

## #Analiza novih komunikacijskih elementov na družbenem omrežju @Twitter

Katerina Pertot, Maja Petrovčič, Nika Strojjan

Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani  
AŠkerčeva cesta 2, 1000 Ljubljana maja.petrovcic86@gmail.com,  
nikica.spika@gmail.com,  
kpertot@gmail.com.

### Povzetek

V pričujočem prispevku so navedeni izsledki analiz novih komunikacijskih elementov na družbenem omrežju Twitter (heštegi, omembe uporabniških imen in izrazov čustev - emojiji in emotikoni), ki smo jih opravili za slovenski in angleški del korpusa spletnih uporabniških vsebin Janes. V prispevku je podrobneje opisan potek analiz ter kriteriji za analize. Analize so bile izvedene na podlagi frekvenčnih seznamov in naključnih vzorcev rabe, ki smo jih pridobili v korpusu Janes.

### #Analysis of New Communication Elements on the @Twitter Social Network

This paper reports on the analysis of new communication elements used on the Twitter social network (hashtags, mentions and emotional expressions - emojis and emoticons). The analysis was conducted with data, acquired from the English and Slovene parts of the corpus of user-generated Slovene called Janes. The analysis procedure and its criteria are also described in this article. The basis for our analysis were frequency lists and random usage samples from the Janes corpus.

## 1 Uvod

Jezik, ki ga uporabljamo na internetu, je drugačen od tistega, ki ga uporabljamo v vsakdanjem življenju. Na internetu smo namreč lahko anonimni, v vsakdanji komunikaciji pa to anonimnost zelo težko dosežemo, sploh če z nekom komuniciramo iz oči v oči. O preiskovanju jezika, ki ga uporabljamo na internetu, je v svoji knjigi pisal tudi David Crystal (2001, str. 5): »Naša naloga je preiskovati lingvistične značilnosti tako imenovane »elektronske revolucije« in podrobneje preučiti, ali se uporaba jezika na internetu res spreminja.«

V pričujočem prispevku se osredotočamo na jezik, ki ga uporabljajo uporabniki družbenega omrežja Twitter<sup>1</sup>, za katerega je značilna vrsta novih komunikacijskih elementov. Na Twitterju ima vsak uporabnik svoje uporabniško ime, ki je lahko njegovo lastno ali pa izmišljeno. Uporabniška imena se uporabljajo tudi v sporočilih pri omenjanju drugih uporabnikov (ang. mention). Za lažje sledenje objavam na Twitterju uporabniki uporabljajo t. i. heštege (ang. hashtag). Za izražanje svojih čustev in trenutnega počutja pa se uporabljajo izrazi čustev – emojiji (ang. emoji) in v zadnjem času vedno bolj pogosti emotikoni (ang. emoticon).

Kljub temu, da so tovrstni elementi v komunikaciji precej novi in so jih uvedli razvijalci družbenih omrežij, so med uporabniki postali zelo priljubljeni in se tudi hitro ustalili v računalniško posredovani komunikaciji. Vendar temeljitih jezikoslovnih raziskav razširjenosti, rabe in pomena omenjenih elementov še ni veliko niti za večje svetovne jezike, še toliko bolj pa to velja za slovenščino, ki je ponudniki aplikacij pri razvoju funkcionalnosti niso imeli v mislih. Zato nas je v okviru raziskave, ki jo predstavljamo v tem prispevku, zanimalo, kako omembe, heštege in čustvene elemente uporabljajo slovenski uporabniki družbenega omrežja Twitter, kadar tvtajajo v

slovenščini, za primerjavo pa smo analizirali tudi njihove objave v angleščini.

Pred analizo smo si postavili 3 hipoteze:

1. V angleškem delu korpusa bodo omembe večkrat uporabljene v tujem jeziku kot v slovenskem.
2. Čustvene izraze (emojije in emotikone) bodo bolj pogosto uporabljale ženske kot moški.
3. Heštegi bodo v obeh delih korpusa večkrat enobesedni kot večbesedni.

## 2 Sorodne raziskave

Uporabniška imena na internetu so lahko drugačna od pravega imena osebe. Na ta način si oseba zagotovi neke vrste anonimnost, gradi svojo identiteto in izkazuje kreativnost. Uporabniška imena na Twitterju je proučeval J. Olivier (Olivier, 2014), ki ugotavlja, da jih je potrebno obravnavati kot posebno podkategorijo psevdonomov, ki se razlikujejo od vzdevkov. Ob tem poudarja, da uporabniška imena na Twitterju zapisujemo kot druga lastna imena, opaža pa tudi nekatere posebne značilnosti: vrstni red pri imenu in priimku je pogosto obrnjen, v imenih se lahko pojavljajo številke, velike začetnice pogosto niso upošteevane, v imenih se uporabljajo tudi drugi simboli. Kadar uporabnike omenjamo (ang. mention) v tvitih, pred uporabniško ime zapišemo znak @, s čimer je imetnik računa o omembi samodejno obveščen (Twitter, 2016). V naši raziskavi smo opazovali, kako omembe uporabniških imen slovenski uporabniki Twitterja uporabljajo v slovenskih in angleških tvitih.

