
KORPUSNA ANALIZA VLOGE OZNAČEVALCEV MEDLEKSEMskih RAZMERIJ V ORGANIZACIJI BESEDILA

V članku predstavljamo označevalce medleksemskih razmerij v metabesedilni vlogi. Označevalci medleksemskih razmerij v metabesedilni vlogi so tisti stalni deli besedil, s pomočjo katerih v besedilu povezujemo elemente pojmovnega sistema, hkrati pa z njimi povezave med elementi pojmovnega sistema tudi komentiramo. V prvem delu članka predstavimo označevalce medleksemskih razmerij, ki smo jih ugotovili za slovenščino, in njihovo uspešnost pri zajemanju dejansko pojmovno povezanih leksikalnih enot iz korpusa, s pomočjo korpusne analize v nadaljevanju pokažemo, da se v vlogi metabesedilnih organizatorjev pojavljajo tipično v strokovnih besedilih, na koncu pa natančneje analiziramo besedilno vlogo izbranih označevalcev in njihovo tipično ubesediljenje.

1 Uvod

Pojmovno medsebojno povezane leksikalne enote v besedilu pogosto nastopajo s podobnimi besedilnimi sopojavnicami, povezujejo pa jih deli besedil, ki jih lahko opazujemo kot stalne besedilne vzorce. Če imamo na voljo dovolj velike količine besedilnih podatkov, lahko tovrstne besedilne vzorce prepoznamo, prav tako pa za posamezne ugotovimo tudi njihovo tipično oziroma prevladujočo besedilno vlogo. Stalni besedilni vzorci tako povezujejo leksikalne enote različnega obsega, ki so na pojmovni ravni medsebojno povezane, bodisi kot pojmovno enakovredne, v razmerju nad- oz. podrejenosti itd.

Z vidika metabesedilne organizacije tako lahko prepoznamo tiste stalne dele besedil, s pomočjo katerih besedilo organiziramo, klasificiramo, interpretiramo oz. ocenjujemo propozicijsko vsebino besedila ali se nanjo odzivamo, vendar k njej ne dodajamo kaj novega (Pisanski 2002: 184). V našem primeru nas zanimajo predvsem besedilni vzorci kot besedilni organizatorji in klasifikatorji, v besedilu na tipičen način namreč povezujejo elemente pojmovnega sistema, hkrati pa z njimi povezave med elementi pojmovnega sistema tudi komentiramo.

2 Metoda in hipoteze

Leksikon določenega jezika je strukturiran kot omrežje pojmov in povezav med njimi, z njihovim ubesediljenjem je besedilo običajno reprezentacija pojmovne realnosti, kot jo glede na predstavljeni besedilni referenčni svet zamejuje posamezno besedilo. Ko torej hočemo v besedilo vpeljati razmerje med posameznimi pojmi ali razložiti razmerja med njimi, pogosto uporabljamo stalne besedilne vzorce, ki smo jih v slovenščini poimenovali označevalci medleksemskih razmerij.¹

Naš namen je najprej predstaviti tovrstne besedilne vzorce in njihovo uspešnost pri zajemanju dejansko pojmovno povezanih leksikalnih enot iz korpusa. Zavedamo se namreč dejstva, da so tovrstni besedilni vzorci razpršeni na različne besedilne vloge, zato predpostavljamo, da vsi ne bodo povezovali leksikalnih enot nedvoumno; zanimalo nas bo torej, kateri so tisti, ki so v besedilih specializirani le za eno vlogo, kateri pa večvlogovni, in tako za pridobivanje medsebojno pojmovno povezanih leksikalnih enot manj uspešni.

Samodejno odkrivanje pojmovnih razmerij med informacijsko bogatimi elementi besedila je pomemben raziskovalni cilj tako pri rudarjenju besedil (angl. *text mining*), iskanju podatkov (angl. *information retrieval*) in odkrivanju znanja (angl. *knowledge discovery*). Vsem trem je skupen cilj, da na podlagi velikih besedilnih zbirk z določenega strokovnega področja, denimo elektronskih izdaj strokovnih revij, z uporabo računalniškojezikoslovnih metod ali strojnega učenja odkrivajo nove, dotlej neznane drobce znanja in še neopaženih povezav med pojmi.

Področje, ki ima še največ pridobiti od takšnih metod, je denimo medicina, kjer je zaradi izjemnega obsega strokovne literature nemogoče, da bi človeško oko obdržalo celovit pregled nad njo in opažalo nove morebitne povezave. Kot je razvidno iz zgodnjih poskusov (Swanson in Smalheiser 1997), je novo znanje mogoče odkrivati že s preprostimi statističnimi metodami računanja korelacij med besedami, ki se pojavljajo v posameznih strokovnih člankih. Njegova odmevna raziskava je denimo pokazala na dotlej neznano vzročno povezavo med pomanjkanjem magnezija in migrenskim glavobolom, kar so kasneje tudi ekperimentalno dokazali.

Bolj jezikoslovno naravnani pristopi uporabljajo leksikalno-skladenjske vzorce, kot je npr. *{Sam BZ}*, **še posebej** *{Sam BZ}*, kar nam pomaga odkrivati razmerje nad- oziroma podpomenskosti, kot pri *Evropske države, še posebej Francija* (Hearst 1998; Finkelstein-Landau in Morin 1999; Agichtein in Gravano 1999).

Sorodno metodo uporabljamo tudi v pričujoči analizi, kjer iskanje medleksemskih vzorcev poteka na podlagi oblikoslovno označenega in lematiziranega korpusa.

¹ Gre za podoben tip metabesedilnih elementov, kot jih v angleščini zasledimo pri kategoriji metabesedil, poimenovanih *code glosses*. V to kategorijo metabesedila namreč sodijo tisti besedilni vzorci, ki se v besedilo vključujejo zaradi predvidevanja tvorca besedila o prejemnikovem vedenju, kot npr. *to je (po)imenovano, z drugimi besedami, to lahko definiramo kot* ipd. (Hyland 2005: 52); v okviru opazovanja metabesedila so bili že prej definirani prav z zornega kota komunikacije med tvorcem besedila in prejemnikom kot tisti, ki pripomorejo premostiti morebitne vrzeli pri prejemnikih besedila in pomagajo razjasniti morebitne pomenske dvoumnosti (Vande Kopple 1984: 85).

