

TERMINOLOGIJA ODNOSOV Z JAVNOSTMI: KORPUS – LUŠČENJE – TERMINOLOŠKA PODATKOVNA ZBIRKA

Nataša LOGAR BERGINC

Univerza v Ljubljani, Fakulteta za družbene vede

Špela VINTAR

Univerza v Ljubljani, Filozofska fakulteta

Špela ARHAR HOLDT

Trojina, zavod za uporabno slovenistiko

Logar Berginc, N., Vintar, Š., Arhar Holdt, Š. (2013): Terminologija odnosov z javnostmi: korpus – luščenje – terminološka podatkovna zbirka. Slovenščina 2.0, 1 (2): 113–138.

URL: http://www.trojina.org/slovenscina2.0/arhiv/2013/2/Slo2.0_2013_2_06.pdf.

V prispevku prikazujemo analizo luščenja eno- in večbesednih terminoloških kandidatov, ki smo ga izvedli za potrebe priprave terminološke podatkovne zbirke odnosov z javnostmi na podlagi korpusa KoRP z luščilnikom LUIZ. Podrobneje se posvečamo dvojemu: (a) izluščenim enobesednim samostalniškimi terminološkimi kandidatom, katerih seznam primerjamo s pogostostnim seznamom samostalnikov v korpusu KoRP in vrednotimo glede na terminološkost, kot sta jo prepoznala dva področna strokovnjaka, ter (b) izluščenim večbesednim nizom z glagolskim in samostalniškim jedrom. Nadgrajeno metodo luščenja in izboljššan prikaz rezultatov smo dopolnili še z analizo priklica. Potrdili oz. ugotovili smo, da je v primerjavi s pogostostnim seznamom terminološki potencial enot v zgornjem delu seznama izluščenih samostalnikov večji, da imajo izluščeni glagolski besedni nizi predvsem kolokacijsko vrednost, ne pa tudi terminološke, in da so najbolj terminološko produktivni vzorci luščenja samostalniških zvez z naslednjo zgradbo: [pridevnik + samostalniik], [pridevnik + in + pridevnik + samostalniik] ter [pridevnik + pridevnik + samostalniik]. Analiza priklica je pokazala predvsem nizko stopnjo strinjanja med obema področnima strokovnjakoma, sicer pa je bil priklic razmeroma visok.

Ključne besede: korpus KoRP, projekt TERMIS, luščilnik LUIZ, luščenje terminov

1 UVOD

Korpusno jezikoslovje je z možnostjo meritev zelo različnih lastnosti jezika prineslo opazen premik zlasti v leksikografiji – pa tudi terminografiji (npr. Atkins, Rundell 2008; Bergenholtz, Tarp 1995; Biber in dr. 1998; Čermák 2011; Halliday in dr. 2004; Hanks, 2008; Leech 1992; McEnery, Wilson 1996; Pearson 1998; Schryver 2003; Sinclair 2004; Teubert, Krishnamurty 2007; pri nas npr.: Krek 2003; Gorjanc 2005; Gorjanc, Krek 2005; Gantar 2007; Gantar 2009; Logar 2007; Vintar 2008). Na osnovi korpusa, ki ima razviden namen, merila gradnje, odločitve snovalcev ter besedilnozvrstno, časovno in drugo zgradbo, lahko področni strokovnjaki in jezikoslovci, ki sodelujejo pri pripravi terminoloških slovarjev ali terminoloških podatkovnih zbirk, svojo interpretacijo, ki bo posledično podana kot jezikovni opis in verjetno tudi predpis, oprejo na podatke, pridobljene iz resnične jezikovne rabe; z umikanjem nuje po introspekciji kot edini možnosti presoje pa se povečata kredibilnost in veljavnost prepoznavanja vseh pojavov – tudi jezikovnih. Pri takem pristopu je ena ključnih prednosti, ki kvalitativno in kvantitativno presega nekorpusno terminografijo, računalniško podprto luščenje terminoloških kandidatov.¹

Pristopov k luščenju terminoloških kandidatov je več, pri skoraj vseh pa gre za kombinacijo jezikoslovnega znanja o naravi terminov ter izrabo matematičnih lastnosti porazdelitve besed in besednih nizov v korpusih (Vintar 2008: 100; Vintar 2009: 346–347 in tam navedena literatura). Za slovenščino je bilo luščenje terminoloških kandidatov preizkušeno že večkrat, npr. za projekt

¹ V zadnjih letih je pri razvoju inteligentnih aplikacij vse bolj razširjena tudi gradnja ontologij, ki se s terminologijo kot poimenovalnim sistemom (in posledično terminografijo) sreča na pojmovni ravni, tj. na ravni strukturiranja znanja – prim. Uschold, King (1995); Corcho in dr. (2003); Buitelaar in dr. (2004); De Nicola in dr. (2009); Kapoor, Sharma (2010).

VoiceTran I in II na področju vojaške terminologije ter na področju računalništva in informatike za Islovar (Vintar, Erjavec 2008; Vintar 2009; Vintar 2010: 47); pa tudi pri pripravi terminološke podatkovne zbirke odnosov z javnostmi. Za slednjega smo že predstavili (Logar, Vintar 2008), da je bilo uspešno zlasti pri pridobivanju večbesednih terminoloških kandidatov, ki je potekalo po kombinaciji statistične in jezikoslovne metode. Na osnovi takrat prepoznanih pomanjkljivosti smo metodo in prikaz rezultatov izboljšali – v nadaljevanju prikazujemo, na kakšen način in s kakšno uspešnostjo.

2 PREIZKUS

Korpus, iz katerega smo luščili terminološke kandidate odnosov z javnostmi, je korpus KoRP. KoRP vsebuje 1,8 milijona pojavníc, je enojezični, sinhroni in statični korpus strokovnih besedil. Na spletu je prosto dostopen vse od svojega nastanka leta 2007. Od aprila 2013 dalje je nameščen na strežniku nl.ijs.si (Erjavec 2013).² Pred luščenjem, ki bo prikazano v nadaljevanju, smo ga ponovno označili z najnovejšo različico označevalnika podjetja Amebis, d. o. o. (Romih, Holozan 2002; Holozan 2006; Arhar Holdt 2011: 22–23, 28–29).³

S pomočjo luščilnika LUIZ (Vintar 2010) smo želeli pridobiti dvoje:

- a) enobesedne terminološke kandidate: samostalnike, glagole, pridevnike in prislove;
- b) večbesedne terminološke kandidate: samostalniške in glagolske besedne zveze.

