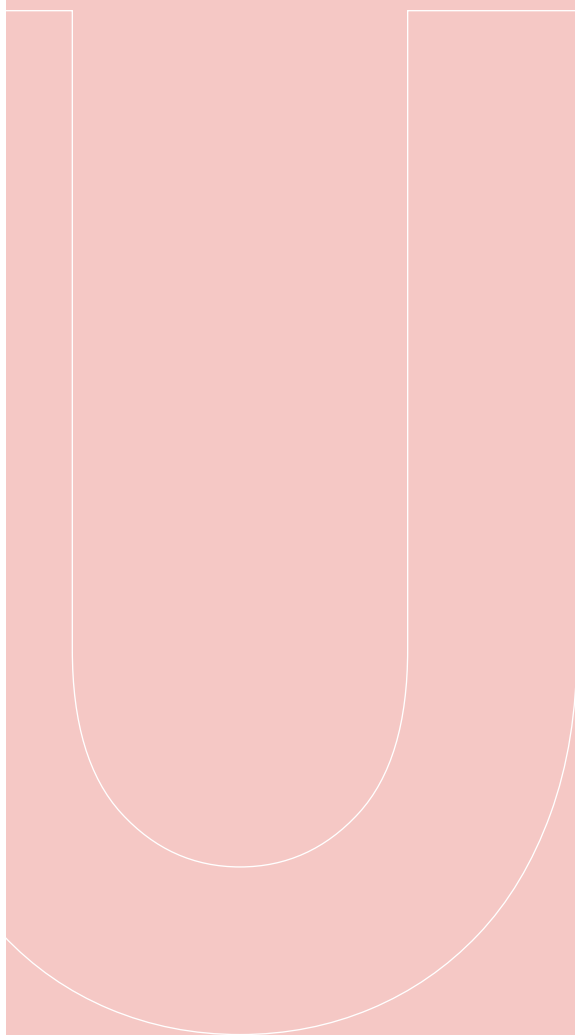


Uvod



Z razmahom svetovnega spleta so skokovito narasle tudi uporabniško generirane vsebine, kot so blogi, forumi in družbena omrežja, ki so v zadnjih dveh desetletjih postale pomemben vir človeškega znanja in mnenj ter katalizator bottom-up komunikacijskih praks, ki prispevajo k demokratizaciji jezika. Posledično narašča tudi potreba po temeljitem, multidisciplinarnem razumevanju tovrstnega sporazumevanja, ki ga pomembno določajo specifične družbene in tehnične okoliščine, v katerih nastaja in zaradi katerih se tudi precej razlikuje od pisnega standarda. Zanj je značilna raba pogovornih in tujejezičnih izrazov, nekanoničnih zapisovalnih praks in skladnje, idiosinkratičnih okrajšav in hiter dotok novega besedišča. Dodatna pomembna značilnost tovrstne visoko participatorne, interaktivne in multimodalne komunikacije so bogati in lahko dostopni (sociodemografski) metapodatki, ki po eni strani odpirajo številne nove priložnosti za raziskovalce, ne samo v korpusnem in računalniškem jezikoslovju, temveč tudi v digitalni humanistiki in družboslovju, po drugi pa raziskovalcem prinašajo nove tehnične, jezikoslovne, pravne in etične izzive.

Medtem ko je bila slovenščina z viri, metodologijo in priročniki za standardni jezik že do sedaj razmeroma dobro opremljena, je do začetka projekta JANES na področju nestandardnega pisnega jezika zevala globoka vrzel, saj nismo imeli ne reprezentativnih korpusov za študije tega segmenta jezika ne orodij za njihovo analizo in obdelavo, prvine nestandardne slovenščine pa so prav tako komajda vključene v jezikovne opise, priročnike in pedagoško prakso. Zato je bil projekt, ki bi zagotovil infrastrukturo in metodologijo za analizo sodobne nestandardne pisne slovenščine, kot se uporablja v novih medijih, v slovenskem okolju že dolgo težko pričakovan.

