

# Primerjava rabe ekspresivnih ločil v tvitih slovenskih uporabnikov in uporabnic

Eneja Osrajnik,\* Darja Fišer,† Damjan Popič†

\*Maribor

eneja.osrajnik@gmail.com

†Filozofska fakulteta Univerze v Ljubljani

darja.fiser@ff.uni-lj.si, damjan.popic@ff.uni-lj.si

## Povzetek

V pričujočem prispevku obravnavamo rabo ekspresivnih ločil, ki jih razdelimo na pravopisna ločila in emotikone, v tvitih slovenskih uporabnikov in uporabnic. Na podlagi podatkov, zbranih v podkorpusu tvitov korpusa Janes (Fišer et al., 2014), analiziramo razlike v rabi ekspresivnih ločil v tvitih moških in žensk ter njihov sentiment. Kot osnova za določanje sentimenta nam služijo avtomatsko pripisane oznake sentimenta tvitov v korpusu, upoštevamo pa tudi kontekst posameznih tvitov. Raziskava je pokazala, da uporabniki obeh spolov uporabljajo več emotikonov kot pravopisnih ločil, ženske emotikone uporabljajo pogosteje kot moški, ti pa pogosteje uporabljajo pravopisna ločila. Raziskava je pokazala tudi, da je glede na kontekst prevladujoč sentiment tako emotikonov kot pravopisnih ločil pri obeh spolih negativen.

## Comparing the use of expressive punctuation marks in Slovenian tweets written by male and female users

The following article deals with the use of expressive punctuation marks, which we divide into orthographic punctuation marks and emoticons, in the tweets of Slovenian male and female users. Based on the data gathered in the corpus Janes (Fišer et al., 2014), specifically in its subcorpus of tweets, we analyse the differences in the usage of expressive punctuation marks and their sentiment in the tweets of Slovenian male and female users. The automatically appointed sentiment designations and the context of each analysed tweet served as the basis for our analysis. We concluded that both genders use emoticons more frequently than orthographic punctuation marks and that female users use emoticons more frequently than male users. On the other hand, male users use orthographic punctuation more frequently than female users. We also established that the sentiment of the overwhelming majority of tweets – both male and female – is negative.

## 1 Uvod

Ena od značilnosti človeške komunikacije je izražanje čustev, načini njihovega izražanja pa se razlikujejo glede na medij, preko katerega sporazumevanje poteka. »V govornjeni komunikaciji čustva izražamo s pomočjo telesne govornice, obrazne mimike, tona glasu, glasnosti, tempa, poudarkov in s pomočjo pavz, v standardni pisni komunikaciji pa jih verbaliziramo« (Crystal, 2001). Upoštevati je treba, da je v tradicionalnem smislu narava javnega pisnega diskurza taka, da se čustva pretežno ne izražajo – vsaj ne odkrito – saj naj bi bil pisni diskurz bil objektivni in nezaznamovan. Kadar pa pisci svoje počutje oz. čustva kljub temu želijo izraziti, v diskurz zavestno vnašajo zaznamovane ekspresivne prvine. Komunikacija, ki poteka preko družbenih omrežij, predstavlja svojevrstno kategorijo pisnega diskurza, saj uporabniki svoja čustva sicer lahko verbalizirajo, vendar jim družbena omrežja omogočajo tudi neverbalno izražanje »temeljnih družbenih drž in osebnih občutij, čustev in čustvenih stanj, odnosov do drugih« (Ule in Kline, 1996: 43). Izražajo jih lahko z emotikoni, torej kombinacijami »različnih znakov, ki jih najdemo na tipkovnici, namenjeni pa so izražanju čustev in nadomeščanju izrazov obraza« (Crystal, 2001: 36), z ločili in s ponavljanjem ločil, ponavljanjem črk v besedah, velikimi tiskanimi črkami, gif, <sup>1</sup> hešteg <sup>2</sup> ipd.

<sup>1</sup> Prevzeto iz angleškega izraza *gif*, gre pa za vrsto animiranih slik.

<sup>2</sup> Slovenjen angleški izraz *hashtag*. Gre za znak #, ki stoji pred ključnimi pojmi sporočila v tuitu, s pomočjo katerih lahko iščemo tvite s podobno vsebino, na primer: *Papa cokolado in je najšrečnejši na svetu :-)* *#lumpi#funny#dog#chocolate*.

V pričujoči raziskavi smo se omejili na ločila, s katerimi uporabniki izražajo svoja čustva in jih v nadaljevanju imenujemo ekspresivna ločila. Neekspresivna ločila (vejica, pika, vprašaj, klicaj, dvopičje, podpičje, narekovaj, vezaj, pomišljaj ipd. v primerih, ko gre za običajno skladijsko ali neskladijsko rabo, ki ne izraža uporabnikovih čustev), smo iz analize izločili. Ekspresivna ločila smo razdelili na emotikone in »pravopisna ločila«. Med pravopisna ločila v prispevku prištevamo tri pike oz. zamolk, vprašaj in klicaj, s katerimi uporabnik izraža svoja čustva, pa tudi kopičenje ene vrste ločila za izražanje večje intenzivnosti čustev. Prvotno smo med pravopisna ločila uvrstili tudi piko (primer ekspresivne rabe pike je *Ha. Ha. Ha. Zelo smešno.*),<sup>3</sup> vendar v korpusu nismo našli primerov ekspresivne rabe pike (treba je upoštevati, da je ekspresivno rabo ločil težko analizirati iz konkordanc), zato je nismo vključili v raziskavo.

