
KONCEPT MREŽNEGA POMENSKEGA SLOVARJA SLOVENSКИH BESED

Prispevek predstavlja možnost povezav in pregledovanja besed po pomenih, ki v tiskanih slovarjih ni možna. Opisane so prednosti in možnosti, ki jih nudi tako organiziran slovar, in predstavljena je delovna verzija mrežno urejenega slovarja slovenskih besed v računalniško dostopni obliki. Namen predstavljenega mrežnega slovarja je uporabniku ponuditi informacije o posameznih besedah ne le v smislu lematizacije (slovarskega gesla, v katerem je navedena besedna vrsta obravnavane besede), ampak tudi pomensko povezane besede širšega in ožjega pomena. V opisani delovni verziji je v pomensko mrežo urejenih nekaj nad 5000 besed, večinoma samostalnikov. Delovna verzija je na voljo uporabnikom na zgoščenci, ki jo zainteresirani bralci lahko dobijo od avtorja na njegovem naslovu.

Ključne besede: mrežni slovar, slovenski slovar, hierarhija pomenov, drevo odločitev, pomenske zveze, osebek/predmet

1 Uvod

Združevanje besed po pomenih nima samo teoretičnega, ampak tudi praktičen pomen. Najpreprostejša možnost uporabe je delo s slovarjem, v katerem so gesla urejena ne le po abecednem redu, ampak so pri vsakem poleg razlage pomena (pomenov) dodane še informacije o geslih, ki sodijo v njegov *širši* pa tudi *ožji* pomenski krog. Na primer pri besedi *slon* je dobro vedeti, da sodi v pomenski krog *sesalcev*, ti pa naprej med *vretenčarje*, ali pa, da sodita v *ožji* oziroma bolj specifičen pomenski krog *afriški* in *indijski slon*. Dandanes posamične informacije o pomenu ali o pripadnosti širšim in ožjim pomenskimi krogom katerega koli gesla

ni težko dobiti. Če drugega ne, je na voljo svetovni splet. Problem nastane, ko potrebujemo informacije o velikem številu besed v zelo kratkem času, tj. takrat, ko ročno iskanje postane neprikladno oziroma nemogoče, kot npr. pri računalniškem razumevanju besedila. Računalnik prebira besede drugo za drugo, opravi besedno in morda tudi stavčno analizo, vendar s tem ne doseže »razumevanja« sporočila. Pri računalniškem »razumevanju« ne gre za razumevanje besedila v dobesednem pomenu besede, ampak za avtomatizirano luščenje osnovnih informacij v smislu »kdo, kdaj, kje, kako in zakaj«. V tem prispevku je opisan model računalniško podprtega slovarja pomenov, ki lahko do informacij o pomenskih povezavah besed dostopa v realnem času. Pri računalniškem razumevanju besedila sta besedna in stavčna analiza v veliko pomoč, saj računalnik sproti, ne glede na oblike, čase, spole, sklone, vide, osebe itd., pretvarja besede v *leme* oziroma v slovarska gesla, tako da je možno njihove pomene poiskati po lematizirani obliki.

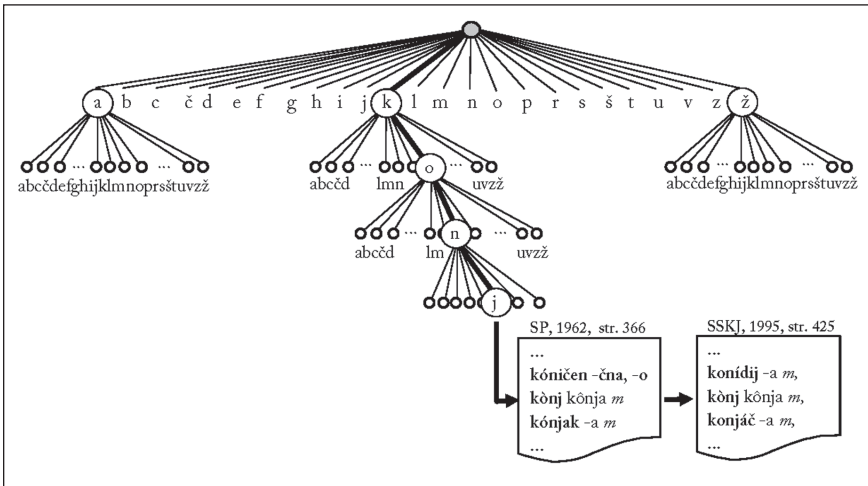
Glavna hipoteza računalniškega razumevanja besedil je predpostavka, da sta nabora pomenskih skupin besed v dveh besedilih z različnima temama različna. Izbira besed je sicer odvisna od avtorja in njegovega sloga, vendar si morajo biti besedila, ki obravnavajo isto ali podobno tematiko, kar se tiče pomenskih skupin besed precej bolj podobna kot besedila, ki obravnavajo povsem različne tematike.

