

## Korpusna analiza nestandardne stave vejice po uvajalnih prislovnih zvezah

Eneja Osrajnik,<sup>†</sup> Darja Fišer,<sup>‡\*</sup> Vojko Gorjanc<sup>‡</sup>

<sup>†</sup> Ulica Pohorskega bataljona 43, 1000 Ljubljana  
eneja.osrajnik@gmail.com

<sup>‡</sup> Oddelek za prevajalstvo, Filozofska fakulteta, Univerza v Ljubljani  
Aškerčeva 2, 1000 Ljubljana

\* Odsek za tehnologije znanja, Institut »Jožef Stefan«  
Jamova cesta 39, 1000 Ljubljana  
darja.fiser@ff.uni-lj.si

### Povzetek

V pričujočem prispevku obravnavamo razširjenost nestandardne stave vejice po uvajalnih prislovnih zvezah (UPZ) v slovenskem akademskem diskurzu, natančneje v diplomskih in doktorskih delih, ter diskurzu uporabniško generiranih spletnih vsebin, in sicer v zasebnih in poslovnih blogih. Proučujemo tudi vpliv atraktorjev, tj. besed v predložni funkciji, ki uvajajo UPZ, in vpliv dolžine UPZ na stavo tovrstne vejice. Na podlagi vzorca 6000 zadetkov korpusa spletne nestandardne slovenščine Janes in korpusa akademske slovenščine KAS smo ugotovili, da je tovrstna stava vejice glede na žanr občutno bolj razširjena v diskurzu uporabniško generiranih spletnih vsebin, znotraj akademskega diskurza pa se pogosteje pojavlja v delih študentov dodiplomskega kot podiplomskega študija. Poleg tega se tovrstna stava vejice najpogosteje pojavlja po kratkih UPZ, njena razširjenost pada z večanjem razdalje med atraktorjem in vejico, najpogostejši atraktorji v obeh vrstah diskurza pa so predlogi *zaradi*, *za*, *po*, *na*, *kljub*, *poleg* in *ob*.

### Corpus analysis of non-standard comma usage after introductory adverbial phrases

This paper addresses the prevalence of non-standard comma usage after introductory adverbial phrases (AIP) in Slovene academic discourse (in undergraduate and PhD theses) and user-generated content (UGC; in private and corporate blogs). We also analyse how the so-called attractors, i.e. words with a prepositional function at the beginning of IAPs, and the length of AIPs impact this type of non-standard comma usage. Based on the comparison of a sample of 6,000 hits in the Janes corpus of non-standard UGC and the KAS corpus of academic texts, it was determined that non-standard comma is considerably more widespread in UGC than in academic discourse, where it is more frequent in undergraduate than postgraduate theses. This type of comma most often occurs after short IAPs, its prevalence declines with the increasing distance between the attractor and the comma, and among the most widespread attractors, regardless of the discourse type, are Slovene prepositions *zaradi* (because), *za* (for), *po* (after), *na* (on), *kljub* (despite), *poleg* (by), and *ob* (by).

## 1. Uvod in namen raziskave

Nestandardna stava vejice, tj. stava v nasprotju s trenutno kodifikacijo, je vedno aktualna tema, saj vejica velja za eno zahtevnejših prvin slovenskega standardnega jezika, »o stavi katere so si bili kritiki najbolj in najpogosteje navzkriž« (Dobrovoljc, 2004: 188). Napake pri stavi vejice se pojavljajo celo v besedilih jezikovnih uporabnikov s formalno jezikoslovno izobrazbo, kar je pokazala raziskava stave vejice v lektoriranih avtorskih besedilih in prevodih v korpusu Lektor (Popič, 2014), kjer je vstavljena vejica najpogostejši lektorski popravek. Poleg tega sta Fišer in Popič (2015) analizirala stavo vejic v uravnoteženem korpusu Kres (Logar et al., 2012) s standardnimi in lektoriranimi besedili ter v korpusu nestandardne spletne slovenščine Janes (Fišer et al., 2014). S primerjavo stave vejic po »predložnih zvezah«, <sup>1</sup> uvedenih zlasti s predlogi *kljub*, *zaradi* in *glede*, sta ugotovila, da je vejica za tovrstnimi zvezami pogosteje stavljena v Janesu kot v Kresu.

To raziskavo je dopolnila pilotna študija o nestandardni stavi vejice v slovenskih tvitih različnih stopenj standardnosti (Popič et al., 2016). Avtorji študije so želeli analizirati, na kakšen način raba vejice v uporabniških spletnih vsebinah odstopa od norme in kako

nanjo vpliva formalnost komunikacije. Rezultati analize so pokazali, da odvečna vejica predstavlja zelo majhen delež primerov nestandardne stave vejice in se največkrat pojavlja v besednih zvezah (npr. v sestavljenih veznikih), v vezalnem ali ločnem priredju, za stavčnim členom in med enakovrednima odvisnikoma.

Odločili smo se, da bomo dozdašnje raziskave nadgradili s primerjavo stave nestandardne »odvečne« vejice po uvajalnih prislovnih zvezah v akademskem diskurzu, in sicer v diplomskih in doktorskih delih, ter diskurzu uporabniško generiranih spletnih vsebin, pri čemer se bomo osredotočili na poslovne in zasebne bloge. Proučevanje »odvečne« vejice je zelo zanimivo, saj tovrstna vejica ne glede na standardnost konteksta opozarja na pomanjkljivo pravopisno znanje piscev besedil – iz psihološkega vidika namreč stava ločil zahteva več navora kot njihovo opuščanje (ibid.: 152).

Cilj pričujočega prispevka je ugotoviti, kakšna je razširjenost nestandardne stave vejice po uvajalnih prislovnih zvezah, ter kako t. i. atraktorji in dolžina uvajalnih prislovnih zvez vplivata na nestandardno stavo vejice v akademskem diskurzu in diskurzu uporabniško generiranih spletnih vsebin.

### 1.1. Vejica po UPZ v slovenski kodifikaciji

Uvajalne prislovne zveze oz. UPZ (Marko in Osrajnik, 2015: 494) so samostalniške zveze na začetku povedi, ki določajo krajevne, časovne in vzročnostne okoliščine,

<sup>1</sup> V pričujočem prispevku jih imenujemo uvajalne prislovne zveze.

uvedene pa so z besedami v predložni funkciji, ki jih Lengar Verovnik (2003) in Žibert (2006) imenujeta atraktorji: »Opazujoč jezikovno prakso, se zdi, da se vejica pisem še posebej rada prikrađa ob stavčnih členih, uvedenih s t. i. lažnimi atraktorji, kot so predlogi kljub, zaradi in glede« (ibid.: 52) (primer [1]).

