

Terminologija v spletnih forumih

Špela Vintar

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

Povzetek

V prispevku raziskujemo rabo terminologije v treh slovenskih spletnih forumih: avtomobilizem.com, med.over.net in kvarkadabra.net. Z uporabo luščilnika terminologije in ročne validacije terminoloških kandidatov skušamo ugotoviti delež specializiranih enot v posameznih forumih in podforumih, posebej pa se posvetimo nestandardnim elementom pri rabi terminologije. Rezultati kažejo precejšnje razlike v specializiranosti forumov, ta pa ne sovpadajo z ravni standardiziranosti. Prikazemo tudi razlike v rabi terminologije glede na tip foruma (moderiran ali nemoderiran) in raven znanja razpravljalca (specialist ali laik).

Terminology in Internet Forums

The paper explores the use of terminology in three Slovene internet forums: avtomobilizem.com, med.over.net and kvarkadabra.net. We employ automatic term recognition and manual validation of term candidates in an attempt to establish the term density of individual subforums. A set of methods is proposed for the recognition of nonstandard terms. Our results reveal large differences between forums in terms of specialisation level, which does not necessarily correlate to the level of standardisation. Finally we explore some properties of terminology depending on the type of forum (active or passive moderation) and on the level of expertise of the forum user (expert vs. lay).

1 Uvod

Spletni forum je spletni prostor, kjer lahko uporabniki v obliki objavljenih sporočil sodelujejo v različnih razpravah (Wikipedija, 17. 8. 2015). Kot eden od precej razširjenih spletnih žanrov ima forum določene značilnosti (prim. Varga, 2011: 32–34):

- Spletni forum običajno združuje vrsto razprav z določenega področja in tako uporabnike povezuje v virtualno skupnost (npr. mama.si, mojpes.net, hribi.net itd.).
- Spletni forumi imajo drevesno strukturo, pri kateri se vrhnje področje deli na posamezne podforume, znotraj njih pa na teme, sestavljene iz niza objav.
- Razprave v spletnih forumih imajo pretežno značilnosti pisne komunikacije in se tipično vsaj nekaj časa arhivirajo.

Zaradi teh značilnosti so spletni forumi – vsaj tisti, ki obravnavajo določeno področje – oblika specializirane komunikacije in so kot taki zanimivi za terminološko analizo, prek njihove hierarhične strukture pa jih lahko uredimo v bolj ali manj specializirane podkorpuse.

Spletni forum lahko s komunikološkega vidika opredelimo kot interaktivno okolje, kjer poteka množično komuniciranje v razmerju eden z mnogimi in/ali kot skupinsko razmerje mnogih z mnogimi, v njem pa potekata javna in zasebna komunikacija (Petrovčič, 2005: 13). Predvsem zaradi anonimnosti in odsotnosti neposrednega vizuelnega stika lahko uporabnik spletnega foruma komunicira drugače kot pri neposredni, fizični komunikaciji, po drugi strani pa forum ni povsem neregulirano spletno okolje, saj danes praktično vsa forumska spletišča eksplicitno določajo pravila vedenja, za njihovo upoštevanje pa skrbi moderator (Plant, 2004: 60).

Najbolj obiskani forumi pri nas (in podobno v tujini) so mesta za izmenjavo izkušenj in znanja o določeni temi, zato jim v nadaljevanju lahko rečemo strokovni spletni forumi.

Namen prispevka je raziskati značilnosti rabe terminologije v takšnih forumih. Raziskava temelji na šestih tematskih podkorpused s treh slovenskih forumov, in sicer med.over.net, avtomobilizem.com in kvarkadabra.net, ki so zajeti v korpusu Janes v0.3. Osrednja vprašanja raziskave so:

- Kolikšen je delež terminoloških enot v posameznem podkorpusu in ali iz tega lahko sklepamo o ravni specializiranosti posameznega foruma?
- Kakšno je razmerje med standardnim in nestandardnim terminološkim besediščem v posameznih forumih?
- Ali so obstoječe metode samodejnega luščenja terminologije uspešne tudi pri luščenju terminologije iz spletnih forumov?
- Kako se raven standardnosti spletnih besedil kaže pri rabi terminologije?