Členjenje besed v pomenske skupine ni enoličen ali lahko določljiv postopek. Besede, če niso tehnični ali strokovni izrazi, imajo navadno več pomenov in se uporabljajo v različnih tematskih in pomenskih povezavah. Vsaka pomenska skupina lahko zajema različno širok krog besed in ima lahko tudi pomenske podskupine. Te se lahko med seboj prekrivajo in/ali delijo naprej v še manjše podskupine. Na drugi strani se pomenske skupine združujejo v večje, širše množice besed s skupnim pomenskim imenovalcem. Na primer beseda *riba* je v pomenski skupini *živali*, ki vsebuje celo vrsto podskupin, kot so *sesalci*, *ptiči*, *izumrle živali*, *domače živali* in podobne, pri čemer se skupina *domače živali* delno prekriva s *sesalci* in deloma s *ptiči*. Če bi dodali še pomensko skupino *domači ljubljenci*, bi lahko med *domače živali* uvrstili tudi *ribe* in še nekaj drugih skupin. Posamezne podskupine so lahko glede notranje sestave med seboj povsem različne. Za slovar pomenov je važno, da nudi uporabniku neko osnovno shemo pomenskih členitev oziroma razporeditev besed v pomenske skupine, hkrati pa omogoči, da si lahko pomenske skupine izbere, sestavi in uredi sam.

Prikaz različnih pomenskih podskupin za vsako geslo ni edina naloga slovarja pomenov. Zelo pomembno je tudi vzporedno zasledovanje in prikazovanje poti, ki vodijo od gesla k različnim pomenskim skupinam, s katerimi je geslo asociirano oziroma pridruženo. Ker imajo nekatere besede veliko različnih pomenov, je pri avtomatski obdelavi besedil potrebno sprotno sporočanje vseh pomenskih povezav za vse besede iz besedila.

2 Hierarhična (drevesna) povezava besed

Slovarji v knjižni obliki so, ne glede na svoj namen, vedno urejeni hierarhično, se pravi v drevesni strukturi. Hierarhija ali drevo je sistem povezav, ki je podoben verižni reakciji. Iz vsake točke (križišča ali vozla) se odcepi več poti. Na koncu vsake poti je vozlel, ki se ponovno razcepi. Ker imamo v slovenščini 25 črk, je slovar razdeljen na 25 poglavij. V hierarhiji oziroma drevesni strukturi jim pravimo *veje*. V vsaki od petindvajsetih začetnih vej slovarja so vse besede, ki se začno z isto črko. In v vsaki so besede abecedno urejene po drugi črki. Abecedna ureditev se nadaljuje v vseh besedah po vseh črkah.



Slika 1: Abecedna ureditev slovarja, ki ima drevesni (hierarhični) dostop do posameznih gesel. Sosednja gesla iskane besede v slovarju večinoma niso pomensko močno povezana. Na sliki sta prikazani obe sosednji besedi gesla konj pri dveh različnih slovarjih.

Abecedna ureditev omogoča hierarhičen oziroma drevesni dostop do vsake besede. Hierarhično iskanje po slovarju kaže **Slika 1**. Začetno razvejišče ima 25, vsako zopet nadaljnjih 25 vej, včasih tudi kakšno manj itd. Črke *q*, *w*, *x*, in *y* smo zanemarili, ker v *Slovarju slovenskega knjižnega jezika* (SAZU 1995) s temi črkami ni besed niti za eno stran. Besed v slovarju ne iščemo zaporedno, ampak se ob vsaki črki besede odločamo, kje nadaljevati glede na to, katera črka sledi prejšnji. Teoretično moramo narediti toliko odločitev, kolikor je v besedi črk. V praksi se proti koncu besede ni treba več odločati in listati po slovarju, ampak poiščemo besedo kar z iskanjem po vrsti.

Drevesna ureditev ima dve značilnosti. Prva je ta, da je v posameznih vejah število besed zelo različno. Zaradi tega najdemo nekatere besede hitro, druge iščemo nekoliko dlje. V vsakem primeru je iskanje relativno hitro, saj v množici skoraj 150 000 gesel najdemo katero koli izmed njih v nekaj sekundah. Druga značilnost pa je v tem, da besede, ki stoje skupaj, večinoma nimajo medsebojnih pomenskih

povezav. V najboljšem primeru je pomenska povezava šibka. Če pogledamo sosednji besedi gesla *konj*: *koničen* in *konjak*, ki ju navaja SP (SAZU 1962: 366) ali *konidij* in *konjač*, kot ju navaja SSKJ (SAZU 1995: 425) (**Slika 1**), vidimo, da je težko govoriti o pomenski povezanosti. Da ne bi naredili krivice, je treba povedati, da so v dobrih slovarjih različni pomeni posameznih besed tudi navedeni. Redkeje so navedene širše ali ožje pomenske skupine, v katere beseda sodi. Tako so v SSKJ pri geslu *konj* navedeni še pomeni *šahovski konj*, konj kot *telovadno orodje* in konj kot *neroden človek*. Pri širših in ožjih pomenskih skupinah bi bilo seveda nespametno od slovarjev pričakovati toliko informacij, kot jih lahko dobimo o posameznih pojmi, stvareh ali bitjih v strokovnih priročnikih ali na svetovnem spletu. Preglednost gesla terja pri knjižni obliki slovarja velike omejitve.