PROJEKT JANES

V nacionalnem temeljnem raziskovalnem projektu »Viri in metode za raziskovanje nestandardne spletne slovenščine« z neformalnim imenom JANES – Jezikoslovna analiza nestandardne slovenščine (J6–6842), ki ga je od 1. 7. 2014 do 1. 2. 2018 financirala Javna agencija za raziskovalno dejavnost Republike Slovenije, pri njem pa je sodelovalo 12 raziskovalcev in 16 študentov s Filozofske fakultete Univerze v Ljubljani in Instituta »Jožef Stefan«, smo zgradili obsežen korpus internetne slovenščine in na podlagi obsežne jezikoslovne analize razvili metode za izboljšanje avtomatskega procesiranja nestandardne pisne slovenščine. Korpus vsebuje tvite, bloge, forume in komentarje na novice in članke v Wikipediji, torej najpomembnejše zvrsti besedil, ki jih uporabniki ustvarjajo na spletu. Korpus je jezikoslovno označen na več ravneh in prosto dostopen preko zmožljivih konkordančnikov, zato je dobrodošel za teoretično in uporabno jezikoslovje.

Na osnovi korpusa smo izvedli vrsto jezikoslovnih raziskav, in sicer primerjavo internetne slovenščine s pisnim standardom in govorom, študijo žaljivega govora

na družbenih omrežjih, poglobljeni analizi večbesednih zvez in terminologije v internetnih besedilih ter identifikacije pomenskih premikov besed. Izvedene raziskave so s kombinacijo naj sodobnejših korpusnih metod in metod s področja računalniškega jezikoslovja omogočile celovit vpogled v segment jezika, ki se hitro spreminja, dobiva vse pomembnejšo vlogo na vseh področjih našega udejevanja ter je bil doslej iz različnih razlogov zanemarjan.

Rezultat projekta so tudi ročno označni korpusi, s pomočjo katerih smo razvili metode za izboljšanje avtomatskega procesiranja nestandardne pisne slovenščine, ter spletni slovar tviterščine,¹ ki je namenjen učiteljem, učencem, jezikoslovcem, leksikografom in širši zainteresirani javnosti. Vsi razviti jezikovni viri so ponujeni v odprti dostop pod licenco Creative Commons, razvita orodja za označevanje pa vključena v spletne delotoke, razvili pa smo tudi prototipni sistem za kontinuirano gradnjo spremljevalnega korpusa. Vsi razviti viri in orodja so objavljeni in se bodo vzdrževali v okviru slovenske raziskovalne infrastrukture CLARIN.²

ODMEVNOST PROJEKTA JANES

Poleg pričujoče monografije smo v okviru projekta objavili 1 tujo monografijo, 1 tematsko številko znanstvene revije, 6 prispevkov v znanstvenih revijah, 8 poglavij v monografskih publikacijah in 50 prispevkov na mednarodnih in domačih znanstvenih konferencah. Vse publikacije so zbrane na projektni strani: <http://nl.ijs.si/janes/publikacije/>.

Na temo projekta smo izvedli 9 vabljenih predavanj na mednarodnih konferencah, poletnih šolah (Nemčija, Italija, Hrvaška) in tujih univerzah (Švica, Hrvaška, Srbija, Italija, Češka) ter opravili 2 enomesečna znanstvena obiska v okviru akcije COST EnELL na Češkem in 1 enomesečni raziskovalni obisk v Italiji. Rezultati projekta so bili nagrajeni s 1. mestom na mednarodni deljeni nalogi CLIN2017 pri normalizaciji nestandardnih besedil (Sang et al. 2017). V okviru projekta sta nastali 2 magistrski nalogi (Reher 2017 in Škrjanec 2017).

Oganizirali smo 2 poletna tabora spletne slovenščine za dijake, 1 poletno šolo za študente, 1 seminar in 1 mednarodni seminar empiričnega jezikoslovja za jezikoslovce, 1 domačo in 4 mednarodne znanstvene delavnice ter 1 domačo, 1 mednarodno znanstveno konferenco in 1 okroglo mizo. Vsi dogodki z gradivi so dostopni na projektni spletni strani: <http://nl.ijs.si/janes/dogodki/>.