Družbeni dejavniki, kot so spol, starost, stopnja izobrazbe, socialno-ekonomski razred in mnogi drugi vplivajo na jezik. Študije o jezikovni rabi na družbenih omrežjih so pokazale, da obstajajo razlike v rabi heštegov med ženskami in moškimi (Cunha, 2014). Poleg tega so v sorodnih raziskavah (Gotti) ugotovili, da je raba heštegov zelo pogosta, saj se v povprečju v več kot 50 % primerov tvitov pojavlja vsaj en heštegi. Gotti je v raziskavi proučeval tudi postavitev heštegov v tvitih. Rezultati so

<sup>1</sup> URL: <https://twitter.com/>

pokazali, da se je kar 87 % heštegov pojavilo v vlogi epiloga oziroma na kocu tvita.

Glede emotikonov in emojijev so sorodne raziskave opredelile emotikone (1) kot indikatorje čustev (npr. vesel ali žalosten), (2) kot nečustvene pomene (npr. kot šala), in (3) in kot znak (npr. :) za ublažitev izreka. S povečano priljubljenost emojijev so začeli raziskovalci raziskovati vlogo emojijev v tekstovni komunikaciji. Ugotovili so, da z njimi ne izražamo zgolj čustev, ampak se uporabljajo tudi za druge namene, kot so vzdrževanje pogovora, igriva interakcija in ustvarjanje nekih značilnosti v komunikaciji, ki so skupne vsem udeležencem (Umashanthi Pavalanathan in Jacob Eisenstein, 2015).

### 3 Zasnova raziskave in metodologija

#### 3.1 Predstavitev korpusa

Analiza, ki jo predstavljamo v tem prispevku, je bila izvedena s pomočjo slovenskega in angleškega dela korpusa Janes Tviti 0.3.4 (Erjavec et al., 2015).

Slovenski del korpusa vsebuje več kot 56 milijonov besed oz. 4 milijone tvitov, ki jih je napisalo okoli 7.600 avtorjev. Med njimi prevladujejo moški (53 %), ki so v korpus prispevali tudi največji delež tvitov in enak delež pojavnic (56 %). Žensk je približno pol manj, in sicer slabih 25 %. Objavile so 27 % tvitov in enak delež pojavnic v korpusu. V podobnem deležu se pojavljajo uporabniki, ki jim ni bilo mogoče pripisati spola (22 %).

Angleški del korpusa vsebuje 8 milijonov besed oz. 1 milijon tvitov, kar je štirikrat manj kot slovenski, vendar je to pričakovano, saj so v korpus Janes zajeti samo uporabniki, ki aktivno tvitajo v slovenščini.

#### 3.2 Potek raziskave

V raziskavi smo analizirali rabo omemb uporabniških imen, heštegov in emotikonov ter emojijev v slovenskem in angleškem delu korpusa Janes Tviti 0.3.4. Za vsakega od obravnavanih komunikacijskih elementov smo izdelali frekvenčne sezname najpogostejših in podrobno analizirali najpogostejših 1000 elementov z vsakega seznama. Za vse tri tipe obravnavanih elementov smo oblikovali čim bolj podobne kriterije jezikoslovne analize, kot so slovensko / tuje, standardno / nestandardno, enobesedno / večbesedno /okrajšano in tematika. Kriterije za posamezen tip analiziranih elementov podrobneje predstavljamo v razdelku 4.

V drugem delu analize smo rabo novih komunikacijskih elementov proučevali za različne družbene skupine uporabnikov, pri čemer smo podkorpuse izdelali s pomočjo metapodatkov, s katerimi je opremljen korpus. Upoštevali smo naslednje parametre: moški / ženske in zasebni / javni račun. Tako v slovenščini kot v angleščini smo za vse tri tipe komunikacijskih elementov v vseh štirih izdelanih podkorpusedih analizah opravili za 100 najpogostejših elementov. Opazovali smo, v čem so si rezultati podobni in v čem se razlikujejo, pri čemer smo vselej upoštevali rabo v sobesedilu.

Zadnji del analize je bil namenjen podrobni analizi skladske, pomenoslovne in pragmatične vloge obravnavanih komunikacijskih elementov v slovenskih in angleških tvitih, ki smo jo izvedli na naključnem vzorcu 100 konkordanc za posamezni podkorpus, izdelan v drugem delu analize.

## 4 Analiza podatkov

### 4.1 Omembe uporabniških imen

Slovenski korpus vsebuje 3,3 milijone oz. 46.974 na milijon omemb uporabniških imen. Relativno gledano jih angleški korpus vsebuje le nekaj več, 595.936 oz. 51.356 na milijon. To pomeni, da se pri tvitanju v slovenščini uporabniki vključujejo v bistveno manjšo družbeno mrežo kot v angleščini, najverjetneje zato, ker je pri tvitanju v slovenščini ta družbena mreža pretežno lokalna, pri tvitanju v angleščini pa mednarodna.