V analizo smo vključili tudi emotikone, torej ustaljene kombinacije ločil z ekspresivno funkcijo, ki izražajo zadržek do dobesednega pomena, do izbrane formulacije ipd. Kot ugotavlja Praprotnik (2003: 14), so »tovrstne simbolne reakcije pravzaprav nujne, saj izjave same na sebi ne pomenijo dovolj in ne kažejo čustvenega stanja govornca/govornice, tako da bi brez fizične ekspresije, ki razkodira specifični kontekst izjave, obstajala velika možnost napačne interpretacije.« Emotikone obravnava tudi Crystal (2008), ki preučuje značilnosti sms-sporočil, (te lahko prenesemo tudi na jezik tvitov), in sicer identificira šest glavnih značilnosti sms-sporočil – uporabo piktogramov, logogramov, inicializmov, izpust

<sup>3</sup> Primer je za večjo nazornost dodala avtorica, saj v korpusu primerov ekspresivne pike nismo našli.

črk in nestandardno črkovanje. Trdi, da je »ena najvidnejših značilnosti pisanja sms-sporočil uporaba posebnega pravopisa – uporaba posameznih črk, števil in tipografskih simbolov, ki predstavljajo besede ali celo /.../ zvoke, ki jih asociiramo z določenimi dejanji. /.../ Kadar grafične simbole uporabljamo v te namene, gre za *logograme* ali *logografe*. Logograme lahko v sms-sporočilih uporabljamo posamično ali jih medsebojno kombiniramo: *s5* spet ali *ju3* jutri.<sup>4</sup>

Pri logogramih je najpomembnejša njihova izgovorjava, ne pa vizualna podoba. Prav v tem se najbolj razlikujejo od grafičnih znakov, imenovanih *emotikoni*<sup>5</sup>/.../, kjer pomen v celoti temelji na obliki simbola (kadar jih prebiramo s strani, z glavo, nagnjeno v levo stran): :-) smeško, ;-)) pomežik, :-@ krik (ali držimo glavo pokonci, na primer v japonskem ali nekaterih drugih vzhodnoazijskih sistemih): (\*o\*) presenečenje, (^\_^) ljubko. Kadar vizualne podobe ali slike predstavljajo predmete ali koncepte, jih imenujemo *piktogrami*, med katere sodijo tudi emotikoni.<sup>6</sup> Emotikoni kot dogovorjene kombinacije grafičnih simbolov torej izražajo koncept posameznih čustev. Po drugi strani ekspresivna pravopisna ločila sicer omogočajo izražanje čustev, vendar se uporabljajo posamično, če jih uporabljamo več skupaj (v zaporedni ponovitvi ali kombinaciji različnih ločil), pa gre za govorcevo odločitev, in ne za družbeno dogovorjeno kombinacijo znakov, ki bi predstavljala določeno čustvo.

V raziskavi opazujemo ekspresivno rabo ločil v jeziku moških in žensk, ki je tudi sicer pogosto obravnavana tema v sociolingvističnih raziskavah. Tannen (1990) na primer ugotavlja, da ženski stil komunikacije izraža več podpore in bolj vzpostavlja odnose, medtem ko je za moško komunikacijo značilno podajanje poročil in visoka stopnja informativnosti, Schler et al. (2006), ki so analizirali 300-milijonski korpus blogov, pa so ugotovili, da blogerke veliko pogosteje pišejo o svojem zasebnem življenju in uporabljajo bolj osebni stil pisanja kot blogerji. Razlike v jeziku moških in žensk je na podlagi korpusa norveških sms-sporočil preučeval tudi Ling (2015) in ugotovil, da še posebej mlajše ženske v sms-sporočilih uporabljajo širši register, pišejo daljša in kompleksnejša sporočila kot moški ter pogosteje uporabljajo ekspresivne oz. čustvene elemente. Ekspresivno rabljena ločila v povezavi s spolom pogosto obravnavajo kot »označevalce razburjenosti« (Waseleski, 2006). Kivran-Swaine et al. (2013) pa raziskujejo povezavo med jezikom, spolom in socialnimi razmerji, ki se kažejo v komunikacijskih vzorcih družbenih medijev, in ugotavljajo, da »ženske uporabljajo več jakostnih prislovov, zaimkov in emotikonov, še posebej v komunikaciji z drugimi ženskami«. Med ločili v moškem in ženskem jeziku pa najbolj izstopa klicaj, saj se ugotavlja, da ga mnogo pogosteje uporabljajo ženske kot moški (Colley in Todd, 2002; Rubin in Greene, 1992; Scates, 1981)«.