Tako eno- kot večbesedne kandidate smo luščili s pomočjo oblikoskladenjskih vzorcev (ti so pri enobesednih enotah sestavljeni le iz enega člena) in terminološke uteži, ki se izračuna na podlagi pogostosti pojavitve v

² http://nl.ijs.si/noske/sl-spec.cgi/first_form?corpname=korp_sl

³ Različica korpusa KoRP, ki je dostopna na prej omenjeni spletni strani, je sicer označena s statističnim označevalnikom Obeliks (Grčar in dr. 2012).

specializiranim korpusu v primerjavi s splošnim korpusom ter frazeološke stabilnosti enote. Slednja je relevantna zgolj pri večbesednih enotah in pomeni, da bo v primeru terminoloških variacij (npr. *merilo uspešnosti delovanja/merilo za uspešnost delovanja*) višjo utež dobila tista enota, ki se v korpusu pojavi večkrat. Skupno smo uporabili 39 oblikoskladenjskih vzorcev, od tega 30 s samostalniškimi jedrom, 9 z glagolskim (Tabela 1, prvi stolpec). Za tako obsežno število vzorcev smo se odločili zaradi večje zanesljivosti ocene tovrstnega pristopa k pridobivanju terminov v slovenščini, smo pa že pred začetkom luščenja predvidevali, da bo približno polovica vzorcev dala le malo ali celo nič terminološko zanimivega gradiva.

Ker se med izluščenimi kandidati pogosto znajdejo tudi lastna imena, ki pa za slovarske namene v veliki meri niso zanimiva, smo vse enote, ki so vsebovale besede z veliko začetnico, umaknili na ločen seznam.

3 ANALIZA REZULTATOV

Rezultat luščenja so bili sezname, na katerih je bilo skupno 47.007 večbesednih enot (brez lastnih imen; število po vzorcih in primeri so v Tabeli 1) oz. 16.190 enobesednih enot (brez lastnih imen; Tabela 2).

Vzorec	Število kandidatov	Primer
Samostalniške zveze		
1. P S*	17.242	<i>lokalna skupnost</i>
2. S S	9.362	<i>vir informacije</i>
3. S S S	932	<i>merilo uspešnosti delovanja</i>
4. P S S	1.670	<i>uraden vir informacije</i>
5. S P S	3.160	<i>dan odprtih vrat</i>
6. S D S	4.370	<i>sporočilo za javnost</i>
7. S D S S	648	<i>orodje za doseganje cilja</i>
8. S D P S	1.174	<i>odnosi z internimi javnostmi</i>
9. P P S	1.398	<i>celostna grafična podoba</i>
10. P D S	621	<i>vodilni v podjetju</i>
11. R P S	618	<i>srednje veliko podjetje</i>

12. S S S S	53	<i>dvig kakovosti življenja otrok</i>
13. S P S S	198	<i>doseganje poslovnih ciljev organizacije</i>
14. P S S S	130	<i>osrednje zanimanje svetovne javnosti</i>
15. P S P S	381	<i>refleksivni model komunikacijskega menedžmenta</i>
16. S S P S	321	<i>model načrtovanja merljivih ciljev</i>
17. S P P S	245	<i>model dvosmernega simetričnega odnosa</i>
18. P P S S	51	<i>upravljano komunikacijsko ravnanje organizacije</i>
19. S S D S	982	<i>strategija odnosov z javnostmi</i>
20. P S D S	971	<i>tržni odnosi z javnostmi</i>
21. R P S S	51	<i>vnapij pripravljeni predlogi vprašanj</i>
22. R P D S	63	<i>tesno povezan s teorijo</i>
23. P D P S	141	<i>značilen za blagovno znamko</i>
24. P D S S	100	<i>potreben za razrešitev konflikta</i>
25. R P P S	51	<i>točno določena ciljna javnost</i>
26. P S in S	445	<i>medijski čas in prostor</i>
27. S S in S	365	<i>trženje izdelkov in storitev</i>
28. S in S S	434	<i>mnenje in stališče javnosti</i>
29. S in P S	367	<i>čast in dobro ime</i>
30. P in P S	463	<i>tiskani in elektronski mediji</i>
Glagolske zveze		
31. R G	3.032	<i>pomembno vplivati</i>
32. G D S	2.208	<i>odgovoriti na vprašanje</i>
33. G R	1.602	<i>delovati neodvisno</i>
34. D S G	878	<i>v nadaljevanju predstavljati</i>
35. G D R	101	<i>biti do sedaj</i>
36. G kot S	98	<i>delovati kot posrednik</i>
37. G kot P	42	<i>biti kot nov</i>
38. D R G	14	<i>od nekdanj spremljati</i>
39. G kot R	13	<i>ocenjevati kot zelo</i>
SKUPAJ	47.007	

Tabela 1: Izluščeni večbesedni terminološki kandidati: število po vzorcih in primeri.

* P = pridevnik, S = samostalnik, D = predlog, R = prislov, G = glagol

Besedna vrsta	Število kandidatov	Primer
1. S	7.379 (z lastnimi imeni: 10.731)	<i>javnost, odnos, organizacija</i>
2. P	4.854	<i>komunikacijski, blagovni, zaposleni</i>
3. R	1.379	<i>veliko, pogosto, vedno</i>
4. G	2.578	<i>vplivati, sporočiti, komunicirati</i>
SKUPAJ	16.190	

Tabela 2: Izluščeni enobesedni terminološki kandidati: število po besednih vrstah in primeri.

Sezname smo podrobneje pregledali in prišli do ocene relevantnosti metode luščenja, dodatno pa izvedli še analizo priklica.

3.1 Enobesedni terminološki kandidati (samostalniki)

Predstavili bomo le analizo vrhnjega dela seznama samostalniških terminoloških kandidatov, ki smo ga ocenjevali z dveh vidikov:

- a) v primerjavi s pogostostnim seznamom samostalnikov iz korpusa KoRP in
- b) glede na oceno terminološkosti, ki sta jo dala dva strokovnjaka s področja odnosov z javnostmi.

Že ob primerjalnem ogledu zgolj prvih 20 izluščenih samostalnikov in prvih 20 samostalnikov po pogostosti (Tabela 3) opazimo nekaj razlik: pri luščenju so med prvimi dvajsetimi poimenovanja *management*, *deležnik*, *model* in *novinar* (v Tabeli 3 krepki tisk), ki jih na pogostostnem seznamu med prvimi dvajsetimi ni; nasprotno pa so na pogostostnem seznamu samostalnikov višje, tj. do dvajsetega mesta, uvrščeni *človek*, *skupnost* in *stran* (v Tabeli 3 ležeči tisk).

Terminološko luščenje: prvih 20 samostalnikov	Pogostost: prvih 20 samostalnikov
javnost	javnost
odnos	odnos
organizacija	organizacija
komuniciranje	podjetje

podjetje	komuniciranje
medij	medij
znamka	znamka
informacija	leto
leto	informacija
cilj	cilj
okolje	primer
management	skupina
področje	okolje
primer	program
skupina	področje
program	človek
deležnik	(zaposleni)*
vloga	vloga
model	skupnost
novinar	stran

Tabela 3: Vrhnji del seznama izluščenih samostalnikov in samostalnikov po pogostosti v korpusu KoRP.

* Pri luščenju med prvimi dvajsetimi, vendar pri pridevnikih.