#### 4.1.1 Tipologija omemb

Pri analizi najpogostejših 1000 omemb v celotnem slovenskem in angleškem korpusu tvitov smo upoštevali tri kriterije:

- ali so omembe v slovenskem ali tujem jeziku,
- ali so omembe enobesedne, večbesedne ali okrajšane,
- ali so omembe napisane v standardnem ali nestandardnem jeziku.

Kot okrajšane smo obravnavali uporabniška imena, ki so vsebovala kratice (*nk\_maribor*, *DARS\_SI*, *HDDOlimpija*), uporabniška imena, kjer je bilo okrajšano ime ali priimek (*MKamarin*, *zzTurk*, *ABratusek*) ter uporabniška imena, kjer se je pojavil okrajšani doktorski naziv ali angleška okrajšava »mr.« (*drVinkoGorenak*, *mr\_foto*, *MrGabbah*).

Kot standardne smo označili omembe, napisane z velikimi ali malimi začetnicami ter s podčrtajem za presledek, kot npr. *MarkoSket*, *savicdomen*, *Petra\_Jansa*. Vse ostale omembe smo označili kot nestandardne.

Rezultati tega dela analize so pokazali, da se tako v slovenskem kot v angleškem delu korpusa pojavlja več omemb slovenskih kot tujih uporabniških imen: v slovenskem korpusu delež znaša 84 %, v angleškem pa 65,8 %. To ni nenavadno, saj so v angleški korpus zajeti tviti slovenskih uporabnikov, kadar ti tvitajo v angleščini in so zaradi tega pričakovano še vedno v precejšnji meri vezani na slovenski prostor.

Pri primerjavi uporabniških imen, ki se pojavljajo v slovenskem in angleškem delu korpusa, smo ugotovili, da se v angleškem delu korpusa pojavljajo uporabniška imena, ki se v slovenskem delu ne pojavljajo, in obratno. V slovenskem korpusu se ne pojavljajo izrazi, ki so vezani na angleško govoreči prostor. To so večinoma imena tujih oseb (*@joerogan*, *@lindseyvonn*) ali časopisov (*@Independent*, *@nytimes*). V angleškem delu korpusa pa se med najpogostejšimi 1000 omembami ne pojavljajo imena slovenskih političnih strank (*@stranka SLS*, *@strankaSD*). Delež omemb, ki se pojavljajo samo v angleškem delu korpusa, znaša 29,8 %, kar pomeni, da se oba frekvenčna seznama prekrivata kar v 70,2 %.

Pri obeh delih korpusa so največkrat uporabljena večbesedna uporabniška imena (65 %), sledijo jim enobesedna (18 %), najmanjkrat pa se pojavijo okrajšana (17 %). Večbesedna uporabniška imena so največkrat sestavljena iz imena in priimka (*@petrasovdat*, *@MatevzNovak*), imen podjetij ali spletnih portalov (*@uporabnastran*, *@RevijaReporter*), identificirali pa smo tudi cele besedne zveze ali stavke (*@MyBlueDragoness*, *@JsSmRenton*). Pri okrajšavah gre največkrat za krajšanje imena ali priimka (*@JJansaSDS*, *@nmusar*) ali ime, v katerem se pojavi kratica

(@RTV\_Slovenija, @strankaSD). Nekajkrat se pojavljajo tudi idiosinkratična kratična ali alfanumerična poimenovanja, katerih pomena brez poznavanja ozadja ni mogoče določiti (@Z3MQP, @z8\_LJ).

Pri proučevanju standardne in nestandardne rabe se je izkazalo, da je v obeh delih korpusa nekoliko pogostejša nestandardna raba (53 %). Kot nestandardno smo tu označili vsa poimenovanja, ki vsebujejo številke (@lenci53, @vinkovasle1), so zapisana v pogovornem jeziku (@abejz\_no, @merineseri) ali pa so nerazumljiva (@DC43, @schoo666).

#### 4.1.2 Analiza omemb pri različnih skupinah uporabnikov

V tem delu analize smo se osredotočili na primerjavo rabe omemb pri različnih skupinah uporabnikov:

- moški in ženski računi in
- zasebni in korporativni računi.

Rezultati so pokazali, da so v slovenskem delu korpusa v vseh kategorijah pogosteje uporabljena uporabniška imena v slovenskem jeziku, medtem ko v angleškem delu korpusa nekatere skupine uporabnikov pogosteje uporabljajo imena v tujem jeziku (npr. moški in ženske s korporativnimi računi). To nakazuje, da slovenski uporabniki pri tvitjanju s korporativnimi računi uporabljajo več omemb zaradi poslovnega sodelovanja s tujimi podjetji in iz tega razloga tudi več tvitajo v angleščini kot uporabniki z zasebnimi računi.