Analiza ekspresivne rabe ločil v slovenskih uporabniških spletnih vsebinah še ni bila opravljena, prav tako ne poznamo primerjav ekspresivne rabe ločil pri moških in ženskah nasploh, zato se jim posvečamo v pričujočem prispevku, v katerem se omejimo na družbeno omrežje Twitter. S pomočjo podatkov, zbranih v podkorpusu tvitov korpusa Janes (Fišer et al., 2014), ugotavljamo, kakšne so razlike v rabi ekspresivnih ločil v tvitih moških in žensk ter kakšen je prevladujoč sentiment – pozitiven ali negativen –<sup>7</sup> identificiranih ekspresivnih ločil. Glede na pregled literature pričakujemo, da slovenske uporabnice Twitterja uporabljajo širši nabor ekspresivnih ločil in da jih uporabljajo pogosteje kot moški, hkrati pa želimo potrditi domnevo, da je prevladujoč sentiment ekspresivnih ločil pozitiven in da se emotikoni v tvitih uporabnikov obeh spolov pojavljajo pogosteje kot pravopisna ločila, saj v nasprotju s pravopisnimi ločili emotikoni poleg veselja in jeze oz. žalosti izražajo tudi sarkazem in ironijo (to lahko izraža tudi na primer navaden smeško), včasih pa le omilijo negativno sporočilo, ki ga spremljajo.

## 2 Priprava podatkov

Za potrebe raziskave smo pregledali seznam uporabnikov družbenega omrežja Twitter, zajetih v korpus Janes (Fišer et al., 2014), in jim določili spol (moški, ženski, nevtrarno). Najprej smo pregledali seznam uporabnikov, pri katerih je bilo mogoče zaznati rabo vsaj 5 prvoosebni glagolskih oblik v pretekliku ali prihodnjiku, ki v slovenščini eksplicitno izražajo spol uporabnika in je zato oznako spola uporabnikom mogoče pripisati avtomatsko glede na prevladujočo uporabljeno glagolsko obliko. Avtomatsko pripisane spole smo nato preverili še ročno, in sicer s pregledom uporabniških profilov uporabnikov na Twitterju, torej njihovega uporabniškega imena, morebine rabe osebnih imen in opis profila, ter tako preverili, ali pripisana oznaka spola drži. Na ta način smo določili spol 5.883 uporabnikom omrežja Twitter. Z avtomatskim postopkom je bil spol pravilno pripisan 5.320 uporabnikom (90 %), pri 563 (10 %) pa smo oznako spola morali popraviti. Med slednjimi je 12 (2 %) takšnih, ki so dobili pripisano napačno oznako zaradi navajanja tvitov drugih uporabnikov Twitterja, v 551 primerih (98 %) pa je šlo za uporabniške račune javnih ustanov (na primer potovalnih agencij, društev, podjetij), ki jim ni mogoče določiti spola, zato smo jim zaradi večje preglednosti ob predhodnem posvetu z avtorji korpusa pripisali oznako nevtrarno.

Nato smo pregledali še datoteko s seznamom 1.708 preostalih uporabnikov Twitterja v korpusu Janes, ki niso uporabili dovolj osebnih glagolskih oblik, s pomočjo katerih bi jim bilo mogoče avtomatsko pripisati spol, in jim ga pripisali ročno. 561 uporabnikom (33 %) smo na ta način spol lahko določili, 1.147 uporabnikom (67 %) pa ne, saj so pisali v izrazito neosebni stilu, iz katerega ni bilo mogoče razbrati spola, ali pa je šlo za uporabniške račune javnih ustanov. Tem smo zato pripisali oznako nevtrarno. Skupno smo torej določili spol 7.591 uporabnikom, med katerimi močno prevladujejo moški

<sup>4</sup> Avtorjeve primere smo za večjo nazornost nadomestili z ustreznimi slovenskimi primeri.

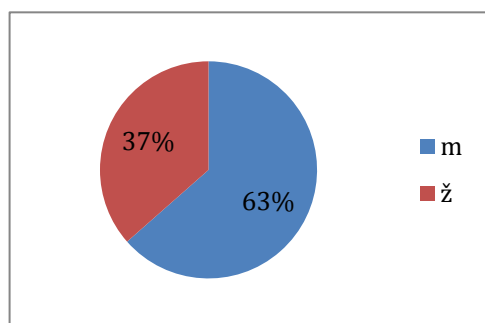
<sup>5</sup> ang. *emoticons*

<sup>6</sup> Iz angleščine prevedla avtorica članka.

<sup>7</sup> Prevzeto iz angleškega izraza *sentiment*, ki pomeni počutje, občutek, razpoloženje (gl. Merriam-Webster).

(53 %), preostali uporabniki pa so enakomerno razporejeni na ženske (24 %) in nevtralne uporabniške račune (23 %).

Na podlagi pripisanih oznak smo izdelali podkorpus tvtov moških in žensk, ki ob upoštevanju zgolj tistih, ki so glede na avtomatsko pripisani oznaki L2 in L3 (Ljubešič et al. 2015) napisani v nestandardni slovenščini, vsebuje 18.207.584 pojavnic. Od tega so jih 11.561.770 (63 %) prispevali moški, 6.645.814 (37 %) pa ženske (glej graf 1).



Graf 1: Delež pojavnic uporabnikov in uporabnic, napisanih v nestandardni slovenščini.

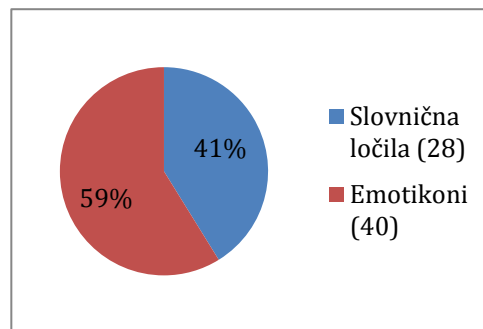
Tvtov moških je v podkorpusu torej dvakrat več kot tvtov žensk. Ker razmerje med številom pojavnic moških in ženskih uporabnikov ni uravnoteženo, pri analizi podatkov upoštevamo relativno frekvenco (normalizirano na milijon pojavnic).