Pregled razlik, ki so se pokazale do stotega mesta tabele, pokaže, da so bili pri terminološkem luščenju višje uvrščeni (torej med prvimi stotimi) *dejavniki*, *izvajanje*, *javnost*, *komunikator*, *manager*, *menedžment*, *tveganje*, *uspešnost*, *vedenje* in *zaupanje*; medtem ko so pri pogostostnem seznamu višje uvrščene besede *kot*, *mesto*, *načrt*, *Publica*, *služba*, *str.*, *svet* in *vlada*. Tudi če pogledamo spremembe mest terminoloških kandidatov med prvimi stotimi mesti, so premiki zanimivi: *management*, ki je pri terminološkem luščenju na 12. mestu, je po pogostosti na 60. mestu, in tako npr. še *deležnik* – 17. (41.), *oglaševanje* – 29. (44.), *manager* – 31. (130.), *marketing* – 36. (59.), *praktik* – 40. (87.), *načrtovanje* – 53. (81.), *komunikator* – 65. (198.), *uspešnost* – 85. (111.), *tveganje* – 97. (122.). Nazadovali pa so pri luščenju samostalniki *človek* – 29. mesto, po pogostosti pa 17. mesto, *stran* – 33. (20.),

sistem – 94. (64.), *podatek* – 96. (71.) ipd.

Strokovnjaka odnosov z javnostmi sta za potrebe priprave terminološke podatkovne zbirke pregledala celotni pogostostni seznam samostalnikov, ki je obsegal skoraj 12.500 enot.⁴ Če kot načeloma potrjene s pogostostnega seznama vzamemo le tiste, ki so s strani obeh strokovnjakov dobili kljukico (druge oznake so bile še: *nekaj manjka*, *prečrtano*, *vprašaj*), ugotovimo, da jih je med prvimi dvestotimi takih 117. Če na enak način pogledamo tudi seznam terminološko izluščenih samostalniških kandidatov, je rezultat nekoliko boljši: med prvimi dvestotimi izluščenimi je dvojno kljukico dobilo 125 kandidatov. Če preverimo npr. še mesta od 700 do 800 na obeh seznamih, zopet ugotovimo rahlo prednost terminološko izluščenega seznama, in sicer v razmerju 42 : 38. Pri mestih od 1.000 do 1.100 pa je uporabljeno luščenje glede na odločitve obeh strokovnjakov v še večji prednosti: 41 : 28.

Povzamemo lahko, da med vrhnjim delom seznama samostalnikov, ki se kot terminološki kandidati kažejo po luščenju, in vrhnjim delom pogostostnega seznama samostalnikov ni zelo velikih razlik, a so te z vidika ocene terminološkosti vedno v prid terminološko izluščenim seznamom, kar pomeni, da je pri naboru enobesednih terminov za geslovník bolj smiselno izhajati iz slednjih.

3.2 Večbesedni terminološki kandidati

Uspešnost samodejnega luščenja večbesednih terminoloških kandidatov smo s pomočjo dveh strokovnjakov s področja odnosov z javnostmi ocenili že pri prvem luščenju, ki smo ga izvedli leta 2007 in v katerega smo zajeli osem samostalniških oblikoskladenjskih vzorcev (prvih osem v Tabeli 1). Za ponazoritev podajamo vrhnji del takratnega seznama, ki prikazuje

⁴ Tovrstno ocenjevanje terminološkosti zgolj po seznamu, brez vpogleda v sobesedilo in brez posvetovanja s še drugimi področnimi strokovnjaki ima omejitve in pomanjkljivosti – pri projektu smo ga naknadno nadgradili z doslednim vpogledom v sobesedilo v korpusu KoRP ter s posvetovanjem z večjo skupino področnih strokovnjakov.

terminološke kandidate, za katere sta oba strokovnjaka potrdila, da gre za termine odnosov z javnostmi (Tabela 4). Skupno oceno uspešnosti takratnega luščenja lahko strnemo v ugotovitev, da je pri prvih 1.000 enotah na seznamu "v drugi polovici /.../ sicer narašča/l/ delež kolokacij, vendar je /bilo/ na tem seznamu veliko terminov" (Logar, Vintar 2008: 13).⁵

Mesto na seznamu pri terminološkem luščenju	Terminološki kandidat
1.	odnosi z javnostmi
4.	korporativna identiteta
5.	odnosi z mediji
6.	komunikacijski menedžment
10.	uglednostni kapital
11.	lokalna skupnost
13.	ciljna javnost
15.	krizni management
17.	deležniška skupina
22.	poslovni komunikator
25.	komunikacijska aktivnost
27.	odnosi z zaposlenimi
31.	komunikacijska kompetenca
32.	komunikacijski program
33.	komunikacijski management

Tabela 4: Prvih 15 izluščenih večbesednih terminoloških kandidatov (luščenje iz leta 2007), potrjenih s strani dveh strokovnjakov odnosov z javnostmi.

V nadaljevanju prispevka izmed vseh 39 na novo luščenih oblikoskladenjskih

⁵ Pri tem smo ločili med (a) večbesednim terminom kot stalno zvezo, poimenovanjem in (b) kolokacijo kot leksikalno in/ali pragmatično povezano ponovljivo sopojavitvijo vsaj dveh leksikalnih enot, ki sta med seboj v neposrednem skladenjskem razmerju (Bartsch, 2004, nav. po Heid, Gouws 2006: 980), pri čemer smo kolokacijo razumeli v ožjem pomenu, tj. kot prosto zvezo. Razmejitev je seveda groba (prim. Vintar, 2003: 74; Erjavec, Vintar, 2004: 104; Logar, Vintar, 2008: 12–14), jo pa ohranjamo tudi v tem prispevku.

vzorcev najprej povzemamo oceno terminološke zanimivosti vzorcev z glagolom (razdelek 3.2.1), tj. zvez glagola s prislovom, glagolskih predložnih zvez in zvez glagola s podrednim veznikom *kot* (zadnjih 9 vzorcev v Tabeli 1; o enobesednih glagolskih terminih, tudi izluščenih, gl. več v Logar, Vintar 2008: 8–9, 11). V razdelku 3.2.2 pa nato sledi še ocena terminološke produktivnosti vzorcev s samostalniškim jedrom; teh je bilo 30 in so bili po obsegu največ štiribesedni.

