

Raba vejice v uporabniških spletnih vsebinah

Damjan Popič,* Darja Fišer,*† Katja Zupan,† Polona Logar‡

* Oddelek za prevajalstvo Filozofske fakultete v Ljubljani
Aškerčeva 2, 1000 Ljubljana

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

† Odsek za tehnologije znanja, Institut »Jožef Stefan«

Mednarodna podiplomska šola Jožefa Stefana

Jamova cesta 39, 1000 Ljubljana

katja.zupan@ijs.si

‡ Oddelek za slovenistiko Filozofske fakultete v Ljubljani

Aškerčeva 2, 1000 Ljubljana

polona.logar1@gmail.com

Povzetek

V prispevku predstavimo pilotno študijo nestandardne stave vejice v slovenskih uporabniških spletnih vsebinah. Najprej predstavimo razvoj tipologije za označevanje nestandardne stave vejice v slovenščini, v nadaljevanju pa predstavimo rezultate analize 500 naključno izbranih tvitov različnih stopenj standardnosti. Analiza je pokazala, da je v slovenskih uporabniških vsebinah problematična predvsem manjkajoča vejica, in sicer med odvisniki in nadrednimi stavki ter pri pri- ter pastavnih strukturah.

Comma Use in User-Generated Content

The article presents a pilot study of non-standard use of the comma in Slovene user-generated content. Initially, the development of the typology for annotating non-standard comma use in Slovene is presented. Afterwards, the results of an analysis of 500 randomly selected tweets (with two varying levels of standardness) are given. The results show that in Slovenian user-generated content comma use is problematic mostly in regard to the missing comma, especially between dependent and independent clauses, and after and before small clauses.

1 Uvod

S primerjalno raziskavo korpusov Janes (Erjavec et al. 2015) in Kres (Logar et al. 2012) je bilo pokazano, da raba vejice na družbenih omrežjih ne peša in da so razlike med rabo vejice v tradicionalnih in novomedijskih besedilih manjše, kot pregovorno velja (Popič in Fišer 2015). Vendar je bila raziskava omejena le na ožji nabor tipičnih atraktorjev vejic (Verovnik 2003, Žibert 2006) in na samo pogostnost ter razporejenost vejic, zato jo želimo nadgraditi s pričujočim prispevkom in pripraviti izhodišče za celovito analizo rabe vejice, predvsem pa razviti čim bolj univerzalen kategorizacijski sistem za označevanje nestandardne rabe vejice in ga preizkusiti na širšem vzorcu.

V pričujoči raziskavi preverimo, v čem raba vejice v uporabniških spletnih vsebinah odstopa od norme. Analiza temelji na označevanju napačno stavljene vejice v naključnem vzorcu 500 tvitov iz korpusa Janes v0.4. Pri tem smo kot merilo upoštevali še stopnjo jezikovne in tehnične standardnosti besedil (Ljubešić et al. 2015) ter v vzorec zajeli po 250 tvitov z naslednjima oznakama: T1L3 (nestandarden jezik, a standardna raba velikih začetnic, presledkov, ločil) in T3L3 (tako tehnično kot tudi jezikovno povsem nestandardni tviti). Ker nas zanima stava vejice v nestandardni slovenščini, tvitov, zapisanih v standardni slovenščini (L1), v analizo nismo zajeli, vendar v prihodnje načrtujemo razširitev raziskave tudi na standardne tvite in druge tipe uporabniških spletnih vsebin, kot so komentarji na novice, forumska sporočila, blogi in pogovorne strani na Wikipediji.

Na podlagi označenega korpusa smo opravili kvantitativno ter preliminarno kvalitativno analizo manjkajočih in odvečnih vejic v uporabniških spletnih vsebinah ter tako dobili vpogled v najbolj problematične kategorije.

2 Metodologija

Označevanje nestandardne stave vejic je potekalo v orodju WebAnno (Eckart de Castilho et al. 2014), uporabili pa smo kategorije, predstavljene v Tabeli 1. V nadaljevanju najprej predstavimo zasnovano in izpopolnjevanje tipologije za označevanje nestandardne rabe vejic, zatem pa opišemo še proces označevanja in delotok.