### 3 Analiza podatkov

V podkorpusih tvtov moških in žensk z avtomatsko pripisanima stopnjama nestandardnosti L2 in L3 (gl. Ljubešič et al., 2015) smo identificirali vsa ločila, ki se v tvitih pojavijo vsaj desetkrat. Ob upoštevanju širšega konteksta, v katerem se ločilo pojavi, smo iz nadaljnje analize izločili neekspresivna ločila, torej ločila v skladijski in neskladijski rabi, s katerimi uporabnik ne izraža svojih čustev. Tako smo npr. med naslednjima primeroma prvega izločili iz analize, saj gre za neekspresivno rabo klicaja, drugega pa vključili, saj klicaj izraža negativno čustvo uporabnika:

*Pridi v četrtek na Medijsko ogledalo!  
Ampak ti si povrhu še ženska, bogtenimarad!*

Iz relevantnih zadetkov smo izdelali seznam 68 ekspresivnih ločil uporabnikov in uporabnic, ki smo jih nadalje razdelili na 28 pravopisnih ločil (od katerih smo jih 27 identificirali v moških tvitih, 22 pa v ženskih) in 40 emotikonov (34 se jih pojavlja v moških tvitih, 33 pa v ženskih), kar prikazuje Graf 2:



Graf 2: Delež pravopisnih ločil in emotikonov med identificiranimi ekspresivnimi ločili.

Pri tem zaporedne ponovitve istega ločila upoštevamo kot različna ločila, saj izražajo različno stopnjo čustev. Tako na primer pravopisno ločilo **!!!!** izraža večjo stopnjo razburjenosti kot **!**, enako emotikon **:)))** izraža večjo intenzivnost čustev kot **:)**. Kot različna ločila upoštevamo tudi različne kombinacije vprašaja in klicaja s ponovitvami enega ali drugega ločila, saj te izražajo različno stopnjo čustev – **?!?!** na primer izraža intenzivnejše čustvo kot **?!** (seveda ob upoštevanju širšega konteksta, v katerem se ločilo pojavi), kar lahko razberemo iz sledečih primerov iz korpusa:

*Z busom le, če je nujno. Ha!\_Kdo pravi, da so novinarji nemočni?  
hahaha.. dobra **!!!!**  
al je lepo, e? **:)**  
Voje ti si res ena #faca **:)))**  
A s' kej Slovenca al nč **?!**  
A niste mogli it na fotra a **?!?!***

Zavedamo se, da nekateri identificirani emotikoni (na primer **:/** in **:-/**) izražajo isto vrsto čustva, vendar nas zanima tudi nabor različnih ekspresivnih ločil uporabnikov in uporabnic, zato smo na seznam uvrstili vse identificirane različice emotikonov.

#### 3.1 Razlike v rabi ekspresivnih ločil pri uporabnikih in uporabnicah

Kot prikazuje tabela 1, smo izmed vseh 68 identificiranih ekspresivnih ločil v podkorpusu moških tvtov identificirali 61 ločil – od tega se uporablja dvakrat več emotikonov (34 oz. 56 %) kot pravopisnih ločil (27 oz. 44 %). V podkorpusu ženskih tvtov smo identificirali 55 različnih ločil, od tega prav tako več emotikonov (33 oz. 60 %) kot pravopisnih ločil (22 oz. 40 %). Pri tem naj spomnimo, da je podkorpus moških tvtov dvakrat večji od podkorpusa ženskih tvtov, zato nekoliko večjega nabora ločil pri moških brez podrobnejše raziskave na enako velikih naključnih vzorcih iz obeh podkorpusov ne gre interpretirati kot znak večje nagnjenosti moških k uporabi bolj raznolikega nabora ekspresivnih ločil.

	Ekspresivna ločila	Pravopisna ločila	Emotikoni
m ločila	61	27 (44 %)	34 (56 %)
ž ločila	55	22 (40 %)	33 (60 %)

Tabela 1: Delež pravopisnih ločil in emotikonov v tvitih uporabnikov in uporabnic.

Tako moški kot ženske uporabljajo več emotikonov kot pravopisnih ločil, razlika med obojimi pa je pri obeh spolih primerljiva (razmerje je okoli 60:40). Ena izmed možnih razlag za večje število emotikonov je, da emotikoni izražajo večji nabor čustev kot pravopisna ločila, lahko so tudi le dodatek k sporočilu, ki omili njegov ton. V nadaljevanju podrobneje predstavimo analizo rabe emotikonov in pravopisnih ločil v moških in ženskih tvitih, popoln popis katerih smo vključili v tabelo 3 na koncu članka.

### 3.1.1 Emotikoni

Moški najpogosteje uporabljajo naslednjih pet emotikonov: :-), :)), :))), :-), ;). Ženske pa :)), :))), ^^, :))), :-))). Zanimivo je, da sta obema spoloma skupna le :) in :))). Čeprav se ostali emotikoni razlikujejo, oba spola najpogosteje uporabljata emotikone, ki izražajo pozitivna čustva, na primer smeške in pomežike. Glede na relativno frekvenco ženske pogosteje od moških uporabljajo 21 (78 %) emotikonov (na primer (:, :))), :-))) ), moški pogosteje uporabljajo 6 (22 %) emotikonov (na primer :-), :-))) ), 13 emotikonov, ki pa jih uporabljajo le moški ali le ženske, pri tem nismo upoštevali. Zaključimo lahko, da uporabnice emotikone uporabljajo pogosteje od uporabnikov.

