

## Analiza krajšanja slovenskih sporočil na družbenem omrežju Twitter

Teja Goli,\* Eneja Osrajnik,† Darja Fišer‡<sup>+</sup>

\* Kropa 48a, 4245 Kropa  
teja.goli@gmail.com

† Šerugova 10, 2000 Maribor  
eneja.osrajnik@gmail.com

‡ Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani  
Aškerčeva 2, 1000 Ljubljana  
darja.fiser@ff.uni-lj.si

<sup>+</sup> Odsek za tehnologije znanja, Institut Jožef Stefan  
Jamova cesta 39, 1000 Ljubljana

### Povzetek

V pričujočem prispevku obravnavamo pogostost in načine krajšanja slovenskih sporočil na družbenem omrežju Twitter. Za analizo smo uporabili vzorčni podkorpus 800 tvitov z različno stopnjo tehnične in jezikovne standardnosti ter v njem označili pojave krajšanja na ravneh zapisa, leksike in skladnje. V njih smo zabeležili skupno 3464 pojavov krajšanja, ki so bili uvrščeni v 32 različnih kategorij. V nestandardnih tvitih je krajšanja bistveno več kot v standardnih, količinsko in tipološko največ se jih pojavlja na ortografski, najmanj pa na skladijski ravni.

### Shortening Phenomena in Slovene Tweets

This paper addresses the frequency and types of shortening phenomena in Slovene tweets. The analysis was carried out on a subcorpus of 800 tweets of various levels of technical and linguistic standardness in which instances of shortening strategies were observed on the levels of spelling, lexis and syntax. We have registered 3,464 instances of shortening strategies that belong to 32 different categories. The results show that shortening strategies are much more common in non-standard tweets and that the highest number and widest range of shortening strategies arise on the orthographic level, whereas they are the lowest on the syntactic level.

## 1 Uvod

Družbeno omrežje Twitter postaja vse popularnejše tudi v Sloveniji. Ena njegovih glavnih značilnosti, po kateri se razlikuje od ostalih družbenih omrežij, je omejena dolžina posameznega sporočila na 140 znakov. Ta omejitev je prevzeta iz protokola SMS-sporočil, ki so omejeni na 160 znakov, pri čemer so pri Twitterju 20 znakov obdržali za zapis uporabniškega imena avtorja tvita (Rogers, 2014).

Med jezikoslovci, kot tudi novinarji in laično javnostjo, že od samih začetkov komunikacije preko SMS-sporočil velja, da je zanje značilno pogosto krajšanje sporočil, kot glavna razloga za to pa navajajo omejitev dolžine sporočila in neergonomske tipkovnice na mobilnih telefonih (Crystal, 2001), podobno naj bi veljalo tudi za visoko interaktivne oblike računalniško posredovane komunikacije, med katere sodijo tviti. Vendar tovrstne raziskave za večino jezikov še niso bile opravljene, tiste, ki so bile izvedene, pa kažejo, da o krajšanju sporočil kot univerzalnem pojavu tovrstne komunikacije ne moremo govoriti, saj se tako količina kot načini krajšanja sporočil od jezika do jezika močno razlikujejo (Bieswanger, 2007).

Po drugi strani je zanimiva ugotovitev, da je krajšanje pogosto tudi v sporočilih, ki ne dosegajo največje dovoljene dolžine (Thurlow, 2003), kar nakazuje na to, da načelo ekonomičnosti zaradi omejene dolžine ni avtorjev edini motiv za krajšanje sporočil, temveč ima tovrstna jezikovna raba tudi družbeno in čustveno razsežnost, s katero izražamo osebno identiteto in izkazujemo pripadnost neki družbeni skupini (Sveningsson, 2001), pa tudi ustvarjamo spontano, lahko vzdušje v komunikaciji (Androutsopoulos in Schmidt, 2001).

Cilj pričujočega prispevka je ugotoviti, ali, v kolikšni meri ter na kakšen način svoja sporočila krajšajo slovenski uporabniki Twitterja, ne glede na to, ali je krajšanje namerno/zavestno, posledica oz. kombinacija drugih dejavnikov ali povsem nenamerno/nezavestno. V naslednjem razdelku predstavimo sorodne raziskave, v 3. razdelku opišemo zasnovi pričujoče raziskave, v 4. razdelku podamo analizo rezultatov, nato pa prispevek sklenemo z zaključki in načrti za prihodnje raziskave.