[1] *Razen ožjih družinskih članov, nihče ne prihaja v otrokovo bližino* (korpus Janes v0.4, podkorpus blogov s komentarji, objave podjetij na platformi rtvslo.si).

Kot pravi Lengar Verovnik (2003: 51), je razlog za tovrstne »odvečne« vejice najbrž možna pretvorba v odvisnik, ki že po definiciji vsebuje povedek in tako zahteva tudi rabo vejice, delno pa uporabnike najbrž zavede tudi stava vejic glede na stavčno fonetiko. Po koncu uvajalnih prislovnih zvez – zlasti daljših – namreč z glasom premolknemo, jezikovni uporabniki pa tako »začutijo potrebo« po stavi vejice. To potrjujejo izsledki pilotne študije iz leta 2015 (Marko in Osrajnik), ki kažejo, da jezikovni uporabniki zaradi kombinacije nepoznavanja slovnčnih pravil in stavčnofonetičnih kriterijev pogosto stavlajo nestandardno vejico po UPZ. Kljub temu pa je tovrstna stava vejice v Slovenskem pravopisu 2001 le površno omenjena pri določilu o stavi vejic pri polstavčnih konstrukcijah: »Samostalniških zvez, nastalih iz odvisnika, z vejico ne ločimo od okolja« (SP, 2001: § 333) (primer [2]).

[2] *Ob slovesni podelitvi bralnih značk osnovnošolcem je govoril tudi šolski ravnatelj* (SP, 2001: § 333).

## 1.2. Akademski diskurz in diskurz uporabniško generiranih spletnih vsebin

V pričujoči raziskavi smo primerjali nestandardno stavo vejice po UPZ v akademskem diskurzu, in sicer v podkorpusih diplomskih in doktorskih del korpusa akademske slovenščine KAS (Erjavec et al., 2016), ter diskurzu uporabniško generiranih spletnih vsebin, in sicer v podkorpusu blogov korpusa spletne nestandardne slovenščine Janes (Fišer et al., 2014). Za to primerjavo smo se odločili, ker nas zanimajo razlike med besedilnimi zvrstmi, ki se pojavljajo v precej različnih kontekstih. Akademski diskurz je relativno homogen, obravnava ozko zastavljeno, pogosto strokovno temo, in nastaja z namenom informiranja ožjega kroga bralcev, pri čemer si avtorji prizadevajo, da bi bila njihova dela sprejeta v določenem krogu bralcev – pogosto gre za strokovnjake na določenem področju (Suomela-Salmi in Dervin 2009: 120). Akademski diskurz mora upoštevati veljavne standarde in zahteve glede jezikovne standardnosti, obsega pa na primer diplomska in magistrska dela, doktorske disertacije, znanstvene članke itd. Po drugi strani diskurz uporabniško generiranih spletnih vsebin zajema vse vrste vsebin, ki jih proizvedejo neplačani uporabniki na spletu, na primer tvite, bloge, forume, komentarje videoposnetkov itd. Jezik uporabniško generiranih spletnih vsebin (zlasti na družbenih omrežjih) pogosto ne upošteva slovnčnih pravil, vključuje pogovorne in regionalne izraze, velikokrat pa prihaja tudi do zatipkanih in pravopisnih napak. V nasprotju z akademskim diskurzom tovrstna besedila niso vezana strogo določenim in natančno predpisanim formalnim standardom za oblikovanje besedil in načeloma ne obravnavajo akademskih tematik.

### 1.2.1. Korpus Kas

Korpus akademske slovenščine KAS zajema 50.793 besedil oz. 1.189.100.198 pojavnic. Obsega diplomska (81 %) in magistrska dela (13 %), doktorske disertacije (1,4 %), znanstvena dela (1,5 %), kot so prispevki na konferencah in izvorni znanstveni članki, ostala besedila (1,4 %), na primer predgovore, spremna besedila in učna gradiva, specialistična dela (1,1 %) in strokovna dela (0,8 %), kot so strokovni članki in monografije.

Pri naši raziskavi smo uporabili različico korpusa KAS-*proto* in se osredotočili na podkorpus diplomskih nalog (81 % korpusa), ki zajema 42.212 besedil oz. 850.937.549 pojavnic, in podkorpus doktorskih disertacij (1,4 % korpusa), ki vsebuje 700 besedil oz. 52.874.876 pojavnic. Izpustili smo podkorpus magistrskih nalog, saj se nam je zdelo zanimivo primerjati dve skrajnosti – pravopisno kompetenco študentov po koncu prve stopnje univerzitetnega študija in na koncu njihove univerzitetne izobrazbe.

### 1.2.2. Korpus Janes

Korpus spletne slovenščine Janes 0.4<sup>2</sup> obsega skupno 9.055.251 besedil oz. 208.261.725 pojavnic in vključuje tvite (83 %), forumska sporočila (9 %), blogovske zapise (4 %), komentarje na spletne novice (3 %) in vsebine pogovornih strani na Wikipediji (1 %). Da bi omejili vrstno raznolikost, smo v raziskavo zajeli le podkorpus blogov (4 % korpusa), ki vključuje izvorne objave in komentarje zasebnih uporabnikov (*private*) in podjetij (*corporate*) z domen *publishwall.si* in *rtvslo.si*. Pri tem smo v raziskavo zajeli le izvorne objave, saj obsegajo v povprečju skoraj desetkrat več besed na besedilo kot pa komentarji. Izvirne objave v podkorpusu blogov obsegajo štiri kategorije in vključujejo 42.030 besedil oz. 18.256.749 pojavnic. Največji delež so prispevali zasebni uporabniki z domene *rtvslo.si* (50,3 %), sledijo pa jim podjetja z domene *publishwall.si* (24,4 %), zasebni uporabniki z domene *publishwall.si* (19,7 %) in nazadnje podjetja z domene *rtvslo.si* (5,6 %).