Podobna metoda se uporablja tudi kot ena od metod za avtomatsko luščenje terminologije iz besedilnih virov oz. kot eden od postopkov, ki v kombinaciji z drugimi v hibridnih modelih omogočajo avtomatsko pridobivanje terminologije za gradnjo terminoloških podatkovnih zbirk (Bowden idr. 1996; Richardson idr. 1998; Meyer idr. 1999).

Preverili bomo tudi, v kolikšni meri so za slovenščino ugotovljeni označevalci medleksemskih razmerij specifični za strokovno besedilo in posledično besedilno povezovanje leksikalnih enot s terminološko vrednostjo. Predpostavljamo, da je gostota pojavitev tovrstnih besedilnih vzorcev večja v strokovnih besedilih.

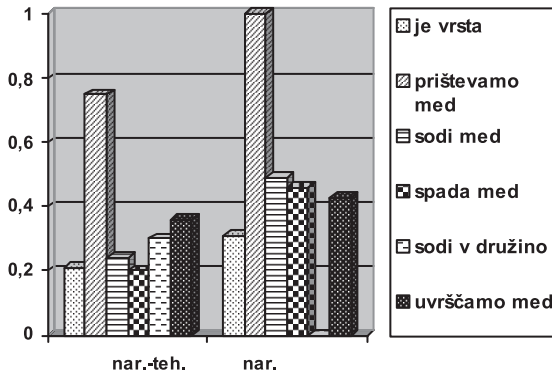
V nadaljevanju bomo posamezne označevalce opazovali glede na njihovo specifično besedilno vlogo in skušali ugotoviti, kako se posamezni tipično obnašajo v besedilu, kot to lahko ugotavljamo s korpusno analizo, in sicer tako korpusa slovenskega jezika *FIDA* (Erjavec idr. 1998) kot nove generacije slovenskega referenčnega korpusa *FidaPLUS* (Arhar in Gorjanc 2007).

3 Označevalci medleksemskih razmerij

Za slovenščino je bila predstavljena metoda prvič uporabljena z namenom ugotavljanja, kateri so besedilni vzorci, ki tipično povezujejo pomensko povezane leksikalne enote, ter kako učinkovito lahko z njihovo uporabo iz korpusa pridobimo podatke o medsebojno povezani terminološki leksiki (Vintar in Gorjanc 2000: 37–44). Na podlagi podkorpusa naravoslovno-tehničnih besedil v korpusu *FIDA* in zgledov v tuji literaturi (Meyer idr. 1999; Pearson 1998: 174–175) so bili za slovenščino ugotovljeni nekateri tovrstni besedilni elementi, ki jih navajamo glede na vlogo v besedilni organizaciji:

- povezovanje ekvivalentnih pojmov: *ali, ali tudi, imenujemo (tudi), imenovan tudi, sinonim, je sinonim za, znan tudi kot, znan tudi pod imenom, je poimenovan, nosi ime ...*
- hierarhično povezovanje pojmov: *je, kot je (na primer), kot je npr., je vrsta, prištevamo med, sodi med, med * sodi, spada med, spada v družino, uvrščamo med, med * uvrščamo, uvrščamo v skupino ...*
- povezovanje pojmov v razmerju del – celota: *ima, ima * dele, je iz, je sestavljen iz, vsebuje ...*

S korpusno analizo je bilo za izbrane tudi ugotovljeno, kako uspešni so pri nedvoumнем zajemanju pojmovno povezane terminološke leksike iz korpusa; tu je bila raziskava omejena na označevalce hierarhičnega razmerja nad- oz. podrejenosti v podkorpusu naravoslovno-tehničnih strokovnih besedil v korpusu *FIDA* (Vintar in Gorjanc 2000), kasneje pa preizkušena uspešnost zajetja pojmovno povezane leksike še z oženjem področja na besedila samo naravoslovnih ved (Gorjanc 2005: 125). Število pojavitev določenega označevalca v besedilih, ko ta nedvoumno poveže leksikalna elementa v razmerju hierarhične nad- oz. podrejenosti, smo delili s številom vseh pojavitev označevalca v besedilu; tako pomeni 1 najvišjo možno, 0 pa najnižjo uspešnost.



Graf 1: Uspešnost zajetja pojmovno povezane leksike z označevalci hierarhično pod- oz. nadrejenih pojmov v podkorpusah naravoslovno-tehničnih in naravoslovnih besedil v korpusu FIDA.

Izkazalo se je, da postane z ožanjem področja zajetje pri nekaterih označevalcih veliko bolj uspešno, najizraziteje pri označevalcu *prištevamo med*, nekaterih pa v strokovno ožjem naboru besedil sploh ne najdemo več, npr. *sodi v družino*; pri večini pa velja, da se je uspešnost zajetja nedvoumnihih medleksemskih povezav precej povečala. Vse to kaže na specializiranost določenih besednih zvez v okviru strokovnega področja.

Kasneje je bila raziskava razširjena tudi na nekatere druge besedilne elemente v vlogi označevalcev medleksemskih razmerij. Izkazalo se je namreč, da v tej vlogi nastopajo ne le besedilni vzorci, ampak je vzorec oblikovan tudi z uporabo neskladenjsko rabljenih ločil, predvsem oklepaja in narekovaja. Tako se je kot eden od pogostejših besedilnih vzorcev za povezovanje ekvivalentnih pojmov pokazal vzorec $\{Sam\}$ $\{(Sam)\}$ oz. $\{Sam/Sam\}$ $\{(Sam/Sam)\}$ (Gorjanc 2005: 123–124).