3.2.1 VEČBESEDNI TERMINOLOŠKI KANDIDATI: GLAGOLSKE ZVEZE

Analiza je pokazala, da le dva od devetih vzorcev dasta nekaj enot, ki bi jih kot celoto lahko vključili kot samostojno geslo v terminološko podatkovno zbirko odnosov z javnostmi, in da je ob teh dveh vzorcih le še eden, ki ima tovrsten potencial – gre za: $[R G]^6$ in $[G R]$ ter $[G D S]$. Natančnejši pregled vseh treh je najprej zajel odstranitev primerov z *biti*, *morati*, *želeti* in *jesti* pri $[G D S]$ (kar je s seznama odstranilo 20 % nerelevantnih primerov) ter pri $[R G]$ oz. $[G R]$ še *moči*, *hoteti*, *smeti*, *imeti*, *začeti*, *postati*, *dobiti*, *iti*; *lahko*, *tako*, *treba*, *mogoče*, *bolj*, *vedno*, *sam*, *potrebno*, *rad*, *nekaj*, *najbolj*, *zato*, *sicer*, pri $[G R]$ pa še dodatno *veliko*, *glede*, *čim* in *vse*. Na ta način so se pri vzorcu $[R G]$ podatki zmanjšali za polovico, pri $[G R]$ pa za 68 %. Potencialnih glagolskih večbesednih terminov je bilo, kot rečeno, le nekaj, in sicer po naši presoji vsi kot zveza glagola *komunicirati* in prislovov *dvosmerno*, *strateško*, *simetrično*, *javno*, *individualno*, *osebno*, *rutinsko* ter *navzven*. Vse ostalo pri teh treh vzorcih (pa tudi pri drugih, čeprav manj) je tako predvsem upoštevanja vreden prikaz kolokacijskega okolja glagolov in samostalnikov, ki so njihov del – če so ti glagoli in samostalniki terminološki, bodo seveda tudi njihovi besedni nizi ob ustrezni pogostosti kot kolokacije prišli v terminološko podatkovno zbirko, sicer pa ne, npr.:

- **anketirati:** *anketirati [vodjo, izvajalca, skupino]; [vzorčno, naknadno] anketirati; anketirati z [vprašalnikom];*

⁶ Zaradi lažje berljivosti vzorce v besedilu zamejujemo z oglatimi oklepaji.

- **informacija:** [seznaniti, razpolagati, početi, povezati] z informacijo; [iti] za informacijo; [soditi] med informacije; [zaupati] glede informacije; [predelati] v informacijo; [pritti] do informacije.

Lahko torej povzamemo, da so v našem primeru z luščenjem glagolskih vzorcev nastali sezname, ki z vidika nabora samostojnih iztočnic niso pomembno dopolnilo terminografskega dela, z vidika vsebine terminološke podatkovne zbirke (podatkov znotraj iztočnic) pa to vendarle lahko so – zlasti če predpostavljamo, da želimo v terminološko podatkovno zbirko vključiti tudi značilno besedilno okolje terminov. Pri slednjem si je sicer mogoče pomagati tudi z naprednimi korpusnimi orodji, ki tovrstne podatke prikazujejo samodejno, taka je npr. aplikacija *Besedne skice* v orodju *Sketch Engine*⁷ (Kilgarriff in dr. 2004; Krek, Kilgarriff 2006; Kilgarriff, Kosem 2012). Analiza je obenem izpostavila seznam glagolov in prislovov, ki bi jih bilo pri eventualnih prihodnjih luščenjih glagolskih zvez v odnosih z javnostmi smiselno avtomatsko izločiti s končnega seznama rezultatov in s tem podatke vnaprej selekcionirati; gre denimo za glagole *biti*, *imeti*, *iti*, *dati*, modalne glagole, deloma fazne glagole ipd. ter del prislovov, predvsem tistih, ki so v jeziku zelo pogosti, pomensko pa zelo splošni (*tako*, *zelo*, *lahko*, *vedno* itd.).

3.2.2 VEČBESEDNI TERMINOLOŠKI KANDIDATI: SAMOSTALNIŠKE ZVEZE

Kot je bilo nakazano že v 2. poglavju, smo za luščenje samostalniških besednih zvez uporabili 30 oblikoskladenjskih vzorcev. Pri vsakem od njih smo ročno pregledali prvih 100 kandidatov na seznamu (ali manj, če je bil seznam krajši od 100 enot). Vse besedne zveze so bile natančno preučene v besedilnem okolju v korpusu KoRP. Enot nismo ocenjevali kot terminoloških zgolj za področje odnosov z javnostmi. Kot termin smo ocenjevali *celo* enoto. Enote, ki smo jih prepoznali kot lastna imena, npr. *Mednarodno združenje poslovnih komunikatorjev*, *Zakon o gospodarskih družbah*, niso bile štete med termine.

⁷ <http://www.sketchengine.co.uk/>

Tabela 5 prikazuje število identificiranih terminov tako, da si vzorci sledijo od najbolj do najmanj terminološko produktivnega.

Vzorec	Število terminov med prvimi stotimi (ali manj, op. 7)	Primeri
1. P S	87	<i>blagovna znamka, lokalna skupnost, neprofitna organizacija</i>
2. P in P S	62	<i>notranja in zunanja javnost, splošni in spodbujeni imidž, funkcionalna in čustvena vrednota</i>
3. P P S	45	<i>integrirano marketinško komuniciranje, nevladna neprofitna organizacija, celostna grafična podoba</i>
4. P S D S	20	<i>tržni odnosi z javnostmi, krizni odnosi z javnostmi, dvosmerni odnosi z javnostmi</i>
5. P S in S	17	<i>borzni posrednik in analitik, strateško načrtovanje in upravljanje, medijski čas in prostor</i>
6. S P S	17	<i>piramida blagovne znamke, upravljanje žgočih problemov, informacija javnega značaja</i>
7. R P S	11	<i>cenovno občutljiva informacija, gensko spremenjena hrana, srednje veliko podjetje</i>
8. S D S	7	<i>odnosi z javnostmi, sporočilo za javnost, konferenca za novinarje</i>
9. S P P S	6	<i>model dvosmernega asimetričnega komuniciranja, model reflektivnega komunikacijskega menedžmenta, model enosmernega simetričnega komuniciranja</i>
10. S S	6	<i>vir informacij, pospeševanje prodaje, študija primera</i>
11. P S P S	5	<i>elokventna artikulacija organizacijskega diskurza, atomski model blagovne znamke, diskurzivni model komunikacijskega menedžmenta</i>

12. P S S	5	<i>situacijska teorija javnosti, uradni vir informacij, diskurzivna obnovitev ugleda</i>
13. S D P S	4	<i>upravljanje s človeškimi viri, orožje za množično uničevanje, upravljanje z javnimi temami</i>
14. S P S S	3	<i>model družbene amplifikacije tveganja, model družbene ojačitve tveganja, model interesnih skupin podjetja</i>
15. S S S	2	<i>pospeševalec reševanja problemov, analiza poročanja medijev</i>
16. P P S S (52) ⁸	1	<i>dvosmerni simetrični model komuniciranja</i>
17. S S in S	1	<i>priročnik vprašanj in odgovorov</i>
18. S S D S	1	<i>orodje odnosov z javnostmi</i>
19. S S P S	1	<i>prizma istovetnosti blagovne znamke</i>
20. P D P S	0	/
21. P D S S	0	/
22. P D S	0	/
23. P S S S	0	/
24. S D S S	0	/
25. R P D S (63)	0	/
26. R P P S (51)	0	/
27. R P S S (51)	0	/
28. S in P S	0	/
29. S in S S	0	/
30. S S S S (53)	0	/
SKUPAJ	301	

Tabela 5: Število terminov med (praviloma)⁹ prvimi 100 enotami s seznamov izluščenih večbesednih terminoloških kandidatov s samostalniškim jedrom.