Na blogovska besedila smo se osredotočili tudi zaradi njihove dolžine – obsegajo namreč daljše število besed na besedilo (v povprečju 71,3) kot drugi zajeti žanri, zato so najbolj primerljiva z besedili v korpusu Kas. Poleg tega so se nam blogi zdeli zanimivi, saj jih Crystal (2011: 20) umešča med dve skrajnosti spletnega oz. internetnega jezika, ki predstavlja kombinacijo govornega in zapisanega jezika. Crystal v eno skrajnost umešča vrste internetnega jezika, ki se ne razlikujejo bistveno od tradicionalnih besedil. Pri znanstvenih besedilih, kamor prištevamo diplomske naloge in doktorske disertacije, sta digitalna in tiskana oblika celo identični. V drugo skrajnost pa uvršča spletne klepetalnice in podobne platforme za neposredno sporočanje (na primer Facebook ali Twitter), kjer je besedilo sicer zapisano, vendar vsebuje določene ključne lastnosti govornega jezika (ibid.), kot so pričakovanje takojšnjega odziva na besedilo, nestalnost besedil (saj se lahko izbrišejo) in sproščeno vzdušje, značilno za pogovor. Obenem glede na jezikovne značilnosti spletnih besedil loči besedila z

<sup>2</sup> Pred objavo tega prispevka je bila objavljena posodobljena različica korpusa 1.0, ker pa so bile vse ročne analize že zaključene, smo v okviru te raziskave ostali pri različici 0.4.

enako zgradbo in grafično oblikovanostjo kot tradicionalni tisk (na primer doktorske disertacije) ter sporočila z omejenim številom znakov in posledično preprostejšo stavčno strukturo (na primer tviti). Med ti skrajnosti umešča bloge, kjer je uporaba različno zapletene zgradbe besedila odvisna od posameznega jezikovnega uporabnika.

## 2. Zasnova raziskave

Za celostno proučevanje pojava nestandardne vejice po UPZ je treba poznati število vseh povedi z UPZ (z nestandardno stavbo vejice oz. z »odvečno« vejico in s standardno stavbo brez vejice v skladu s trenutno kodifikacijo) v posameznem podkorpusu. Tovrstne povedi smo v posameznih podkorpusih izluščili z iskalnim ukazom CQL (Corpus Query Language) v konkordančniku Sketch Engine (Kilgariff et al., 2004), nato pa pregledali izluščene zadetke v obliki besednih nizov oz. konkordanc.

### 2.1. Iskanje zadetkov v podkorpusih

Ker zaradi dolžine in zapletenosti ni bilo mogoče oblikovati enotnega iskanega ukaza, s katerim bi hkrati izluščili zadetke z vejico in brez nje, ob tem pa ne bi vseboval tudi velikega števila nerelevantnih zadetkov, smo skupno število povedi z UPZ in nestandardno stavbo vejice poiskali z dvema ločenima iskalnima ukazoma. S prvim iskalnim ukazom smo zajeli vse povedi z nestandardno stavbo vejice po UPZ z dolžino do vključno 7 besed po atraktorju, z drugim iskanjem pa povedi z nestandardno stavbo vejice po UPZ z dolžino 8 besed in več po atraktorju. Nato smo sešteli število tovrstnih povedi iz obeh iskanj in tako dobili skupno število vseh zadetkov z nestandardno stavljeno vejico ne glede na dolžino UPZ v posameznem podkorpusu. Pri tem smo drugi iskalni ukaz zastavili tako, da smo z njim obenem zajeli tudi povedi z UPZ in standardno stavbo vejice.

#### 2.1.1. Prvo iskanje

S prvim iskalnim ukazom smo torej v podkorpusih izluščili vse zadetke z UPZ dolžine od 2 do vključno 7 besed po atraktorju, ki jim sledi nestandardno stavljena vejica (v nadaljevanju »zadetki z vejico znotraj 2–7 besed po atraktorju«). V ukazu smo določili, da najkrajša izluščena UPZ vsebuje vsaj dve besedi po atraktorju, saj smo v praksi opazili, da tovrstnim zvezam z le eno besedo po atraktorju večinoma sledi odvisnik (npr. *Zaradi tega, ker*), takšni primeri pa niso relevantni za našo raziskavo. Omejitev na sedem besed pa je poljubna, saj smo UPZ z osmimi besedami in več zajeli v drugem iskanju.

```
[word="[:upper:].+" & tag="D.*"] [word!="," & tag!="G.*" & word="*[:lower:].*"] [...] {2,7} [word="," ] [tag!="V.*" & word!="zakaj | naj | kar | pa | ki | kjer" ] within <s/>
```

Slika 1: Iskalni izraz za prvo iskanje

Primer [3] zadetka z UPZ s petimi besedami po atraktorju, ki ji sledi nestandardno stavljena vejica:

[3] *Za iskrico dvoma o medsebojnem sporazumevanju, ni ostalo več prostora* (korpus Janes v0.4, podkorpus blogov s komentarji, objave zasebnih uporabnikov na platformi publishwall.si).

#### 2.1.2. Drugo iskanje

Z drugim iskalnim ukazom smo poiskali vse relevantne povedi, v katerih se morebitna vejica pojavlja šele po 8 besedah in več po atraktorju (v nadaljevanju »zadetki z UPZ brez vejice znotraj 2–7 besed po atraktorju«).

Predvidevali smo, da bomo s tem iskalnim ukazom izluščili dve kategoriji zadetkov. Prva obsega vse zadetke z nestandardno stavbo vejice po UPZ, ki jih nismo izluščili v prvem iskanju – torej zadetke z UPZ dolžine 8 besed in več, ki jim sledi nestandardno stavljena vejica. Število tovrstnih zajetih zadetkov smo prišteli k relevantnim zadetkom, zajetim s prvim iskanjem, in tako dobili skupno število vseh zadetkov z nestandardno stavljeno vejico ne glede na dolžino UPZ v posameznem podkorpusu.

```
[word="[:upper:].+" & tag="D.*"] [word!="," & tag!="G.*" & word="*[:lower:].*"] [...] {2,7} [word!="," ] [tag!="V.*" & word!="zakaj | naj | kar | pa | ki | kjer" ] within <s/>
```

Slika 2: Iskalni izraz za drugo iskanje

Primer [4] zadetka z UPZ z osmimi besedami, ki ji sledi nestandardno stavljena vejica:

[4] *Ob navideznem izčrpanju vseh zalog novih konstruktivnih političnih idej, ljudstvo pesimistično razpoloženje, nezaupanje v prihodnost in demoralizacijo preganja tako, da si svoje frustracije in ponižanje nacionalnega ponosa zdravi z uspehi športnikov, ki jih sili /.../* (korpus Janes v0.4, podkorpus blogov s komentarji, objave zasebnih uporabnikov na platformi publishwall.si).