2.1 Tipologija

Za opis nestandardne (ne)rabe vejice smo razvili obsežno tipologijo, pri tem pa smo se oprli na aktualni jezikovni predpis (Slovenski pravopis 2001, Pravila), z nekaj dopolnitvami in prilagoditvami. Za podporo sorodnim raziskavam in vpogled v sistem označevanja pri pričujoči raziskavi smo smernice za označevanje objavili na spletu.¹ Pri sestavi kategorizacije smo se primarno držali načela, da naj bo ta čim bolj univerzalna (da torej lahko s čim bolj podobnimi kategorijami opiše čim več primerov nestandardne rabe vejice (tj. odvečne vejice, manjkajoče vejice itd.)). Seveda to ni bilo povsem izvedljivo, zato tipologija vsebuje tudi kategorije, vezane predvsem na specifične primere rabe (tako denimo kategorija Stavčni člen opisuje predvsem nestandardno »stavčnočlensko« rabo vejice, ki je vezana zgolj na odvečno vejico), povečini

¹ Smernice za označevanje so na voljo na spletni strani <http://nl.ijs.si/janes/wp-content/uploads/2014/09/vejice-smernice.pdf>.

pa so kategorije vendarle univerzalne in namenjene opisu katerega koli segmenta jezika – tipologija torej ni namenjena zgolj opisu nestandardne stave vejice v spletni rabi, temveč je bil namen sestavljalavcev prav ta, da lahko zadosti opisu katerega koli žanra ali besedilnega tipa.

Poleg jezikovnosistemskega opisa stave vejice smo se pri pripravi tipologije oprli tudi na že opravljene empirične, izgradivne preglede stave vejice v različnih okoljih, in sicer v spletnem (Popič in Fišer 2015), šolskem (Logar in Popič 2015, Kosem et al. 2012) ter v okolju poklicnih piscev (Popič 2014). Z začetno različico tipologije smo označili testni nabor tvitov in na podlagi pridobljenih spoznanj tipologijo dopolnjevali. Bistvene razlike med začetno in končno tipologijo so se nanašale predvsem na nestandardno odvečno vejico, pri kateri tipična razdelitev glede na jezikovnosistemski opis odpove, saj temelji na skladenjski razčlenitvi – odvečnost vejice pa temelji ravno na tem, da uporabnik skladenjske razčlenitve ne (pre)pozna. Tipologijo podajamo v nadaljevanju, brez ponavljajočih se podkategorij – pri vseh segmentih skladenjske rabe vejice smo namreč uvedli še podkategoriji levo- in desnosmerne vejice (Korošec 2003), saj v tem prepoznavamo različne (potencialne) vzgibe in razloge za nestandardno stavo vejice.

Manjkajoča/odvečna vejica/napačen znak
1 Skladenjska
1.1 Priredje
1.1.1 Vežalno <i>@xxx jst sm se že zdavni odločla, da bo prej! se je pa tut že zdavni okol obrnu, in prehiteva po rasti..tko, da držim pesti! @xxx</i>
1.1.2 Stopnjevalno <i>@xxx tip nikakor ne sodi v odbore kaj sele v parlament...ta je bolj za debate pr'Kovac...tam je bolj po njegovem stilu...</i>
1.1.3 Ločno <i>@xxx @xxx @xxx Zlata, men se odpre le uradna stran RTV. Kaj moram nardit, oz za katero oddajo gre?</i>
1.1.4 Protivno <i>@xxx Razumem ampak point rapa na začetku je bil predvsem v tem kakšno zgodbo boš povedal btw ma Kendrick tud bangerja...</i>
1.1.5 Vzročno <i>@xxx za urgenco #UKC in ZD zelo strinjam..lah bi še mal fasado poštimal pa kak oddelk dogradil drgač pa super ja..#domzale</i>
1.1.6 Posledično <i>@xxx not even close.. ampak vseen dobr vedet :) .. zdey bom ene dva tedna na zlo low budget tko da se ne branim marelic :)</i>
1.1.7 Pojasnjevalno <i>@xxx jao, to so nardil možu, ko je bil v ZDA in to pred startom dirke. Na srečo je šla ekipa dol kasneje in da sem bila pooblaščena.</i>
1.2 Odvisnik
1.2.1 Osebkov <i>@xxx pri moji ghetto mikrovalovki tm enih 20sek.mal ščekirej umes. je dost da je topla, za čez šmorn glih kul. :) @xxx</i>
1.2.2 Predmetni <i>@xxx neki sm vidu sproti, drugač pa ne vem kaj dogaja in kaj je to..:) razen komada, pa da je "snovalac" le tega otvoril zadevo..</i>
1.2.3 Krajevni