V obeh delih korpusa se v vseh kategorijah večkrat pojavljajo večbesedne kot enobesedne omembe uporabniških imen. Analiza je pokazala, da v slovenskem delu korpusa večbesedne omembe uporabniških imen največkrat uporabljajo ženske z zasebnim računom (63 %), najmanjkrat pa moški s korporativnim računom (52 %). V angleškem delu korpusa pa večbesedne omembe imen največkrat uporabljajo ženske s korporativnim računom (73 %) in najmanjkrat moški s korporativnim računom (59 %).

V povprečju je v obeh delih korpusa uporabljenih več standardnih (52 %) kot nestandardnih (48 %) oblik uporabniških imen.

#### 4.1.3 Analiza omemb v sobesedilu

V tem delu smo na vzorcu 100 naključnih pojavitev omemb v korpusu analizirali rabo omemb v sobesedilu. Za analizo smo izbrali naslednje kriterije:

- ali se omemba uporablja na začetku, na sredini ali na koncu tvita,
- ali se omemba uporablja skladijsko ali neskladijsko,
- ali tvit vsebuje eno samo omembo ali več in
- kaj uporabniško ime označuje (osebo, kraj, spletni portal, podjetje, skupino, politično stranko ali oddajo).

Glede na rezultate analize prvega kriterija smo ugotovili, da se omembe v obeh delih korpusa v povprečju uporabljajo na začetku tvita (51 %). Večinoma gre za omembe uporabniških imen s to funkcijo, da bi jih opozorili na določeno temo, ki zadeva te uporabnike ali pa so pri njej že prej sodelovali. V slovenskem delu korpusa jih na začetku največkrat uporabljajo moški z zasebnimi računi (82 %), najmanjkrat pa ženske s korporativnimi računi (40 %). V angleškem delu korpusa pa je bila

situacija malo drugačna: omembe na začetku tvita največkrat uporabljajo ženske z zasebnimi računi (54 %) in najmanjkrat moški z zasebnimi računi (23 %).

Skladijska analiza je pokazala, da so v obeh delih korpusa omembe pogosteje rabljene skladijsko. Pri angleškem delu so zaradi skladnje angleškega jezika skladijsko rabljene vse omembe, medtem ko se v slovenskem delu pojavlja tudi neskladijska raba, katere delež znaša 5 %. V slovenskem delu korpusa se neskladijska raba pojavlja predvsem takrat, ko pred omembo uporabniškega imena stojijo predlogi in bi bilo omembo potrebno pregibati, kar v tehničnem smislu ni mogoče, saj aplikacija v tem primeru omembe ne prepozna (*tale me je spomnila na @BesedaDneva, še malo pa se bo začel letošnji festival Morja in sonca v @avditorij Portorož*). Pojav bo zanimivo opazovati v daljšem obdobju, ko bomo lahko ugotovili, ali so se uporabniki tehnični omejitvi prilagodili z jezikovno rabo, ki se težavi uspešno izogne ali pa se je zaradi tehničnih okoliščin začela spreminjati slovenska skladnja in pred tradicionalno neprvosklonskimi predlogi dopušča tudi imenovalnik.

Tako v slovenskem kot tudi v angleškem delu se v povprečju večkrat pojavlja ena omemba uporabniškega imena na tvit (57 %). Pri več omembah na tvit se je izkazalo, da uporabniki v obeh delih korpusa največkrat uporabljajo 2 omembi na tvit (24 %), največje število omemb v tvitu pa je 8.

V obeh delih korpusa uporabniška imena največkrat označujejo osebe, sledijo jim spletni portali (@YouTube, @BesedaDneva), podjetja in politične stranke. Pri tem je zanimivo, da stranke omenjajo le moški uporabniki v slovenskem delu korpusa, ki imajo zasebni račun.

## 4.2 Hešteg

V tem razdelku predstavljamo rezultate analize heštegov, ki so zbrani v slovenskem in angleškem korpusu Janes Tviti v.0.3.4. Heštegovi so na družbenem omrežju Twitter ena izmed najbolj priljubljenih elementov izražanja. V slovenskem delu korpusa se jih pojavi dober milijon oz. 15.319 na milijon, v angleškem pa 450 tisoč oziroma 38.735 na milijon, kar je relativno gledano več kot 2,5 krat pogosteje kot v slovenskih tvitih. Ker je slovenski del korpusa tvitov šestkrat večji od angleškega, je razumljivo, da vsebuje dvakrat več različnih heštegov (209.629) kot angleški (114.943).

### 4.2.1 Tipologija heštegov

Pri heštegih smo pri analizi frekvenčnega seznama 1000 najbolj pogostih pojavnic opazovali naslednje:

- tema (šport, politika)
- enobesedni, večbesedni ali okrajšava,
- slovenski ali tuji jezik in
- standarden ali nestandarden

Kot nestandardne smo označili vse heštege, pri katerih avtorji niso zapisali lastnih in zemljepisnih imen z veliko začetnico, kjer so uporabljali tujke kot npr. #Sochi2014. Če sta se pojavili dve pojavnici, kot sta #PLTS in #plts, smo kratico, napisano z majhnimi črkami, označili kot nestandardno. V to kategorijo spadajo tudi vsi pogovorni izrazi, kot so #gotofje, #fuzbal, #butale, vse večbesedne zveze, kot sta #TvitajmoZaNase in #NaDanašnjiDan, pri katerih so vse naslednje komponente napisane z veliko

začetnico in vse besedne zveze, pri katerih avtor ni uporabil šumnikov.