Ker ta iskalni ukaz prepoveduje prisotnost vejice znotraj 2–7 besed po atraktorju, smo zajeli še eno kategorijo zadetkov – povedi s standardno stavbo vejice po UPZ. Gre za povedi, skladne s trenutno slovensko kodifikacijo, zato smo pričakovali, da bodo predstavljale precejšen delež zadetkov tega iskanja.

Primer [5] zadetka z UPZ s standardno stavbo vejice:

[5] *Po mnenju za migracije pristojnega grškega ministra Ioanisa Muzalasa se meja z Makedonijo za te migrante ne bo več odprla, poroča nemška tiskovna agencija dpa* (korpus Janes v0.4, podkorpus blogov s komentarji, objave zasebnih uporabnikov na platformi publishwall.si).

### 2.2. Vzorci za analizo

Ker so izbrani podkorpusi preveliki, da bi pregledali vse dobljene zadetke, smo pri vsakem iskanju pregledali vzorec 250 naključnih dedupliciranih zadetkov v vsaki izmed štirih kategorij izvornih objav Janesovega podkorpusa blogov (objave zasebnih uporabnikov in objave podjetij z domene rtvslo.si ter objave zasebnih uporabnikov in objave podjetij z domene publishwall.si – skupaj 1000 zadetkov), 1000 naključnih zadetkov v Kasovem podkorpusu diplomskih nalog in 1000 naključnih zadetkov v Kasovem podkorpusu doktorskih disertacij. V skupnem seštevku obeh iskanj smo torej ročno pregledali 6000 zadetkov.

Omeniti je treba, da smo med objavami podjetij z domene rtvslo.si pri prvem iskanju analizirali le 240, pri drugem pa 187 zadetkov, saj smo jih toliko izluščili z ukazom CQL. Ker je vzorec pregledanih povedi v drugem iskanju manjši, je tudi število identificiranih povedi z nestandardno vejico v drugem iskanju sorazmerno manjše.

### 3. Analiza rezultatov in razprava

Z ročnim pregledom smo v obeh iskanjih izluščili zadetke z UPZ (s standardno in nestandardno stavo vejice). Čeprav smo si prizadevali oblikovali kar najbolj učinkovita iskalna ukaza, smo med zadetke zajeli številne nerelevantne zadetke brez UPZ, na primer naslove poglavij in navajanje virov (primer [6]), ter zadetke s standardno stavljeno vejico po UPZ, pri čemer prislovni zvezi sledi odvisnik, vrinjen stavek ali polstavčna struktura (primer [7]). Pojavljali so se tudi primeri povedi, ki se začnejo z naštevanjem, zato so posamezne naštevalne enote ločene z vejico, kot to narekuje trenutna kodifikacija (primer [8]).

[6] *Iz knjige nasvetov, 09.11.15* (korpus Janes v0.4, podkorpus blogov s komentarji, objave zasebnih uporabnikov na platformi rtslo.si).

[7] *Na podlagi rezultatov, dobljenih pri izvajanju encimske esterifikacije D,L-MK v SC CO2 pri 7,5 MPa in 35 °C, smo v nadaljevanju izvedli encimsko esterifikacijo z razmerjem substratov 1:3,6, brez molekularnih sit in s hitrostjo mešanja 700 obr/min* (korpus KAS-proto, podkorpus doktorskih disertacij).

[8] *Ob današnji strukturi, strokovni usposobljenosti in številu gradbenih inšpektorjev, finančni podhranjenosti gradbene inšpekcije in ob upoštevanju prepovedi zaposlovanja upravičeno dvomimo, da bodo inšpekcijske službe kos tej pomembni nalogi* (korpus Janes v0.4, podkorpus blogov s komentarji, objave podjetij na platformi rtslo.si).

Tovrstne zadetke smo zanemarili, saj niso relevantni za našo raziskavo. V nadaljevanju smo podrobneje analizirali identificirane relevantne primere povedi z nestandardno in standardno stavo vejice po UPZ, pri čemer smo analizo razdelili na tri dele: (1) določanje razširjenosti nestandardne stave vejice po UPZ v posameznih podkorpusih, (2) analiza razdalje (tj. števila besed) med atraktorjem in nestandardno vejico ter (3) analiza atraktorjev, po katerih uporabniki stavljajo nestandardno vejico.

#### 3.1. Razširjenost nestandardne vejice po UPZ

Prvi del analize smo razdelili na dva koraka. V prvem smo z enostavnim sklepnim računom ocenili število vseh zadetkov s standardno in nestandardno stavo vejice po UPZ v posameznem podkorpusu. To pomeni, da smo v

pričujočem prispevku celotno število tovrstnih zadetkov v podkorpusu ocenili po naslednji enačbi:

$$n_{(ne)standardnih} = n_{zadetkov} \cdot \frac{n_{identificiranih}}{n_{pregledanih}}$$

Slika 3: Sklepni račun za ocenitev števila vseh zadetkov s standardno in nestandardno stavo vejice po UPZ v posameznem podkorpusu.

Ob tem:

–  $n_{(ne)standardnih}$  označuje ocenjeno število primerov s standardno (*Ocenjeno št. s standardno stavo vejice* v tabeli 2) oz. nestandardno stavo vejice (*Ocenjeno št. z nestandardno stavo vejice* v tabelah 1 in 2) po UPZ v posameznem podkorpusu,

–  $n_{zadetkov}$  označuje število vseh izluščenih zadetkov v posameznem podkorpusu (npr. 139.576 v Kasovem podkorpusu diplomskih nalog; gl. tabelo 1),

–  $n_{identificiranih}$  označuje število identificiranih primerov z nestandardno stavo vejice po UPZ v vzorcu npregledanih (npr. 400 v Kasovem podkorpusu diplomskih nalog; gl. tabelo 1),

–  $n_{pregledanih}$  označuje število pregledanih zadetkov v posameznem podkorpusu (npr. 1000 v Kasovem podkorpusu diplomskih nalog).

Rezultate izračuna za ocenitev števila identificiranih primerov s standardno in nestandardno vejico po UPZ v podkorpusih predstavljamo v tabelah 1 in 2, pri čemer:

– »KAS Dipl« označuje Kasov podkorpus diplomskih nalog,

– »KAS Dr« označuje Kasov podkorpus doktorskih disertacij,

– »RTV Z« označuje objave zasebnih uporabnikov bloga rtslo.si v korpusu Janes,

– »RTV P« označuje objave podjetij z bloga rtslo.si v korpusu Janes,

– »PUB Z« označuje objave zasebnih uporabnikov bloga publishwall.si v korpusu Janes in

– »PUB P« označuje objave podjetij z bloga publishwall.si v korpusu Janes.