Med njimi posebno mesto zaseda serija izobraževanj JANES Ekspres, ki smo jo s finančno podporo ARRS za promocijo slovenske znanosti v tujini izvedli v Lju-

¹ <http://nl.ijs.si/janes/viri/slovarcek-tvitterscine/>

² <http://www.clarin.si/>

bljani, Zagrebu in Beogradu. Na izobraževanjih smo domačim in tujim kolegom predstavili smernice za ročno označevanje učnih korpusov za nestandardni jezik ter platformo za označevanje WebAnno. Rezultat teh izobraževanj so 3 ročno označeni zlati standardi za označevanje, lematizacijo in normalizacijo nestandardnega jezika za slovenščino, hrvaščino in srbsščino. S tem smo metodo, razvito v okviru projekta, preizkusili in prenesli na dva druga jezika, ki smo ju obogatili s pomembnimi novimi jezikovnimi viri.

Projekt je bil širši javnosti predstavljen v 10 radijskih oddajah, 2 televizijskih in 2 časopisnih intervjujih ter v 1 prispevku v reviji. Odzivi v medijih so zbrani na projektni strani: <http://nl.ijs.si/janes/publikacije/odzivi-v-medijih/>.

POMEN PROJEKTA JANES

Projekt JANES je postavil pomemben mejnik v slovenskem teoretičnem, uporabnem in računalniškem jezikoslovju, saj obsežnih raziskav na področju razvoja in uporabe jezikovnih virov in orodij, prilagojenih za nestandardno pisno slovenščino, do sedaj praktično ni bilo. Rezultati projekta JANES pomembno prispevajo k boljšemu poznavanju značilnosti nestandardne internetne slovenščine, ki z razmahom informacijskih tehnologij postaja vse pomembnejši in pogost način pisne komunikacije. Vrsta raziskav, ki smo jih izvedli v projektu, je kontrastivno, tako kvalitativno kot kvantitativno, osvetlila različne ravni takega jezika, kar je uporabno v nadaljnjih raziskavah in aplikacijah na področju leksikologije in sociolingvistike ter usvajanja in poučevanja jezika. Poleg tega smo v projektu intenzivno uporabljali in razvijali inštrumentarij korpusnega jezikoslovja (korpusnoprimerjalne študije, metode za luščenje kolokacij, pomenskih premikov itd.), kar bo koristilo vsem nadaljnjim korpusnim raziskavam slovenskega jezika.

Projekt je zagotovil tudi razvoj računalniških metod za identifikacijo in luščenje ciljnih besedil s spleta, normalizacijo, oblikoskladenjsko označevanje in lematizacijo nestandardnega jezika ter profiliranje avtorjev in identifikacijo pomenskih premikov, ki so pomembni ne samo za slovenščino, temveč so izrazito aktualne tudi mednarodno, saj se s podobnimi izzivi spopadajo tudi drugi jeziki.

Projekt je omogočil izdelavo kvalitetnih virov nestandardnega jezika, kar je najbolj oprijemljiv rezultat projekta in vključuje obsežen, sodoben in jezikoslovno označen korpus, serijo ročno označenih učnih množic oz. zlatih standardov za jezikoslovne in jezikovnotehnološke raziskave in razvoj ter leksikalno bazo oz. spletni slovarček nestandardne slovenščine. Vsi ti izdelani viri prvič omogočajo celovit vpogled v značilnosti in razvoj spletnega jezika ter odpirajo možnosti za številne nove raziskave s področja jezikoslovja in jezikovnih tehnologij za

slovenščino, in to ne samo projektnim partnerjem, temveč vsem slovenskim in tujim raziskovalcem.

V projektu smo veliko pozornosti posvetili implementaciji in promociji dobrih znanstvenih praks, ki jih v slovenskem prostoru, posebej na področju jezikoslovja, še močno primanjkuje. Projekt je za razliko od introspektivnega jezikoslovja ali jezikoslovnih analiz na napaberkovanih primerih uporabe promoviral empirično podprto korpusno jezikoslovje, jeziko(slo)vne podatke strukturiral po mednarodnih priporočilih in ne v nedokumentiranih ad-hoc formatih, najbolj pomembno pa je, da so viri, izdelani v okviru projekta, dostopni pod licenco Creative Commons, kar omogoča odprto diseminacijo izdelanih podatkov in s tem njihovo maksimalno izkoriščenost, saj mdr. omogoča preverljivost rezultatov, preprečuje dvojno financiranje raziskav in spodbuja gospodarski napredek. S tem smo bistveno pripomogli k preseganju prevladujočega stanja v Sloveniji, kot ga opisujejo Štebe in dr. (2013): »Kljub nekaterim zametkom [omogočanja dostopa do raziskovalnih podatkov] je to področje kritično podhranjeno zaradi prevladujoče kulture zapiranja in monopoliziranja podatkov.«