V slovenskem delu korpusa so bili prvi trije najpogostejši heštegi *#junaki*, *#plts*, *#slochi*, v angleškem delu pa *#Slovenia*, *#Ljubljana*, *#slovenia*.

V slovenskem delu se je največ heštegov nanašalo na temo športa (20 %), sledi politika (11 %). Na veliko razliko pri zastopanosti tematik v heštegih smo naleteli v angleškem delu korpusa, v katerem smo zasledili bistveno višji delež heštegov s področja podjetništva in izjemno nizko število političnih heštegov (zgolj 3 %). Na podlagi tega sklepamo, da lokalno politično dogajanje ni tema, o kateri bi slovenski uporabniki Twitterja razpravljali v angleščini.

Tako v slovenskem (53 %) kot tudi angleškem (60 %) korpusu prevladujejo enobesedni heštegi, s 32 % jim sledijo večbesedne zveze v slovenskem delu (*#ligaprvaokov*, *#nočnastraža*) in v angleškem 29 % (*#NBABallot*) ter kratice (*#EU*, *#lol*), ki v slovenskem delu predstavljajo 15 %, v angleškem 11 %. Ti rezultati kažejo na potrebo po še posebej ekonomičnem izražanju v heštegih.

Rezultati so pokazali, da je velika večina slovenskih heštegov (62 %) rabljenih v tvitih, napisanih v standardni slovenščini. Zdi pa se presenetljivo, da je raba šumnikov v slovenskem korpusu zelo nizka, saj zgolj 5 % heštegov vsebuje šumnike. V angleškem korpusu je odstotek heštegov, napisanih v standardni angleščini, še višji (72 %), kar nakazuje, da slovenski uporabniki angleščino uporabljajo za formalnejše oblike komuniciranja kot slovenščino.

#### 4.2.2 Analiza heštegov pri različnih skupinah uporabnikov

V drugem delu analize smo rabo heštegov primerjali v izbranih podkorpusih, ki smo jih ustvarili glede na metapodatke, s katerimi je opremljen korpus Janes Tviti v.0.3.4, in sicer ženske (zasebno, javno) in moški (zasebno, javno).

V podkorpusih slovenskih tvitov moških in žensk (moški-zasebno, moški-javno, ženske-zasebno, ženske-javno) so rezultati pokazali, da so ženske s korporativnim računom večkrat uporabile večbesedne heštege (48 %), moški pa enobesedne (45,5 %). V angleškem delu so najpogostejše enobesedne zveze, ki so jih največkrat uporabile ženske s korporativnim računom (60 %).

Največji razkorak opazimo pri primerjavi med moškimi in ženskami glede na rabo standardnih in nestandardnih rabo heštegov v slovenskem delu, ki kaže, da moški nekoliko večkrat uporabljajo standardne heštege (56 %), predvsem tisti s korporativnimi uporabniškimi računi (61 %). Poleg tega ženske z zasebnimi računi najmanjkrat pišejo heštege v standardni slovenščini (54 %). Medtem pa oboji pogosteje uporabljajo heštege v standardni angleščini, a največkrat ženske s korporativnimi računi z 75 %.

Ugotovili smo tudi, da moški z zasebnim računom uporabljajo več (11 %) političnih heštegov kot ženske (tako imetnice zasebnih kot javnih računov), poleg tega je v obeh moških podkorpusih 26 % heštegov, največ pri uporabnikih z zasebnim računom z 30 %, ki se nanašajo na šport. V angleškem delu je zanimivo, da ženske uporabnice v angleškem podkorpusu s korporativnim uporabniškimi računi niso uporabile niti enega športnega heštega, najdemo pa veliko število heštegov s področja

podjetništva (*#ecommerce*, *#Google*) in marketinga (*#SocialMediaMarketing*, *#contentmarketing*). V nasprotju z moškimi uporabniki, ki ne glede na zaseben (27 %) ali uraden (16 %) račun, uporabljajo športne heštege. Predvidevamo lahko, da uporabnice ne opravljajo služb, ki bi zahtevale uporabo športnih tegov, in na uradnih računih uporabljajo le besedišče s svojega delovnega področja, medtem ko se moški z uradnim računom opredeljujejo tudi s svojimi pristoječnimi mislimi in mnenji oz. to od njih zahteva tudi narava njihovega dela (*#srcebije*, *#mismomaribor*).

#### 4.2.3 Analiza heštegov v sobesedilu

V zadnjem delu analize naključnega vzorca 100 konkordanc nas je zanimalo, kje v tvitu uporabniki heštege najpogosteje uporabljajo, ali so heštegi rabljeni stavčno oz. nestavčno ter kakšno pragmatično funkcijo v tvitu opravljajo.