<i>@xxx o/ mel sem srečo. Če nebi blo nesreče jutraj na Ac bi imel avto tam kjer je bilo meter vode. Sm se pa peljal do Ajdušne</i>
1.2.4 Časovni <i>@xxx Vse živo ... trust me !! :) Preden pa jaz grem na Švedsko pa tebe verjetno čaka kofi pri Slamiču hehe :) Kaj pravš ? ;)</i>
1.2.5 Načinovni <i>Tud macki so po lastnikih. K rai se tko igra da lezi na istem mestu pa zamahne s tacko sam takrt ko igraca pride cist do njega #lenoba</i>
1.2.6 Vzročni <i>Fakt št. 1: Skor sm zgubila glavo na tekočih stopnicah ker sm oprezala za enim tipom. Fakt št. 2: V torbici nosim cedevito. #normalday</i>
1.2.7 Namerni /
1.2.8 Pogojni <i>@xxx @xxx @xxx @xxx @xxx @xxx ja ce bote organiziral v 'otrokuprijaznim' uram pa tud jz pridem :)</i>
1.2.9 Dopustni <i>@xxx je dobra in rešuje življenje poštenim ljudem, ki drgač ne bi imeli možnosti, mi je vseen če tud kak opravičil napiše @xxx</i>
1.2.10 Prilastkov <i>drgač pa ivan ni nestrpen do srbov , bosancev in hrvatov. ma enga prjatla k ma prjatla k ma prjatla k ma vodovodarja iz prijedora #truestory</i>
1.3 Polstavek <i>Jst bi tud najdu kovanec vreden veliko denarja. Kje je drgač sploh logika? Gor piše neki, vrednost pa veliko večja. Težave modernega sveta.</i>
1.4 Stavčni člen <i>@xxx poznam komb.proja+kneipp+franck za "belo kavo". sem rad pil, ko sem bil mali:) za muckefuck, pa sem slišal 3 dni nazaj :D</i>
1.5 Besedna zveza <i>Medtem, ko vsi brenčite o stricih, MK in BP jaz razmišljam le o #Gaza. In razmišljam kdaj je Twitter postal just another social network.</i>
1.6 Pri-, pa- in dostavek ter izpostavek <i>sej sm nasla v smeteh pol ful dobr res haha jao</i> 👋👋
2 Neskladenjska
2.1 X → vejica /
2.2 Vejica → X /
2.3 Tipkarska napaka /
2.4 Drugo (gl. Tabela 6)

Tabela 1: Pregled skladenjskih in neskladenjskih oznak po kategorijah.

Kot lahko vidimo, kategorije pretežno temeljijo na razdelitvi, podani v jezikovnem predpisu, dodani sta kategoriji Stavčni člen in Besedna zveza, pri neskladenjski rabi vejice pa smo želeli čim podrobneje razčleniti, za katero vrsto nestandardne stave gre, zato so vključene še štiri podkategorije. Za neskladenjsko rabo vejice je rezerviran tudi parameter »napačen znak«, saj v tovrstnih primerih kategoriji manjkajoče in odvečne vejice nista (nujno) relevantni.

V določeni meri smo jezikovnosistemske kategorije tudi združevali (gl. kategorijo 1.6), in sicer na mestih, kjer drobljenje informacij po naših pričakovanjih ne bi bistveno izboljšalo podatkov o nestandardni rabi vejice, bi pa nesorazmerno povečalo število kategorij.

2.2 Označevanje

Med procesom označevanja je bilo dvojno označenih 500 tvitov. Pri označevanju je bilo mogoče za posamezni primer izbrati le eno kategorijo, večkategorialnih oznak – npr. ko je utemeljitev za stavo vejice lahko tako levo- kot desnosmerna – nismo dopuščali. Če je bilo mogočih več interpretacij, smo izbrali kategorijo, ki je bila v danem kontekstu bolj povedna oz. verjetnejši razlog za napačno stavo vejice, pri čemer smo favorizirali odvisnike. Pogost primer je npr. oziralni odvisnik sredi tro- ali večstavčnih povedi, kjer pogosto umanjka levosmerna vejica odvisnika.