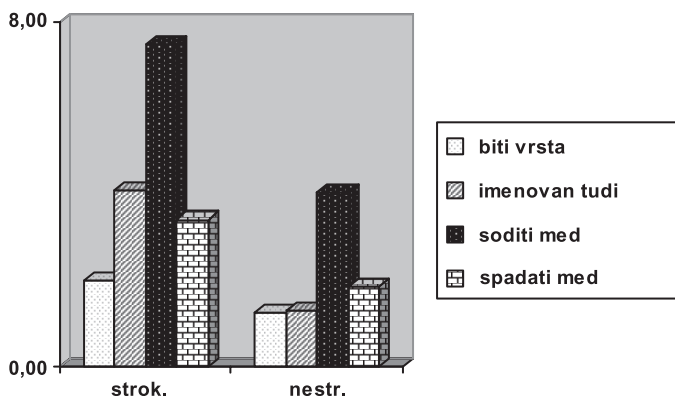
3.1 Označevalci medleksemskih razmerij glede na zvrst besedila

Kot rečeno, je eno od vprašanj, ki smo si ga zastavili v izhodišču, za katera besedila so označevalci medleksemskih razmerij najznačilnejši. Predpostavljamo namreč, da gre za besedilne elemente, ki so značilnejši za strokovna besedila. Za ugotavljanje prisotnosti označevalcev v besedilih glede na zvrst smo naključno izbrali označevalce *biti vrsta*, *imenovan tudi*, *spadati med* in *soditi med* ter ugotovili njihovo pogostost v besedilih različnih podkorpusev v korpusu *FidaPLUS*. Ker podkorpuse glede na število pojavnic v njih (namerno) niso obsegovno identični, smo izračunali pogostost pojavitev izbranih označevalcev glede na število pojavnic v podkorpuseu.

Najprej nas je zanimala njihova prisotnost v umetnostnih besedilih glede na neumetnostna. Ugotovili smo, da je pogostnost pojavljanja v podkorpuseu umetnostnih glede na podkorpuse neumetnostnih v korpusu *FidaPLUS* zanemarljiva. Omenjeni označevalci se namreč v vseh umetnostnih besedilih pojavijo le 623-

krat, v neumetnostnih 56.100-krat, njihova relativna pogostnost glede na število pojavnic v obeh podkorpusih pa je v umetnostnih skoraj za 100-milijonkrat nižja kot v neumetnostnih.

V nadaljevanju nas je v podkorpusu neumetnostnih besedil zanimalo, kako pogosto se omenjeni označevalci pojavljajo v strokovnih besedilih glede na nestrokovna. Tu smo izračunali število pojavitev posameznega označevalca glede na število pojavnic v podkorpusu strokovnih oziroma nestrokovnih besedil.



Graf 2: Razmerje med relativno pogostostjo označevalcev medleksemskih razmerij v podkorpusu strokovnih in nestrokovnih besedil v korpusu FidaPLUS (relativna pogostnost označevalca v 10^{-5}).

Potrdira se je naša predpostavka, da so tovrstni označevalci značilni za strokovna besedila, saj v strokovnih besedilih nanje naletimo v povprečju še enkrat pogosteje kot v nestrokovnih, pri čemer je razlika največja pri označevalcu *imenovati tudi*, ki se v strokovnih besedilih pojavlja več kot trikrat pogosteje kot v nestrokovnih, najmanjša pa pri označevalcu *biti vrsta*, ki je v strokovnih besedilih 1,6-krat pogostejši kot v nestrokovnih. Zanimivo je, da je razmerje pogostnosti v strokovnih glede na nestrokovna besedila tako pri označevalcu *spadati med* kot *soditi med* identično (1,8-krat večja prisotnost v strokovnih kot v nestrokovnih) kljub sicer precejšnji razliki v njuni pogostnosti v besedilih sploh, ki je izrazito v prid označevalcu *soditi med*.

3.2 Označevalci medleksemskih razmerij v organizaciji strokovnega besedila

Kot primere, kakšne podatke o tipičnem ubesediljenju terminologije lahko pridobimo iz korpusa, si oglejmo razmerja pojmovne enakovrednosti, hierarhične nad- in podrejenosti ter meronimije. Za nadaljnjo analizo smo izbrali tiste, za katere je bilo ugotovljeno, da v slovenskem korpusu dajejo boljše rezultate (Vintar in Gorjanc 2000; Gorjanc 2005).

3.2.1 Povezovanje istega pojma z različnimi termini

Ker naj bi bil sam pojmovni sistem določenega strokovnega področja organiziran tako, da bi se izogibal pojmovno enakovrednim izrazom, torej primerom, ko ima isti pojem dva ali celo niz terminov, v strokovnem besedilu ne bi pričakovali pogostih povezav terminov z istim pojmom. Četudi v terminološkem sistemu obstajajo ekvivalentni pari oz. nizi, bi v besedilu pričakovali njihovo izločanje, saj bi lahko delovali kot besedilni stilem, kar je v nasprotju s težnjo strokovnega besedila po natančnosti in jasnosti sporočila. Vendar se tovrstne besedilne vezi tudi v strokovnem besedilu pojavljajo, in sicer po eni strani v primeru, ko gre za novo še neustaljeno terminologijo, po drugi pa v primeru metaterminološkega navezovanja (Gorjanc 1999: 153–154). Tu gre za tipično organizacijo strokovnega besedila, kjer se pojasnjujejo razmerja znotraj terminološkega sistema in so tako tudi predvidljiva lastnost predvsem tistih strokovnih besedil, ki pojasnjujejo terminološka razmerja tudi širši, ne le strokovni javnosti.