Pričakujemo, da bodo rezultati raziskave uporabni pri razvoju orodij za računalniško obdelavo spletnih besedil, obenem pa bodo prinesli nova spoznanja o jezikoslovnih značilnostih specializirane komunikacije na spletu.

2 Sorodne raziskave

S sociološko-komunikološkega vidika so spletni forumi zanimivi kot družbena okolja, v katerih se med člani izmenjujejo izkušnje, prepričanja in znanje, ob tem pa se gradi tudi posameznikova pripadnost skupnosti. Kimmerle et al. (2011) tako raziskujejo konstruiranje znanja v spletnem forumu alternativne medicine in ugotavljajo, da pri opazovanem spletnem forumu prepričanja uporabnikov tvorijo holistično ideologijo, ki prerašča okvire dejstev in iz virtualne skupnosti ustvarja parareligiozno bratovščino.

Podrobno se s procesi prenašanja znanja v spletnih forumih ukvarja Varga (2011) v svoji doktorski disertaciji, pri čemer s postopki analize diskurza raziskuje strategije prepričevanja, razlaganja in podajanja navodil v različnih strokovnih spletnih forumih. Skozi analizo pokaže, kako se

identiteta razpravljalcev samodejno vzpostavi skladno z njihovo stopnjo obveščenosti in s skupnim ciljem izmenjati znanje.

Petrovčič (2005) v svojem diplomskem delu raziskuje deliberativnost v razpravljalnih forumih, pri tem pa izhaja iz dveh nasprotujočih si izhodišč. Nekateri avtorji namreč (Pavlik, 1994; citirano po Petrovčič, 2005: 5) zagovarjajo stališče, da »odsotnost normativne strukture in relativna anonimnost posameznikov razpravljalcem ponujata večjo svobodo za javno izražanje argumentiranih mnenj, medtem ko so drugi avtorji [...] nasprotnega mnenja, saj trdijo, da sta vizualna anonimnost razpravljalcev in pomanjkljiva normativna strukturiranost pogovorov glavni vzrok za kratkotrajnost in plitkost stikov med razpravljalci ter podajanje emocionalno nabitih nekritičnih mnenj.« To izhodišče je posredno relevantno tudi za jezikoslovno analizo forumov, saj se uporabnikova virtualna identiteta kaže primarno skozi njegov način izražanja.

Bolj jezikoslovno usmerjenih raziskav forumov je več, a se nobena nam znana raziskava ne usmerja prav v jezikoslovne značilnosti terminološkega besedišča. Montero et al. (2007) denimo ugotavljajo, da visoka pojavnost modalnih glagolov v forumskih besedilih ta približuje govorjenemu diskurzu. Za naš prostor sta relevantni še raziskava Nataše Jakop (2008), ki opazuje (ne)spoštovanje pravopisnih pravil v slovenskih spletnih forumih, čeprav brez kvantitativne analize, in magistrsko delo Nataše Zupančič (2009), ki izvede celovito korpusno analizo jezikovnih značilnosti šestih spletnih forumov.

Zupančič (ibid.) analizira različne ravni diskurza na spletnih forumih in predstavi kvantitativne rezultate v zvezi z ortografskimi posebnostmi, frazeologijo, skladijskimi značilnostmi in pragmatiko, posebej se posveti tudi vprašanju jezikovnega izbiranja na spletnih forumih. Sklepne ugotovitve skladijske in ortografske analize za izbrano gradivo pokažejo, da jezik spletnih forumov sicer v marsičem odstopa od standardne slovenščine, vendar v njem še vedno prevladujejo značilnosti zapisane besede. Za našo raziskavo so izsledki Nataše Zupančič uporabni zgolj posredno, saj je v njej uporabila gradivo z zelo splošnih forumov (diva.si, lunin.net, mobisux.com, mojtrener.com, razprave.com in siol.net), pri izboru objav pa se je omejila na teme, ki se ukvarjajo z resničnostnim šovom Bar.