Tako pri rabi pravopisnih ločil kot emotikonov je opaziti pogosto uporabo zaporedne ponovitve istega ločila. Med emotikoni je več takih z zaporedno ponovitvijo istega ločila – teh je 21 (52 %), nekaj manj pa je emotikonov brez ponovitev – teh je 19 (48 %). Emotikoni s ponovitvami so variacije sledečih emotikonov: :, :-), :(, :- (in ;), glede na intenzivnost izražene ekspresivnosti pa se ponavlja število oklepajev oz. zaklepajev. Med emotikoni vsebuje največje število ponovitev smeško z 9 ponovitvami zaklepaja, ki smo ga identificirali v podkorpusu ženskih tvitov.

V izboru najpogosteje rabljenih emotikonov obeh spolov sta emotikona :) in :))).

### 3.1.2 Pravopisna ločila

Moški najpogosteje uporabljajo naslednja pravopisna ločila: ..., ?, !, !!!, ????, ženske pa: !, ?, ..., !!!, ??? . Zanimivo je, da se izbor petih najpogosteje rabljenih pravopisnih ločil pri obeh spolih prekriva, razlikujejo se le po pogostosti rabe. Glede na relativno frekvenco ima 7 (33 %) pravopisnih ločil več pojavitev v tvitih uporabnic (na primer !, ?!), 14 (67 %) ločil ima več pojavitev v tvitih uporabnikov (na primer !?!, ??), 7 pravopisna ločil pa uporabljajo le moški, zato smo jih iz tega dela analize izključili. Analiza kaže, da uporabniki pravopisna ločila uporabljajo pogosteje od uporabnic.

Pravopisna ločila lahko razdelimo v tri kategorije:

- pravopisna ločila brez zaporednih ponovitev istega ločila (v korpusu smo identificirali 2 taki ločili, in sicer vprašaj in klicaj),
- zaporedne ponovitve pike, vprašaja in klicaja (v korpusu smo identificirali 21 takih ločil) in
- kombinacije vprašaja in klicaja (v korpusu smo identificirali 5 takih ločil).

Ugotovili smo, da vprašaj in klicaj brez zaporednih ponovitev pogosteje uporabljajo uporabniki moškega kot ženskega spola, vendar smo identificirali več pravopisnih ločil z zaporednimi ponovitvami istega ločila kot brez njih, in sicer gre za ponovitve pik, vprašajev in klicajev. Med ločili z zaporednimi ponovitvami istega ločila tako moški kot ženske najpogosteje uporabljajo večpičja (... , ..., .....), in sicer smo identificirali 12 večpičij z več kot tremi zaporednimi pikami, pri čemer moški in ženske najpogosteje uporabljajo večpičje s štirimi zaporednimi pikami, zaporedje z devetimi pikami uporabljajo le ženske, zaporedja z dvanajstimi, s trinajstimi in štirinajstimi zaporednimi pikami pa le moški. Najdaljše večpičje torej vsebuje štirinajst ponovitev in se pojavi v podkorpusu moških tvitov.

Identificirali smo tudi 4 različne zaporedne ponovitve vprašaja (tako moški kot ženske najpogosteje uporabljajo tri zaporedne ponovitve vprašaja, največje identificirano število ponovitev je pet, to ločilo pa uporabljajo le moški) in 4 različne zaporedne ponovitve klicaja (tako moški kot ženske najpogosteje uporabljajo tri zaporedne ponovitve, največje identificirano število ponovitev klicaja je pet, ločilo pa moški uporabljajo pogosteje od žensk). Pri tem se zavedamo, da se je (še posebej pri več kot treh zaporednih ponovitvah pike) uporabnik morda zatipkal, vendar bi v takšnem primeru dodal le eno ali dve »odvečni« ločili. Zaporedna ponovitev na primer sedem pik pa kaže, da jih je uporabnik napisal namenoma, da poudari intenziteto svojih čustev.

Analizirali smo tudi kombinacije vprašaja in klicaja ter ugotovili, da oba spola najpogosteje uporabljata kombinaciji ?! in !?, moški redko uporabljajo tudi !?! in !?!!, ženske pa sploh ne.

### 3.1 Sentiment ločil

Sledi analiza ekspresivnosti pravopisnih ločil in emotikonov, kjer smo klasificirali 60 naključnih pojavitev posameznega ločila. Če se je ločilo v korpusu pojavilo manj kot 60-krat, smo pregledali vse pojavitve.

Avtomatsko pripisane oznake sentimenta tvitov, v katerih se pojavijo identificirana ekspresivna ločila, so nam služile kot smernica za določanje sentimenta ločil. Za končno določitev pozitivnega ali negativnega sentimenta ločil smo ročno pregledali tvite, v katerih se identificirana ekspresivna ločila pojavijo, saj je njihov sentiment odvisen od širšega konteksta. Klicaju ali vprašaju lahko na primer določimo tako pozitiven kot negativen sentiment, odvisno od konteksta, v katerem se pojavljata.