Postavitev heštegov lahko razdelimo v tri kategorije:

1. začetek oz. uvod: hešteg napoveduje temo tvita;
2. sredina: v tem primeru so pogosto rabljeni stavčno; najpogosteje kot predmet, osebek, povedek in prislovno določilo kraja;
3. konec oz. epilog: hešteg je na koncu tvita, da še dodatno zaznamuje temo tvita.

Podobno kot v sorodnih raziskavah so tudi naši rezultati pokazali, da se večina heštegov pojavi na koncu tvita: 72 % heštegov v slovenskem in 61 % v angleškem korpusu. V slovenskem delu jih je kar 83 % je uporabljenih nestavčno, 80 % pa z namenom, da dodatno zaznamuje temo tvita. Podobno so pokazali rezultati iz angleškega dela, kjer prevladuje nestavčna raba heštegov (81 %).

Največ stavčno uporabljenih heštegov najdemo v slovenskem podkorpusu moški-zasebno, in sicer 24 %; med njimi jih je 70 %, ki se jih pojavi na sredini. Med angleškimi korpusi največ stavčno rabljenih heštegov najdemo v angleškem podkorpusu ženske-javno (34 %), med njimi jih je 83 %, ki se jih pojavi na sredini. Tako kot v angleškem kot v slovenskem delu so heštegi na sredini največkrat uporabljeni kot predmet, na drugem mestu kot osebek in na zadnje prislovno določilo kraja.

Poleg tega smo ugotovili, da uporabniki v angleščini uporabljajo več heštegov v enem tvitu (78 %) kot v slovenščini (60 %). Iz tega lahko sklepamo, da se uporabniki v slovenščini tekoče izražajo v polnih povedih, medtem ko pri tvitanju v angleščini pogosteje nizajo heštege ob krajših izjavah.

#### 4.3 Emotikoni in emojiji

V analizi smo obravnavali tudi emotikone in emojije. Z obema želi uporabnik izraziti določeno čustvo, ki bi ga v komunikaciji v živo izrazil z intonacijo, gestami ali obrazno mimiko in bi ga v pisni računalniško posredovani komunikaciji težje izrazil z besedami. V celotnem korpusu smo opazili, da v primerjavi z moškimi ženske uporabljajo več emojijev (ženske rel.frek. 180,8, moški 74,8). Tudi v angleščini so emojiji pogostejši pri uporabnicah (rel.frek. 210) kot pri uporabnikih (rel.frek. 56,7).

### 4.3.1 Tipologija emotikonov in emojijev

Pri emotikonih in emojijih smo za analizo frekvenčnega seznama 1000 najpogostejših pojavnic izbrali naslednje tri kriterije:

- pomen čustvenega izraza,
- enostaven ali sestavljen čustveni izraz in
- pozitiven ali negativen čustveni izraz.

Kot sestavljene smo upoštevali emotikone, ki vsebujejo več kot 2 znaka, zaradi majhne velikosti angleškega korpusa v njem ni 1000 različnih emotikonov in emojijev, zato smo v analizo vključili vse.

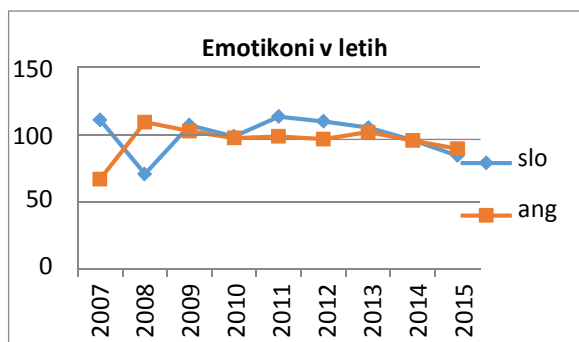
**Emotikoni:** Rezultati te analize so pokazali, da so emotikoni v slovenskem in angleškem jeziku po pomenu primerljivi, saj prednjačijo emotikoni s pozitivnim sporočilom, najpogostejši (5,4 %) so izrazi smeha (npr. :-) ali :) ). Na drugem mestu so emotikoni, ki izražajo žalost (: ( in :( ) (4,1 %). Pri kriteriju sestavljen/enostaven opazamo, da v obeh jezikih prevladujejo sestavljeni emotikoni (88 % v angleščini, 80 % v slovenščini).

**Emojiji:** Na prvo mesto (v slovenščini 13,6 % in v angleščini 12,1 %) se v obeh jezikih uvrščajo emojiji, ki upodabljajo naravo (🌸, 🍷), na drugem mestu (v angleščini 5,1 % ter slovenščini 2,9 %) jim sledijo emojiji, ki zaznamujejo hrano (🍷, 🍷). Izmed emojijev, ki označujejo čustva, so v obeh jezikih najpogostejši (v angleščini 2,6 % in v slovenščini 2,1 %) tisti, ki izražajo ljubezen (❤️, ❤️).