Ne glede na to, koliko informacij je dodano posameznemu geslu, ostaja drevesni ureditvi njena druga slabost, da gesla, ki so razvrščena po abecedi, med seboj niso povezana. Kaj mislimo z izrazom *povezna gesla*? Za pojasnilo bomo ponovno vzeli geslo *konj* in navedene pomene *domača žival*, *šahovska figura* in *telovadno orodje*. Najprej je očitno, da to niso slovarska gesla v uveljavljenem smislu, ampak dvobesedni opisi skupin, ki v *širšem* pomenu zajamejo pomen gesla. V prvi skupini so še *krave*, *pujsi*, *psi*, *mačke* itd., v drugi so *kralj*, *kraljica*, *trdnjava*, *lovec* itd. in v tretji *bradlja*, *drog*, *koza*, *krogi* itd. Za slovar bi bilo zelo nenavadno, če bi bile pri geslu, ki poimenuje večjo skupino predmetov ali osebkov, npr. *perutnina*, *mesec* ali *poklic*, naštete vse besede, ki jih geslo združuje (npr. vse, kar je perutnina, pri prvem, imena mesecev pri drugem, vsi poklici pri tretjem itd.). To je po eni strani škoda, po drugi pa bi vpeljava naštevanja vseh pomensko sorodnih besed ali celo dvobesednih gesel pripeljala do preobsežnih slovarjev in do ponavljanja opisov. Dvo- ali večbesedna gesla niso prepovedana, so pa toliko nedefinirana oziroma toliko različnih vrst, da se jih vsi slovarji izogibajo.

3 Slovarska mreža ali mrežno urejen slovar

Pravkar opisana težava povezovanja in razvrščanja besed in gesel v pomenske skupine se da s pomočjo računalniških programov rešiti tako, da pri tem ne izgubimo osnovnih informacij, ki jih slovarji nudijo. Pri računalniško vodenih slovarjih število gesel, ne glede na to, ali so eno- ali večbesedna, nima praktično nikakršne vloge. Računalniki omogočajo dostop do posameznega eno- ali večbesednega gesla neposredno, ne da bi iskali po hierarhiji zaporedja črk. Neposreden dostop pomeni, da računalnik iz gesla, ki ga išče, *izračuna* naslov, na katerem je v računalniškem spominu geslo shranjeno. Ker so za računalnik tako črke kot številke izražene binarno (z ničlami in enicami), je slovarsko geslo kot binarno številko treba z matematično pretvorbo samo spraviti v okvir naslovov, ki so določeni za shrambo slovarja. Pri tem je povsem vseeno, ali gre pri geslu za eno ali več besed.

Najpomembnejša sprememba pri organizaciji slovarja, ki jo predlagamo, je uvedba povezav med posameznimi gesli. V hierarhično zasnovanem slovarju so gesla povezana samo z abecednim redom. Celotna informacija o vsakem geslu je na mestu, kjer je geslo v slovarju shranjeno. Drugih povezav ni. Pri hierarhiji vodi od vrha do vsakega gesla ena sama pot. Prav tako gre nazaj od gesla proti vrhu ista pot v obrnjeni smeri (**Slika 1**). V trenutku, ko med gesli, ki so shranjena na različnih mestih, uvedemo povezave, se hierarhija pretvori v mrežo. Vsako geslo (eno- ali večbesedno) predstavlja v mreži vozle. Značilnost mreže je v tem, da je do večine vozlov možno priti po več poteh. Prav tako je možno iz vozla oditi po različnih povezavah.

Če geslo *konj* povežemo z geslom *sesalec*, tega pa z geslom *žival*, smo ustvarili povezavo (zanko), ki omogoči, da računalniški slovar dá informacijo, da je konj sesalec in žival, hkrati pa tudi informacijo, da je med živalmi skupina sesalcev, med katerimi je tudi konj. Takoj je očitno, da mora h geslu *žival* voditi toliko povezav, kolikor različnih živali je navedenih v slovarju, h geslu *sesalec* pa vodijo le povezave od tistih gesel, ki opišejo sesalce. Obratno vodijo od gesla *konj* le tiste povezave, ki dajejo temu geslu različne pomene. Očitno je tudi, da so skupine, h katerim vodi več povezav, splošnejše od tistih z manj povezavami.