Podkorpus	Št. vseh zadetkov	Št. pregledanih zadetkov	Št. identificiranih primerov z nestandardno stavo vejice (delež glede na št. pregledanih zadetkov)	Ocenjeno št. z nestandardno stavo vejice
KAS Dipl	139.576	1.000	400 (40,0 %)	55.830
KAS Dr	5.423	1.000	254 (25,4 %)	1.377
RTV Z	1.759	250	86 (34,4 %)	605
RTV P	240	240	102 (42,5 %)	102 <sup>3</sup>
PUB Z	844	250	89 (35,6 %)	300
PUB P	753	250	80 (32,0 %)	241

Tabela 1: Rezultati iskanja povedi z vejico znotraj 2–7 besed po atraktorju.

<sup>3</sup> V tem primeru ne gre za ocenitev, temveč dejansko število primerov, saj smo pregledali vse zadetke.

Podkorpuz	Št. vseh zadetkov	Št. pregledanih zadetkov	Št. identificiranih primerov z nestandardno stavo vejice (delež glede na št. pregledanih zadetkov)	Ocenjeno št. z nestandardno stavo vejice	Št. ident. primerov s stand. stavo vejice (delež glede na št. pregledanih zadetkov)	Ocenjeno št. s standardno stavo vejice
KAS Dipl	199.194	1.000	67 (6,7 %)	13.346	778 (77,8 %)	154.972,9
KAS Dr	91.523	1.000	42 (4,2 %)	3.844	836 (83,6 %)	76.513,2
RTV Z	883	250	23 (9,2 %)	81	142 (56,8 %)	501,5
RTV P	187	187	21 (11,2 %)	21	118 (63,1 %)	118,0
PUB Z	440	250	25 (10,0 %)	44	139 (55,6 %)	244,6
PUB P	832	250	18 (7,2 %)	60	169 (67,6 %)	562,4

Tabela 2: Rezultati iskanja povedi brez vejice znotraj 2–7 besed po atraktorju.

V drugem koraku tega dela analize smo na podlagi dobljenih rezultatov ocenitve izračunali še relativni delež nestandardne stave vejice v posameznih podkorpuzih.

$$\text{razširjenost nestand. vejice} = \frac{n_{\text{nestandardnih}}}{n_{\text{nestandardnih}} + n_{\text{standardnih}}}$$

Slika 4: Enačba relativnega deleža za izračun razširjenosti nestandardne stave vejice v podkorpuzih.

Tako smo dobili odstotek povedi z nestandardno stavo vejice po UPZ znotraj vseh izluščenih zadetkov v posameznem podkorpuzu. Rezultati izračuna razširjenosti nestandardne stave vejice po UPZ so predstavljeni v tabeli 3:

	Št. primerov s stand. vejico	Št. primerov z nest. vejico	Razširjenost nest. vejice po UPZ
KAS Dipl	154.972,9	69.176,4	30,9 %
KAS Dr	76.513,2	5.221,4	6,4 %
RTV Z	501,5	686,3	57,8 %
RTV P	118,0	123,0	51,0 %
PUB Z	244,6	344,5	58,5 %
PUB P	562,4	300,9	34,9 %

Tabela 3: Razširjenost nestandardne stave vejice po UPZ v posameznem podkorpuzu.

Kot je razvidno iz rezultatov, se nestandardna vejica najpogosteje pojavlja v objavah zasebnih uporabnikov blogovskih besedil (58,5 % na blogu publishwall.si in skoraj 58 % na blogu rtvslo.si), nekoliko manj je razširjena v objavah podjetij bloga rtvslo.si (51 %), zanimiv pa je precej nižji delež razširjenosti v objavah podjetij na publishwall.si (skoraj 35 %).

Po pričakovanjih se nestandardna stava vejice najredkeje pojavlja v podkorpuzu doktorskih disertacij (nekaj več kot 6 %), presenetljivi pa so rezultati za

podkorpuz diplomskih nalog, v katerih njena razširjenost znaša skoraj 31 %. Na podlagi teh rezultatov lahko torej sklepamo, da študenti dodiplomskega študija nestandardno vejico stavijo precej pogosteje kot študenti podiplomskega študija. Oboji so sicer svoje izobraževanje na področju slovnice in pravopisa končali že v srednji šoli (razen študentov jezikoslovnih smeri), vendar imajo študenti po koncu študija več izkušenj pri pisanju formalnih besedil in morda zato vejico stavijo redkeje, zelo verjetno pa je tudi, da primere nestandardne stave vejice v njihovih besedilih prestežejo mentorji ali lektorji.

Poleg tega lahko sklepamo, da je nestandardna stava vejice po UPZ občutno bolj razširjena v besedilih korpusa Janes kot korpusa Kas. To smo tudi pričakovali, saj morajo pisci besedil v okviru akademskega diskurza strogo upoštevati pravila knjižne slovenščine, medtem ko pisci uporabniško generiranih spletnih vsebin jezikovni standardnosti posvečajo manj pozornosti. V slednjih je pogosta prisotnost odvečne vejice vendarle zanimiva, saj je znano, da uporabniki v uporabniško generiranih besedilih pogosto opuščajo črke in ločila ter uporabljajo okrajšave oz. simbole (Fišer et al., 2018: 125), da prihranijo čas in prostor, zato bi morda prej pričakovali opuščanje sicer obveznih vejic.

### 3.2. Razdalja med atraktorjem in vejico

V drugem delu analize smo proučevali razdaljo (število besed) med atraktorjem in nestandardno vejico v posameznih podkorpuzih. Za pridobitev uravnoteženih rezultatov analize obeh iskanj smo v drugem delu analize (v prvem delu to ni bilo potrebno, saj analiza temelji na deležih) število povedi z nestandardno vejico v večjem vzorcu sorazmerno zmanjšali, in sicer smo ga pomnožili z razmerjem med vzorcema (187/240). Tako smo izračunali število povedi z nestandardno vejico, kot če bi tudi v prvem iskanju pregledali vzorec 187 povedi.