### 1.1 Sorodne raziskave

Fenomeni krajšanja v t. i. novih medijih so v zadnjih desetletjih postali priljubljen predmet raziskav v številnih jezikih. Pri tem prednjačijo predvsem analize SMS-sporočil, e-pošte in spletnih klepetalnic (npr. Crystal 2001, Thurlow 2003), zadnje čase pa dobivajo pozornost tudi družbena omrežja. Alis in Lim (2013) z analizo 229 milijonov ameriških objav na Twitterju med 2009 in 2012 ugotavljata, da tviti v ameriškem prostoru postajajo vse krajši in kompaktnjši tako na sintaktični kot na leksikalni ravni. Poleg krajšanja zaznata porast žargonizmov in priložnostnih tvorjenk, zaradi katerih kljub krajšanju ne pride do izgube informacij.

Gouws et al. (2011) se posvečajo analizi in avtomatskemu označevanju nestandardnega angleškega besedišča avstralskih, hongkonških, britanskih in ameriških uporabnikov Twitterja. Med najpogostejšimi pojavi krajšanja identificirajo nadomeščanje večznakovnega niza z enim znakom, transformacije s pripono, izpuščanje samoglasnikov, nadomeščanje s predpono, nadomeščanje »you« z »u«, izpuščanje zadnje črke, združevanje besed ter nadomeščanje »th« z »d«.

Beiswenger (2007) primerja leksikalne redukcije, kot so inicializmi, kmitve, skrajšane oblike glagolov, črkovno-številčni homofoni, fonetični zapisi in logogrami, v nemških in angleških SMS-sporočilih. Ugotovi, da je v nemških SMS-sporočilih šestkrat manj redukcij kot v angleških, pri čemer Nemci pogosteje uporabljajo inicializme, Angleži pa vse ostale oblike krajšav.

V slovenskem prostoru krajšanje v SMS-sporočilih opazuje Jarnovič (2007), ki s svojo anketno raziskavo ugotavlja, da je jezikovna podoba SMS-sporočil odvisna predvsem od dojemanja SMS-ov s strani uporabnikov. Rezultati njene raziskave so pokazali, da je uporaba razpeta med tehnično in tehnološko prostorsko omejenostjo na 160 znakov, razumljivostjo sporočila ter ekspresivnim izražanjem. Po njenih ugotovitvah izpuščanje presledkov ter uporaba različnih ustaljenih in spontanah krajšav služita predvsem ustvarjalnosti pri izražanju in nista zgolj pojav, ki je odvisen od prostorskih omejitev. Ta sklep utemelji z dejstvom, da anketiranci z osnovno in srednjo izobrazbo prikazujejo bolj raznoliko uporabo krajšav kot ljudje z višjo in visoko izobrazbo.

V elektronski komunikaciji se krajšanja pri proučevanju rabe ločil dotakne tudi Michelizza (2012), ki zaznava neskladenjsko rabo pike in dvopičja, ki se pojavljata pri zapisu končnic internetnih domen ali formatov datotek (.exe, de:wiki) in prehajajo tudi v pridevniško rabo. Omeni tudi pogosto rabo stičnega tropičja in zvezdic, ki nakazuje izpust dela besede in predstavlja t. i. grafični evfemizem.

## 2 Zasnova raziskave

V pričujoči raziskavi smo analizirali pogostost in načine krajšanja slovenskih sporočil na družbenem omrežju Twitter. Za analizo smo uporabili vzorčni podkorpus tvitov, ki je že bil ročno označen na ravni stavčne segmentacije, tokenizacije in normalizacije nestandardnih besed (Čibej et al., 2016) ter obsega 800 tvitov z različno stopnjo tehnične in jezikovne standardnosti (Ljubešić et al., 2015), kot je podano v tabeli 1, pri čemer je bil vzorec nestandardnih tvitov (L3T3) za namene zagotavljanja jasnosti in doslednosti smernic za označevanje dvojno označen. Izmed vseh pregledanih tvitov je bilo 21 tvitov označenih kot nerelevantnih za raziskavo.

Vrsta tvitov glede na stopnjo standardnosti	Število tvitov
tehnično in jezikovno standardni (T1L1)	200
tehnično standardni, jezikovno nestandardni (T1L3)	200
tehnično nestandardni, jezikovno standardni (T3L1)	200
tehnično in jezikovno nestandardni (T3L3)	200
Skupaj	800

Tabela 1: Število označenih tvitov pri posamezni stopnji tehnične in jezikovne standardnosti.

Označevanje pojavov krajšanja je potekalo v orodju WebAnno, prosto dostopni označevalski spletni platformi, ki omogoča označevanje besedila na več ravneh (Eckart de Castilho et al., 2014).

## 3 Tipologija

Pojave krajšanja smo opazovali na ravneh zapisa, leksike in skladnje. Vse nivoje smo razdelili na najmanj dve in največ štiri podkategorije (gl. tabele 3, 4 in 5).