Kot kaže tabela, sta se za najbolj terminološko zanimiva izkazala vzorca $[P S]$ ter $[P in P S]$, ki sta skupno dala več kot 60 % terminov. Vzorec $[P in P S]$ z

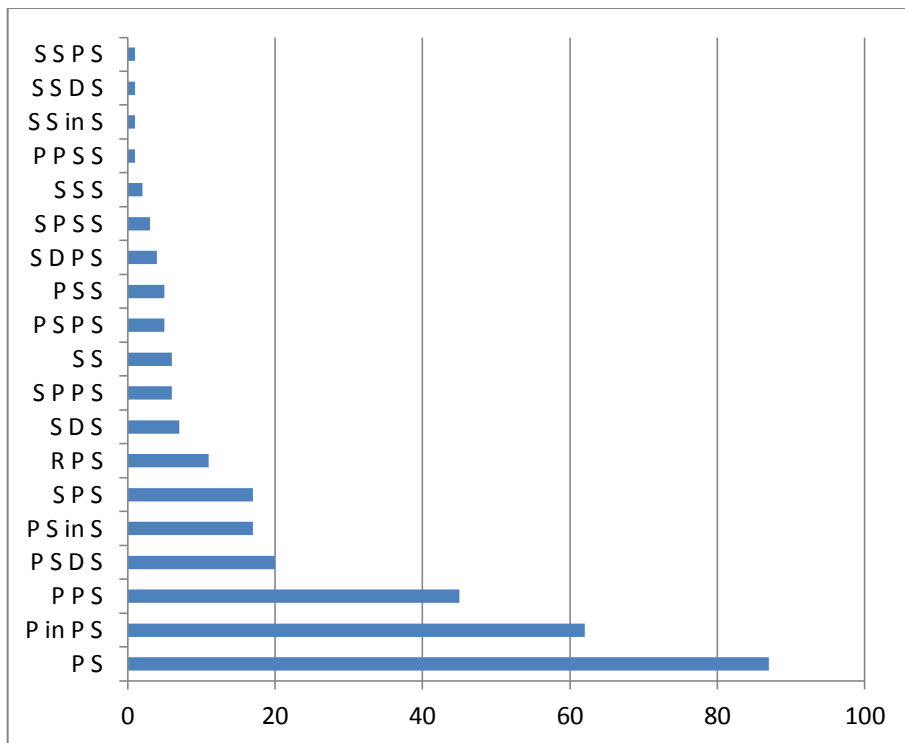
⁸ Številka v oklepaju pomeni število vseh kandidatov, izluščenih po tem vzorcu (seznam je bil torej krajši od 100 enot).

⁹ Izjema je bil namreč vzorec $[P P S S]$, katerega seznam enot je bil krajši, obsegal je le 52 terminoloških kandidatov.

rezultati tipa *notranja in zunanja javnost* je v resnici razširjeni vzorec *[P S]*, zato je mogoče – če gledamo zgolj zgradbo termina – rezultate videti kot združeno celoto (kot vzorec za luščenje pa mora ostati tak, kot je).¹⁰ Sledi jima vzorec *[P P S]*, pri katerem je bila kot terminološko relevantna ocenjena slaba polovica kandidatov. Na četrtem mestu je z 20 % vzorec *[P S D S]*, ki pa ga prevladujoče krepí osnovni termin te stroke, gre namreč za primere kot *marketinški odnosi z javnostmi* (med vsemi 20 termini je drugačna le *plačana objava v medijih*). Po 17 terminov sta dala še vzorca *[P S in S]* ter *[S P S]*. V nadaljevanju z 11 termini sledi vzorec, ki ga začenja prislov *[R P S]*, s primeri kot *cenovno občutljiva informacija*, *tržno naravnano podjetje*, *dobro obveščeni državljan*. 11-odstotni obseg se glede na običajno izpuščanje tega vzorca iz razvida tipičnih jezikovnosistemskih lastnosti strokovnih besednih zvez v slovenščini (Vidovič Muha 1988; Vidovič Muha 2000: 68–70) ne zdi zanemarljiv. Sledi 12 vzorcev, ki so med prvimi stotimi enotami na seznamu (ali manj, če je bil seznam krajši) dali le po enega in vedno manj kot 10 terminov (Slika 1 kaže število terminov pri vzorcih, ki so dali vsaj en termin), nato pa še 11 vzorcev, ki niso dali nobenih terminološko zanimivih rezultatov.

Obseg vzorcev je bil že ob načrtu luščenja ocenjen kot zelo širok; predvidevali smo, da bodo vzorci v polovici primerov dali malo ali celo nič terminov, čeprav velja imeti v mislih dejstvo, da smo luščenje izvedli na manj kot dva milijona besed obsegajočem korpusu. Predvidevanje se je v našem primeru potrdilo: tretjina samostalniških vzorcev (11) je pri pregledu prvih 100 kandidatov dala ničeln terminološki izplen, pridružimo jim lahko še 7 vzorcev, ki so dali manj kot 5 terminov, kar je skupaj dobra polovica samostalniških vzorcev. Med skupno izluščenimi in pregledanimi 3.000 različnimi večbesednimi vzorci s samostalniškim jedrom (vrhnji del seznamov) smo tako dobili skupno 301 termin.

¹⁰ Terminograf pa jih nato seveda (lahko) razdruži.



Slika 1: Število terminov med (praviloma) prvimi 100 enotami s seznamov izluščenih večbesednih terminoloških kandidatov s samostalniškim jedrom, ki so dali vsaj en termin.

3.3 Priklic

Kot je razvidno iz dosedanje razprave, se je natančnost luščenja pri posameznih vzorcih močno razlikovala, različne evalvacije iz preteklih preizkusov pa kažejo, da se pri osnovnem naboru oblikoskladenjskih vzorcev natančnost luščenja za prvih 100 kandidatov giblje med 70 in 90 % (Vintar in Erjavec 2008; Vintar 2010). V nasprotju z merjenjem natančnosti, ki je razmeroma enostavno (pregledati moramo le vrh seznama predlaganih kandidatov in ugotoviti, koliko je pravih terminov), pa je ugotavljanje priklica bistveno težje.

S priklicem pri jezikovnih tehnologijah merimo sposobnost sistema, da v

množici podatkov prepozna zadovoljiv odstotek iskanih primerov. Z drugimi besedami: pri našem preiskusu nas je zanimalo, koliko terminov, ki se v besedilu pojavijo, ostane neizluščenih. V ta namen smo izbrali znanstveni članek s področja odnosov z javnostmi in prosili dva področna strokovnjaka (ne ista strokovnjaka, kot sta pregledovala sezname), naj v članku označita vse terminološko relevantne izraze (Slika 2).

I.