V tabeli 4 predstavljamo rezultate analize stave nestandardne vejice glede na razdaljo med atraktorjem in vejico v posameznih podkorpuzih:

Razdalja med atraktorjem in vejico	Kas Dipl	Kas Dr	RTV P	RTV Z	PUB Z	PUB P	Skupaj (delež)
2	86	46	23,4	32	22	26	235,4 (19,9 %)
3	89	57	22,6	18	21	18	225,6 (19,0 %)
4	97	50	11,7	19	16	14	207,7 (17,5 %)
5	50	40	10,9	10	11	18	139,9 (11,8 %)
6	45	36	4,7	3	6	9	103,7 (8,8 %)
7	33	25	6,2	4	4	4	76,2 (6,4 %)
8	15	13	10	7	5	9	59 (5,0 %)

9	18	10	3	8	3	7	49	(4,1 %)
10	9	3	3	3	4	1	23	(1,9 %)
11	5	6	0	3	2	3	19	(1,6 %)
12	7	4	1	0	2	1	15	(1,3 %)
13	4	2	1	1	0	2	10	(0,8 %)
14	4	0	2	1	1	0	8	(0,7 %)
15	0	2	1	0	0	0	3	(0,3 %)
16	2	0	0	0	0	0	2	(0,2 %)
17	2	1	0	0	0	0	3	(0,3 %)
18	0	0	0	0	0	0	0	(0,0 %)
19	0	0	0	0	0	2	2	(0,2 %)
20	0	1	0	0	0	0	1	(0,1 %)
21	0	0	0	0	1	0	1	(0,1 %)
Skupaj:	467	296	100,5	109	98	114	1.184,5	(0,1%)

Tabela 4: Število primerov glede na razdaljo in podkorpus ter delež vsote primerov posameznih razdalj glede na število vseh primerov.

Na podlagi izsledkov dozdajšnjih raziskav (Marko in Osrajnik, 2015) glede pogostejše nestandardne stave vejice po daljših UPZ zaradi stavčnofonetičnih kriterijev smo pričakovali, da bo nestandardna stava vejice naraščala z večanjem števila besed med atraktorjem in vejico, po določeni razdalji pa začela upadati. Ugotovili pa smo, da v splošnem njena stava z večanjem razdalje precej enakomerno pada. Ob tem smo več kot polovico (skoraj 60 %) primerov identificirali v UPZ z 2–4 besedami med atraktorjem in vejico. Najkrajše UPZ (z 2 besedama med atraktorjem in vejico) predstavljajo kar 20 % vseh identificiranih primerov nestandardne stave vejice, temu tesno sledi stava vejice po 3 (19,2 %) in 4 besedah (17,5 %), šele nato se pojavi nekoliko večji preskok do deleža stavljenih vejic po 5 besedah (11,8 %). Po UPZ z 10 besedami med atraktorjem in vejico pa se v podkorpusih pojavljajo le še posamezni primeri povedi z nestandardno stavo vejice.

Stava nestandardne vejice enakomerno pada tudi po posameznih podkorpusih. Manjše izjeme so Kasov podkorpus diplomskih nalog, kjer se največ primerov z vejico pojavi po UPZ s 4 besedami, podkorpus doktorskih disertacij, kjer je vejic po UPZ s 3 in 4 besedami nekoliko več kot vejic po UPZ z 2 besedama, in zasebne objave v Janesovem podkorpusu blogov publishwall.si, kjer je število primerov z vejico po UPZ s petimi besedami nekoliko večje od števila nestandardnih vejic po UPZ s štirimi besedami.

Preskoki med posameznimi kategorijami Janesovega podkorpusa blogov so nekoliko večji, kar je razumljivo, saj smo raziskavo opravili na manjših vzorcih. Zanimivo je, da se kar v polovici podkorpusov pojavi poved z UPZ, ki vsebuje vsaj 20 besed med atraktorjem in vejico. Primera najdaljše UPZ (dolžine 21 besed) z nestandardno stavo vejice smo identificirali v Kasovem podkorpusu diplomskih nalog (primer [9]) in v objavah podjetij Janesovega podkorpusa blogov publishwall.si (primer [10]):

[9] *Ob tej razdelitvi zahodne Evrope na Evropsko gospodarsko skupnost (EGS) in Evropsko združenje za prosto trgovino (EFTA) konec petdesetih let prejšnjega stoletja, se je Danska skupaj z Združenim kraljestvom, Švedsko, Norveško, Švico, Avstrijo in Portugalsko pridružila zvezi EFTA* (korpus KAS-protokol, podkorpus diplomskih nalog).

[10] *Ob napovedih načrtovanja dveh plinskih terminalov v italijanskem delu Tržaškega zaliva in po začetku predvidenih postopkov za izdajo potrebnih soglasij italijanske vlade, je vlada Republike Slovenije, na pobudo zainteresirane /.../* (korpus Janes v0.4, podkorpus blogov s komentarji, objave podjetij na platformi publishwall.si).

#### 4. Atraktorji

V tretjem delu analize smo podrobneje analizirali atraktorje v zadetkih z nestandardno stavo vejice in primerjali atraktorje v povedih z nestandardno in standardno stavo vejice v korpusih KAS (torej skupno v vseh štirih podkorpusih blogov) in Janes (torej skupno v podkorpusih diplomskih in doktorskih nalog).

##### 4.1. Atraktorji v povedih z nestandardno stavo vejice

Pri podrobni analizi atraktorjev v povedih z nestandardno vejico (iz obeh korpusov) smo ugotovili, da vsi identificirani atraktorji spadajo med predloge, edina izjema je oziralni zaimek *kakršnihkoli*. Identificirane atraktorje smo glede na skladiščno oz. pomensko razmerje, ki ga vzpostavijo v UPZ, razdelili v 16 kategorij: vzročnost (*zaradi, ob, od, vsled*), rezultat (*do*), izbor (*od, med, izmed*), odnosnost (*pod*), predmetnost/opredelitev (*kakršnihkoli, za, pri, ob, med, glede, pod, znotraj, nad*), odsotnost/izvzemanje (*brez, razen*), premikanje/sprememba (*iz, onkraj*), časovnost (*po, za, na, pri, ob, od, pred, med, skozi, do, okrog*), prostorskost/položaj (*na, pred, znotraj, sredi, izza*), stališče/vir (*po, na, iz, skozi*), primerjava/nasprotje/zamenjava (*po, za, na, namesto*), nasprotovanje (*zoper, proti*), namen (*za*), način/sredstvo (*na, med, skozi, preko*), dopustnost (*kljub, navkljub*) in dodajanje/naštevaje/stopnjevanje (*poleg, razen, zraven*).