Med navedenimi označevalci sta za potrebe korpusne analize s postopki, ki jih uporabljamo, označevalca *ali* in *ali tudi* nezanimiva, saj sta besedilno preveč razpršena na različne besedilne vloge, tako da so rezultati zajetja dveh terminov z istim pojmom zelo slabi. Drugače pa je pri označevalcih, kot so *imenovan tudi* oz. *imenujemo tudi* in *znan kot*.

v astronomiji *teleskop z veliko lečo,	imenovan tudi	objektiv, ki zbira svetlobo z nebesnega objekta in jo
Slika: Kukulcanov tempelj,	imenovan tudi	El Castillo, v Chichén Itzáju, Jukatan, Mehika
oksidant lahko uporabljajo plin CO (ogljikov oksid, včasih	imenovan tudi	ogljikov monoksid) in namesto vodika metanol, ki v
edino naravno jezero. Za matični kras oziroma Kras,	imenovan tudi	Komenski kras, pa je značilna odsotnost površinskih vod.
slovenska beseda. Malo manj znana vrsta je lugat,	imenovan tudi	kukutini. Ta vampir je zelo podoben drugim balkanskim vrstam
bil na položaju izvršnega direktorja. Capellas je bil sočasno	imenovan tudi	za predsednika družbe in izvoljen v upravni odbor. Glavna
za uspeh teoretično največ možnosti folkler C-2,	imenovan tudi	America. C-2 je bila vojaška označba folklerja
bolezen. Povzročajo jo virus iz skupine herpes virusov,	imenovan tudi	Epstein- Barr virus (EBV). Za prenos
Povzročajo jo virus iz skupine herpes-virusov,	imenovan tudi	Epstein- Barr virus (EBV). Za prenos
zanimalo, kako na menstrualno kri reagira črni medved,	imenovan tudi	baribal, ki je bolj krotke narave in v ZDA
učinkovitejše in cenejše osnovne plošče kakor Intelov nabor 82440FX (imenovan tudi	Natoma). V primerjavi z Intelovim naborom, ki
Vzorec z zelo majhnimi črnimi pikicami,	imenovan tudi	pršenje<<, na peresih pernega plašča pri kokoškah
posnetkov istega območja nastane t. i. interferogram,	imenovan tudi	cefranka. Na sliki je območje Fort Irwin v Kaliforniji
računalniško podprt prostorski podatkovni sistem (PPS), pogosto	imenovan tudi	geografski informacijski sistem (GIS). Odločitve bodo morale
sistemu. Sodobna tehnologija – prostorski podatkovni sistem (imenovan tudi	geografski informacijski sistem – GIS) – daje
samodejnim navijanjem. Poganjal ga je kaliber Heuer 12,	imenovan tudi	Chronomatic, ki so ga z združenimi močmi razvile družbe
vratič, Chrysanthemum parthenium, sorta 'Schneeball',	imenovan tudi	materine drobtinice.
pravi kupec, Filip II., vojvoda Orleanski (imenovan tudi	Regent), ki ga je kupil za cesarski dvor
ima svoje slabosti in prednosti. Polje RAID 0 (imenovan tudi	striping<<) pomeni tako organizacijo, da operacijski
3D, 0,18-mikronski Celeron (ponekod	imenovan tudi	Celeron II; Celeron I je pač stari 0,
zraka. Naberejo ga pod trd prvi par kril,	imenovan tudi	pokrovke.
Vitamin B1,	imenovan tudi	iamin, je verjetno najbolj znan med šestimi vitamini skupine
Vitamin B2,	imenovan tudi	riboflavin, je pravzaprav deležen najmanj pozornosti med vitamini.
ga torej povzgnili v vitamin." Koencim Q10,	imenovan tudi	ubikinon, je dejansko v vseh človeških celicah; tam
Za strukturno ravnotežje skrbi pisanolistni klorofit,	imenovan tudi	zelenčica ali kosmuljka.
lockheed F-104 G starfighter, v Nemčiji	imenovan tudi	leteča krsta(1). Idejni oče projekta je

Zgled 1: *Del konkordančnega niza z jedrom konkordance imenovan tudi iz korpusa FidaPLUS.*

Označevalca *imenujemo tudi* in *imenovan tudi* zelo pogosto izločita prave ekvivalente, npr. *dušikov oksid – smejalni plin, vitamin B1 – tiamin, vitamin B2 – riboflavin, dostop do kode – doseg procedure, spletna stran – HTML dokument* ipd. Hkrati pa se izkaže, da ta dva označevalca v besedilu dejansko povezujeta ne le leksikalne ekvivalente, ampak termin in njegovo parafrazo, ki je terminološki opis, npr. *trdine so snovi v trdnem agregatnem stanju; železnata tla so tla, bogata predvsem z železovimi spojinami* ipd. Prav pri tovrstnih besedilnih povezavah pa se izkaže, da se besedila gradijo z vmesnim besedilnim navezovanjem, tako da poimenovanje in parafraza nista v stičnem položaju; v takih primerih konkordančni niz za zajetje obeh elementov ne zadostuje.

Prav tako so značilni pravadninsko gibčni in drzni morski in kopenski kuščarji, ki so doma izključno na Galapaškem otočju. Videti so zelo nevarni, vendar so naravnost mili. Črni morski kuščarji živijo kot edina vrsta kuščarjev na svetu večinoma v morju. Ko iščejo alge in haloge se potapljaajo tudi do deset metrov globoko. Odvečno sol izločajo iz telesa s posebnimi žlezami v nosnicah. Samci morskih kuščarjev (dolgi so več kot meter) se radi v trumah sončijo na toplih skalah. V obdobju parjenja se okrasijo s sijoče pisanimi grebeni in lakotnicami. Nekoliko daljši, ampak precej redkejši legvani imajo rumenkastorjave maroge. Ta samotarski kuščar, **imenovan tudi** žlezoglavi legvan, je v preteklosti zelo trpel zaradi podivjanih mačk in psov.

