (npr. *zej* za *zdaj*). Drugo razliko pogojuje prostorska oddaljenost udeležencev, kar se kaže skozi prevladujoči obliki svojilnih in kazalnih zaimkov med deiktičnimi izrazi v spletnih žanrih (npr. *tale, tvoj*). Tretja opazna točka razhajanja med govorom in analiziranimi spletnimi žanri je potek načrtovanja in tvorjenja besedil, zaradi česar se razlikujejo predvsem ekspresivni izrazi (*eee* v govoru, *lol* v spletnih žanrih).

V nadaljnjih raziskavah nameravamo preučiti večji nabor ključnih besednih oblik, podrobneje raziskati ostale nestandardne kategorije (izgovoru podoben zapis, neformalni izrazi) in preveriti možnost uporabe metapodatkov v korpusu Gos za potrebe profiliranja avtorjev spletnih besedil.

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.

6 Literatura

- Svenja Adolphs in Ron Carter. 2013. Spoken corpus Linguistics. From monomodal to multimodal. Routledge.
- David Crystal. 2007. How language works. Penguin Books.
- Darja Fišer, Tomaž Erjavec, Ana Zwitter Vitez in Nikola Ljubešić. 2014. JANES se predstavi: metode, orodja in viri za nestandardno pisno spletno slovenščino. V: Proceedings of the 17th International Multiconference Information Society - IS 2014, str. 56–61, Ljubljana.
- Susan Herring. 2002. Interactional coherence in CMC. Journal of computer-mediated communication.
- Roman Jakobson. 1976. Six leçons sur le son et le sens. Minuit.
- Adam Kilgarriff, Pavel Rychly, Pavel Smrz in David Tugwell. 2004. The Sketch Engine. Proceedings EURALEX 2004, str. 105–116, Lorient.
- Nataša Logar Berginc in Simon Krek. 2012. New Slovene Corpora within the Communication in Slovene Project, Prace Filologiczne, 63, str. 197–207.
- Nikola Ljubešić, Darja Fišer, Tomaž Erjavec, Jaka Čibej, Dafne Marko, Senja Pollak in Iza Škrjanec. 2014. Predicting the level of standardness of text in user generated content, Proceedings of the RANLP15 Conference. Hissar, Bulgaria.
- Mary-Annick Morel in Laurent Danon Boileau. 1998. Grammaire de l'intonation. Ophrys.
- Karen Tracey in Jessica S. Robles. 2013. Everyday talk, Second Edition: Building and Reflecting Identities. Guilford Press.
- Alison Wray. 2005. Formulaic Language and the Lexicon. Cambridge University Press.
- Ana Zwitter Vitez in Darja Fišer. 2015. From mouth to keyboard: the place of non-canonical written and spoken structures in lexicography. str. 250–267, Proceedings of the eLex 2015 conference, United Kingdom.