V omenjeni raziskavi predstavljene značilnosti spletne slovenščine so bile torej pridobljene na podlagi korpusa splošnih, ne pa strokovnih forumov, kar ne pomeni, da v določeni meri ne držijo tudi za naše gradivo. Predvsem se ugotovitve o leksikalnih, skladijskih in frazeoloških značilnostih forumskih besedil nikjer ne dotikajo terminologije, poleg tega je komunikacijski okvir precej drugačen: če je pri strokovnih forumih temeljni namen razpravljalcev izmenjava znanja, je pri splošnih forumih –

še posebej pri razpravah o resničnostnem šovu – v ospredju izmenjava mnenj.

Pri naši raziskavi tako ohranjamo metodologijo korpusnoterminološke analize, ki temelji na samodejnem luščenju terminologije iz oblikoskladenjsko označenih besedil, dodali pa smo nekaj novih metod za raziskovanje nestandardnosti terminologije.

3 Predstavitev gradiva

Vir gradiva, uporabljenega v raziskavi, je korpus spletne slovenščine Janes v0.3 (Fišer et al., 2015), ki vsebuje 29 odstotkov besedil s spletnih forumov, kar je dobrih 47 milijonov pojavnic. V korpusu so zajeti trije spletni forumi, in sicer avtomobilizem.com, med.over.net in kvarkadabra.net, ki se precej razlikujejo po tematskih področjih, številu uporabnikov in objav, skupno pa jim je, da gre pri vseh treh v najširšem smislu za področno specializirane forume. Tako se forum avtomobilizem.com s skoraj 40.000 uporabniki razglša za »največjo avtomobilistično skupnost v Sloveniji«, spletišče med.over.net se je izvorno vzpostavilo kot vrsta moderiranih zdravstvenih posvetovalnic, sčasoma pa so se jim pridružile še manj specializirane razpravljalske skupine, forum kvarkadabra.net pa je prostor za razprave o znanosti, predvsem astronomiji, fiziki, biologiji in filozofiji, katerega skupnost šteje nekaj čez 4 tisoč uporabnikov.

Potencialno pomemben parameter pri različnih (pod)forumih je aktivno oz. pasivno moderiranje. Čeprav imajo vsa tri foruma spletišča moderatorje, ki domnevno bdijo nad uporabniki in njihovim spoštovanjem pravil uporabe foruma, pa so le zdravstvene posvetovalnice na med.over.net forumi z aktivnim moderiranjem, se pravi razprave med nestrokovnjaki in strokovnjakom, običajno zdravnikom specialistom. Tudi med njimi so glede dinamike moderatorja razlike, ki jih lahko ugotovimo zgolj s prebiranjem posameznega podforumu. Pri nekaterih zdravstvenih posvetovalnicah je namreč v naslovu podforumu naveden moderator, ki pa se oglašča redko in razpravo prepušča laičnim uporabnikom, pri drugih pa so nizi objav strukturirani strogo v obliki vprašanj in odgovorov in torej potekajo v obliki dialogov med pacientom in zdravnikom, laični komentarji na »tujo« temo pa niso dovoljeni.

Podkorpus	Število objav	Število pojavnic	Aktivno moderiran
Avtomobilizem.Styling	21.634	866.974	ne
Avtomobilizem.SamSvojMojster	23.030	953.037	ne
Med.Over.Net.Ginekologija	1724	259.991	da
Med.Over.Net.Kontracepcija	5470	403.991	ne
Med.Over.Net.Hujšanje	2410	255.034	ne
Kvarkadabra.Astronomija	1601	107.365	ne

Tabela 1: Osnovni podatki o izbranih podkorpusih.