Organizacija, ki uspešno rešuje svoje probleme z okoljem, dobi iz odnosa z delničarji, porabniki, zaposlenimi, vlado ter lokalno skupnostjo prednosti, saj ji le-ti lahko dajo podporo ali pa jo onemogočijo. Vsaka organizacija, ki dobro odgovori na pritiske svojih »konstituentov«, ima večjo verjetnost, da dobi njihovo podporo, saj nenehno tekmuje za omejeno javno podporo (Verčič, Grunig 1995, 36). Odnosi z javnostmi lahko pomagajo organizacijam s tem, da prepoznajo najbolj pomembne komponente organizacijskega okolja ter da z uporabo strateško vodenih komunikacij »zgladijo« organizacijske odnose z okoljem. Po Higinu je namreč uspešen strateški management tisti, ki »spremlja dosežke organizacijskega poslanstva, ki se 'slučajno' pokrivajo z upravljanjem in vodenjem odnosa organizacije s svojim okoljem«.

II.

Organizacija, ki uspešno rešuje svoje probleme z okoljem, dobi iz odnosa z delničarji, porabniki, zaposlenimi, vlado ter lokalno skupnostjo prednosti, saj ji le-ti lahko dajo podporo ali pa jo onemogočijo. Vsaka organizacija, ki dobro odgovori na pritiske svojih »konstituentov«, ima večjo verjetnost, da dobi njihovo podporo, saj nenehno tekmuje za omejeno javno podporo (Verčič, Grunig 1995, 36). Odnosi z javnostmi lahko pomagajo organizacijam s tem, da prepoznajo najbolj pomembne komponente organizacijskega okolja ter da z uporabo strateško vodenih komunikacij »zgladijo« organizacijske odnose z okoljem. Po Higinu je namreč uspešen strateški management tisti, ki »spremlja dosežke organizacijskega poslanstva, ki se 'slučajno' pokrivajo z upravljanjem in vodenjem odnosa organizacije s svojim okoljem«.

Slika 2: Del besedila, v katerem sta strokovnjaka za odnose z javnostmi podčrtala izrazje, ki sta ga prepoznala kot terminološkega.

Iz obeh podčrtanih člankov smo nato izpisali označene termine in opazovali dvoje:

a) v kolikšni meri se mnenji obeh strokovnjakov o terminološkosti ujemata (Tabela 6);

b) v kolikšni meri se nabor "človeško" izbranih terminov ujema s samodejno izluščenim seznamom (Tabela 7).

Prvi strokovnjak	184
Drugi strokovnjak	261
Presek	109
Unija	415
Ujemanje med strokovnjakoma (IAA)	0,26

Tabela 6: Ujemanje pri oceni terminološkosti med obema strokovnjakoma.

Kot vidimo, sta strokovnjaka v istem besedilu označila zelo različno število terminov. Slika 2 npr. kaže, da sta oba podčrtala *problem, odnos z delničarji/porabniki/zaposlenimi/vlado/lokalno skupnostjo, podporo, javno podporo* in *organizacijski odnos*; samo eden od njiju pa besede oz. zveze *konstituent, pritisk, strateško vodena komunikacija, organizacijsko poslanstvo* in *odnos*. V besedni zvezi *komponenta organizacijskega okolja* je bila različno določena meja termina: en podčrtovalec je izbral celo zvezo, drugi le njen zadnji del (*organizacijsko okolje*). V celoti je bil presek med izboroma le 109 terminov. Če njuna izbora združimo, dobimo unijo v velikosti 415 terminov.¹¹

V nadaljevanju smo merili priključitev na obeh seznamih, na preseku in uniji. Pri merjenju priključitve preverjamo, ali se termin, ki ga je označil strokovnjak, v točno takšni obliki pojavi tudi med izluščenimi termini, pri čemer vzamemo

¹¹ Obseg neujemanja je sicer pričakovan in bi bil skoraj gotovo še večji, če bi bilo podčrtovalcev še več. Prim. npr. Pearson 1998: 26–28; Vintar 2003: 74; Erjavec, Vintar 2004: 104.

najprej celoten seznam izluščenih kandidatov, nato pa prvih 10.000 in 5.000 enot.

	Priklic na preseku	Priklic na uniji
Vsi izluščeni terminološki kandidati (63.179)	0,93	0,85
Prvih 10.000	0,84	0,72
Prvih 5.000	0,75	0,63

Tabela 7: Priklic izluščenih terminoloških kandidatov.

Rezultati priklica so bili dobri. Od 109 terminov, ki sta jih označila oba strokovnjaka, jih sistem ni izluščil le 7, in sicer gre v treh primerih za angleške izraze (*issue management, press clipping, cluster analiza*), v dveh primerih za petbesedne enote, naši vzorci pa niso presegali dolžine štirih besed (*pragmatična raven odnosov z javnostmi, odzivni razvojni model strateškega načrtovanja*), in v enem primeru za vzorec [*P P*], česar pri izhodiščni izbiri vzorcev nismo predvidevali (*dvosmerni asimetrični*). Kot je pričakovati, priklic pada sorazmerno z "rezanjem" števila kandidatov, vendar dosega pri prvih 5.000 kandidatih (kar je manj kot 10 % celotnega seznama izluščenih) še vedno 0,75 (oz. 0,63), kar je dober rezultat.

4 SKLEP

Pri vsakem ročnem prepoznavanju terminov v besedilu naletimo vsaj na štiri težavne točke (prim. Sliko 2 in Tabelo 6), ki so: meja med terminološko in splošno leksiko, razmerje med terminologizacijo in determinologizacijo, termini več strok ter terminološke kolokacije. Zato je še toliko bolj pomembno, da imamo objektivni kazalnik večje oz. manjše potencialne terminološkosti. Programu, ki deluje na podlagi statističnih izračunov in vnaprej danega, pri tokratnem preizkusu precej obsežnega nabora oblikoskladenjskih vzorcev, smo dali prav tako nalogo: da v korpusu besedil odnosov z javnostmi prepozna strokovno izrazje. S terminološko utežjo, katere

izhodišče je razmerje relativnih pogostosti besed v specializiranem korpusu in splošnem korpusu – v našem primeru KoRP in FidaPLUS – smo dobili več seznamov; v vrhu enega od njih so npr. samostalniki *javnost*, *odnos*, *organizacija*, *komuniciranje*, *podjetje*, *medij* in *znamka*. Pred njihovo dokončno vključitvijo v geslovnik je bila še vedno potrebna analiza s strani področnih strokovnjakov ter jezikoslovcev, a tak pristop kakovostno vendarle močno presega zgolj individualne in večkrat hipne odločitve posameznikov, ki se iskanja terminov lotijo ročno.