Mreža povezav nima prioritete smeri, zato so vse povezave, ki vodijo iz vozla ali k njemu, enakovredne. Če mreži določimo prioriteto smer, lahko za vsako povezavo med dvema vozla povemo, ali pelje navzgor, navzdol ali vodoravno. Na primer pri ribiški mreži, ki visi z enim vozlom pritrjena na vrvi, sta smeri navzdol in navzgor, po katerih prehajamo od enega vozla do drugega, očitni. Prav zaradi tega, ker so pomenske skupine bolj ali manj splošne, je v slovarsko mrežo treba uvesti razlikovanje smeri od posebnega k splošnemu in obratno. Ko pride računalnik do gesla (vozla), mora pregledati različne povezave in če smer do vsake povezave ni določena, ne more vedeti, katere povezave vodijo k splošnejšim in katere k ožjim pomenskimi skupinam. Zaradi tega je v slovarski mreži pomenov potrebno izbrati najvišje mesto – *vrh* ali *vrhnji vozle*, ki ustreza korenu hierarhične ureditve. Z izbiro vrhnjega vozla je določena prioriteta smer mreže. Prioritetna smer vodi od gesla proti vrhu k skupinam, ki združujejo bolj splošne opise, obratna smer pa *proč* od vrha pomenov, k bolj natančnim podskupinam z vedno bolj specifičnim pomenom. S tem je ohranjena informacija, kaj je ožji in kaj širši pomen gesla in kako si pomenske skupine s širšim ali z ožjim pomenom sledijo.

Zaradi različno organiziranih in različno širokih pomenskih skupin je bilo nujno uvesti večbesedna gesla. Npr. geslo *žival* je enobesedno, a je izredno splošno, saj zajema vse živali, navedene v slovarju. Pri razdelitvi take pomenske skupine na ožje pomene moramo uvesti podskupine. Te so lahko enobesedne, npr. *sesalec*, *ptič*, *riba*, *dvoživka* itd., ali večbesedne, npr. *domača žival*, *divja žival* itd. Dvobesedna gesla so možna kjer koli v mreži. V mrežnem slovarju so gesla enakomerno razpršena po celotni mreži. Zaradi številnih povezav med gesli (pomenskimi skupinami) ta v mrežnem slovarju ne morejo biti urejena po abecedi kot pri hierarhiji, kjer

je abecedni red *edina* povezava med dvema gesloma. Njihovo mesto si lahko predstavljamo v središču za njih najbolj značilne pomenske skupine, fizično pa je lahko kjer koli. Računalnik najde geslo neposredno po vnosu in ga z vsemi pomenskimi povezavami sporoči uporabniku. Manjši problem je z iskanjem manj znanih dvo- ali večbesednih gesel. Če je le-to splošno znano, ga bomo našli, če pa ne vemo, da obstaja, ga ni mogoče iskati. Ker želimo imeti v mrežnem slovarju čim več pomenskih povezav, se manj znanih gesel nabere veliko. Večbesedna gesla oziroma skupine, npr. *priimek znane osebe* ali *sestavni del vozila*, bomo v mrežnem slovarju našli ob »potovanju« po mreži oziroma po računalniškem zaslonu, ki prikazuje povezave v dovolj široki okolici gesla.

Usmerjena mreža je kompromis med hierarhijo in čisto mrežo, ki nima usmeritve. Njena slabost je bolj zapleten sistem vodenja informacij o vsakem vozlu (geslu). Skupaj z informacijo o povezavi, tj. na katere vozle oziroma gesla je posamezen vozec povezan, je treba voditi še informacijo, ali vodi povezava proti vrhnjemu vozlu (tj. proti splošnejšim skupinam) ali navzdol po mreži, tj. proti bolj specifičnim, ožje definiranim pomenskim skupinam. Na tem mestu se v računalniško problematiko in tehnično izvedbo rešitve za opisano organizacijo ne bomo spuščali.

Poleg bolj zapletenega vodenja informacij o povezavah med gesli in pomenskimi skupinami je pri pomenskem slovarju še težava, ki pa je k sreči zaznavna samo pri grajenju. Vse povezave med gesli in pomenskimi skupinami je namreč treba vnašati ročno, kar je zamudno. Če k temu prištejemo še dejstvo, da so načela, po katerih naj bi tvorili, imenovali in členili pomenske skupine, zelo individualna in ni nekih splošnih pravil, je razumljivo, zakaj takih slovarjev še nimamo.