Projekt je bil tudi izrazito interdisciplinaren, saj je združeval znanstvenike s humanistične in naravoslovne institucije s področja jezikoslovja in računalništva. S tem je projekt prispeval k povezovanju dveh tradicionalno ločenih strok, ki pa sta obe potrebni za razvoj sodobnega empirično osnovanega jezikoslovja in jezikovnih tehnologij. Glede na to, da je metodologija izgradnje virov in orodij jezikovno neodvisna, so pristopi uporabni, in tudi že bili uporabljeni, za sorodne jezike (hrvaščina, srbsščina), ki tovrstnih virov in orodij še nimajo, kar rezultatom daje pomembno večjezično razsežnost.

Razviti viri, orodja in metode bodo omogočili prenos znanj na vsa področja, ki uporabljajo spletne vsebine, ki jih ustvarjajo uporabniki. Zaradi odprtosti in zapisa leksikalne baze je ta tako avtorskoppravno kot tehnično primerna za vključitev v druge leksikalne vire, kot npr. v načrtovani novi slovar sodobnega slovenskega jezika, s čimer bo integrirana v širši slovarski projekt, od katerega bodo imeli korist vsi govorci slovenskega jezika. Predlagani projekt je ponudil tako kontrastivne raziskave kot konkretne in dostopne vire nestandardne pisne slovenščine, ki jih je mogoče uporabiti pri pouku oz. načrtovanju pouka slovenščine na osnovni, srednješolski in univerzitetni ravni, pa tudi pri učenju slovenščine kot tujega jezika.

Glede na to, da razviti viri in orodja niso omejeni le na nekomercialno rabo, bodo neposredno uporabni tudi za slovenska podjetja, ki se ukvarjajo z informacijsko-komunikacijskimi tehnologijami, npr. v povezavi s semantičnim spletom, poizvedovanjem po informacijah, rudarjenjem po besedilih ali povzemanjem besedil, in ki bodo vedno pogosteje v svoje produkte želela vključevati

tudi obdelavo slovenskega jezika, pri čemer se bodo morala spopadati tudi ali celo predvsem z nestandardnim jezikom, saj količina takih besedil bliskovito narašča.

Ob neposredni koristi za podjetja, ki bodo lahko bolje procesirala slovenski jezik, bodo sledile tudi koristi za družbo, saj bo govorcem slovenščine omogočen dostop do produktov, ki bolje podpirajo slovenski jezik. S tem bomo pomagali zmanjšati e-izključenost govorcev, ki so pogosto priklenjeni na tujejezične aplikacije, slovenščini pa omogočili boljše funkcionalnost in razvoj v digitalni dobi, kar je tudi izpostavljeno v Resoluciji o nacionalnem programu za jezikovno politiko 2014–2018.

VSEBINA MONOGRAFIJE

Monografija *Viri, orodja in metode za analizo spletne slovenščine* v štirinajstih poglavjih združuje vse pomembnejše raziskave, ki so se zvrstile v projektu JANES, tem pa se s prispevki pridružujejo tudi nekateri domači in tuji kolegi, ki se pri svojem delu posvečajo tej raziskovalni tematiki, a pri projektu formalno niso sodelovali. Poglavja smo organizirali v dva večja sklopa: v prvem delu monografije predstavimo izdelavo virov in orodij, v drugem pa nanizamo metodološko heterogene raziskave posameznih vidikov spletne slovenščine, ki so bile na razvitih virih opravljene na različnih ravneh jezikovnega opisa.

1. del: Viri in orodja za analizo spletne slovenščine

V prvem poglavju **Tomaž Erjavec**, **Nikola Ljubešić** in **Darja Fišer** predstavijo korpus slovenskih spletnih uporabniških vsebin Janes, ki vsebuje tvite, spletne forume, novice in uporabniške komentarje nanje, uporabniške in pogovorne strani na Wikipediji ter blogovske zapise in komentarje nanje. Opišejo postopek zajema in obdelave besedil za vsakega od vključenih virov, predstavijo avtomatske in ročne postopke za obogatitev korpusa z dragocenimi metapodatki ter oblikovanje in objavo korpusa, prav tako pa podajo tudi kvantitativno analizo zgrajenega korpusa, ki nudi vpogled v naravo zajetega gradiva.