Pojavi krajšanja na nivoju zapisa se v osnovi delijo na izpuščanje, opuščanje in nadomeščanje. Pri *Izpuščanju* smo se osredotočali na vsakršno krajšanje zapisa besede na ravni črk, ne glede na razlog. Pri tem ločujemo izpuste samoglasnikov, izpuste soglasnikov in kombinacije izpustov samoglasnikov in soglasnikov, ki se nadalje ločijo glede na to, ali so izpuščeni na začetku (*mam*), na koncu (*rajš*) ali sredi (*tko*) posamezne besede. Krajšanje sporočil z opuščanjem presledkov in ločil uvrščamo v ločeno kategorijo, ki jo zaradi boljše preglednosti poimenujemo *Opuščanje*, pri čemer ne želimo vzbujati vtisa, da je neuporaba presledkov in ločil zavestnejša od izpuščanja črk. Opuščanje presledkov pri ločilih in simbolih ter med besedami, tj. združevanje besed, je označeno ločeno. Pri opuščanju ločil so upoštevana končna ločila, pike za okrajšavami ter deli večdelnih ločil, tj. treh pik, oklepajev in narekovajev. Kot nadomeščanje je opredeljena uporaba drugih (krajših) nizov črk, zapis s številčnimi homofoni, substitucija domačih črk s tujejezičnimi in obratno (tu smo upoštevali le tiste črke, ki niso del slovenske abecede), zapis z logogrami ter nadomeščanje leksemov z emotikoni, emojiji ali piktogrami.

Na leksikalni ravni smo delno izhajali iz tipologije Markusa Bieswangerja (2007) in upoštevali vse pojave, pri katerih izbrana oblika obsega manj znakov kot polno razvezana beseda oz. besedna zveza. V osnovi smo krajšave razdelili na ustaljene in neustaljene. Med prve spadajo kratice (*EU*, *LOL*, *FDV*, *KPK*), sledijo jim ustaljene okrajšave s piko ali brez (*npr.*, *slo.*), pri čemer je manjkajoča pika označena na nivoju zapisa, posebej pa so označeni tudi ustaljeni simboli, formule in krnjene besede (*š*, *EUR*, *kg*). Med neustaljene krajšave so uvrščene žanrskospecifične krajšave, ki so nesklonljive (*rd*, *lp*, *Lj*, *btw*), sledi jim kategorija krajšanih neologizmov (*mata*, *appi*, *Zoki*), tretja neustaljena skupina krajšav pa so priložnostne krajšave, ki jih posamezni avtorji tvitov ustvarijo ad hoc in se izgovarjajo črkovno (*VD*, *pr.*).

Zaradi pestrih možnosti interpretacije pri označevanju sintaktičnih elips smo na skladijski ravni označevali zgolj tiste izpuste izbranih besednih vrst, ki niso posledica sicer običajnih izpustov zaradi sobesedila, temveč evidentni izpusti zaimkov, predlogov, samostalnikov in glagolov (*na trstenjakovi dobra knjiznica bojda*). Izpusti glagolov so dodatno razdeljeni na podkategorijo pomožnih in glavnih glagolov.

Na vsaki ravni smo dodali še kategorijo Drugo, če bi med označevanjem naleteli na pojave, ki ne spadajo v nobeno od navedenih kategorij. Kadar smo pri isti besedi opazili več kot en pojav krajšanja, smo označili vse. Izjema je krajšanje na leksikalni ravni, saj v takih primerih nismo označevali izpuščanja in nadomeščanja črk na nivoju zapisa.

## 4 Potek označevanja

Označevanje je potekalo v treh fazah. Za vzpostavitev tipologije in smernic za označevanje je bilo najprej dvojno označenih petdeset testnih tvitov, ki so bili izbrani naključno in niso bili uporabljeni za raziskavo. Zatem je bilo za testiranje izdelane tipologije in smernic dvojno

označenih 100 jezikovno in tehnično nestandardnih tvitov (T3L3) ter 100 jezikovno standardnih in tehnično nestandardnih tvitov (T3L1). Za dokončno uskladiitev tipologije je sledila faza kuriranja oznak omenjenih tvitov, ki je odpravila nesoglasja med označevalkama in s katero je bilo označevanje dokončno poenoteno. Nadaljnjih 600 tvitov iz preostalih kategorij tehnične in jezikovne standardnosti (T1L1 ter T1L3) sta označevalki označili enojno in brez kuriranja, sta pa med označevanjem komunicirali in skupaj razreševali morebitna vprašanja.

Enega od najbolj problematičnih vidikov označevanja so predstavljale narečne besede, saj se je izkazalo, da je težko ločevati med rabo nestandardne leksike in krajšanjem. Nestandardnih besed (*šipa, ketna, čuza* ipd.) nismo dojemali kot pojav krajšanja, čeprav so krajše od svojih standardnih ustreznice, medtem ko smo kot krajšanje razumeli npr. nestandardne redukcije končnic besed (*jedu* namesto *jedel*).