Da je mrežna organizacija pomenov potrebna in tudi zaželena, se nakazuje že v navadnih slovarjih s tem, ko so pri geslu velikokrat kot razlaga naštetih nekateri člani iste pomenske skupine ali pa je navedena informacija, v katero širšo pomensko skupino pomenov geslo sodi, na primer s tem da je povedano, da je geslo *rastlina* ali *mineral* ipd. Pri organizaciji mrežnega slovarja je treba upoštevati tudi prekrivajoče se pomenske skupine. Na primer v geslu (pomenski skupini) *poklic* morajo biti povezave na vse poklice in med njimi tudi na gesli *lovec* in *kmet*. Ti dve gesli pa sta prav tako povezani na pomensko skupino oziroma dvobesedno geslo *šahovska figura*. Podobnih primerov je veliko. Nove povezave lahko v mrežnem slovarju delamo kadar koli in na katero koli geslo oziroma pomensko skupino, ne glede na to, kolikokrat in kam je bilo geslo povezano že pred tem. Problemi prekrivajočih se pomenskih skupin se pojavljajo pri prikazovanju gesel z velikim številom povezav na računalniškem zaslonu, saj se mora uporabnik sproti odločati, koliko posameznih gesel želi videti in v katero smer bo šel s pregledom naprej.

4 Opis manjšega slovarja pomenskih povezav 5000 besed

Trenutno je mrežni slovar, ki vsebuje približno 5000 besed, sestavljen tako, da gredo iz najvišjega mesta, ki se imenuje *vrh pomenov*, štiri veje. Ker je vrh pomenov najvišja točka mreže, gredo vse povezave od njega samo »navzdol«. V tem je povsem podoben začetni razdelitvi navadnega slovarja, ki ima 25 vej. Omenjene štiri povezave so usmerjene proti štirim skupinam:

1. *osebek/predmet*,
2. *opis predmeta/osebka*,
3. *povedek* in
4. *opis povedka*.

Skupina *Osebek/predmet* trenutno vsebuje večino gesel in pomenskih skupin. V drugi skupini so predvideni pridevniki. Razdeljena je v tri podskupine: 2.1 *Kakovostni*, 2.2 *Vrstni* in 2.3 *Svojilni pridevniki*. Trenutno je v njej le okrog 100 testnih primerov. Pri pridevniki je poleg omenjene členitve predvidena tudi povezava na ustrezen samostalnik oziroma ustrezne samostalnike. V tretji skupini so predvideni glagoli, ki naj bi bili urejeni podobno kot v sistemu, ki ga predvideva Levinova (Levin 1993: II. del), vendar prilagojeni slovenskemu besedotvorju, v četrti skupini pa so prislovi. Ker je težišče dela na členjenju samostalnikov, sta tretja in četrta skupina trenutno prazni. Glede na raznolikost pomenskih povezav v najrazličnejših besedilih je dana tudi možnost, da lahko poleg omenjenih štirih skupin dodamo še kakšno, če bi to zahteval bolj podroben in natančen pomenski opis posameznih stavčnih delov. Trenutno je v pomenski podatkovni bazi v prvi skupini (*osebek/predmet*) okrog 5000 samostalnikov. Ta skupina je razdeljena na sedem pomenskih podskupin:

- 1.1 *lastno ime*,
- 1.2 *živa narava*,
- 1.3 *neživa narava*,
- 1.4 *naravni produkt*,
- 1.5 *človeški izdelek*,
- 1.6 *pojem* in
- 1.7 *zaimek*.

Vsaka skupina se deli še naprej. Skupina 1.1 *Lastno ime* je razdeljena na sedem podskupin različnih vrst lastnih imen. Skupina 1.2 *Živa narava* vsebuje deset pomenskih podskupin, in sicer 1.2.1 *Človek*, 1.2.2 *Žival*, 1.2.3 *Rastlina*, 1.2.4 *Celični nivo*, 1.2.5 *Skupina ljudi*, 1.2.6 *Skupina živali*, 1.2.7 *Skupina rastlin*, 1.2.8 *Del telesa*, 1.2.9 *Del rastline* in 1.2.10 *Mitološko bitje*. Vsaka podskupina se deli naprej toliko časa, dokler v njej niso samo samostojna gesla, ki ne zahtevajo nadaljnje členitve na ožja pomenska področja.

Katero koli geslo lahko postane pomenska skupina, saj je pri vsakem vozlu v slovarju predvidenih po 1000 možnih povezav oziroma nadaljevanj v obe smeri. V začetku gradnje pomenskega slovarja v nobeni skupini ni niti približno tolikšnega

števíla gesel. Se pa že v opisanem primeru kaže pri skupinah *moških*, *ženskih* in *geografskih imen*, da bi bila ta omejitev lahko dosežena. Novo pomensko skupino vpeljemo takrat, ko ob dodajanju novega gesla ne najdemo primerne obstoječe skupine, ali pa takrat, ko ocenimo, da je obstoječo skupino smiselno razdeliti na dve manjši.