V drugem poglavju **Jaka Čibej**, **Špela Arhar**, **Tomaž Erjavec** in **Darja Fišer** predstavijo družino ročno označenih korpusov, ki so bili izdelani za potrebe učenja jezikovnotehnoloških orodij za izboljšanje stavčne segmentacije, tokenizacije, normalizacije, oblikoskladenjskega označevanja in lematizacije, prav tako pa tudi za natančnejše proučevanje jezikovnih pojavov v slovenski računalniško posredovani komunikaciji, kot so skladenjske značilnosti, pojave krajšanja, raba vejice, preklapljanje med jeziki in regionalnih prvin.

V tretjem poglavju **Nikola Ljubešić, Tomaž Erjavec** in **Darja Fišer** predstavijo razvoj orodij za avtomatsko procesiranje nestandardne slovenščine, s katerimi merijo stopnjo standardnosti besedil, izboljšujejo stavčno segmentacijo, izvajajo normalizacijo nestandardno zapisanih besed in oblikoskladenjsko označevanje ter v besedilih razpoznavajo imenske entitete. Z vrednotenjem rezultatov, dobljenih z razvitimi orodji, pokažejo znatno povečanje kvalitete procesiranja nestandardnega jezika, prav tako pa tudi nakažejo možnosti prenosa razvitih orodij na druge jezike.

V četrtem poglavju **Matej Martinc, Senja Pollak** in **Ana Zwitter Vitez** predstavijo vgradnjo razvitih orodij za procesiranje spletnih vsebin v okolje ClowdFlows, ki omogoča lažji in hitrejši zajem, obdelavo in analizo lastnih korpusov tudi brez programerskega znanja. Delotok preizkusijo na leksikalni, besednovrstni in skladenjski analizi pozitivnih in negativnih komentarjev festivala Evrovizija.

2. del: Raziskave spletne slovenščine

V petem poglavju **Darja Fišer, Maja Miličević** in **Nikola Ljubešić** obravnavajo značilnosti nestandardnega zapisa besed v slovenskih tvitih z opazovanjem pretvorb iz standardnega v nestandardni zapis besed treh različnih vrst transformacij: izpust, dodajanje in zamenjavo črk. Največ transformacij identificirajo med polnopomenskimi besedami, najpogostejše pa so tiste, do katerih prihaja pri slovničnih besedah. Najpogostejši so izpusti, predvsem samoglasnikov, do česar največkrat prihaja na koncu besed, s čimer se neformalna komunikacija v tvitih približuje govoru.

V šestem poglavju **Damjan Popič** in **Darja Fišer** proučita rabe vejice v slovenskih tvitih, kjer ju zanima, v kolikšni meri je stava v tovrstnih vsebinah rabljena v skladu z jezikovnim standardom in na katerih mestih imajo uporabniki s stavo vejice največ težav. Za opis rabe vejice razvijeta tipologijo ter smernice za označevanje, ki ju preizkusita z pomočjo dveh označevalcev. Ugotovita, da tudi v nestandardni računalniško posredovani komunikaciji raba standardne vejice prevladuje. Pri veliki večini nestandardne rabe vejice gre za izpuste, ki ne izkazujejo posebnosti medija, identificirata pa tudi namerne izpuste vejice, ki so rezultat tehničnih in družbenih okoliščin tovrstnega komuniciranja.

V sedmem poglavju **Jaka Čibej** predstavi analizo regionalnih jezikovnih različic v spletni slovenščini, za kar v skladu z razvito tipologijo ročno označi tvite uporabnikov iz različnih slovenskih regij. Višjo zastopanost nestandardnih prvih zazna pri uporabnikih iz gorenjske in dolenjske regije, te pa so najpogostejše in najbolj sistematično izkazane v omejenem naboru zaprtih besednih vrst, predvsem v

obliki izpustov samoglasnikov. Nestandardnega besedišča je bilo v označenem vzorcu za tretjino, nestandardnih oblikoslovnih prvin pa zelo malo.