Po koncu označevanja je kuratorica pregledala vseh 500 dvojno označenih tvitov in pri neskladjih med označevalcema sprejela dokončno odločitev. Po podatkih orodja WebAnno je Cohenov koeficient ujemanja med označevalcema kappa znašal 0,57, kar je srednje dobro ujemanje. Zelo dobrega ujemanja zaradi težavnosti problema, sprotnega vzpostavljanja tipologije in smernic ter neizkušenosti označevalcev ni bilo mogoče pričakovati – tudi zato smo se odločili za dvojno označevanje. Neskladja med označevalcema so bila najpogostejša predvsem pri naslednjih segmentih:

- ločevanje posameznih medmetov znotraj pastavka;
- prepoznavanje vezalnosti, protivnosti in posledičnosti veznikov *pa* ter *in*;
- obravnava smeškov in drugih metajezikovnih struktur, tipičnih za uporabniške vsebine;
- prepoznavanje vrst odvisnikov, zlasti v primerih, pri katerih so vezniške strukture zapisane nestandardno.

Pri prvih treh problematičnih mestih smo se odločili za princip minimalne intervencije, zato se je kuratorica izogibala posegov v izvorno stavo vejic, razen če je poseg nedvoumno predvidevalo določeno pravopisno pravilo ali kontekst, medtem ko je bila pri zadnjem problematičnem torišču potrebna vsebinska odločitev. Za čim ustrežnejšo obravnavo težavnih segmentov smo sprejeli natančne smernice, da bo označevanje v prihodnje enostavnejše in bolj usklajeno.

Skupno je bilo pri 500 tvitih označenih 405 mest, kar pomeni v povprečju 0,8 oznake na posamezni tvit.

3 Analiza

V nadaljevanju predstavimo izsledke analize nestandardne rabe vejic v uporabniških spletnih vsebinah. Najprej predstavljamo kumulativne rezultate po posameznih kategorijah, ki jih podajamo v Tabeli 2, v nadaljevanju pa se osredotočimo na posamezne (pod)kategorije, zlasti tiste, ki so glede na rezultate empirične raziskave v uporabniških spletnih vsebinah še posebno problematične.

Kategorija	Odveč	Manjka	Nap. znak
1 SKLADENJSKA	19 (4,7 %)	382 (95,3 %)	0
1.1 Priredja	8 (2 %)	35 (8,7 %)	0
1.1.1 Vezalno	6	8	0
1.1.2 Stopnjevalno	0	2	0

1.1.3 Ločno	2	0	0
1.1.4 Protivno	0	13	0
1.1.5 Vzročno	0	1	0
1.1.6 Posledično	0	10	0
1.1.7 Pojasnjevalno	0	1	0
1.2 Odvisniki	2 (0,5 %)	230 (57,4 %)	0
1.2.1 Osebkov	0	26	0
1.2.2 Predmetni	1	89	0
1.2.3 Krajevni	0	1	0
1.2.4 Časovni	0	20	0
1.2.5 Načinovni	1	11	0
1.2.6 Vzročni	0	16	0
1.2.7 Namerni	0	0	0
1.2.8 Pogojni	0	25	0
1.2.9 Dopustni	0	1	0
1.2.10 Prilastkov	0	41	0
1.3 Polstavek	0	2 (0,5 %)	0
1.4 Stavčni člen	3 (0,7 %)	0	0
1.5 Besedna zveza	6 (1,5 %)	0	0
1.6 Pa-, pri- do- in izpostavek	0	115 (28,7 %)	0
2 NESKLADENJSKA	0	3 (75 %)	1 (25 %)

Tabela 2: Pregled skladenjskih in neskladenjskih oznak po kategorijah. Pri nadkategorijah so podani deleži oznak znotraj skladenjskega in neskladenjskega segmenta.