Infrardeče sevanje, *elektromagnetno valovanje z valovnimi dolžinami med 106 in 103m, ki so nekoliko daljše od valovnih dolžin vidne svetlobe. V spektru je med rdečo svetlobo in mikrovalovi, sevajo pa ga segreta telesa (najbolj takrat, če rdeče ali belo žarijo). Včasih ga **imenujemo tudi** toplotno *sevanje. Energija fotonov v tem delu spektra je primerljiva z energijo nihanja atomov ali molekul v trdnih snoveh. Ko sevanje zadane ob snov, lahko odda svojo energijo nihajočim atomom in jim s tem poveča *amplitudo nihanja. Zato ta vrsta sevanja pogosto segreva telesa, ki jih obseva. Tudi Zemlja se deloma segreva zaradi infrardečega sevanja s Sonca, ki prodre skozi ozračje. Danes pogosto slikajo površje Zemlje iz zraka z infrardečim filmom s pomočjo barvnih filtrov. Na takih infrardečih posnetkih se lepo razločijo vsa področja, ki različno sevajo, na primer različne vrste rastlinstva, kamnite in vodne površine, pa tudi nekatere druge značilnosti pokrajine.

Gre za v strokovnih besedilih predvidljivo navezovanje med deli besedil in imensko zvezo s kazalnim zaimkom v anaforični funkciji. Tako kot je v prvem primeru zgoraj, lahko pride tudi do anaforičnega navezovanja, kjer jedro anaforične besedne zveze poimenuje del besedila, s katerim vzpostavlja besedilno vez; hkrati v vzpostavitvijo besedilne vezi pa se nanosnik tudi poimenuje (Gorjanc 1999: 149).

odzivnik na določeno kodo in oddani signal (med letalci **znan kot** squawk<< krik) bo radar prepoznal, kontrolor
 Nekaj vrst ustvarni pojav, ki je zaradi kričeče barve **znan kot** rdeča plima<<, obenem pa proizvedejo toksine,
 >O). Dušikov oksid je morda najbolj **znan kot** smejalni plin<<, vendar je hkrati izredno močan
 I> (dolgovrati, dolgorepi zgornjejujski zavropod, bolj **znan kot** brontozaver), <I>Brachiosaurus<I>
 električni terapijo kot posebno disciplino, zaradi česar je tudi **znan kot** oče in utemeljitelj elektroterapije. V njej je za stimulacijo
 27, ko so oplenili mesto Rim, dogodek je **znan kot** sacco di Roma.
 a sočasnih EM-polj na človeka je še manj **znan kot** vpliv ene vrste EM-polja. Zato je prevladujoče
 socialno-ekonomskih in verskih razlik. Ta odnos je **znan kot** slavni kompromisni belgijski stil<<.
 Malta ima enodomni parlament, **znan kot** Predstaviški dom. Brez bonusnih<< sedežev ima trenutno

sovjetskega tipa. Poskus reformiranja komunističnega sistema leta 1968, **znan kot** Praška pomlad<<, se je ponesrečil, ker
 ki so ga tedaj poimenovali isletin, danes pa je **znan kot** inzulin. Po zaslugi inzulina, ki ga sintetično proizvajajo
 znatni kot spremembe magnetnega polja. Ta pojav in je **znan kot** magnetna nevihta, ki traja nekaj ur. Delci s
 prihodnosti grozi zaradi aidsa. Njegov način delovanja je postal **znan kot** soglasna diplomacija<<.
 padla v četrtek, 24. oktobra, ki je **znan kot** "črni četrtek". Zlom so v naslednjih dneh
 rojen leta 1836, umrl 1891.) in postal **znan kot** "veseli monarh", saj je imel zelo rad

Zgled 2: *Del konkordančnega niza z jedrom konkordance znan kot iz korpusa FidaPLUS.*

Tudi označevalec *znan kot* daje podobne rezultate kot *imenovan tudi*, prav tako lahko opazimo podobno ubesediljenje, izpostavlja tako razmerje med leksikalnimi ekvivalenti kot tudi med termini in opisi; te povezave pa so prav tako pogosto nestične z uporabo vmesnega anaforičnega navezovanja. Konkordančni niz izpostavi še en zanimiv tip metaterminološkega navezovanja, tj. navezovanje izimenskega termina na področju naravoslovno-tehničnih strok (Košmrlj Levačič 1998: 143–145). Pri tovrstnem povezovanju se kot najpogostnejši označevalec pojavlja sicer označevalec *nosi ime*:

John MacAdam, torej. Že ko mu je bilo 14 let, se je z rodne Škotske odselil naravnost k ne ravno ubožnemu stricu v New York. Tam ga je res čakala Amerika in po tem, ko si je v stričevi računovodski hiši že pri sedemindvajsetih ustvaril celo premoženje, se je leta 1783 vrnil v svoje škotsko mesto Ayr. Tam se je začel ukvarjati z gradnjo cest in v zgodovino se je zapisal kot inženir, po katerem danes **nosi ime** makadam. S svojimi cestami je zaslovel daleč naokrog. Zaradi kakovosti cest, ki jih je gradil iz gramoza, vezanega z zablatenim peskom na utrjeno podlago, so ga Angleži kmalu nagradili z delovnim mestom glavnega nadzornika vseh metropolitanskih angleških cest. Njegove ceste so bile tudi nekoliko konveksne, da je voda čim hitreje odtekla in ni pronicala v podlago. Do konca 19. stoletja je bila večina glavnih evropskih cest makadamiziranih oziroma narejenih po njegovem receptu.