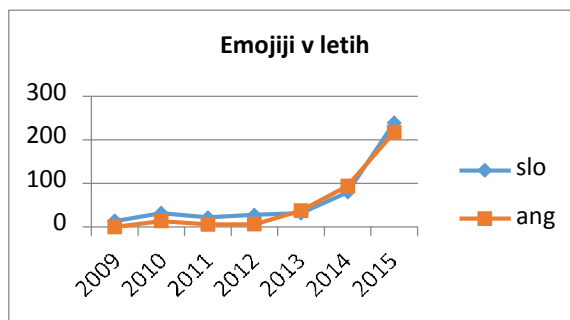
Vse emojije smo obravnavali kot enostavne, saj predstavljajo en sam Unicode znak.

Pri analizi celotnega korpusa smo se osredotočili na tisti kriterij, ki se nam je zdel najbolj zanimiv, in sicer na kronološko analizo čustvenih izrazov.

Kot kažeta Grafa 1 in 2, raba emojijev stalno narašča, še posebej strmo v zadnjih treh letih, medtem ko raba emotikonov v obeh jezikih več ali manj stagnira oz. celo počasi upada.



Graf 3: Emotikoni v letih



Graf 4: Emojiji v letih

### 4.3.2 Analiza emotikonov in emojijev pri različnih skupinah uporabnikov

V tem delu analize smo v izbranih podkorpusih, ki smo jih ustvarili glede na metapodatke ženske (zasebno, javno) moški (zasebno, javno), primerjali enake kriterije kot v razdelku 2.3.1 (pomen, pozitiven/negativen, sestavljen/enostaven).

V analizo smo vključili 50 najpogostejših pojavnic.

**Moški z javnimi računi,** ki tvitajo v slovenščini, uporabljajo več **emotikonov**, ki ponazarjajo ljubezen (6 %), a tudi več, ki jih ponazarja žalost v primerjavi z angleščino (16 % žalost v slovenščini, medtem ko je teh v angleščini le 6 %). Pri analizi **emojijev** v tem podkorpusu pa smo ugotovili, da je v slovenščini uporabljenih kar 14 % emojijev, ki zaznamujejo naravo, v angleščini pa samo 6 %. Pri **moških z zasebnimi računi** v slovenskem jeziku smo ugotovili, da je pogostost emojija, ki ponazarja žalost 6-odstotna, v angleščini pa znaša 2 %. Pri zasebnih računih smo opazili tudi razliko med moškimi in ženskami. Pri emotikonih je na primer razvidno, da uporabljajo moški, ki tvitajo v slovenskem jeziku več emotikonov za žalost v primerjavi z ženskami (16 % proti 12 %), a tudi več emotikonov, ki ponazarjajo smeh (46 % proti 36 %). V angleščini pa obratno, saj so emotikoni za smeh pogostejši pri ženskah (50 % proti 44 %). Za emojije vidimo, da izmed tistih, ki uprizarjajo ljubezen so v slovenščini bolj popularni pri ženskah kot pa ne pri moških (12 % proti 8 %) in prav tako emojiji za smeh.

**Ženske z zasebnimi računi** uporabljajo v angleškem jeziku več emotikonov za žalost (16 % proti 12 %) in več emotikonov, ki ponazarjajo ljubezen (10 % angleščina, 2 % slovenščina).

### 4.3.3 Analiza emotikonov in emojijev v sobesedilu

V tem delu analize smo za izbrane 4 podkorpuse v obeh jezikih na naključnem vzorcu 100 konkordanc opazovali, ali so emojiji in emotikoni v tvitih rabljeni namesto ločila ali namesto besede ter ali stavek omilijo ali ga podkrepijo. V zadnji kategoriji smo ločevali emojije in emotikone, ki podkrepijo stavek tako, da smo opazovali kontekst, npr.:

- 1) *ženitnega posrednika se pa ne grem :D*
- 2) *Lučkeee, juhej :)))*

- 1) *Tiho bodi, ves da nisem tebe mislila!! ☹*
- 2) *Pr šepanju mi gre najbolj na živce to, da sm tak počasna – bolj k bolečina, res. ☹*

Pri št. 1) smo upoštevali, da emotikon/emoji omili stavek, pri št. 2) pa, da ga podkrepi.

Pogostost uporabljanja **emotikonov** namesto ločil je večja pri moških z zasebnimi računi, ki tvitajo v slovenščini, kot pa v angleščini (52 % proti 48 %). Več emotikonov omili stavek v angleščini, kot v slovenščini (40 % proti 32 %). Če pogledamo ženske uporabnice, lahko opazimo, da ženske z zasebnimi računi v obeh jezikih uporabljajo skoraj enako število emotikonov namesto ločil (59 % v slovenščini, 58 % v angleščini). Moški z zasebnimi računi uporabljajo veliko več krepilnih emotikonov v slovenščini kot v angleščini (68 % proti 45 %), medtem ko jih ženske z zasebnimi računi uporabljajo malo več v angleščini (63 % proti 61 %). Večina uporabnikov ne nadomešča besed z emotikoni.