##### 4.2. Atraktorji v korpusih KAS in Janes

Glede na razširjenost posameznih predlogov v obeh korpusih smo ugotavljali, kateri najpogosteje uvajajo povedi s standardno in nestandardno stavo vejice. Tako smo preverili, ali se v povedih s standardno oz. nestandardno stavo vejice tipično pojavlja določen nabor besed v predložni funkciji – tako bi namreč lahko potrdili,

da gre za močne »atraktorje«, ki govorce »zavajajo« k stavi nestandardne vejice.

Rezultati primerjave so predstavljeni v tabeli 5, zaradi prostorske omejitve pa navajamo le 10 najbolj razširjenih

besed v predložni funkciji, popoln seznam pa je objavljen v Osrajnik (2018).

KAS				Janes			
Nestandardna stava vejice (delež)		Standardna stava vejice (delež)		Nestandardna stava vejice (delež)		Standardna stava vejice (delež)	
zaradi	16,9 %	za	19,6 %	po	20,9 %	po	22,2 %
za	13,4 %	pri	16,3 %	zaradi	11,5 %	na	18,1 %
pri	11,7 %	na	16,0 %	na	10,4 %	za	15,1 %
po	10,9 %	po	10,1 %	za	9,9 %	zaradi	7,7 %
na	9,7 %	zaradi	10,0 %	ob	7,4 %	ob	5,8 %
kljub	9,4 %	ob	4,8 %	kljub	5,0 %	pri	4,0 %
poleg	9,3 %	kljub	4,5 %	poleg	4,7 %	kljub	3,7 %
ob	5,5 %	med	4,3 %	od	4,1 %	med	3,5 %
med	1,8 %	poleg	3,7 %	pri	3,8 %	od	3,3 %
pred	1,8 %	iz	3,4 %	namesto	3,6 %	poleg	2,8 %

Tabela 5: 10 najbolj razširjenih predlogov, ki uvajajo UPZ s standardno in nestandardno stavo vejice, in njihov delež v izbranih podkorpusih korpusov KAS in Janes.

V povedih z UPZ in nestandardno stavo vejice obeh korpusov smo identificirali skupno 34 besed v predložni funkciji, pri čemer se 22 predlogov pojavi v obeh korpusih, 3 se pojavijo le v KAS-u (*znotraj, izmed, vsled*), 9 pa le v Janesu (*zraven, zaradi, nad, proti, onkraj, sredi, kakršnihkoli, okrog, izza, zoper*). V povedih z UPZ in standardno stavo vejice pa se pojavi skupno 31 različnih besed v predložni funkciji, in sicer se jih 21 pojavlja v obeh korpusih, 5 smo jih našli le v KAS-u (*znotraj, izmed, konec, vzdolž, zraven*), 5 pa le v Janesu (*čez, okoli, kraj, onstran, mimo*).

Po primerjavi najpogostejših predlogov v zadetkih s standardno in nestandardno stavo vejice – podrobneje smo analizirali predloge z vsaj 5-odstotno razširjenostjo v korpusu – smo ugotovili, da se v KAS-u tako v povedih s standardno kot z nestandardno stavo vejice najpogosteje pojavljajo atraktorji *za, pri, na, po, zaradi*, v povedih z nestandardno stavo vejice pa tudi predlogi *kljub, poleg* in *ob*. V Janesu pa so tako v povedi s standardno kot z nestandardno stavo vejice med najbolj razširjenimi atraktorji predlogi *po, na, za, zaradi* in *ob*, v povedih z nestandardno stavo vejice pa je precej pogosti tudi predlog *kljub*.

Zavedati se je treba, da so omenjeni predlogi nasploh pogosto rabljeni v slovenščini, saj z njimi izražamo najrazličnejša razmerja in pomene, kot so vzročnost (npr. s predlogom *zaradi*), predmetnost (npr. s predlogom *za*), časovnost (npr. s predlogoma *pri* in *ob*), primerjavo (npr. s predlogom *na*), stališče (npr. s predlogom *po*), dopustnost (s predlogom *kljub*) in naštevanje (npr. s predlogom *poleg*). Ker so že nasploh pogosto rabljeni, je težko trditi, da imajo določeni predlogi vlogo atraktorja, ki uporabnike »zavaja« k stavi nestandardne vejice po UPZ, zlasti kadar so primerljivo razširjeni tako v povedih s standardno kot z nestandardno stavo vejice oz. so v nekaterih primerih celo pogostejši v povedih s standardno kot z nestandardno stavo vejice (npr. predlogi *za, pri, na* in *po*). Lahko pa na osnovi občutno večje razširjenosti določenih predlogov v povedih z nestandardno stavo vejice sklepamo, da je tovrstna stava vejice verjetnejša v povedih z določenimi

predlogi, kot so *zaradi* (skoraj 7 % večja razširjenost v povedih z nestandardno kot s standardno vejico v KAS-u in skoraj 4 % večja razširjenost v povedih z nestandardno kot s standardno vejico v Janesu), *kljub* (skoraj 5 % večja razširjenost v povedih z nestandardno kot s standardno vejico v KAS-u in skoraj 1,5 % večja razširjenost v povedih z nestandardno kot s standardno vejico v Janesu), *poleg* (5,6 % večja razširjenost v povedih z nestandardno kot s standardno vejico v KAS-u in skoraj 2 % večja razširjenost v povedih z nestandardno kot s standardno vejico v Janesu) in *ob* (nekaj manj kot 1 % večja razširjenost v povedih z nestandardno kot s standardno vejico v KAS-u in nekaj več kot 1,5 % večja razširjenost v povedih z nestandardno kot s standardno vejico v Janesu).

## 5. Zaključek

Analiza v pričujočem prispevku je pokazala, da je nestandardna stava vejice po UPZ pogostejša v diskurzu uporabniško generiranih spletnih vsebin kot v akademskem diskurzu. Najbolj razširjena je v blogovskih objavah zasebnih uporabnikov, ki uporabljajo najbolj sproščen jezik, najredkeje se pojavlja v doktorskih disertacijah, opazna pa je tudi precejšnja razlika med razširjenostjo vejice v doktorskih in diplomskih delih. Čeprav je stava vejice v slednjih relativno pogosta, je kljub temu redkejša kot v besedilih korpusa Janes.