## 5 Analiza rezultatov in diskusija

Tabela 2 prikazuje pojave krajšanja po analiziranih nivojih, ki smo jih identificirali v različno standardnih tvitih.

Nivo	Stopnja stand.	Št.	%
Zapis (86,92 %)	T1L1	185	5,34 %
	T1L3	781	22,55 %
	T3L1	716	20,67 %
	T3L3	<b>1329</b>	<b>38,37 %</b>
Leksika (11,61 %)	T1L1	85	2,45 %
	T1L3	114	3,29 %
	T3L1	74	2,14 %
	T3L3	<b>129</b>	<b>3,72 %</b>
Skladnja (1,47 %)	T1L1	12	0,35 %
	T1L3	12	0,35 %
	T3L1	<b>17</b>	<b>0,49 %</b>
	T3L3	10	0,29 %

Tabela 2: Število posameznih identificiranih pojavov krajšanja.

Vsaj en pojav krajšanja smo identificirali v 89,4 % analiziranih tvitov. Vsega skupaj smo zabeležili 3464 pojavov krajšanja, ki so bili uvrščeni v 32 različnih kategorij. Za tri kategorije, ki smo jih predvideli v fazi oblikovanja tipologije, v analiziranem vzorcu nismo identificirali nobenega primera. Vse tri sodijo na ortografski nivo, in sicer nadomeščanje s številčnimi homofoni, nadomeščanje z logogrami in nadomeščanje z emotikoni/emojiji/piktogrami, kar je zanimivo, saj so ravno te v medijih pogosto izpostavljene kot značilni pojavi jezika v računalniško posredovani komunikaciji. Z oznako \$0, ki predstavlja nerelevantne tvite za analizo, smo označili 21 tvitov. Ti med drugim vključujejo avtomatsko generirane tvite, tujejezične tvite ipd. V enem izmed primerov smo identificirali celo tvit, ki je v celoti napisan v bohoričici.

Največ pojavov krajšanja najdemo v tehnično in jezikovno nestandardnih tvitih (T3L3), saj smo v njih identificirali kar 43 % vseh redukcij. Sledijo jim tehnično standardni in jezikovno nestandardni tviti (T1L3), kjer smo

zabeležili dobro četrtino (26 %) vseh pojavov krajšanja. Tehnično nestandardni in jezikovno standardni tviti (T3L1) vsebujejo 23 % redukcij, najmanj krajašanja (8 %) pa najdemo pri jezikovno in tehnično standardnih tvitih (T1L1).

Glede na vrhni nivo v tipologiji močno prevladujejo redukcije na ortografski ravni, saj predstavljajo 87 % vseh pojavov krajšanja. Na leksikalnem nivoju smo zabeležili dobrih 11,5 %, na skladenjskem pa nekaj manj kot 1,5 % odstotka vseh pojavov krajšanja. 722 pojavnic v vzorcu je prejele vsaj dve oznaki, pri 21 pa smo zabeležili po tri. Maksimalno število oznak na pojavnico je bilo štiri, najdemo pa jih pri dveh primerih (*nardil* (oznake OIKV, OOLK, OISV in OOPB) in *rtvsl* (oznake LNŽ, OOLO, OOPB in LUZ)).

### 5.1 Ortografski nivo

Kot je razvidno iz tabele 2, so pojavi krajšanja najpogostejši na nivoju zapisa. Motivacija za izpuščanje samoglasnikov in soglasnikov je pogosto približevanje zapisu, ki sledi govornemu obliki besede, pri opuščanju presledkov in ločil je precej tudi nenamernega izpuščanja, kot so tipkarske napake, do namernega izpuščanja pa prihaja tako zaradi osebnega sloga in izražanja identitete uporabnika kot tudi tehničnih okoliščin, pri čemer razlogi niso vedno enoznačni.

Identificirano krajšanje na ravni zapisa je podrobneje predstavljeno v tabeli 3. Najpogostejše je opuščanje presledkov pri ločilih (29,76 %), čemur sledi izpuščanje samostalnikov na koncu besede (17,77 %), opuščanje končnega ločila (14,48 %) v stavkih, izpuščanje samostalnika na sredi besede (14,31 %) in nadomeščanje daljših nizov črk s krajšimi nizi (6,91 %). Omeniti velja tudi, da se presledek pri ločilih najpogosteje opušča za vejico, nekoliko redkeje pa za piko ali tropičjem na koncu povedi.