V osmem poglavju **Darja Fišer** in **Nikola Ljubešić** opišeta pristop za polavtomatsko prepoznavanje pomenskih premikov, ki temelji na primerjavi semantičnih profilov besed v korpusu tvitov in v referenčnem korpusu. Za vrednotenje rezultatov razvijeta tipologijo pomenskih premikov, pri čemer kandidate preverita s pomočjo ročne, korpusno podprte leksikografske analize. Na tak način identificirata največ pomenskih premikov, ki se v tvitih pojavijo zaradi dnevnih dogodkov in neformalnih sporočanjških okoliščin.

V devetem poglavju **Špela Arhar Holdt** analizira skladenjske značilnosti računalniško posredovane slovenščine na ravni besednega reda, ki jo izvede s pomočjo skladenjsko ročno označenega vzorca z odvisnostnim sistemom JOS, prilagojenim specifikam računalniško posredovane slovenščine, kot so novomedijski komunikacijski elementi (emojiji, sklici na uporabniška imena, heštegji), tujejezične prvine, jezikovni fragmenti in nestandardna raba ločil. S kategorizacijo označenih besednorednih značilnosti osvetli doslej še neraziskan segment skladnje računalniško posredovane komunikacije in nakaže možnosti nadaljnjih raziskav besednega reda v slovenščini, s primerjavo ujemanja oznak različnih označevalcev pa izpostavi tudi zelo različno zaznavo (ne)zaznamovanosti in (ne)standaradnosti besednega reda pri govorcih.

V desetem poglavju **Ana Zwitter Vitez** in **Darja Fišer** s triangulacijo korpusov Janes, Gos in Kres preverita prisotnost prvin govornega jezika v računalniško posredovani komunikaciji na ravni ključnih besednih oblik in ključnih besed. Ugotovita, da so na besednovrstni ravni besedila računalniško posredovane komunikacije bistveno bližje pisnim besedilom kot govornemu diskurzu. Na ravni besedišča so posebej zanimivi za govor specifični elementi interakcije z drugimi udeleženci, ki so najpogosteje prisotni v tvitih in komentarjih.

V enajstem poglavju **Mija Michelizza** obravnava rabo ključnikov v slovenskih tvitih, ki so primarno namenjeni kategoriziranju, lahko pa opravljajo tudi različne komunikacijske vloge. Ugotavlja, da čeprav ključniki predstavljajo novejši jezikovni element, ki v računalniško posredovani komunikaciji izstopa, ti večinoma ostajajo znotraj svoje primarne kategorizacijske vloge, tisti, ki opravljajo komunikacijske vloge, pa so pogosteje skladenjsko vpeti v besedilo. Med identificiranimi fenomeni izstopajo ključniki, ki izkazujejo povezanost objavljanja tvitov s spremljanjem drugih medijev.

V dvanajstem poglavju **Špela Reher** in **Darja Fišer** proučita preklapljanje med jeziki v slovenskih tvitih. Analiza temelji na ročno označenem vzorcu tvitov, za kar tudi razvijeta lastno označevalno shemo. Ugotovita, da preklapljanje ni redek pojav, da med jeziki preklpov izrazito prevladuje angleščina, da do preklpov

pogosteje prihaja znotraj stavkov, da so preklopi v enaki meri eno- in večbesedni in da vključujejo tudi slovnične besedne vrste. Kodno preklapljanje opravlja številne diskurzivne funkcije, kot so referenčna, ekspresivna in poudarjalna, glede na semantično polje pa so preklopi pogosto povezani s popularno kulturo, zlasti TV-oddajami, športom, hrano in Twitterjem.

V trinajstem poglavju **Helena Dobrovoljc** obravnava spremembe pri pozdravljanju, poslavljanju in naslavljanju v pisemskem sporazumevanju v digitalnem okolju. Osvetli spontane in naučene spremembe pisemskega diskurza, ki prodirajo v vse tipe elektronskega pisemskega sporazumevanja in prikazujejo, kako so se v različnih obdobjih spreminjala družbena razmerja in piščev odnos do pisemskega sporazumevanja.