Kot lahko razberemo iz Tabele 2, so oznake med kategorijami razporejene zelo heterogeno, zelo neenakomerno pa so posejane tudi oznake glede na vrsto skladenjskega razmerja. Ta heterogenost je značilna za obe ravni tehnične standardnosti tvitov, ki sta zajeti v analiziranih podatkih.

Kot lahko vidimo, je bilo v 500 tvitih zanemarljivo malo (4) primerov napačno stavljene neskladenjske vejice, prav tako pa se je pri izbrnem vzorcu pokazalo, da je v uporabniških spletnih vsebinah skorajda zanemarljiv pojav odvečna vejica (19 primerov), saj je velika večina (382 primerov) oznak vezana na manjkajočo vejico. Pri tem je treba poudariti, da zaradi majhnega vzorca te ugotovitve ni primerno posploševati. Velika večina popravkov se nanaša na skladenjsko rabo vejice (401 od skupno 405 primerov oz. 99 %). Glede na rezultate empirične analize lahko vidimo, da sta še posebno akutni kategoriji manjkajočih vejic pri odvisnikih (230 primerov oz. 57 % vseh skladenjskih oznak) in pastavčnih strukturah (115 primerov oz. 28,7 % vseh skladenjskih oznak).

3.1 Manjkajoča vejica

Kot lahko vidimo v Tabeli 2, je stava vejice z vidika manjkajoče vejice problematična predvsem pri odvisniških in pristavčnih strukturah, zato se v nadaljevanju nekoliko podrobneje posvetimo prav temu kategorijama. V Tabeli 3 je podan prikaz najpogostejših odvisniških vrst z označeno manjkajočo vejico.

Odvisnik	Število oznak
Predmetni (D)	80 (19,8 %)
Prilastkov (D)	29 (7,1 %)
Osebkov (D)	24 (5,9 %)
Vzročni (D)	16 (4,0 %)
Pogojni (D)	13 (3,2 %)
Časovni (D)	13 (3,2 %)
Pogojni (L)	12 (2,9 %)

Prilastkov (L)	12 (2,9 %)
Načinovni (D)	10 (2,5 %)
Predmetni (L)	9 (2,2 %)
Časovni (L)	7 (1,7 %)
Osebkov (L)	2 (0,5 %)
Načinovni (L)	1 (0,2 %)
Krajevni (D)	1 (0,2 %)
Dopustni (D)	1 (0,2 %)

Tabela 3: Pregled manjkajočih vejic pri odvisnikih po frekvenci in deležu znotraj skladske kategorije.

Kot lahko vidimo, je od skupno 230 manjkajočih vejic pri odvisnikih edina zares problematična kategorija desnosmerne vejice pri predmetnih odvisnikih, pri katerih najdemo več kot tretjino primerov z manjkajočo vejico. Tudi na splošno se je izkazalo, da je desnosmerna vejica precej bolj akutna kot levosmerna, saj zaseda šest najpogostejših mest z manjkajočo vejico (175 od skupno 230 manjkajočih vejic). Edini odvisniški vrsti, pri katerih smo zaznali nekaj več manjkajočih levosmernih vejic, sta kategoriji prilastkovih in vzročnih odvisnikov. Za slednje lahko sklepamo, da pogosto začenjajo povedi, pri prilastkovih odvisnikih pa uporabniki pogosto pozabijo skleniti oziralni odvisnik.

	Odvečna	Manjkajoča
Levosmerna	/	79 (19,7 %)
Desnosmerna	/	36 (9,0 %)

Tabela 4: Pregled oznak pri pa-, pri-, do- in izpostavnih strukturah po frekvenci in deležu skladske oznak.

Pri pristavkih, dostavkih, izpostavnih in pastavkih je problematična predvsem vejica na začetku, relativno pogosta pa je tudi nestandardna stava brez vejice na koncu stavka. V številnih primerih gre za nestandardno stavo vejice pri denimo medmetih (npr. *hahaha*), ki jih – tako domnevamo – uporabniki spletnih omrežij ne dojemajo kot pastavčne tvorbe, temveč kot metajezikovno pojavnost oz. verbalizirane emotikone. Zelo pogosta je tudi pojavnost nestandardne rabe vejice pri členkovnih pastavkih na začetku in koncu stavka, zlasti v navezavi s sklicem na uporabniško ime. Te rabe v korpusu sicer nismo označevali kot odmik od standarda, predvsem zaradi velike pogostosti in zaradi tehničnih zahtev uporabe Twitterja, saj lahko stavljenje vejice v sklice povzroči težave s samim sklicem. V vsakem primeru pa način uporabe sklicev potrjuje domnevo, da uporabniki izražajo metajezikovne prvine z uporabo jezikovnih sredstev.