Nivo1	Nivo2	Nivo3	Primer	Pogostost
Izpušč.	začetek b.	samostaln.	<i>mam</i>	1,89 %
		soglasnik	<i>lej</i>	0,20 %
		oboje	<i>koj</i>	0,07 %
	sredina b.	samostaln.	<i>bedn</i>	14,31 %
		soglasnik	<i>današnega</i>	1,13 %
		oboje	<i>kera</i>	1,13 %
konec b.	samostaln.	<i>anglesk</i>	17,77 %	
	soglasnik	<i>sam</i>	1,13 %	
	oboje	<i>lah</i>	1,93 %	
Opuš.	presl.	pri ločilih	<i>prepričano,da</i>	<b>29,76 %</b>
		med besedami	<i>inče tood njihni</i>	3,49 %
	ločil	konč.	<i>Ti to iz rokava streš</i>	14,48 %
		pike za okr.	<i>Slo</i>	1,43 %
		večbes.	<i>Kajmak in marmelada..</i>	3,42 %
		ločila		
Nadom.	daljših nizov s krajšimi s številčnimi homofoni		<i>ponuju</i>	6,91 %
	domačih črk s tujejezičnimi		<i>explozij</i>	0,53 %
	tujih črk z domačimi		<i>Tviter</i>	0,37 %
	z logogrami		-	0,00 %
	z emotikoni		-	0,00 %
drugo			<i>EPPja, 13incni</i>	0,46 %

Tabela 3: Delež posameznih kategorij krajšanja na nivoju zapisa.

Kategorije nadomeščanja na nivoju zapisa so zanimive tudi zato, ker smo v tipologiji predvideli več vrst krajšanja, kot smo jih z analizo gradiva identificirali. Nadomeščanj s številčnimi homofoni, logogrami in emotikoni/emojiji/piktogrami namreč v pregledanem vzorcu tvitov nismo našli, vendar predvidevamo, da bi se te kategorije pojavile, če bi označili večji vzorec tvitov.

Identificirali pa smo tako nadomeščanje domačih črk s tujejezičnimi in tujih črk z domačimi kot tudi nadomeščanje daljših nizov črk s krajšimi, vendar se je pri nadomeščanju z domačimi črkami izkazalo, da tako prilagojen zapis v večini primerov ni krajši od izvirnega (*stori* namesto *story*, *world* namesto *world*). Pri nadomeščanju domačih s tujimi črkami so novi izrazi dejansko krajši od izvirmih, zanimivo pa je, da je v večini primerov (v 9 primerih od 11 identificiranih) niz črk »ks« zamenjan s tujejezično črko »x« (npr. *expert* namesto *ekspert*).

## 5.2 Leksikalni nivo

Na leksikalnem nivoju smo zabeležili dobrih 11 % vseh pojavov krajšanja.

Nivo1	Nivo2	Primer	Pogostost
ustalj. krajšave	začetnice/kratice	<i>BDP</i>	34,83 %
	okrajšave	<i>št., slo.</i>	11,69 %
neustal. krajšave	simboli, formule, krnjene besede	<i>€, +</i>	14,93 %
	žanrsko-specifične	<i>Tw.</i>	12,44 %
drugo	krajšani neologizmi	<i>appi, Zoki</i>	9,45 %
	priložnostne	<i>sod[nik]</i>	12,94 %
		<i>e-sožalje</i>	3,73 %

Tabela 4: Delež posameznih kategorij krajšanja na leksikalnem nivoju.

Najpogostejša je raba ustaljenih krajšav (34, 83 %) z začetnicami ali kraticami, kamor spadajo imena strank (*SDS, PS*), državnih tvorb (*EU, ZDA*) in osebna lastna imena (*JJ*). Te zelo variirajo, saj med 140 označenimi krajšavami najdemo kar 104 različne. 27 različnih krajšav s simboli se v analiziranih besedilih pojavi 60-krat. Najpogostejša je uporaba znaka +, ki največkrat nadomešča veznik *in*, ter črke *x*, ki je uporabljena namesto besede *krat*.

Med najzanimivejše identificirane pojave krajšanja sodijo neustaljene krajšave, ki imajo izrazit strateški značaj časovnega, prostorskega oz. tehničnega krajšanja sporočil, zanimive pa so tudi z vidika kreativnosti uporabnikov in pestrosti jezika v elektronskih medijih. Tej kategoriji sledijo neustaljene priložnostne in žanrske krajšave, ki se pojavijo 50-krat. Priložnostne krajšave smo definirali kot ad hoc krajšanja, pri kateri avtor uporabi skovanko, ki jo lahko razumemo zgolj s pomočjo konteksta, v katerem se pojavi. Zabeležili smo 47 različnih priložnostnih krajšav, od tega je najpogostejši *PV* (v pomenu *predsednik vlade*), ki se trikrat. Med najzanimivejšimi so npr. *Klinca<sup>TM</sup>* (Univerzitetnega kliničnega centra), *upravič* (upravičiti) ter *odl.* (odločitev). Tudi neustaljene žanrske krajšave se pojavijo 50-krat. Kot te razumemo vse tiste krajšave, ki se

izgovarjajo po posameznih črkah in so tipične za elektronska besedila. Med najbolj značilne sodijo npr. *rt, btw, fb, lj in jbt*.