Najbrž bo vsakdo hitro ugotovil, da zanimivi in dragoceni okrasni kamen tanzanit **nosi ime** po Tanzaniji, in ne bo se motil. Lepe neznane modre kristale so odkrili leta 1967 v Tanzaniji jugovzhodno od Kilimandžara, vendar so se znanstveniki skoraj deset let prepirali, nekateri so ga pesniško imenovali modri zaklad Afrike<<, drugi so ga imeli za safir, tretji za cordierit. Šele leta 1976 ga je znano draguljarsko podjetje Tiffany iz New Yorka končno le imenovalo po nahajališču, kajti šele takrat so zanesljivo ugotovili, da je mineral po kemijski sestavi zoisit, ki mu modro barvo daje element vanadij. Po kemijski sestavi je tanzanit kalcijev aluminijev silikat, ki ima trdoto od 6 do 6 po Mohsovi trdotni lestvici, kar je edina njegova slaba lastnost, saj se brušene ploskve zato lahko opraskajo in izgubijo sijaj. Če kamne ogledujemo z več strani, zaradi pleohroizma lahko opazimo tri barve: vijoličasto, modro in rjavo ali rumeno. Primerki, ki jih kopljejo v edinem rudniku tanzanita na svetu v kraju Arusha v tanzanijskem orovju Merelani, so sivi, zeleni, rumeni ali rjavi.

Kot smo že omenili, so pogosti označevalci terminoloških ekvivalentov tudi ločila v svoji neskladenjski vlogi, predvsem narekovaj in oklepaj; v besedilu zaznamujeta ekvivalente tako, da se v narekovaju ali oklepaju pojavi ekvivalent, ki je manj

pogost, še ne ustaljen ali tujejezični (Gorjanc 1996: 256–257). Tudi v korpusu smo skušali poiskati tovrstne pare s pomočjo omenjenih dveh ločil, a se je izkazalo, da je predvsem oklepaj mnogofunkcijski, tako da analiza ne da zelenih rezultatov. Če pa iskanje zožimo le na podkorpus naravoslovnih besedil (koda Cobissa *Naravoslovne vede*) in na stični položaj dveh samostalnikov, so rezultati vzpodbudni.

gensko zdravljenje lahko na široko odprlo vrata popravljanju človekovega dednega	materiala (genoma)	in krojenju zelenih lastnosti naših potomcev. V bistvu je
zlijeta. Pri tem se združita materin in očetov dedni	material (DNK)	za nastanek novega živega bitja, ki bo združevalo lastnosti
(Intelligenten Pulse Decodierung), kar pomeni inteligentno dekodiranje	impulzov (signalov)	V NF-dela vezja signal obdeluje mikrokontroler
kosov lahke, 2 mm debele balze. Nepremične dele	repov (stabilizatorje)	in krmila povežemo s kratkimi koščki bakrene žice 0
pa bi bil naboj lahko skoncentriran samo v sestavnih delih	protona (kvarkih)	od katerih bi se elektroni včasihodbili pod zelo
obrazna mimika, premiki rok, nog in drugih delov	telesa (motorika)	kašljanje in vzdihovanje ipd.
dalnega leta 1953 odkrila in objavila, da je deoksiribonukleinska	kislina (DNK)	osrednji nosilec genetskih informacij pri človeku
vrste iz te skupine vsako leto še po nekaj deset	gnezd (srakoper)	druge pa le eno samo (belorepec, klinčak
molekularnem nivoju spol določa t. i. testis determinirajoči,	faktor (TDF)	ki je genetsko kodiran na kratkem kraku spolnega kromosoma
ni močede potisniti niti rezila noža. Pačakuti, deveti	inka (vladar)	je leta 1438 ukazal prenoviti svojo
kjer te ogrožene vrste domujejo. Svetovni sklad za divje	živali (WWF)	je leta 1990 objavil razvrstitvevmočij, kjer so kategorije
na letališčih. V Kip svobode spustijo le še do	vznožja (nog)	Višji del kipa so zaprli za obiskovalce
možganskih celic. Te redke možganske celice s svojimi dolgimi	izrastki (aksoni)	ki služijo kot topovske cevi, vbrizgajo dopamin v
SIM in centru za preverjanje naročnikov, ter naključno določena	številka (RAND)	Postopek preverjanja se izvede pred vsako prijavo mobilnega terminala

Zgled 3: *Del konkordančnega niza, pridobljen z iskalnim pogojem {Sam} {(Sam)} iz podkorpusa naravoslovnih besedil v korpusu FidaPLUS.*

Ko tak konkordančni niz še ročno pregledamo, tako da nam ostanejo primeri z nedvoumnimi ekvivalenti, se izkaže, da se pri oklepaju kot označevalcu pojmovne ekvivalentnosti v besedilu največkrat pojavijo leksikalizirani ekvivalenti, npr. *rdeče krvno telesce – eritrocit, karbolna kislina – fenol, odnašanje prsti – erozija* ipd., hkrati tudi povezave termina in njegovega besednozveznega opisa, npr. *kvark – sestavni del protona, genom – človekov dedni material* ipd., prav tako pa so pogosti tudi primeri, kjer se kot eden od ekvivalentov v paru pojavlja besedilna aktualizacija, npr. *čas ženitve – spomladi*. Besedilni vzorec se tako izkaže kot učinkovit za zajemanje pojmovno ekvivalentnih parov iz besedila. Po večini pri njih ne gre za neustaljeno terminologijo, kot smo na začetku predvidevali, ampak za ustaljeno, ekvivalenti so predvsem v razmerju dveh terminov, kjer je eden v paru prevzet, drugi pa tvorjen v slovenščini, in pari, ker je eden od terminov kratica.

3.2.2 Hierarhično povezovanje pojmov

Razmerje nad- in podrejenosti, v okviru slovarske urejevalnosti nad- in podpomenskosti, je urejevalno slovarsko načelo tudi v smislu dejanske hierarhizacije sveta (Vidovič Muha 2000: 174). Tudi za terminologijo je izjemno aktualno, saj je eno temeljnih urejevalnih načel pojmovnega sistema.

Pri tovrstnih označevalcih se predvsem označevalec nad-/podrejenega razmerja *med* * *spadajo* pokaže kot učinkovit pri zajemanju celotnega niza podpomenk (Vintar in Gorjanc 2000), gre torej za označevalec, pri katerem bi lahko rekli, da je v veliki meri specializiran za izločanje vzporednih podrejenih terminov:

Med **leptone** spadajo **elektron, mion, delec tau** in njihovi pripadajoči **nevtrini** ter **antidelci vseh naštetih delcev**.