Zaradi stavčnofonetičnih kriterijev bi sicer pričakovali, da je nestandardna stava vejice pogostejša po daljših UPZ, vendar smo ugotovili, da njena stava z večanjem razdalje med atraktorjem in nestandardno vejico enakomerno pada tako v splošnem kot tudi po posameznih podkorpusih. Več kot polovica primerov nestandardne stave vejice se pojavi po UPZ z razdaljo 2–4 besed med atraktorjem in vejico, kar 20 % vseh identificiranih primerov pa predstavljajo najkrajše UPZ (z 2 besedama med atraktorjem in vejico). Ugotovili smo tudi, da so najpogostejši atraktorji v obeh proučevanih korpusih predlogi *za, pri, na, po, zaradi, kljub, poleg* in *ob*, pri čemer so atraktorji *zaradi, kljub, poleg* in *ob* bolj razširjeni v povedih z nestandardno stavo vejice.

Rezultati te raziskave so potrdili ugotovitve dosedanjih študij, da imajo uporabniki slovenskega jezika težave pri stavi vejice po UPZ tako v akademskem diskurzu kot v diskurzu uporabniško generiranih spletnih vsebin. S tem smo še dodatno opozorili na potrebo po podrobnejši pojasnitvi tovrstne stave vejice v kodifikacijskih priročnikih, pa tudi pri poučevanju slovnice in pripravi učnih gradiv.

Poleg tega smo v pričujočem prispevku ugotovili, da določene besede v predložni funkciji zares igrajo vlogo atraktorjev, ki jezikovne uporabnike »zavajajo« k nestandardni stavi vejice po UPZ. Za dokončno potrditev obstoja tovrstnih atraktorjev bi bilo treba raziskavo v prihodnosti izvesti na večjem vzorcu povedi, zanimivo pa bi bilo raziskavo razširiti še na magistrske naloge in tako dopolniti proučevanje tovrstne stave vejice v zaključnih pisnih delih vseh treh univerzitetnih študijskih stopenj. Poleg tega bi bilo dragoceno analizirati še druge sorodne besedilne žanre, kot so različna specialistična (tj. strokovna in tehnična) besedila. Glede na rezultate analize, da se nestandardna vejica najpogosteje stavi po krajših UPZ, pa bi v raziskavo lahko zajeli tudi krajša uporabniško generirana besedila, kot so tviti.

### Zahvala

Avtorji se zahvaljujejo anonimnim recenzentom za koristne pripombe. Raziskavo, opisano v prispevku, je podprl projekt ARRS J6-7094 »Slovenska znanstvena besedila: viri in opis«.

### 6. Literatura

- Andrea L. Berez in Stefan Th. Gries. 2009. In defense of corpus-based method: a behavioral profile analysis of polysemous *get* in English. Moran, Steven, Darren Tanner in Michael Scanlon (ur.): *Proceedings of the 24th Northwest Linguistics Conference*. University of Washington Working Papers in Linguistics vol. 27. Seattle, WA, Department of Linguistics, str. 157–166.
- David Crystal. 2011. *Internet Linguistics: A student guide*. London, New York, Routledge.
- Helena Dobrovoljc. 2004. *Pravopisje na Slovenskem*. Ljubljana, Založba ZRC SAZU.
- Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Nataša Logar in Milan Ojsteršek. 2016. Slovenska akademska besedila: prototipni korpus in načrt analiz. V: T. Erjavec (ur.) in D. Fišer (ur.). *Zbornik konference Jezikovne tehnologije in digitalna humanistika*. Ljubljana, Filozofska fakulteta, Univerza v Ljubljani, Slovenija, str. 58–64.
- Darja Fišer, Tomaž Erjavec, Ana Zwitter Vitez in Nikola Ljubešić. 2014. Janes se predstavi: metode, orodja in viri za nestandardno slovenščino. V: T. Erjavec in J. Žganec Gros (ur.). *Language technologies: proceedings of the 17th International Multiconference Information Society – IS 2014*. Ljubljana, Slovenija. Institut »Jožef Stefan«, str. 56–61.
- Darja Fišer in Damjan Popič. 2015. Vejica je mrtva, živela vejica. *Simpozij Obdobja 34*. Ljubljana, Znanstvena založba Filozofske fakultete Univerz v Ljubljani, str. 609–618.
- Darja Fišer, Maja Miličević Petrović in Nikola Ljubešić. 2018. Zapisovalne prakse v spletni slovenščini. Fišer, Darja (ur.): *Viri orodja in metode za analizo spletne slovenščine*. Ljubljana, Znanstvena založba Filozofske fakultete Univerze v Ljubljani, str. 124–139.
- Tina Lengar Verovnik. 2003. Vejica premalo, vejica preveč (2). *Pravna praksa* 22(21): 51.
- Igor Locatelli. 2011. Randomizacija in velikost vzorca. Predstavljeno 9. 12. 2011 na Fakulteti za farmacijo Univerze v Ljubljani.
- Nataša Logar, Miha Grčar, Marko Brakus, Tomaž Erjavec, Špela Arhar Holdt in Simon Krek. 2012. *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. Ljubljana, Trojina, zavod za uporabno slovenistiko.
- Dafne Marko in Eneja Osrajnik. 2015. Slovenska vejica in stavčna fonetika – zakaj ne? *Simpozij Obdobja 34*. Ljubljana, Znanstvena založba Filozofske fakultete Univerze v Ljubljani, str. 493–501.
- Eneja Osrajnik. 2018. *Nestandardna stava vejice po uvajalnih prislovnih zvezah v slovenščini*. Magistrsko delo, Ljubljana, Filozofska fakulteta (v tisku).
- Damjan Popič. 2014. *Korpusnojezikoslovna analiza vplivov na slovenska prevodna besedila* (doktorska disertacija). Ljubljana, Filozofska fakulteta.
- Damjan Popič, Darja Fišer, Katja Zupan in Polona Logar. 2016. Raba vejice v uporabniških spletnih vsebinah. *Proceedings of the Conference on Language Technologies & Digital Humanities, September 29th – October 1st, 2016 Faculty of Arts, University of Ljubljana, Ljubljana, Slovenia*. Ljubljana, Slovenija. Filozofska fakulteta Univerze v Ljubljani, str. 149–153.
- SP 2001 = *Slovenski pravopis* (Jože Toporišič et al.). Ljubljana, Založba ZRC, ZRC SAZU.
- Eija Suomela-Salmi in Fred Dervi. 2009. *Cross-linguistic and crosscultural perspectives on academic discourse*. Amsterdam: John Benjamins Publishing Company.
- Živa Žibert. 2006. *Slovenska vejica: balast ali skladišna nujnost slovenskega knjižnega jezika?* Diplomsko delo, Ljubljana, Fakulteta za družbene vede.