Za namene naše terminološke raziskave smo iz omenjenih forumskih spletišč izbrali 6 podforumov, ki smo jih iz korpusa Janes izluščili s pomočjo metabesedilnih oznak; pri vseh korpusnih besedilih je namreč v oznaki besedila zapisano ime foruma, podforum, teme in avtorja, kar močno olajša ustvarjanje podkorpusov. S foruma avtomobilizem.com smo tako izbrali podforum Styling o lepotnih posegih na vozilih ter Sam svoj mojster o popravilih, s spletišča med.over.net smo vključili zdravstveno posvetovalnico ABC ginekologije in porodništva, ki jo aktivno moderira zdravnik specialist, forum Kontracepcija, kjer uporabnice in uporabniki o tej temi razpravljajo pretežno brez moderatorja, in forum Hujšanje, ki je prav tako moderiran zgolj pasivno in je po naših predvidevanjih tudi najmanj terminološki. S spletišča Kvarkadabra smo izbrali le podforum O svetu za Luno, kjer so zbrani prispevki s področja astronomije. Tabela 1 predstavlja izbrane podkorpuse z osnovnimi korpusnimi podatki, dodan je še podatek o aktivnem moderiranju.

4 Delež in značilnosti terminologije v podkorpisih

V nadaljevanju smo iz vseh šestih podkorpusov izluščili ključne besede ter eno- in večbesedne terminološke kandidate, pridobljene sezname pa smo ročno pregledali in med prvimi 500 izluščenimi kandidati označili termine. Nato smo se posebej posvetili nestandardnim terminološkim elementom, nazadnje pa nas je zanimala še raba terminologije v moderiranem forumu, kjer lahko primerjamo objave, ki jih je napisal strokovnjak, s tistimi, ki so jih prispevali laični uporabniki.

4.1 Luščenje terminoloških kandidatov

Korpus Janes je oblikoskladenjsko označen (Fišer et al., 2015), kar pomeni, da lahko za samodejno luščenje terminologije uporabimo obstoječe orodje LUIZ, ki je bilo doslej uporabljeno že na številnih področjih (Vintar, 2010; Vintar in Fišer, 2011). Luščenje namreč temelji na oblikoskladenjskih vzorcih, ki so ponavadi samostalniške zveze, ti pa se v nadaljnji obdelavi razvrščajo na bolj ali manj terminološke. Na rezultate luščenja močno vpliva kakovost oblikoskladenjskih oznak, v korpusu Janes pa je označevalnih napak precej.

S programom LUIZ smo iz podkorpusov luščili izključno samostalniške besedne zveze, dolge od ene do štirih besed. Izračun terminološkosti temelji na primerjavi pogostosti med specializiranim in referenčnim korpusom; kot referenčni korpus smo uporabili Gigafido (Logar Berginc et al., 2012). Za izračun natančnosti smo pri vseh seznamih izluščenih kandidatov pregledali prvih 500 enot in ročno označili terminološko relevantne izraze. Ob tem je treba poudariti, da je bila presoja nujno subjektivna, saj se pri označevanju nismo posvetovali s področnimi strokovnjaki. Prav tako med termine nismo uvrščali splošnega besedišča, ki je se v določenem podforumu pogosto pojavlja zaradi tematskega okvira, a brez strokovno specifične reference. Tako se denimo pri podforumu Hujšanje kot terminološki kandidati pojavijo izrazi v zvezi s prehrano (*jogurt, zajtrk, obrok*), ki jih ne moremo obravnavati kot specializirane enote, čeprav so tematsko ključne za obravnavani forum. V kontekstu naše raziskave lahko natančnost razumemo kot izmerjeno raven specializiranosti posameznega podforumu.

Pri večini seznamov se pokaže, da je približno polovica izluščenih enot večbesednih, najproduktivnejši vzorec pa je pridevnik in samostalnik. Ostalih besednovrstnih kombinacij je bistveno manj, tri- in večbesedne enote so redke. To opažanje ni presenetljivo, saj se je dvobesedni vzorec pridevnik in samostalnik pokazal za najproduktivnejšega tudi v drugih raziskavah (Logar Berginc in dr., 2013).