Med **enokaličnice** spadajo **palme, trave, bambusi, kukavice, perunike in lilije**.

Med **gliste** spadajo številni zajedavci in škodljivci, kot so talne glistice, rudarska glista, pljučna glista in številne **filarije**, ki pri sesalcih povzročajo elefantijazo.

Med **kožkrilce** spadajo še **čmrlji, ose, sršeni in mravlje**.

3.2.3 Pojmi v razmerju del – celota

Pri tovrstnem razmerju se ponekod uporablja tudi pojem kvazi hiponimija, saj gre za povezavo podobnega tipa kot pri hierarhični pod- oz. nadrejenosti, ki pa je ne moremo jezikovnosistemske definirati, saj v tovrstne relacije vstopajo običajni leksikalni elementi predmetnega sveta z urejevalnim načelom del – celota (Malmkjær 1996: 301). Ker je svet, ki nas obdaja, zgrajen po tem načelu, je predvsem za naravoslovno-tehnične vede to eno temeljnih urejevalni načel v terminologiji.

, ki so značilne za *notranje planete. Led	je sestavljen iz	vodikovih spojin s kisikom, dušikom in ogljikom.
Lidar	je sestavljen iz	oddajnika svetlobe, sprejemnika sipane svetlobe, detektorja za
5 ns in imajo energijo 6 mJ. Sprejemnik lidarja	je sestavljen iz	treh paraboličnih zrcal z goriščno razdaljo 41 cm,
, da bo delo pravočasno končano. Vsak 3D lik	je sestavljen iz	poligonov. Čim več je poligonov, tem natančneje
S-Line	je sestavljen iz	dveh paketov. Športni paket, tega je imel
pari listov ravno tako nasprotno nameščeni. Drugi par listov	je sestavljen iz	treh lističev, tretji par iz petih in tako
varnost podatkov in hitrejšo izdelavo varnostnih kopij. Paket LiveVault	je sestavljen iz	posebnega strežnika LiveVault Server, ki s posebnimi agenti
,6 in 1,8 grama loja. Loj	je sestavljen iz	maščob, beljakovin in odmrlih žlezlih celic. Loj
dolžine 5,6 m in razpona kril 4 m	je sestavljen iz	osmih glavnih delov, ki jih med seboj spajajo
dolžine 5,6 m in razpona kril 4 m	je sestavljen iz	osmih glavnih delov, ki jih med seboj spajajo
Oracle BPEL Process Manager	je sestavljen iz	treh delov. Prvi in glavni del je seveda
Mozaik severne polarne kape Marsa	je sestavljen iz	391 slik, ki jih je posnel orbitalni del
Matrix	je sestavljen iz	štirih delov. V uporabniškem delu opravljamo vse operacije
barvo premorejo do 410 kubičnih centimetrov prostora. Tiskalni mehanizem	je sestavljen iz	tiskalne glave, podobne kakor pri brizgalnikih, rezervorja
zdrujevalni osi obe sili spet združita. Mehanski del menjalnika	je sestavljen iz	štirih kompletov zobnikov, ki omogočajo štiri stopnje hitrosti
Merilnik	je sestavljen iz	osnovne naprave oziroma izzvalne posode v aluminijastem kovčku in
Mozaik fotografij površine Merkurja	je sestavljen iz	desetih posnetkov Marinerja 10. Nastale so malo po
Mezon	je sestavljen iz	enega kvarka in enega antikvarka.
(slika 5-10). Najpreprostejši svetlobni mikroskop	je sestavljen iz	dveh leč. Prvo imenujemo objektiv, drugo pa
Voziček firme Miller	je sestavljen iz	nosilne prečke in valja za verige. Debla se

tekočine, na katerega delujejo nekateri fizikalni zakoni. Model	je sestavljen iz	približno 1000 trikotnikov, ki so povezani v 100
ali pa so razviti posebej za uporabnika. Vsak modul	je sestavljen iz	enega ali več manjših funkcionalnih sklopov imenovanih granule.
IP modul	je sestavljen iz	PHY ethernet 1. level gonilnika, poznanega iz
terabitov (10 trilijonov bitov). EJ-MOSFET	je sestavljen iz	dveh delov.
vsebuje dva H-mosta. En H-most	je sestavljen iz	štirih NPN-tranzistorjev. Ti delujejo kot stikala
na vlaklu in navitiji na progi. Pogonski linearni motor	je sestavljen iz	superprevodnih magnetov na vozilu, ki ustrezajo rotorju v
v bližnji preteklosti nemogoče. Sistem za usmerjanje potiska motorja	je sestavljen iz	treh oglikovih deflektorjev, ki tehtajo skupaj 47 kg

Zgled 4: *Del konkordančnega niza z jedrom konkordance je sestavljen iz iz korpusa FidaPLUS.*

Nabor označevalcev tovrstnih relacij je v primerjavi z drugimi sorazmerno majhen, hkrati pa se izkaže, da so najpogostnejši označevalci tudi zelo uspešni pri zajemanju terminologije v tovrstnem razmerju. Pri besedni zvezi *je sestavljen iz se* v podkorpusu naravoslovno-tehničnih besedil pokaže, da je njena besedilna vloga v glavnem omejena na vzpostavljanje povezav med leksikalnimi elementi, ki so v medsebojnem razmerju del – celota.

4 Sklep

Postopek zajemanja pojmovno medsebojno povezanih terminoloških enot se je pokazal kot sorazmerno uspešen, ko na tak način pridobivamo podatke iz (pod)korpusa strokovnih besedil. Pokazali smo, da so označevalci medleksemskih razmerij značilni prav za strokovna besedila. Tu se pojavljajo v vlogi organizatorjev leksikalnih elementov, ki so povezani kot del pojmovnega sistema določenega strokovnega področja, obenem pa z njimi povezave med elementi pojmovnega sistema tudi razlagamo in komentiramo. Avtomatsko pridobivanje podatkov o pojmovno povezanih leksikalnih enotah iz korpusa je pri posameznih označevalcih medleksemskih razmerij različno uspešno, pri tistih, kjer je uspešnost višja, lahko z nadaljnjo analizo natančneje opazujemo njihovo vlogo v organizaciji strokovnih besedil, saj s predstavljenimi postopki iz korpusa pridobimo dovolj relevantnih podatkov za nadaljnje opazovanje njihove besedilne vloge in tipičnega ubesediljenja.