Posebej smo analizirali moška in ženska pravopisna ločila ter emotikone. Za ponazoritev navajamo primer pravopisnega ločila, ki smo mu glede na avtomatsko pripisano oznako sentimenta tvita in širšega konteksta besedila določili pozitiven sentiment (prvi primer), in primer emotikona, ki smo mu določili negativen sentiment (drugi primer):

*Košiiiiiiiiir!!! Srebrna medalja, osma za Slovenijo!  
Sramota! :/*

Rezultate analize sentimenta ekspresivnih ločil prikazuje tabela 3 (v prilogi). Največjo razliko med deležema negativnih in pozitivnih sentimentov je opaziti pri pravopisnih ločilih, saj jih ima med 27 pravopisnimi ločili v tvitih moških kar 24 (89 %) negativen sentiment, le 3 (11 %) pa pozitivnega. Podobno ima med 22 pravopisnimi ločili v tvitih žensk kar 17 (77 %) ločil negativen in le 5 (23 %) pozitiven sentiment. Razlika med deležem pozitivnih in negativnih sentimentov je nekoliko manjša pri emotikonih – med 34 emotikoni v tvitih moških jih ima 21 (62 %) negativen sentiment, 13 (38 %) pa pozitivnega. Med 33 ženskimi emotikoni jih ima 18 (55 %) negativen sentiment, 15 (54 %) pa pozitivnega, kar tudi prikazuje tudi tabela 2:

Spol	Pravopisna ločila		Emotikoni	
	-	+	-	+
m	24 (89 %)	3 (11 %)	21 (62 %)	13 (38 %)
ž	17 (77 %)	5 (23 %)	18 (55 %)	15 (54 %)

Tabela 2: Deleži ekspresivnih ločil glede na sentiment in spol uporabnikov.

Zdi se presenetljivo, da se največ ekspresivnih ločil pojavlja v negativnem kontekstu, vendar je treba upoštevati, da ekspresivna ločila v negativnih tvitih velikokrat ne izražajo le jeze ali žalosti, temveč tudi sarkazem in ironijo (tudi na primer navaden smeško), včasih pa tudi le omilijo negativno sporočilo, ki ga spremljajo.

#### 4 Sklep

Za komunikacijo na družbenih omrežjih je značilno, da osebe, s katero komuniciramo, ne vidimo ali slišimo. Pomanjkanje telesne govorice in obrazne mimike ter nadsegmentnih zvočnih lastnosti diskurza za izražanje čustev uporabniki družbenih omrežij nadomestijo z uporabo ekspresivnih ločil – emotikonov in pravopisnih ločil. Naša analiza je pokazala, da med identificiranimi ekspresivnimi ločili prevladujejo emotikoni (59 % vseh identificiranih ekspresivnih ločil), manj pa je pravopisnih ločil (41 %). Tudi med ločili, ki jih uporabljajo le moški ali le ženske, večino predstavljajo emotikoni (56 % v moških in 60 % v ženskih tvitih), manj pa je pravopisnih ločil (44 % v moških in 40 % v ženskih tvitih). Moški uporabljajo nekoliko večji nabor ekspresivnih ločil kot ženske, vendar je to morda rezultat neuravnoteženosti korpusov moških in ženskih tvitov. Za potrditev tega zaključka bi torej morali raziskavo ponovno izvesti na enako velikih vzorcih iz obeh korpusov.

Dokazali smo tudi, da uporabniki ne glede na spol pogosteje uporabljajo emotikone kot pa pravopisna ločila. Večjo pogostost emotikonov lahko razložimo z dejstvom, da lahko s pravopisnimi ločili uporabniki družbenih omrežij izražajo precej omejen izbor čustev – izražajo lahko le osnovna pozitivna in negativna čustva (kot sta veselje ali jeza), ne morejo pa izražati žalosti, ironije, sarkazma ipd. Izražanje te vrste kompleksnih čustev jim

omogočajo emotikoni, ki pogosto igrajo zgolj vlogo dodatka k besedilu, saj podkrepijo njegovo sporočilo ali ga omilijo. Analiza rabe emotikonov v tvitih uporabnikov in uporabnic je pokazala, da uporabniki pogosteje uporabljajo emotikone z zaporedno ponovitvijo istega ločila (52 %) kot pa emotikone brez ponovitve ločila (48 %). Najpogosteje rabljeni emotikoni s ponovitvami so variacije sledečih emotikonov: :, :-), :(, :-( in :), glede na intenzivnost izražene ekspresivnosti pa se ponavlja število njihovih oklepajev oz. zaklepajev.

Podrobna analiza pravopisnih ločil pa je pokazala, da pravopisna ločila brez zaporednih ponovitev istega ločila (torej enojna vprašaj in klicaj) predstavljajo 7 % vseh identificiranih pravopisnih ločil, 18 % pravopisnih ločil predstavljajo kombinacije vprašaja in klicaja, uporabniki pa najpogosteje uporabljajo pravopisna ločila z zaporednimi ponovitvami istega ločila – ta predstavljajo 75 % vseh pravopisnih ločil.

Po pričakovanjih smo glede na število ekspresivnih ločil v tvitih uporabnikov in uporabnic ob upoštevanju neuravnoteženosti števila pojavnic v podkorpusu moških in ženskih tvitov ugotovili, da uporabnice emotikone uporabljajo bistveno pogosteje od uporabnikov (med emotikoni, skupnimi obema spoloma, se jih 78 % pogosteje pojavi v tvitih uporabnic, le 22 % pa pogosteje v tvitih uporabnikov). Presenetljivo pa je, da se pravopisna ločila pogosteje pojavljajo v tvitih moških (67 %) Zaključimo lahko, da slovenski uporabniki Twitterja za izražanje čustev pogosteje uporabljajo pravopisna ločila, uporabnice pa emotikone.