S 47 pojavitvami in 22 različicami jim sledijo ustaljene okrajšave, pri čemer uporaba pike za njimi ni vplivala na njihovo (ne)ustaljenost, saj smo manjkajočo piko označili na nivoju zapisa. Med žanrske krajšave med drugim sodijo *cca, dr., itd. in oz.* Na tem nivoju smo beležili še neologizme, v katere smo uvrstili tiste krajšave, ki se izgovarjajo kot besede, in jih je v analiziranem vzorcu najmanj (9,45 %). Sem spadajo tvorbe, kot so *alko, app* in *simultanka* ter različni vzdevki, kot denimo *Zoki* in *Bojči*. Našteli smo jih 38, med njimi je 28 različnih. Od priložnostnih krajšav se ločijo tudi po tem, da so večinoma razumljive tudi brez sobesedila.

## 5.3 Skladenjski nivo

Določanje izpustov oz. krajšanja na skladenjskem nivoju je že samo po sebi problematično, saj je nemalokrat težko oceniti ločnico med namernim krajšanjem in elipsami, ki pripomorejo k večji koherentnosti besedila. Iz tega razloga smo pri označevanju upoštevali le take izpuste besednih vrst, ki niso bili posledica »običajnega« izpusta zaradi sobesedila. Za boljše predstavbo v nadaljevanju navajamo primer izpusta pomožnega glagola, ki smo ga označili na besedi pred identificiranim izpustom, v tem primeru na pojavnici *včeraj*:

*Včeraj [SGG] bil na kuhančku v Ljubljani.*

Nivo1	Nivo2	Primer	Pogostost
izpust glag.	pomož.	<i>uf, zdej mi ze vec stvari jasnih hehe thx za info /.../</i>	60,78 %
	glavnega	<i>Bi blo treba tiralico?</i>	13,73 %
izpust zaimka		<i>...strinjam popolnoma...</i>	3,92 %
izpust predloga		<i>ocene /.../ se izkažejo predvsem [SD] politično motivirane</i>	3,92 %
izpust samostalnika		<i>/.../ uhhh ova je lejpa ka ma na boki [SS] od vseh knjig...</i>	7,84 %
drugo		<i>/.../ kjer lahko lajkaš to stran, ne pa vas morm dodat za frenda /.../</i>	9,80 %

Tabela 5: Delež posameznih kategorij krajšanja na skladenjskem nivoju.

Kot vidimo v tabeli 2, so pojavi krajšanja na skladenjskem nivoju najredkejši, najverjetneje zato, ker skladnja ne omogoča podobne kreativnosti, kot jo zaznamo na leksikalni ravni (npr. uporaba ustaljenih ali neustaljenih krajšav) ali na nivoju zapisa (npr. izpuščanje določenih črk ali nadomeščanje nizov črk), hkrati pa je bistvenega pomena za razumevanje besedila, zato lahko izpuščamo le določene skladenjske elemente.

Najpogostejši pojav krajšanja na skladenjskem nivoju (gl. tabelo 5) je izpust pomožnega glagola *biti*, ki predstavlja kar 60,78 % pojavov krajšanja na skladenjskem nivoju. Temu s 13,73 % sledi izpust glavnega glagola, kar je razumljivo, saj izpust pomožnega glagola načeloma ne onemogoča razumevanja sporočila, pri izpustu glavnega

glagola pa se lahko bistvo sporočila izgubi in je tako razumevanje oteženo. Še redkeje se pojavlja izpust samostalnika, ki že precej otežuje razumevanje besedila. Najredkeje pa se s 3,92 % pojavljata kategoriji izpusta zaimka in predloga, ki ne vplivata bistveno na razumevanje sporočila besedila.

## 5.4 Drugo

Med označevanjem smo našli tudi na nekaj primerov, ki jih nismo mogli umestiti v nobeno izmed osnovnih kategorij. Za take primere smo na vseh nivojih uporabili kategorijo *Drugo*.

Na nivoju zapisa smo sem prišteli označili tiste besede, v katerih je manjkal vezaj (npr. *euprava*, *90tih*, *rtvja*). Skupno se je nabralo 14 takih oznak.

Na leksikalnem nivoju smo pod *Drugo* uvrstili npr. besedi *Desusovec* in *ex*, v zvezi *ex politični zapornik*.