Obe primerjavi enobesednih seznamov, tj. primerjava seznama terminološko izluščenih samostalnikov in pogostostnega seznama samostalnikov ter ocena terminološkosti obeh seznamov s strani strokovnjakov odnosov z javnostmi, sta ponovno potrdili prednost izluščenega seznama. Razlike so morda na prvi pogled majhne, vendarle pa so zelo relevantne, saj potrjujejo ravno tisto, kar nas je še posebej zanimalo: občutljivost terminološke uteži. V zvezi s slednjo v nadaljevanju raziskave načrtujemo še en preizkus, in sicer preverbo vpliva strokovnega področja na večjo oz. manjšo uspešnost rezultatov luščenja v razmerju do terminološke uteži. Ta hip se namreč zdi, da je družboslovno področje, kakršno so odnosi z javnostmi, ki obstajajo na presečišču menedžmenta, marketinga in komunikologije (Gruban 1998: 25), zaradi delne tematske prekrivnosti z referenčnimi korpusi (časopisi in revije, deloma tudi knjige z besedili o aktualnih dogodkih, gospodarstvu, poslu, financah ipd.) za luščenje, ki v uteži vključuje ravno tovrstno primerjavo, večji izziv kot katero drugo (naravoslovno, tehnično) specializirano področje ali podpodročje.

Odločitev za luščenje glagolskih vzorcev, sploh v tolikšnem obsegu, se jezikoslovnemu bralcu morda zdi presenetljiva. Presodili smo, da je potencialnim večbesednim glagolskim terminom kot manj obravnavanemu

delu terminologije¹² vredno dati tovrstno pozornost ter z veliko količino podatkov in statistično podprto potrditi ali ovreči njihovo terminološko relevantnost. Z dopuščanjem morebitne specifičnosti katere od strok si na podlagi preizkusa dovoljujemo posplošitev, da z luščenjem glagolskih oblikoskladenjskih vzorcev, kakršnega smo izvedli, za slovenščino ne dobimo relevantnih seznamov z večbesednimi glagolskimi termini (prim. tudi Arhar Holdt 2011: 121–125). Ob možnosti opazovanja tipičnega besedilnega okolja glagolov s katerim drugim orodjem je luščenje glagolskih vzorcev bolj smiselno povsem opustiti.

Veliko boljši kot pri glagolskih vzorcih so bili rezultati luščenja vzorcev s samostalniškimi jedrom. Med njimi sta vzorca [*P S*] ter [*P in P S*] v vrhnjem delu seznama v več kot 60 % dala gradivo, ki je neposredno uporabno za geslovník. Preden posplošimo prikazana razmerja med produktivnostmi vzorcev na celotno slovensko terminologijo, bi bilo dobro povsem enako metodo preizkusiti še na nekaj drugih strokah, se je pa znova potrdilo, da je za slovensko večbesedno terminologijo najbolj tipična zveza vrstnega pridevnika in samostalnika (Vidovič Muha 2000: 26). Odprto zaenkrat ostaja vprašanje, ali je smiselno še povečati dolžino vzorcev na pet, šest ali več besed. V naši analizi smo se ustavili pri štiribesednih, vendar tej meji nismo pripisovali dokončnosti. Dve petbesedni enoti sta kot termina prepoznala področna strokovnjaka, kar kaže, da v odnosih z javnostmi taki termini so. Ena pot do njih je povečanje obsega vzorcev luščenja, druga pa analiza besedilnega okolja.

Analiza priklica je potrdila naše prepričanje, da smo pri izboljšavah luščilnika ter pri širjenju seznama vzorcev na pravi poti, saj je samodejno luščenje pri terminih, ki sta jih označila oba strokovnjaka, doseglo kar 93-odstotni priklic. Po drugi strani pa se skozi nizko ujemanje med obema strokovnjakoma jasno

¹² Žagar Karer (2011: 32–33) sicer opozarja, da se glagoli v večji ali manjši meri pojavljajo skoraj na vseh področjih, in navaja še nekaj slovenskih avtorjev, ki jih obravnavajo, ni pa izrecna, ali to velja tudi za večbesedne glagolske termine.

kaže subjektivnost same definicije terminološkosti ter posledično zahtevnost zastavljenih ciljev. Kompleksnost slednjih je dobro povzel Sager (1998/99), ko je zapisal, da so termini pravzaprav samo besede s specifično funkcijo, ali drugače rečeno (in kot smo nakazali že zgoraj), termini se formalno z ničimer ne ločijo od drugih besed. Če jih želimo samodejno pridobiti iz strokovnih besedil, moramo za to seveda vseeno oblikovati določena – formalna, saj drugačnih še nismo sposobni računalniško obdelovati – merila, ta pa so dejanskim značilnostim terminološkega inventarja v besedilih vselej lahko le grob približek.

ZAHVALA

Raziskava, predstavljena v prispevku, je nastala v okviru projekta *Terminološke baze podatkov kot osnova strokovnih znanj: model za sistematizacijo terminologij* (<http://www.termis.fdv.uni-lj.si/>), ki jo je po pogodbi št. 1000-11-274193 financira Javna agencija za raziskovalno dejavnost Republike Slovenije ter sofinancerja Pristop, d. o. o., in Gospodarska zbornica Slovenije. Projekt so podprli tudi sponzorji: Elektro Ljubljana, d. d., Mercator, d. d., Pošta Slovenije, d. o. o., in Zavarovalnica Maribor, d. d.

LITERATURA

- Arhar Holdt, Š. (2011): *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Atkins, B. T. S., in Rundell, M. (2008): *The Oxford Guide to Practical Lexicography*. New York: Oxford University Press.
- Bergenholtz, H., in Tarp, S., ur. (1995): *Manual of Specialised Lexicography*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Biber, D., Conrad, S., in Reppen, R. (1998): *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Buitelaar, P., Eigner, Th., in Declerck, T. (2004): *OntoSelect: A Dynamic*

- Ontology Library with Support for Ontology Selection. *Proceedings of the Demo Session at the International Semantic Web Conference*. Hirošima.
- Corcho, O., Fernández-López, O., in Ómez-Pérez, A. (2003): Methodologies, Tools and Languages for Building Ontologies: Where is their Meeting Point?. *Data & Knowledge Engineering*, 46 (1): 41–64.
- Čermák, F., ur. (2011): *Korpusová lingvistika*. Praga: Nakladatelství lidové noviny, Ústav Českého národního korpusu.
- De Nicola, A., Missikoff, M., in Navigli, R. (2009): A Software Engineering Approach to Ontology Building. *Information Systems*, 34 (2): 258–275.
- Erjavec, T., in Vintar, Š. (2004): Korpus kot podpora slovarju informacijskega izrazja slovenskega jezika. *Uporabna informatika*, XXII (2): 97–106.
- Erjavec, T. (2013): Korpusi in konkordančniki na strežniku nl.ijs.si. *Slovenščina 2.0*, 1 (1): 24–49.
- Gantar, P. (2007): *Stalne besedne zveze v slovenščini: korpusni pristop*. Ljubljana: Založba ZRC, ZRC SAZU.
- Gantar, P. (2009): Leksikalna baza: vse, kar ste vedno želeli vedeti o jeziku. *Jezik in slovstvo*, 54 (3/4): 69–94.
- Gorjanc, V. (2005): *Uvod v korpusno jezikoslovje*. Domžale: Založba Izolit.
- Gorjanc, V., in Krek, S., ur. (2005): *Študije o korpusnem jezikoslovju*. Ljubljana: Krtina.
- Grčar, M., Krek, S., in Dobrovoljc, K. (2012): Obeliks: statistični oblikoskladenjski označevalnik in lematizator za slovenski jezik. V T. Erjavec, J. Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*: 89–94. Ljubljana: Institut Jožef Stefan.
- Gruban, B. (1998): Izobraževanje v odnosih z javnostmi. V B. Gruban, D.