Gotovo ostaja veliko povezav, ki jih na predstavljeni način nismo uspeli zajeti. Predlagani model je samo eden od možnih postopkov pridobivanja podatkov o metabesedilni organizaciji, kjer kot organizatorji besedila nastopajo označevalci medleksemskih razmerij. Iskanje besedilnih vzorcev, ki bi zajeli tudi tiste termine, ki nastopajo v manj predvidljivem besedilnem okolju in v nestičnem besedilnem položaju, na katera smo opozorili, se bo lažje nadgradilo takrat, ko bo korpus oblikoslovno bolj nedvoumno označen in bo omogočal večjo kombinacijo iskanj tudi glede na skladišne vzorce. Prav ugotavljanje tipičnih skladišnih vzorcev v korpusu in njihova uporaba pri zajemanju terminološko relevantnih korpusnih enot se je pri vzporednih korpusih že pokazala kot sorazmerno uspešna (Vintar 1999: 167).

Literatura

Agichtein, E. in Gravano, L., 1999: Snowball: Extracting relations from large plain-text collections. *Columbia University Computer Science Department Technical Report CUCS-033099*.

Arhar, Š. in Gorjanc, V., 2007: FidaPLUS – nova generacija slovenskega referenčnega korpusa. *Jezik in slovnstvo* 2. 95–110.

Bowden, P. R., Halstead, P. in Rose, T. G., 1996: Extracting Conceptual Knowledge from Text Using Explicit Relation Markers. *Proceedings of EKAW-96*. Nottingham: University of Nottingham. 147–162.

Čermák, F., 2004: Besedilna uvajalna sredstva rečenic in drugih stalnih besednih zvez. *Slavistična revija* 1. 71–88.

Erjavec, T., Gorjanc, V. in Stabej, M., 1998: Korpus FIDA. *International Multi-Conference Information Society – IS'98*, 6–7 October 1998. Ljubljana: Institut Jožef Stefan. 124–127.

Finkelstein-Landau, M. in Morin, E., 1999: Extracting semantic relationships between terms: Supervised vs. unsupervised methods. *International Workshop on Ontological Engineering in the Global Information Infrastructure*. 71–80.

Gorjanc, V., 1996: Terminologija novejših naravoslovno-tehničnih strok (Ob primeru računalništva in jedrske fizike). Vidovič Muha, A. (ur.): *Jezik in čas*. Ljubljana: Znanstveni inštitut Filozofske fakultete. 15–260.

Gorjanc, V., 1999: Kohezivni vzorec matematičnih besedil. *Slavistična revija* 2. 139–159.

Gorjanc, V., 2005: *Uvod v korpusno jezikoslovje*. Domžale: Izolit.

Hearst, M. A., 1998: Automated discovery of Wordnet relations. Fellbaum, Ch. (ur.): *WordNet: An Electronic Lexical Database*. Cambridge (MA): MIT Press. 131–151.

Hylland, K., 2005: *Metadiscourse*. London, New York: Continuum.

Košmrlj Levačič, B., 1998: Izimensko strokovno izrazje. Humar, M. (ur.): *Slovensko naravoslovno-tehnično izrazje*. Zbornik referatov s posvetovanja o slovenskem naravoslovno-tehničnem izrazju. Ljubljana: Založba ZRC SAZU. 139–146.

Malmkjær, K., 1996: *The Linguistic Encyclopedia*. London, New York: Routledge.

Meyer, I., Mackintosh, K., Barriere, C. in Morgan, T., 1999: Conceptual sampling for terminological corpus analysis. Sandrini, P. (ur.): *Proceedings of TKE '99*. Dunaj: TermNet. 256–267.

Pearson, J. 1998: *Terms in Context*. Amsterdam: John Benjamins.

Pisanski, A., 2002: Analiza nekaterih metabesedilnih elementov v slovenskih znanstvenih člankih v dveh časovnih obdobjih. *Slavistična revija* 50/2. 183–197.

Pisanski, A., 2005: *Konvencije rabe metabesedilnih elementov*. Doktorska disertacija. Ljubljana: Univerza v Ljubljani, Filozofska fakulteta.

Richardson, S. D., Dolan W. B. in Vanderwende L., 1998: MindNet: acquiring and structuring semantic information from text. *Microsoft Research Technical Publications (MSR-TR-98-23)*. Dostopno na <ftp://ftp.research.microsoft.com/pub/tr/tr-98-23.doc> [13. 8. 2007].

Saeed, J. I., 1998: *Semantics*. Oxford (UK), Cambridge (USA): Blackwell.

Swanson, R. in Smalheiser N. R., 1997: An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence* 91. 183–203.

Vande Kopple, W. J., 1985: Some exploratory discourse on metadiscourse. *College Composition and Communication* 36. 82–94.

Vidovič Muha, A., 2000: *Slovensko leksikalno pomenoslovje. Govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete.

Vintar, Š. in Gorjanc, V., 2000: Identifying semantic markers for Slovene. *Strani jezici* 1–2. 37–44.

Vintar, Š., 1999: Računalniško podprto iskanje terminologije v slovensko-angleškem vzporednem korpusu. *Uporabno jezikoslovje* 7–8. 156–169.

Korpusa

Korpus slovenskega jezika FIDA. <<http://www.fida.net>>. (Dostopno 6. 8. 2007.)

Referenčni korpus slovenskega jezika FidaPLUS. <<http://www.fidaplus.net>>. (Dostopno 6. 8. 2007.)