Med prvimi 500 enotami ima največ terminoloških enot podforum Ginekologija (edini z aktivno moderacijo), ta forum izstopa tudi po visokem številu večbesednih, se pravi višje specializiranih enot pri vrhu seznama (*nosečnost, plod, porod, UZ, vstava; nuhalna svetlina, medenična vstava, morfolologija ploda, maternični vrat*). Po deležu terminoloških enot sledijo Astronomija, vendar z manj večbesednimi (*planet, zvezda, sonce, vesolje, gravitacija, galaksija, teleskop, asteroid; črna luknja, spektralna črta, nevtronska zvezda*), Sam svoj mojster (*motor, volan, feltna, odbijač, vijak, platišče; prestavna ročica, sprednji odbijač, zadnja hauba, sesalni kolektor*), Kontracepcija (*tabletk, menstruacija, krvavitev, izcedek, kondom, obroček; spolni odnos, kontracepcijska tabletk, urgentna kontracepcija*), Styling (*homologacija, spojler, ledica, blatnik, števec; dimenzija gum, homologacijski kartonček, jekleno platišče*) in nazadnje Hujšanje, kjer je bilo terminološko relevantnih izrazov zgolj 15 odstotkov (*dieta, prehrana, GI, jedilnik, post; ogljikovi hidrati, način prehranjevanja, ločevalna dieta*).

Podkorpus	Št. izluščenih	Št. večbesednih	Natančnost (N=500)
Avtomobilizem.Styling	12.551	5.591	0,39
Avtomobilizem.SamSvojMojster	14.042	6.366	0,58
Med.Over.Net.Ginekologija	3.768	1.720	0,68
Med.Over.Net.Kontracepcija	5.145	3.280	0,52
Med.Over.Net.Hujšanje	4.037	1.464	0,15
Kvarkadabra.Astronomija	3.145	1.064	0,60

Tabela 2: Luščenje terminologije iz podkorpusov.

Posebna značilnost spletnih forumov so kratice, ki se uveljavijo znotraj določenega foruma oz. podforumu. Precej kratic smo opazili v podforumih med.over.net (*M – menstruacija, ZM – zadnja menstruacija, G – ginekolog, CR – carski rez, UZ – ultrazvok, KT – kontracepcijske tablete, GK – ginekološka klinika* itd.), manj v avtomobilizmu (*CZ – centralno zaklepanje*) in še manj na kvarkadabra.net. Kratice so povečini terminološke in predstavljajo učinkovito sredstvo za večjo ekonomičnost komunikacije, obenem pa igrajo vlogo kot povezovalni členi forumske skupnosti, ki z uporabo kratic izraža pripadnost.

4.2 Nestandardni elementi

V kontekstu terminologije je pojem nestandardnosti problematičen, kajti na mnogih specializiranih področjih se uporabljajo izrazi, ki odstopajo od besedotvornih načel splošne slovenščine, so prevzeti, vsebujejo nebesedne elemente ali simbole, pa zato niso nestandardni. V raziskavi nas je predvsem zanimalo, v kolikšni meri uporabniki forumov pri posredovanju znanja uporabljajo žargonsko terminologijo, sicer tipično za govorjeni diskurz, in v kolikšni meri pri zapisovanju izrazov prihaja do namernih ali nenamernih odstopanj.

Za ugotavljanje nestandardnosti smo uporabili tri metode:

- ročno označevanje nestandardnih terminoloških izrazov med prvimi 500 izluščenimi kandidati; pri tem moramo poudariti, da smo se omejili na termine in nestandardnih splošnih izrazov nismo upoštevali. V tabeli je zapisan odstotek nestandardnih enot od vseh terminološko relevantnih enot med prvimi 500 izluščenimi kandidati (ns_MAN),
- luščenje samostalniških enot, ki se z visoko relativno pogostostjo pojavljajo v podkorpusu, a imajo zelo majhno (<10) pogostost v referenčnem korpusu Gigafida (ns_GF), v tabeli je zapisano število takšnih enot,
- luščenje nestandardnih terminoloških kandidatov s pomočjo leksikona besednih oblik Sloleks (Arhar, 2009); iz prvotno izluščenega seznama kandidatov smo izluščili tiste, pri katerih se vsaj ena od besed ne pojavi v Sloleksu (ns_SL), v tabeli je zapisan odstotek od vseh izluščenih terminoloških kandidatov.