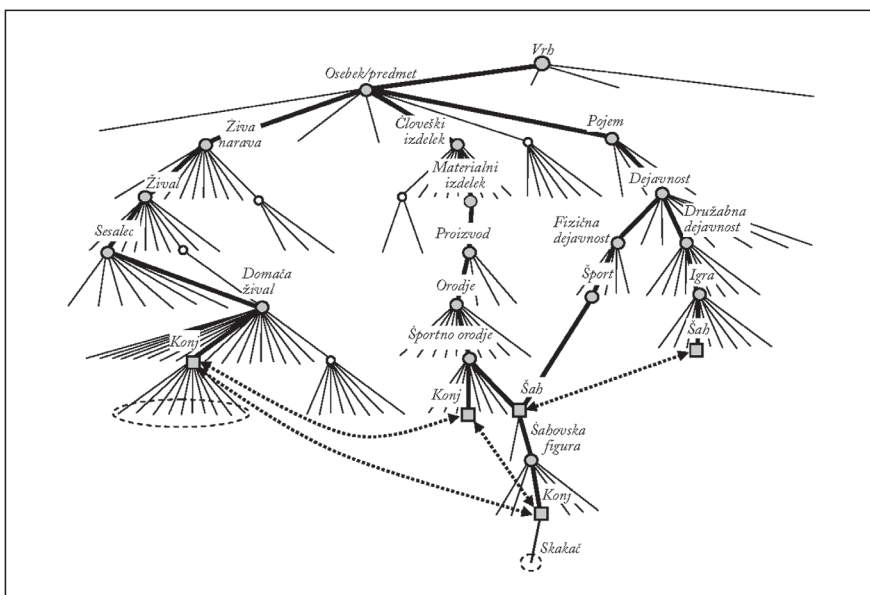
Z vsako razdelitvijo skupine na podskupine se število možnih povezav in število poti, ki vodijo od ene pomenske skupine do druge, večá. Z dodajanjem gesel se struktura slovarja pomenov spreminja, tako, da se število povezav večá, povprečna dolžina poti med geslom in vrhom pa se spreminja zelo malo. Skoraj vsako geslo je namreč možno uvrstiti v neko že obstoječo pomensko skupino. Šele ko vnašamo geslo, ki ga ne moremo uvrstiti v nobeno od obstoječih skupin, in moramo zaradi tega narediti novo, se podaljša tudi povprečna pot med geslom in vrhom. Razumljivo je, da se v začetku tvorjenja slovarja število različnih pomenskih skupin hitro večá, kasneje, ko je pomenski slovar že v veliki meri oblikovan, pa število novih skupin narašča počasneje.

Trenutno je v slovar pomenov vključenih nekaj nad 300 pomenskih skupin in podskupin z zelo različnimi števili gesel. Najštevilnejša pomenska skupina je 1.1.1 *Ime osebe*, ki vsebuje 336 imen, od tega 127 v skupini 1.1.1.1 *Moško ime* in 130 imen v skupini 1.1.1.2 *Žensko ime*. Razliko do 336 besed vsebujejo štiri manjše skupine: 1.1.1.3 *Ime za oba spola*, 1.1.1.4 *Priimek znane osebe*, 1.1.1.5 *Moški* in 1.1.1.6 *Ženski vzdevek*. V povprečju vsebujejo pomenske skupine med 10 in 20 gesel. Navadno lahko že pri skupini, ki šteje več kot deset gesel, pričakujemo, da je možno nadaljnje strukturiranje, kar pomeni razdelitev na nekaj ožjih (se pravi manj številčnih) pomenskih skupin. Pri sedanji količini besed in povezav je povprečno število pomenskih skupin *med* geslom in vrhom manj kot deset. Najdaljša pot je dolga štirinajst povezav. Z večanjem števila povezav med posameznimi pomeni se večá število *različnih* poti med geslom in vrhom pomenskega slovarja. Za analizo besedila je pomembno, da mrežni slovar odkrije in predstavi uporabniku vse možne poti, ki vodijo preko najrazličnejših pomenskih skupin do vrha.