3.2 Odvečna vejica

Vseh primerov odvečne vejice od skupno 401 skladske oznake v zbirki podatkov je bilo 19, kar pomeni manj kot 5 odstotkov vseh oznak, njihova porazdelitev po kategorijah pa je naslednja:

Oznaka	Število
Vežalno priredje	6
Besedna zveza	6
Stavčni člen	3
Ločno priredje	2

Predmetni odvisnik	1
Načinovni odvisnik	1

Tabela 5: Pregled odvečnih vejic glede na kategorijo po frekvenci in deležu skladske oznak.

Kot lahko vidimo, je bila vejica največkrat odvečna v besednih zvezah (denimo med sestavljenimi vezniki), med strukturami v vežalnem ali ločnem priredju, za stavčnim členom in med enakovrednima odvisnikoma. Sodeč po skupnem številu označenih nestandardnih mest, bi pričakovali večje število odvečnih vejic, predvsem stavčnočlenskih, zlasti glede na raziskave na standardnojezikovnem gradivu (prim. Popič 2014). Preden lahko zaključimo, da je neskladska raba vejice pretežno manjkajoča, ker je to ena od strategij krajšanja sporočil v tovrstnem načinu komuniciranja in ker je vejico psihološko in tehnično lažje izpustiti kot pa napisati po nepotrebem, bi bilo nujno treba opraviti analizo na večjem vzorcu besedil tega žanra.

3.3 Neskladska vejica

Kot smo pokazali v predhodnih poglavjih, so bili v celotnem naboru podatkov označeni zgolj štirje primeri nestandardne neskladske vejice, ki so se pojavili v treh tvitih. Te izpisujemo v celoti v Tabeli 6.

Setamo po Lj in pride Zoki mim, se ustavi pa da Emanuelu petko. Tamalmu nc jasn. Recem(,) to je Zoki kralj(,) in se tamal zadere: Zoki kralj! :)
@xxx Ja bi mogla tud jst naletet na dobrega :) ampak ima res ogromno folka sfaljenega (ne ostri do 2.8). Je, kar je.
@xxx "Maybe yes, maybe no, maybe go home(,)" so radi rekli Šerpe na odpravi na Annapurno, kadar smo jih vprašali, če bo vreme OK

Tabela 6: Pregled primerov z označeno napačno rabljeno neskladsko vejico.

Kot lahko vidimo, so od tega tri vejice v resnici skladske, in sicer gre za tri primere soredja (1. in 3. tvit; v oklepaju so podane manjkajoče vejice na mestih, ki so bila označena tudi med označevanjem), ki ga namenoma nismo vključili v tipologijo, saj nismo pričakovali primerov premege govora. Pri drugem primeru (krepko je tiskana pojavnica, ki je bila označena med označevanjem) pa je dejansko uporabljena neskladska vejica, vendar lahko v tem primeru oznako problematiziramo, saj gre – kolikor lahko razberemo iz sobesedila – za navedbo odprtosti zaslonke fotografske leče, ki se tipično podaja s piko. Podobno kot pri nekaterih drugih kategorijah smo glede na dosedanje raziskave, izvedene na standardnem gradivu, pričakovali bistveno več primerov nestandardne rabe neskladske vejice.

4 Sklep

V prispevku smo predstavili rezultate empirične analize nestandardne rabe vejice v uporabniških spletnih vsebinah v slovenščini, raziskavo pa smo izvedli na naboru 500 tvitov. Glavni namen raziskave je bil začrtati nadaljnje raziskave stave vejice v slovenščini, zlasti v primerjavi s

standardnojezikovnim gradivom, v okviru te zasnove pa predvsem določiti, v kolikšni meri raba vejice na Twitterju odstopa od jezikovnega standarda.