Rezultate na koncu primerjamo še z mero tehnične (std_tech) in jezikovne (std_ling) nestandardnosti, ki sta jo razvila Erjavec in Ljubešič (Erjavec, 2015) za namene projekta Janes, v tabeli podajamo povprečno vrednost za podkorpus in standardno deviacijo.

Podkorpus	ns_MAN	ns_GF	ns_SL	std_ling	stdev	std_tech	stdev
Avtomobilizem.Styling	31,1	184	27,0	1,7	0,6	1,8	0,5
Avtomobilizem.SamSvojMojster	20,7	165	21,8	1,7	0,5	1,8	0,5
Med.Over.Net.Ginekologija	3,0	21	10,2	1,5	0,2	1,5	0,3
Med.Over.Net.Kontracepcija	5,7	32	11,5	1,5	0,3	1,4	0,4
Med.Over.Net.Hujšanje	8,0	18	10,6	1,7	0,4	1,6	0,5
Kvarkadabra.Astronomija	6,6	9	11,6	1,5	0,4	1,4	0,4

Tabela 3: Mere nestandardnosti.

Iz Tabele 3 je razvidno, da pri vseh uporabljenih metodah po nestandardnosti izstopata oba podforumu avtomobilizma, najbliže standardu pa je aktivno moderirani forum ginekologije. Za forum avtomobilizma je predvsem značilna raba žargonskih poimenovanj in/ali nestandardnih zapisov za dele vozila (*auspuh/avspuh, felga/feltna, špegu/špegl, števc, spojlerček, žmigovec/žmigavec, radkapa/ratkapa, žarnca, lampa, akumolator*), orodja in postopke (*šrauf/šravf, šprickit, šmirgl, valar, strokovna ugradnja, pedenanje, ličarija*). To pa ne pomeni, da ljubitelji avtomobilizma ne uporabljajo tudi kanonične terminologije, saj je pri mnogih od omenjenih izrazov standardna različica pogostejša od pogovorne (*šipa*: 312, *šajba*: 23). Za avtomobilizem je značilno tudi prepletanje registrov, ko avtor v objavi uporabi nekaj pogovornih in nekaj kanoničnih izrazov:

»Torej, ali se čuje turbina (ker jo počas hudič jemlje), ali je prežgalo dihtungo med turbino in sesalno cevjo (oranžna dihtunga, ki stane cca 12€ – dokaj pogost fail...«

Naši rezultati se skladajo tudi s splošnima merama nestandardnosti, razen pri forumu Hujšanje, kjer pa

moramo upoštevati, da se naša analiza usmerja v ugotavljanje nestandardnosti terminologije in ne tudi drugih jezikovnih značilnosti. Ker je v forumu Hujšanje specializiranega izrazja zelo malo, to tudi ne izstopa po nestandardnosti.

Pri bolj specializiranih podforumih se izkaže, da metoda primerjave s splošnimi referenčnimi viri, kot sta Gigafida in Sloleks, ni učinkovita za ugotavljanje nestandardnosti, saj se med žargonskimi, napačno zapisanimi in pogovornimi izrazi znajdejo tudi visokospecializirani izrazi, ki jih v splošnih referenčnih virih ni. To smo opazili predvsem pri medicinskem izrazju, kjer bi bilo za natančnejše luščenje nestandardnih izrazov poleg splošnih virov nujno uporabiti še medicinski leksikon.