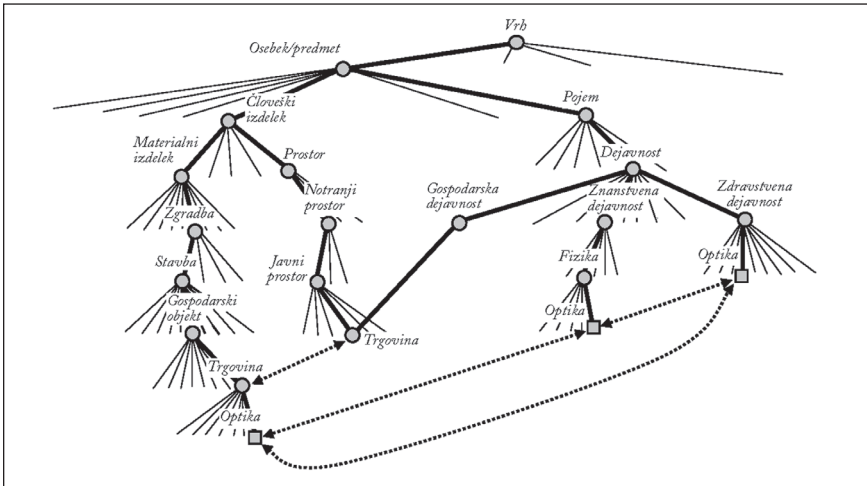
Na **Sliki 2** je prikazana obstoječa mreža povezav gesla *konj* na vse pomenske skupine. Ker je celotna mreža slovarja preobsežna, je prikazan samo majhen del mreže tistih povezav in vozlov, ki so za ta primer pomembne. To so pomenske skupine, ki leže na vseh poteh med geslom *konj* in vrhom *pomenov*. Zaradi nazornosti je mreža prikazana kar se da podobno hierarhiji oziroma drevesu. *Konj* je slovarsko geslo (lema) in je hkrati tudi večja pomenska skupina. Od gesla *konj* vodi pet povezav. Tri gredo proti širšim pomenskim skupinam, tj. navzgor proti vrhu pomenov, dve pa navzdol v ožji pomenski področji, se pravi proč od vrha. Prve tri povezave smo že omenjali: *domača žival*, *šahovska figura* in *športno orodje*. Na **Sliki 2** so povezave gesla *konj* v teh treh skupinah označene z debelo črtkano črto. V prvi navzdol usmerjeni skupini so zbrana gesla, ki jih lahko v širšem pomenu označimo kot *konj-domača žival*, npr: *vranec*, *pram*, *lipicanec*,

šimel, kljusa, kasač, galoper, povodnik itd. Iz naštetih gesel je razvidno, da bi se ta skupina v bodoče lahko naprej delila v ožje pomenske skupine npr.: *konj-barva*, *konj-pasma*, *konj-gibanje* ipd. V drugi navzdol usmerjeni skupini je eno samo geslo *skakač*.

Vseh pet poti vodi preko različnih pomenskih skupin proti skupini *osebek/predmet* in od nje na *vrh*. Pot običajnega pomena besede *konj* gre preko petih pomenskih skupin (*domača žival*, *sesalec*, *žival*, *živa narava* in *osebek/predmet*) do vrha. Pomen besede *konj* v smislu telovadnega orodja gre preko skupin *športno orodje*, *orodje*, *proizvod*, *človeški izdelek* in *osebek/predmet*. Pot besede *konj* v pomenu šahovske figure je za dve pomenski skupini daljša od prejšnjih dveh in ima poleg tega tri variante, ker sreča na poti geslo *šah* s tremi pomeni: (a) *šah* kot *športno orodje*, (b) *šah* kot *šport* in (c) *šah* kot *igra*. V celoti zajame pomenski slovar ob besedi *konj* na petih različnih poteh proti vrhu 19 pomenskih skupin. Pri podajanju besed v ožjih pomenskih skupinah (smer navzdol) pa poda dve že zgoraj omenjeni skupini. Na **Sliki 2** sta gesli *konj* in *šah* narisani, kot da bi bili na različnih mestih, v resnici pa sta to samo dva vozla v mreži, vsak s tremi povezavami navzgor. Šah ima v povezavi navzdol dve gesli: *šahovska figura* in *šahovnica*.



Slika 2: Primer mrežnih povezav pomenskih skupin v mrežnem slovarju. Geslo *konj* v mreži ni na treh (potemnjeni kvadrati), ampak le na enem mestu, kar simbolizirajo črtkane povezave. Pomenske skupine, povezane z geslom *konj*, so označene z večjimi temnimi krogi. Ožji skupini gesla *konj* sta označeni s črtkanima ovaloma. Mrežni slovar poda pri pomenskih povezavah za geslo *konj* tudi tri pomenske povezave za geslo *šah* (*igra*, *šport* in *športno orodje*).



Slika 3: Prikaz povezav, ki so v mrežnem slovarju za pomen gesla optika. Od gesla optika (kvadratek) vodi do vrha mreže pet različnih poti (odebeljene črte) preko 16 različnih pomenskih skupin (potemnjeni krogi).

Kot drugi primer je na **Sliki 3** prikazanih pet poti različnih povezav pomenov od gesla *optika* do vrha mrežnega slovarja. Glavna delitev je pri samem geslu, saj je *optika* opredeljena najprej kot veja *fizike*, potem kot *zdravstvena dejavnost* in nazadnje kot *trgovina* s specifičnimi izdelki. Nazadnje omenjena veja se pri prehodu skozi pomensko skupino *trgovina* razdeli na tri poti, ker ima ta v slovarju tri pomena. Prva gre preko trgovine kot *stavbe* oziroma *materialnega izdelka*, druga preko trgovine kot *javnega prostora*, tretja pa preko trgovine kot *gospodarske dejavnosti*. Pomenska skupina *dejavnost* je ena najboljšežnejših skupin v mrežnem slovarju, saj trenutno obsega sedemnajst različnih vrst dejavnosti. Poleg naštetih treh na **Sliki 3** (*gospodarska*, *znanstvena* in *zdravstvena dejavnost*) sta na **Sliki 2** označeni tudi *fizična* in *družabna dejavnost*.