V nasprotju z našimi prvotnimi pričakovanji pa je analiza sentimenta pri emotikonih in pravopisnih ločilih pokazala, da s pravopisnimi ločili uporabniki obeh spolov najpogosteje izražajo negativna čustva. To velja tudi za emotikone, čeprav je razlika med deleži pozitivnih in negativnih sentimentov pri njih precej manjša kot pri pravopisnih ločilih. To se morda na prvi pogled zdi nekoliko presenetljivo, vendar je treba upoštevati, da ekspresivna ločila v negativnih tvitih velikokrat ne izražajo le jeze ali žalosti, temveč tudi kompleksna čustva, kot sta sarkazem in ironija, včasih pa le omilijo ton sporočila, ki ga spremljajo.

Zaradi razlike v velikosti obeh analiziranih podkorpusov bi za potrditev teh zaključkov morali izvesti podrobnejšo raziskavo na enako velikih naključnih vzorcih iz obeh podkorpusov. Prav tako bi bilo zanimivo raziskavo razširiti še na forume, komentarje, spletne novice in bloge.

#### 5 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.

#### 6 Literatura

- Teja Antončič. 2006. *Jezikovne in stilne posebnosti novih medijev: primer spletnih klepetalnic*. Diplomsko delo, FDV, Ljubljana.
- Michael Argyle. 1988. *Bodily Communication*. London: Methuen, druga izdaja.
- Brittney G. Chenault. 1998. *Computer-Mediated Communication and Emotion: Developing Personal Relationship Via CMC*. *Computer-Mediated*

- Communication Magazine*. Dostopno na: <http://www.december.com/cm/mag/1998/may/chenault.html> (3. julij 2015).
- Ann Colley in Zazie Todd. 2002. Gender-linked differences in the style and content of e-mails to friends. *Journal of Language and Social Psychology*, 21(4): 380–392.
- David Crystal. 2001. *Language and the Internet*. University Press, Cambridge.
- David Crystal. 2008. *Txtng: The Gr8 Db8*. Oxford University Press.
- Darja Fišer et al., 2015: Gradnja in analiza korpusa spletne slovenščine JANES (Zbornik simpozija Obdobja – v tisku).
- Darja Fišer, Tomaž Erjavec, Ana Zwitter Vitez, Nikola Ljubešić. 2014. JANES se predstavi: metode, orodja in viri za nestandardno pisno spletno slovenščino. Zbornik Devete konference Jezikovne tehnologije. Ljubljana: Institut Jožef Stefan.
- Tomaž Erjavec in Darja Fišer. 2013. Jezik slovenskih tvitov: korpusna raziskava. V: A. Žele, ur., *Družbena funkcijskost jezika: (vidiki, merila, opredelitve)*, Obdobja 32, str. 109–116. Ljubljana. Filozofska fakulteta.
- Seth Grimes. 2008. *Sentiment analysis: A Focus on Applications*. BeyeNETWORK – Global coverage of the business Intelligence Ecosystem. Dostopno na: <http://www.b-eye-network.com/view/6897> (27. september 2015).
- Funda Kivran-Swaine, Sam Brody in Mor Naaman. 2013. *Effects of gender and tie strength on twitter interactions*, 18(9). Dostopno na: <http://firstmonday.org/ojs/index.php/fm/article/view/4633/3746> (27. september 2015).
- Simona Kranjc. 2003. Jezik v elektronskih medijih. V: A. Vidovič Muha, ur., *Slovenski knjižni jezik – aktualna vprašanja in zgodovinske izkušnje*, Obdobja 20, str. 435–446. Ljubljana. Filozofska fakulteta.
- Robin Lakoff. 1975. Language and Woman's Place. *Language in Society*, 2: 45–80.
- Rich Ling. 2015. The Sociolinguistics of SMS: An Analysis of SMS Use by a Random Sample of Norwegians. V: R. Ling in P. Pedersen, ur., *Mobile communications: Renegotiation of the social sphere*, str. 335–349, London: Springer.
- Nikola Ljubešić, Darja Fišer, Tomaž Erjavec, Jaka Čibej, Dafne Marko, Senja Pollak in Iza Škrjanec. 2015. Predicting the level of text standardness in user-generated content. *10th International Conference on Recent Advances in Natural Language Processing: Proceedings of RANLP 2015 Conference*, 7.–9. September 2015, Hissar, Bolgarija: 371–378.
- Merriam-Webster. Dostopno na: <http://www.merriam-webster.com/dictionary/sentiment> (25. oktober 2016).
- Mija Michelizza. 2014. Slovenščina v elektronskih medijih. *Razpotja*, 15. Dostopno na: <http://www.razpotja.si/slovenscina-v-elektronskih-medijih/> (26. septembr 2015).
- Róisín Parkins. 2012. Gender and Emotional Expressiveness: An Analysis of Prosodic Features in Emotional Expression. *Griffith Working Papers in Pragmatics and Intercultural Communication* 5(1): 46–54.
- Sarah Pedersen in Caroline Macafee. 2007. Gender differences in British blogging. *Journal of Computer-Mediated Communication*, 12(4): 1472–1492.
- Tadej Praprotnik. 2003. Skupnost, identiteta in komunikacija v virtualnih skupnostih. ISH, Ljubljana.
- Donald Rubin in Kathryn Greene. 1992. Gender-typical style in written language. *Research in the Teaching of English*, 26(1): 7–40.
- Carol Scates. 1981. *A Sociolinguistic Study of Male/Female Language in Freshman Composition*. Neobjavljena doktorska disertacija. University of Southern Mississippi, Hattiesburg, Mississippi.
- Jonathan Schler, Moshe Koppel, Shlomo Argamon in James Pennebaker. 2006. Effects of age and gender on blogging. V: N. Nicolov, F. Salvetti, M. Liberman in J. H. Martin, ur., *Zbornik konference AAAI Spring Symposium on Computational Approaches for Analyzing Weblogs*, str. 199–206, AAAI Press.
- SP 2001 – Slovenski pravopis. Ljubljana: SAZU – ZRC SAZU – Založba ZRC. § 226.
- Deborah Tannen. 1990. *You Just Don't Understand*. New York, Ballantine.
- Mirjana Ule in Miro Kline. 1996. *Psihologija tržnega komuniciranja*. FDV, Ljubljana.
- Tina Verovnik. 2003. Analiza jezikovne kakovosti besedil v vladnem spletu. V: A. Lukšič in T. Oblak, ur., *S poti v digitalno demokracijo*, 147–158. Dostopno na: <http://odkw.fdv.uni-lj.si/eknjige> (3. julij 2015).
- Carol Waseleski. 2006. Gender and the Use of Exclamation Points in Computer-Mediated Communication: An Analysis of Exclamations Posted to Two Electronic Discussion Lists. V: G. Gay, ur., *Journal of Computer-Mediated Communication*, 11, str. 1012–1024.
- Diane F. Witmer in Sandra Lee Katzman. 2006. On-line smiles: Does gender make a difference in the use of graphic accents? *Journal of Computer-Mediated Communication*, 2 (4). Dostopno na: <http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.1997.tb00192.x/full> (26. september 2015).