Med nestandardnimi izrazi s foruma kvarkadabra.net je večina, če odštejemo izrazje brez šumnikov, angleških poimenovanj za fizikalne in druge pojave, nekaj pa je tudi zatipkank in hudomušnih tvorjenk (*nafitana krivulja, hobistična pamet, sključitev, kvazinaravoslovec, complex lajf*).

4.3 Primerjava objav strokovnjaka in nestrokovnjakov

Ker se je iz doslej predstavljenih rezultatov potrdilo, da aktivno moderirani forum Ginekologija izstopa tako po deležu terminologije, ravni specializiranosti in standardnosti, nas je zanimalo, ali je v interakciji med laično skupnostjo in strokovnjakom razlika v rabi terminologije med enimi in drugim. Forum ABC Ginekologije in porodništva že vsa leta njegovega obstoja moderira mag. Stanko Pušenjak, specialist ginekolog in porodničar, ki je v našem podkorpusu avtor 850 objav. Iz podkorpusa smo tako v eno skupino uvrstili zgolj njegove objave, v drugo pa vse ostale, in ju primerjali.

	Ginek.SP	Ginek.Drugi
Št. objav	850	874
Št. pojavnic	130.616	129.388
Št. izluščenih	1.835	2.388
ns SL	5,0	12,7

Tabela 4: Strokovnjak in nestrokovnjaki.

Delež nestandardnih enot po metodi primerjave s Sloleksom je v objavah specialista precej nižji (5,0) kot v objavah laikov (12,7), ročni pregled pa pokaže, da je v izluščenih terminih večina specializiranih medicinskih izrazov, ki se v Sloleksu ne pojavijo (*spermiogram, luteinska cista, Metronidazol, endometritis*). V objavah specialista zasledimo le tri enote, ki bi jih lahko označili za nestandardne zaradi napačnega zapisa (*lumbrikant, stegnjenica, ovulacija*). Nasprotno je v objavah laikov nestandardnih zapisov prek 300, uvrstimo pa jih lahko v tehnično nestandardne (pretežno zapisane brez šumnikov) in jezikovno nestandardne, napačno zapisane terminološke enote (*Dobroston/Dabresten, amnoreja, menstruacija, abortivna tableta, aglutinacija semenčic, iscedek, endometiroza, gljivična okužba*). Nasploh je med laičnimi objavami opaziti veliko variabilnost pri zapisovanju specializiranih enot (*anastetik / anstetik / anastazistka; celarblue / clerblu; ejakucija / ejakuiral; endometiroza / endometrioza / ednometrioza; epizotomija / epiziotomija; genikologinja / ginekloginja / ginekologinja* itd.).

Žal je bil vzorec premajhen, da bi izvedli še analizo kolokacij; predvidevamo namreč, da nestrokovnjaki določene termine uporabljajo v drugačnih sintagmah kot strokovnjaki. Tak primer smo zabeležili pri laični rabi izrazov *tableta/tabletka/zdravilo* v povezavi z glagolom *jesti*; med objavami zdravnika specialista te kolokacije ni najti, temveč se *zdravila/tablete uživa, jemlje* ali *uporablja*.

5 Sklep

V raziskavi smo želeli raziskati značilnosti rabe terminologije v spletnih forumih, pri tem pa smo uporabili luščilnik terminologije kot pripomoček za ugotavljanje ravni specializiranosti posameznega foruma. Analiza je pokazala, da so med obravnavanimi forumi glede ravni specializiranosti precejšnje razlike, najbolj specializiran pa je bil aktivno moderirani forum zdravstvene posvetovalnice, pri katerem zdravnik specialist odgovarja na vprašanja pacientov. Terminološko bogata sta tudi oba podforuma s področja avtomobilizma, hkrati pa sta ta podforuma tudi najbolj nestandardna z uporabo številnih pogovornih izrazov in področno specifičnih okrajšav.

Ker je pri strokovnih forumih osnovni cilj udeležencev izmenjava znanja, ni presenetljivo, da je na splošno ta besedilni žanr s terminološkega vidika zanimiv, jezikovni register določenega foruma pa se oblikuje glede na različne zunajjezikovne parametre, kot so predznanje/izobrazba razpravljalcev, regularnost foruma in strokovnost tematskega področja.