- Verčič in F. Zavrl (ur.): *Preskok v odnose z javnostmi*: 25–44. Ljubljana: Pristop.
- Halliday, M. A. K., Teubert, W., Yallop, C., in Čermáková, A. (2004): *Lexicology and Corpus Linguistics: An Introduction*. London: Continuum.
- Hanks, P. (2008): *Lexicology*. New York: Routledge.
- Heid, U., in Gouws, R. (2006): A Model for a Multifunctional Dictionary of Collocations. *Proceedings of the XII EURALEX International Congress*: 979–988. Torino.
- Holozan, P. (2006): Dodatne dvoumnosti zaradi popustljivosti analizatorja pri analizi slovenskih stavkov. V T. Erjavec in J. Žganec Gros (ur.): *Jezikovne tehnologije*: 146–149. Ljubljana: Institut Jožef Stefan.
- Kapoor, Bh., in Sharma, S. (2010): A Comparative Study Ontology Building Tools for Semantic Web Applications. *International Journal of Web & Semantic Technology*, 1 (3): 1–13.
- Kilgarriff, A., Rychlý, P., Smrz, P., in Tugwell, D. (2004): The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*: 105–116. Lorient.
- Kilgarriff, A., in Kosem, I. (2012): Corpus Tools for Lexicographers. V S. Granger in M. Paquot (ur.): *Electronic Lexicography*: 31–55. Oxford: Oxford University Press.
- KoRP, korpus besedil odnosov z javnostmi*. Dostopno prek:
http://nl.ijs.si/noske/sl-spec.cgi/first_form?corpname=korp_sl (10. maj 2013).
- Krek, S. (2003): Sodobna dvojezična leksikografija. *Jezik in slovtvo*, XLII (1): 45–60.
- Krek, S., in Kilgarriff, A. (2006): *Slovene Word Sketches*. Dostopno prek:
http://nl.ijs.si/is-ltco6/proc/12_Krek.pdf (28. maj 2013).

- Leech, G. (1992): *Corpora and Theories of Linguistic Performance*. V J. Svartvik (ur.): *Directions in Corpus Linguistics*: 105–122. Berlin: Mouton de Gruyter.
- Logar, N. (2007): *Korpusni pristop k pridobivanju in predstavitvi jezikovnih podatkov v terminoloških slovarjih in terminoloških podatkovnih zbirkah: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- Logar, N., in Vintar, Š. (2008): Korpusni pristop k izdelavi terminoloških slovarjev: od besednih seznamov in konkordanc do samodejnega luščenja izrazja. *Jezik in slovnstvo*, LIII (5): 3–17.
- McEnery, T., in Wilson, A. (1996): *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Pearson, J. (1998): *Terms in Context*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Romih, M., in Holozan, P. (2002): Infrastruktura za razvoj jezikovnih tehnologij – korpus FIDA in sistem ASES. V T. Erjavec in J. Žganec Gros (ur.): *Jezikovne tehnologije*: 166. Ljubljana: Inštitut Jožef Stefan.
- Sager, J. C. (1998/99): In Search of a Foundation: Towards the Theory of the Term. *Terminology*, 5 (1): 41–57.
- Schryver, G. de (2003): Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16 (1): 143–199.
- Sinclair, J. (2004): *Trust the Text: Language, Corpus and Discourse*. London, New York: Routledge.
- Sketch Engine*. Dostopno prek: <http://www.sketchengine.co.uk/> (20. april 2013).
- Teubert, W., in Krishnamurty, R., ur. (2007): *Corpus Linguistics: Critical Concepts in Linguistics*. London, New York: Routledge.
- Uschold, M., in King, M. (1995): *Towards a Methodology for Building Ontologies*. Edinburgh: University of Edinburgh.

- Vidovič Muha, A. (1988): Nekatere jezikovnosistemske lastnosti strokovnih besednih zvez. V B. Pogorelec (ur.): *24. seminar slovenskega jezika, literature in kulture: Zbornik predavanj*: 83–91. Ljubljana: Oddelek za slovanske jezike in književnosti Filozofske fakultete.
- Vidovič Muha, A. (2000): *Slovensko leksikalno pomenoslovje: Govorica slovarja*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Vintar, Š. (2003): *Uporaba vzporednih korpusov za računalniško podprto ustvarjanje dvojezičnih terminoloških virov: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- Vintar, Š. (2008): *Terminologija: terminološka veda in računalniško podprta terminografija*. Ljubljana: Znanstvena založba Filozofske fakultete, Oddelek za prevajalstvo.
- Vintar, Š. (2009): Samodejno luščenje terminologije – izkušnje in perspektive. V N. Ledinek, M. Žagar Karer in M. Humar (ur.): *Terminologija in sodobna terminografija*: 345–356. Ljubljana: Založba ZRC, ZRC SAZU.
- Vintar, Š. (2010): Bilingual Term Recognition Revisited: The Bag-of-Equivalents Term Alignment Approach and its Evaluation. *Terminology*, 16 (2): 141–158.
- Vintar, Š., in Erjavec, T. (2008): iKorpus in luščenje izrazja za Islovar. V T. Erjavec in J. Žganec Gros (ur.): *Jezikovne tehnologije*: 65–69. Ljubljana: Institut Jožef Stefan.
- Žagar Karer, M. (2011): *Terminologija med slovarjem in besedilom*. Ljubljana: Založba ZRC, ZRC SAZU.

TERMINOLOGY OF THE PUBLIC RELATIONS FIELD: CORPUS – AUTOMATIC TERM RECOGNITION – TERMINOLOGY DATABASE

The article describes an analysis of automatic term recognition results performed for single- and multi-word terms with the LUIZ term extraction system. The target application of the results is a terminology database of Public Relations and the main resource the KoRP Public Relations Corpus. Our analysis is focused on two segments: (a) single-word noun term candidates, which we compare with the frequency list of nouns from KoRP and evaluate termhood on the basis of the judgements of two domain experts, and (b) multi-word term candidates with verb and noun as headword. In order to better assess the performance of the system and the soundness of our approach we also performed an analysis of recall.

Our results show that the terminological relevance of extracted nouns is indeed higher than that of merely frequent nouns, and that verbal phrases only rarely count as proper terms. The most productive patterns of multi-word terms with noun as a headword have the following structure: [adjective + noun], [adjective + *and* + adjective + noun] and [adjective + adjective + noun]. The analysis of recall shows low inter-annotator agreement, but nevertheless very satisfactory recall levels.

Key words: Slovene terminology, KoRP public relations corpus, TERMIS project, LUIZ term extraction system

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 2.5 Slovenija.

This work is licensed under the Creative Commons Attribution ShareAlike 2.5 License Slovenia.

<http://creativecommons.org/licenses/by-sa/2.5/si/>