Na skladijski ravni smo zabeležili 5 nepredvidenih pojavov krajšanja, ki so najverjetneje nenamerni izpusti oz. napake, saj zanje nismo našli nobene druge razlage. Primer:

@Cvetlicarna Kak FB profil mate to? A ne bi raj naredili en page, kjer lahko lajkaš to stran, ne pa [SD] vas morm dodati za frenda (kar ne dela)?

Iz naslednjega navedenega primera lahko razberemo, da je tvit na koncu odrezan, kar smo označili kot skladijski izpust, manjka pa tudi končno ločilo na nivoju zapisa. Tak izpust smo ugotovili pri dveh primerih:

punce :) vasja danc na bazenu v sgjo od 3 ure dalje :) tk da lahka pridete ko ma tako zeljo vas s [SX] <http://t.co/qvk6jeMeBc>

## 6 Zaključek

V pričujoči raziskavi smo analizirali pogostost in načine krajšanja slovenskih sporočil na družbenem omrežju Twitter. Analiza je pokazala, da izmed vseh 800 analiziranih tvitov le nekaj nad 10 % zapisov ni vsebovalo nobenega krajšanja. Trend krajšanja je med slovenskimi uporabniki Twitterja torej zelo pogost. Največ krajšanj najdemo v tehnično in jezikovno nestandardnih tvitih (T3L3), kjer smo identificirali 43 % vseh redukcij, najmanj krajšav (8 %) pa najdemo pri jezikovno in tehnično standardnih tvitih (T1L1).

Skupno smo zabeležili 3464 pojavov krajšanja, ki so bili uvrščeni v 32 kategorij. Pri 21% pojavnic smo hkrati identificirali dve ali več pojavov krajšanja, maksimalno število oznak na pojavnico je bilo štiri, ki smo jih označili v dveh primerih. Glede na nivoje, ki smo jih določili v tipologiji, močno prevladujejo pojavi krajšanja na nivoju zapisa, ki predstavljajo kar 87 % vseh pojavov krajšanja. Na tem mestu velja omeniti tudi, da ortografski nivo vključuje največ različnih vrst krajšanja. Na leksikalnem nivoju smo zabeležili približno 11,5 %, na skladijskem pa nekaj manj kot 1,5 % odstotka vseh krajšanj. Najpogostejši tip krajšanja je opuščanje presledkov pri ločilih, ku mu sledi izpuščanje samoglasnikov na koncu besede, najredkeje pa sta rabljena izpusta predloga in zaimka.

Z vidika kreativnosti uporabnikov so najzanimivejše neustaljene krajšave na leksikalnem nivoju, ki smo jih razdelili na priložnostne in žanrsko specifične. Pri priložnostnih krajšavah avtor uporabi lastno skovanko, ki

jo razumemo zgolj s pomočjo sobesedila npr. *KlincaTM* (Univerzitetnega kliničnega centra, žanrsko specifične krajšave pa se izgovarjajo po posameznih črkah, so nesklonljive in so značilne za elektronska besedila (*jbt*, *btw*, *fb*, *lj*, *rt*). Med neustaljene krajšave spadajo tudi neologizmi, kot so *alko*, *app* in *simultanka* ter različni vzdevki, kot denimo *Zoki* in *Bojči*. Od priložnostnih krajšav se ločijo po tem, da so razumljive tudi brez sobesedila.

Z analizo smo prišli do jasne ugotovitve, da so slovenski uporabniki zelo nagnjeni h krajšanju besedil in da najpogosteje posežejo po krajšanju zapisa besedila (izpusti presledkov, ločil), manj pogosto pa posežejo po leksikalnih in skladijskih redukcijah. To je razumljivo, saj od naštetih treh kategorij posegi na nivoju zapisa najmanj vplivajo na razumevanje besedila.

Precej manj očitna je motivacija za redukcije. Razen očitnih tehničnih bližnjic z izpuščanjem presledkov in večdelnih ločil v analiziranem vzorcu namreč ni bilo mogoče ugotoviti, ali do krajšanja v sporočilih na Twitterju prihaja zaradi zavestnih odločitev uporabnikov, ki zaradi prostorske omejitve želijo skrajšati število znakov, za osebni slog oz. nenamerne izpuste.

V prihodnje bi bilo zanimivo raziskavo dopolniti z analizo drugih žanrov v korpusu Janes in s primerjavo pojavov krajšanja slovenskih uporabnikov, kadar tvitajo v različnih jezikih. Poleg tega bi lahko še primerjali pojave krajšanja v kratkih in dolgih tvitih, s čimer bi dobili vpogled v vpliv tehničnih okoliščin komuniciranja na objavljena sporočila. Druga zanimiva razsežnost je primerjalna analiza krajšanja v sporočilih, zapisanih na mobilnih napravah, s tistimi, ki so ustvarjena na osebnih računalnikih. Prav tako bi bilo pojave krajšanja zanimivo opazovati v različnih časovnih obdobjih, s čimer bi preverili, ali krajšanje narašča oz. upada ter ali postaja bolj oz. manj homogeno. Nenazadnje pa bi lahko raziskavo razširili tako, da bi se posvetili tudi daljšanju v zapisu, ki je usklajen z glasovno realizacijo besed pri prevzemanju iz jezikov z veččrkji (na primer q – ku in x – ks nasproti ch – č).