V nadaljnjih raziskavah se želimo podrobneje posvetiti pojmovnim razmerjem, ki se pojavljajo znotraj določene strokovne razprave, in odkrivanju tipičnih struktur, ki so značilne za bolj ali manj regulirane komunikacijske situacije posredovanja znanja (npr. pojmovne strukture vprašanj in pozvedb, odgovorov in razlag, navodil in napotkov itd.). Prav tako želimo sistematično raziskati terminološko variabilnost.

Jezikoslovne značilnosti rabe terminologije v spletnih forumih, pa tudi drugih spletnih žanrih, so relevantne z več vidikov. Z računalniško obdelavo bi iz njih lahko pridobivali znanje in izkušnje spletnih uporabnikov in jih uporabili za razvoj inteligentnih aplikacij, Z odkrivanjem terminoloških variacij bi lahko izboljšali spletne iskalnike in razpoznavalnike govora. V nekoliko širšem kontekstu foruma kot virtualne skupnosti za posredovanje znanja pa bodo vsekakor še naprej pomembne tudi komunikološke, socio- in psiholingvistične raziskave, ki nam pomagajo razumeti oblikovanje spletnih identitet in njihovega vpliva na komunikacijo.

6 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.

7 Literatura

- Špela Arhar. 2009. Učni korpus SSJ in leksikon besednih oblik za slovenščino. *Jezik in slovstvo*, 54/3-4, 43-56.
- 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. Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.
- Nataša Logar Berginc, Špela Vintar in Špela Arhar Holdt. 2013. Terminologija odnosov z javnostmi: korpus – luščenje – terminološka podatkovna zbirka. *Slovenščina* 2.0, 1 (2): 113-138.
- Darja Fišer, Tomaž Erjavec, Jaka Čibej in Nikola Ljubešić. 2015. Gradnja in analiza korpusa spletne slovenščine JANES. Obdobja 2015 (v tisku).
- Nataša Jakop. 2008. Pravopis in spletni forumi – kva dogaja? *Slovenščina med kulturami, Zbornik Slavističnega društva Slovenije* 19, 315-327.
- Joachim Kimmerle, Kristina Abels, Katharina Becher, Anna Beckers, Annette Haussmann, Ansgar Thiel in U. Cress. 2011. Construction of health knowledge in an alternative medical community of practice: Hermeneutic analysis of a web forum. V *Connecting research to policy and practice: Proceedings of the 9th computer supported collaborative learning conference*, vol. 1. 1-8.
- 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 RANLP – Recent Advances in Natural Language Processing, Hissar, Bulgaria.
- Begoña Montero, Frances Watts, and Amparo García-Carbonell. Discussion forum interactions: Text and context. *System* 35.4 (2007): 566–582.
- Andraž Petrovčič. 2005. Deliberativnost komuniciranja v spletnih forumih. Diplomsko delo. Univerza v Ljubljani, Fakulteta za družbene vede.
- Robert Plant. 2004. Online Communities. *Technology in Society*, 26 (1), 51–65.
- Christina Varga. 2011. Knowledge Transmission in Cyberspace. Discourse Analysis of Professional Web Forums as Internet Subgenre. Doktorska disertacija, UPF Barcelona.
- Špela Vintar. 2010. Bilingual term recognition revisited: the bag-of-equivalents term alignment approach and its evaluation. *Terminology*, 16 (2), 141–158.
- Špela Vintar in Darja Fišer. 2011. Enriching Slovene WordNet with domain-specific terms. Translation: computation, corpora, cognition, vol. 1, no. 1, str. 29–44. <http://www.t-c3.org/index.php/t-c3/article/view/4>.
- Nataša Zupančič. 2009. Korpusna analiza slovenskega jezika na spletnih forumih. Magistrsko delo, Univerza v Ljubljani, Filozofska fakulteta.