Pri uporabi emojijev pa smo ugotovili, da jih moški z javnimi računi v angleškem jeziku bistveno pogosteje uporabljajo namesto ločil kot v slovenščini (62 % v angleščini in 46 % v slovenščini). Razlika pri moških in ženskah z javnimi računi je tudi pri emojijih, ki omilijo stavek, saj jih moški v slovenščini uporabljajo v večjem številu kot v angleščini (12 % proti 9 %), ženske pa obratno (10 % proti 12 %). Pri ženskah z zasebnimi računi, ki tvitajo v angleščini, je pogostost emojijev, ki podkrepijo stavek, veliko večja v primerjavi s slovenščino (90 % angleščina, 73 % slovenščina), pri ženskah z javnimi računi pa smo opazili ravno obratno (79 % slovenščina, 77 % angleščina).

## 5 Zaključek

V prispevku smo opravili podrobno analizo novih komunikacijskih elementov na družbenem omrežju Twitter v slovenskem in angleškem korpusu Janes Tviti v.0.3.4, ki vsebuje tvite slovenskih uporabnikov. Opazovali smo rabo omemb, heštegov ter emotikonov in emojijev ter jo primerjali med jezikoma ter med ženskimi in moškimi uporabniki, ki tvitajo v zasebne ali službene namene. Omembe se tako v slovenskem kot angleškem delu korpusa pogosteje pojavljajo na začetku tvita. V obeh delih korpusa prevladuje nestandardna oblika, omembe so po večini večbesedne. V angleškem delu korpusa se pojavljajo omembe, ki se v slovenskem delu ne in ravno obratno.

Hešteg se tako v slovenskem kot v angleškem delu korpusa v veliki večini pojavljajo na koncu tvita z namenom, da dodatno zaznamujejo temo tvita, in so večinoma standardni. Pri nestandardnih v slovenščini

izstopa izjemno nizka uporaba šumnikov. Najpogostejša tema heštegov je v slovenskem in angleškem delu korpusa šport. V slovenskem delu sledi politika. Na veliko razliko pri zastopanosti tematik v heštegih smo naleteli v angleškem delu korpusa, v katerem smo zasledili bistveno višji delež heštegov s področja podjetništva in izjemno nizko število političnih heštegov). Moški uporabljajo pretežno enobesedne, ženske pa večbesedne heštege.

V korpusu Janes so najpogosteje uporabljeni emotikoni in emojiji, ki upodabljajo naravo, hrano in čustva, predvsem pozitivna, njihova prevladujoča vloga pa je, da okrepijo izjavo. Medtem ko raba emotikonov počasi upada, raba emojijev strmo narašča. Po obeh tipih čustvenih izrazov v obeh jezikih pogosteje posegajo ženske, ki jih velikokrat uporabljajo namesto končnih ločil.

Z našo raziskavo smo potrdili in ovrgli nekatere naše hipoteze:

1. Omembe uporabniških imen so bile v angleškem delu korpusa večkrat uporabljene v slovenskem jeziku. S tem je bila naša hipoteza ovržena.
2. Ženske uporabljajo več čustvenih izrazov kot moški. S tem je bila naša hipoteza potrjena.
3. Heštegji so bili v obeh delih korpusa večbesedni. Tudi ta hipoteza je bila potrjena.

V prihodnje nameravamo raziskavo razširiti s primerjavo rabe novih komunikacijskih elementov v tvitih uporabnikov, ki prihajajo iz angleško govorečih držav. S tem bomo preverili, v kolikšni meri na njihovo rabo v angleščini vpliva materni jezik, kultura in konvencije računalniško posredovane komunikacije različnih govornih področij.

## 6 Literatura

- David Crystal. 2001. Language and the Internet. Cambridge University Press. The Edinburgh Building.
- Evandro Cunha. 2014. He Votes or She Votes? Female and Male Discursive Strategies in Twitter Political Hashtags. <http://1.usa.gov/1P0WiR7>
- Fabrizio Gotti. Hashtag Occurrences, Layout and Translation: A Corpus-driven Analysis of Tweets Published by the Canadian Government. <http://bit.ly/1S0BvOl>
- Tomaž Erjavec. Darja Fišer. Nikola Ljubešić. Razvoj korpusa slovenskih spletnih uporabniških vsebin Janes. <http://bit.ly/1OBwO9t>
- Jako Olivier. 2014. Twitter usernames: exploring the nature of online South African nicknames. <http://bit.ly/1Qj5scM>
- Umashanthi Pavalanathan and Jacob Eisenstein. 2015. Emoticons vs. Emojis on Twitter: A Causal Inference Approach. School of Interactive Computing, Georgia Institute of Technology. Atlanta, GA 3030. <http://arxiv.org/pdf/1510.08480v1.pdf>
- Twitter glossary. <https://biz.twitter.com/en-gb/glossary>