## 7 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. Posebna zahvala za pomoč pri oblikovanju tipologije označevanja gre Špeli Arhar Holdt, Jaki Čibeju, Tomažu Erjavcu, Damjanu Popiču in Katji Zupan.

## 8 Literatura

- Christian M. Alis in May T. Lim. 2013. Spatio-Temporal Variation of Conversational Utterances on Twitter <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077793#s3>.
- Jannis Androutopoulos in Gurly Schmidt. 2001. SMS-Kommunikation: Ethnografische Gattungsanalyse am Beispiel einer Kleingruppe. V: Meer, D., ur., *Zeitschrift für Angewandte Linguistik*. Bd. 36, Frankfurt/Main, 49–80.
- Markus Bieswanger. 2007. abbrevi8 or not 2 abbrevi8: A contrastive analysis of different space and time-saving strategies in English and German text messages. V: *Texas Linguistics Forum*, volume 50.



- David Crystal. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.
- Jaka Čibej, Darja Fišer in Tomaž Erjavec. (V tisku). Normalisation, Tokenisation and Sentence Segmentation of Slovene Tweets.
- Nicola Döring. 2002. Kurzm. wird gesendet – Abkürzungen und Akronyme in der SMS-Kommunikation. V: *Muttersprache - Vierteljahresschrift für deutsche Sprache*, 112 (2), 97–114.
- Richard Eckart de Castilho et al. 2014. WebAnno: a flexible, web-based annotation tool for CLARIN V: *Proceedings of the CLARIN Annual Conference (CAC) 2014*. CLARIN ERIC, October 2014. <https://www.clarin.eu/content/papers-posters-and-demos-cac2014>.
- Stephan Gouws et al. 2011. Contextual Bearing on Linguistic Variation in Social Media. V: *Proceedings of the Workshop on Language in Social Media (LSM 2011)*. Portland, Oregon. 20–29. <http://dl.acm.org/citation.cfm?id=2021113>
- Ylva Hård af Segerstad. 2002. Use and Adaptation of Written Language to the Conditions of Computer-Mediated Communication. Göteborg: Göteborg University.
- Susan C. Herring. 2001. Computer-mediated discourse. V: D. Schiffrin, D. Tannen, & H. Hamilton, ur., (pp.*The Handbook of Discourse Analysis* 612-634). Oxford: Blackwell.
- Urška Jarnovič. 2007. Diskurzivne značilnosti SMS-ov. *Jezik in slovstvo*, 52 (2): 61–79. Ljubljana: Slavistično društvo Slovenije.
- Mojca Kompara. 2009. Prepoznavanje krajšav v besedilih. V: Peter Weiss, ur., *Jezikoslovni zapiski* 15, št. 1–2. 95–112. Založba ZRC, Ljubljana.
- Nikola Ljubešič, Darja Fišer, Tomaž Erjavec, Jaka Čibej, Dafne Marko, Senja Pollak in Iza Škrjanec. 2015. Predicting the level of text standardness in user-generated content. Zbornik konference RANLP 2015, 7.–9. september 2015, str. 371–378, Hissar, Bolgarija.
- Mija Michelizza. 2012. Ločila in druga pisna znamenja v elektronskih besedilih. V: Nataša Jakop, Helena Dobrovoljc, ur., *Pravopisna stikanja: razprave o pravopisnih vprašanjih*. 151–162. Založba ZRC, Ljubljana.
- Richard Rogers. 2014. Debanalising Twitter. The transformation of an object of Study. V: Katrin Weller et al., ur., *Twitter and Society*. IX–XXVI. Peter Lang Publishing, Inc., New York.
- Peter Schlobinski et al. 2001. Simsen. Eine Pilotstudie zu sprachlichen und kommunikativen Aspekten der SMS-Kommunikation. *Networx* 22. Retrieved July 1, 2006, from <http://www.medienSprache.net/networx/networx-22.pdf>.
- Malin Sveningsson. 2001. Creating a Sense of Community: Experiences from a Swedish Web Chat. The TEM Institute, Dept. of Communication Studies. Linköping, Linköping University: 250.
- Crispin Thurlow. 2003. Generation Txt? The sociolinguistics of young people's text- messaging. *Discourse Analysis Online*, 1 (1).