## 7 PRILOGE

Priloga 1: Tabela 3

Emotikon	Št. pojavitev		Sentiment med naključnimi 60 zadetki				Pravopisno ločilo	Št. pojavitev		Sentiment med naključnimi 60 zadetki			
	m	ž	Negativen		Pozitiven			m	ž	Negativen		Pozitiven	
			m	ž	m	ž				m	ž	m	ž
(:	30	51	19	32	11	19	...	122.251	59.983	48	48	12	12
:))	5.591	5.585	27	31	33	29	....	10.349	6.318	44	31	16	29
:))	2.027	2.705	21	17	26	43	.....	1.903	1.273	42	16	18	44
((:(	0	12	0	7	0	5	.....	792	351	39	42	21	18
:)))	584	574	38	21	22	39	.....	306	247	45	38	15	22
:))))	252	162	27	45	33	15	.....	165	103	47	38	13	22
:))))	80	56	32	21	28	35	.....	136	59	46	41	14	18
:))))))	36	19	15	4	21	15	.....	0	32	0	26	0	6
:))))))	18	18	7	12	11	6	.....	37	16	28	10	9	6
:))))))	0	11	0	5	0	6	.....	25	16	16	12	12	4
:~)	13.906	225	12	24	48	36	.....	24	0	20	0	4	0
:~))	1.946	239	35	18	25	42	.....	15	0	6	0	9	0
:~)))	308	517	25	32	35	18	.....	14	0	8	0	6	0
:~))))	107	49	38	13	22	26	?	141.724	75.542	42	50	16	10
:~))))	47	22	25	26	22	34	??	397	121	55	53	5	7
:~))))))	15	0	10	0	5	0	???	427	146	59	45	1	15
:(((	148	158	56	55	4	5	????	80	31	49	29	11	2
:(((	46	155	43	59	3	1	?????	22	0	16	0	6	0
:((((	14	38	14	31	0	7	!	110.498	77.202	44	34	17	26
:((((	11	0	11	0	0	0	!!	400	168	32	16	28	44
:(-(	551	0	57	0	3	0	!!!	696	191	27	21	33	39
:(-(	50	12	44	12	6	0	!!!!	158	48	25	23	35	25
:(-(	10	0	10	0	0	0	!!!!!	38	21	22	6	17	15
:/	26	18	23	16	3	2	?!	152	129	52	47	8	13
:~/	322	0	56	0	4	0	!?	125	18	48	17	12	1
:('	33	44	31	42	2	2	?!?	39	22	35	22	4	0
:')	0	52	0	19	0	33	!?!	17	0	15	0	2	0
(=	21	47	15	31	6	16	?!?!	15	0	15	0	0	0
=)	13	0	5	0	8	0							
:)	684	461	23	22	37	38							
:))	0	10	0	4	0	6							
(;	20	14	9	8	11	6							
^^	364	1.170	26	16	34	44							
^.^	11	34	4	23	7	11							
>_<	0	23	0	14	0	9							
>.<	55	125	39	48	16	12							
^_^	46	13	27	3	19	10							
*.*	11	21	5	8	6	13							
.*	10	0	3	0	7	0							
>=	0	15	0	11	0	4							

Tabela 3: Podatki o ekspresivnih ločilih, številu njihovih pojavitev in njihov sentiment glede na spol uporabnikov.