Rezultati pretežno pritrujejo dosedanjim raziskavam, obenem pa izpostavljajo nekaj novih težišč pri tej problematiki, in sicer lahko vidimo, da je nestandardna raba vejice na Twitterju vezana predvsem na skladiščno rabo. Pri tem se kot najbolj problematični izkazujeta kategoriji pristavnih struktur in odvisnikov (z manjkajočo vejico). Medtem ko lahko pri pristavnih strukturah del razloga za pogostnost nestandardne stave iščemo pri strukturi Twitterja kot medija in pri novem razumevanju tovrstnih struktur, zlasti pastavkov, pa ugotavljamo, da odsotnost vejice pri predmetnih odvisnikih izhaja iz težav uporabnikov pri prepoznavanju skladiščnih struktur – to pa je v slovenščini univerzalna težava, ki ni vezana na interaktivnost ali formalnost medija. Raziskava je dala precej presenetljive rezultate z vidika rabe odvečne vejice, saj se ta pojavlja precej redkeje kot v primerljivih študijah na standardnojezikovnem gradivu, kar nakazuje na potrebo, da se pričujoča raziskava obogati še z analizo drugih stopenj standardnosti tvitov in drugih vrst uporabniških vsebin, zajetih v korpusu Janes, ter jih neposredno primerja s standardnojezikovnim gradivom. Glede na izsledke analize lahko sklenemo tudi, da bo tovrstne raziskave mogoče opraviti s tipologijo, predstavljeno v pričujočem prispevku.

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) in programa Mladi raziskovalec (37487), ki ju financira ARRS.

6 Literatura

- Richard Eckart de Castilho, Chris Biemann, Iryna Gurevych in Seid Muhie Yimam. 2014. WebAnno: a flexible, web-based annotation tool for CLARIN. V: *Zbornik letne konference CLARIN (CAC) 2014*, Soesterberg, Nizozemska.
- Tomaž Erjavec, Darja Fišer in Nikola Ljubešić. 2015. Razvoj korpusa slovenskih spletnih uporabniških vsebin Janes. V: *Zbornik konference Slovenščina na spletu in v novih medijih*, str. 20–26, Ljubljana. Znanstvena založba Filozofske fakultete.
- Tomaž Korošec. 2003. K pravilom za skladiščno vejico v Slovenskem pravopisu 2001. *Slavistična revija*, 51(2): 247–266.
- Iztok Kosem, Mojca Stritar, Sara Može, Ana Zwitter Vitez, Špela Arhar Holdt in Tadeja Rozman. 2012. *Analiza jezikovnih težav učencev: korpusni pristop*. Trojina, zavod za uporabno slovenistiko, Ljubljana.
- 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. V: *Zbornik konference RANLP 2015, 7.–9. september 2015*, str. 371–378, Hisar, Bolgarija.
- Nataša Logar in Damjan Popič. 2015. Vejica: rezultati anketne raziskave med dijaki in študenti. *Jezikoslovni zapiski*, 21(2): 45–59.
- Nataša Logar Berginc, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek. 2012. *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida*

in ccKRES: gradnja, vsebina, uporaba. Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede Univerze v Ljubljani.

Damjan Popič. 2014. *Korpusnojezikoslovna analiza vplivov na slovenska prevodna besedila*. Doktorska disertacija, Filozofska fakulteta Univerze v Ljubljani.

Damjan Popič in Darja Fišer. 2015. Vejica je mrtva, živela vejica. V: *OBDOBJA 34: Slovnica in slovar – aktualni jezikovni opis*, str. 609–618, Ljubljana. Znanstvena založba Filozofske fakultete.

Tadeja Rozman, Mojca Stritar in Izток Kosem. 2012. Šolar – korpus šolskih pisnih izdelkov. V: *Empirični pogled na pouk slovenskega jezika*. Trojina, zavod za uporabno slovenistiko, Ljubljana.

Slovenski pravopis. 2001. *Pravila*. Ur. Jože Toporišič et al. Založba ZRC in ZRC SAZU, Ljubljana.

Tina Verovnik. 2003. Vejica premalo, vejica preveč (2. del). *Pravna praksa*, 22(21): 51.

Živa Žibert. 2006. *Slovenska vejica: balast ali skladišna nujnost slovenskega knjižnega jezika?* Diplomsko delo, Fakulteta za družbene vede Univerze v Ljubljani.