V štirinajstem poglavju **Iza Škrjanec**, **Nada Lavrač** in **Senja Pollak** predstavijo avtomatsko razpoznavo spola avtorjev in avtoric slovenskih blogovskih zapisov, za kar preizkusijo model z ročno zgrajenimi pravili in model, zgrajen z metodami strojnega učenja. Kot najuspešnejši se izkaže model strojnega učenja, ki je naučen na unigramih pojavnic s pomočjo metode podpornih vektorjev. Analiza najbolj informativnih značilnik tega modela pokaže, da se besedila blogerk in blogerjev razlikujejo po uporabljenih slovničnih in slogovnih sredstvih ter glede na tematiko.

ZAHVALA

Raziskave, opisane v monografiji, so bile opravljene v okviru nacionalnega temeljnega projekta »Viri, orodja in metode za raziskovanje nestandardne spletne slovenščine« (J6-6842, 2014–2017), ki ga financira ARRS.

Urejanje te monografije je bil velik, a slasten zalogaj. Za vse napake in nekonsistentnosti, ki so se prithotapile vse do končne verzije, sem kriva urednica, za kar se opravičujem vsem prizadetim. Monografije pa seveda ne bi bilo brez izjemno predanega dela vseh sodelujočih avtorjev in avtoric, za kar se jim iskreno zahvaljujem: Špela Arhar Holdt, Darja Fišer, Nikola Ljubešič, Maja Miličević Petrović, Jaka Čibej, Helena Dobrovoljc, Tomaž Erjavec, Nada Lavrač, Matej Martinc, Mija Michelizza, Senja Pollak, Damjan Popič, Špela Reher, Iza Škrjanec, Ana Zwitter Vitez.

Posebna zahvala gre tudi recenzentom, ki so prispevke skrbno in natančno pregledali ter s konstruktivnimi pripombami pomembno prispevali k višji znanstveni kvaliteti monografije: Maja Bitenc, Polona Gantar, Vojko Gorjanc, Monika Kalin Golob, Petra Kralj Novak, Simon Krek, Tina Lengar Verovnik,

Nataša Logar, Dunja Mladenić, Marko Robnik Šikonja, Tadeja Rozman, Marko Stabej, Darinka Verdonik, Reinhild Vandekerckhove, Jana Zidar Forte. Za prevode rokopisov, ki so zaradi mednarodne avtorske zasedbe nastali v angleščini, se zahvaljujem Dafne Marko in Lei Anžur. Za neumorno lektorsko in korektorsko prečesavanje monografije sem neizmerno hvaležna Damjanu Popiču, za profesionalno oblikovanje in neusahljiv vir dobre volje pa Juretu Preglauu.

Darja Fišer

Ljubljana, 30. marec 2018

Literatura

- Tjong Kim Sang, Erik, Marcel Bollmann, Remko Boschker, Francisco Casacuberta, Feike Dietz, Stefanie Dipper, Miguel Domingo, Rob van der Goot, Marjo van Koppen, Nikola Ljubešić, Robert Östling, Florian Petran, Eva Petersson, Yves Scherrer, Marijn Schraagen, Leen Sevens, Jörg Tiedemann, Tom Vanallemeersch in Kalliopi Zervanou, 2017: The CLIN27 Shared Task: Translating Historical Text to Contemporary Language for Improving Automatic Linguistic Annotation. *Computational Linguistics in the Netherlands Journal* 7/1. 53–64.
- Škrjanec, Iza, 2017: *Gender-based analysis of Slovene user-generated content*. Master thesis. Jožef Stefan Postgraduate School.
- Reher, Špela, 2017: *Slovenščina na prepihu: kodno preklapljanje v objavah slovenskih uporabnikov Twitterja*. Magistrsko delo. Filozofska fakulteta Univerze v Ljubljani.
- Resolucija o Nacionalnem programu za jezikovno politiko 2014–2018*. http://www.mk.gov.si/fileadmin/mk.gov.si/pageuploads/Ministrstvo/Zakonodaja/2013/Resolucija_-_sprejeto_besedilo__15.7.2013_.pdf.
- Štebe, Janez, Sonja Bezjak in Sanja Lužar, 2013: *Odprti podatki: načrt za vzpostavitev sistema odprtega dostopa do raziskovalnih podatkov v Sloveniji*. Ljubljana: Fakulteta za družbene vede.