Na koncu bi omenili še, da je v pomenski skupini *osebek/predmet* predvidena tudi skupina osebnih zaimkov (pomenska skupina 1.7), ker ti lahko nastopajo v stavku kot osebek.

4 Primer uporabe

Mrežni pomenski slovar ima možnih veliko različnih uporab. V prvi vrsti gre za široka tematska iskanja po slovenskem korpusu besedil, ki se sedaj formira v okviru projekta *Sporazumevanje v slovenskem jeziku* (Amebis, d.o.o. v konzorciju 2008). Iskanja po korpusih zasledujejo dva cilja. Prvi je iskanje pogostosti posameznih fraz, izrazov, besednih oblik, novosti, narečnih posebnosti, podobnih jezikovnih oblik itd., drugi cilj, ki pride do polnega izraza šele, ko korpus obsega zadosten izbor primernih besedil v nekem jeziku, je identifikacija besedil ali večjih delov

(odstavkov, poglavij) z dano tematiko. Iskanje besedil z določeno tematiko je pogosta dejavnost iskalnikov po svetovnem spletu, novinarskega poizvedovanja in iskanj po različni strokovni literaturi, ki je v računalniško berljivi obliki na voljo v vedno večjem obsegu. Morda velja posebej omeniti možnost tematskega iskanja v obstoječi zakonodaji, ki je že sedaj dostopna v računalniško berljivi obliki.

Konkordančna iskanja lahko izvajamo z različnimi iskalniki, ki so na voljo na spletu, npr. konkordančno iskanje v *Novi besedi* (<<http://bos.zrc-sazu.si>>) ali dostopni kako drugače (Zupan in Čeh 2008), *FidaPLUS* (<<http://www.fidaplus.net>>), *IMS Corpus Workbench* (<<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench>>), *WordSmithTM* (<<http://www.lexically.net/wordsmith/>>) ipd. Glavna slabost omenjenih iskalnikov je v omejitvi nabora besed in oblik, s katerimi dovoljujejo iskanje po korpusu. Večina je omejena na nekaj deset besed oziroma njihovih oblik in na nekatere dopolnilne pogoje pri iskanju, kot so oblike IN/ALI in IN/IN, ko gre za iskanje več besed hkrati. Strategija njihovega iskanja je primerjava besed in pogojev iskanja, ki jih predpiše uporabnik s tekočo besedo pregledovanega besedila. Če se pogoji ujemajo, iskalnik sporoči zadetek. Strategija iskanja s pomočjo mrežnega slovarja pomenov je drugačna. Uporabniku ni treba navesti vseh besed, ki naj jih računalnik išče, ampak navede samo pomensko skupino in računalnik bo avtomatično našel vse tekste, ki vsebujejo katero koli besedo iz izbrane skupine, v kateri koli slovnični obliki. Uporabnik z enim ukazom aktivira primerno pomensko skupino, po potrebi tudi več. Če išče npr. besedila s tematiko plemstva v Sloveniji, lahko aktivira pomenski skupini *plemiški naziv* in *geografsko ime v Sloveniji*. Posledično bodo vse besede korpusa, ki sodijo v ti dve pomenski skupini, aktivirale števec zadetkov. Uporabniku ni treba skrbeti za to, ali je navedel vse nazive od barona, baronice in baronese, preko grofa, grofice, kneza in kneginje do cesarja, ali če je morda pozabil na princeso. Mrežni slovar bo našel in označil vsa besedila, katerih besede bodo šle po poti, na kateri leži pomenska skupina *plemiški naziv*. Glede na pogostost zadetkov lahko besedilo z dovolj visoko frekvenco uporabnik izloči kot potencialni zadetek ali pa ga vsaj bežno pregleda. V primerjavi s tem, da bi moral vsako besedo korpusa primerjati s stotinami, lahko tudi tisoči besed, ki so zajete v pomenske skupine, predstavlja iskanje z mrežnim slovarjem časovni prihranek glede na navadno konkordančno iskanje po besedah, za najmanj faktor 100. To pa se pri velikih korpusih že močno pozna.

Kot je bilo že omenjeno, je pomenski slovar nujno potrebno opremiti z dodatnimi programskimi pomagali, s katerimi lahko uporabnik iz obstoječih pomenskih skupin oziroma z lastnim izborom zgradi specifične pomenske skupine in jih nato pri delu stalno uporablja.

5 Sklep

Prikazana je zasnova in delovanje prototipnega mrežnega pomenskega slovarja, ki združuje slovarska gesla v pomenske skupine, organizirane v usmerjeno mrežo

povezav. Trenutno je v obstoječi mreži približno 5000 besed, povezanih v okrog 300 pomenskih skupin. Besedni zaklad pomenskega slovarja večamo s hitrostjo od 500 do 1000 besed na mesec. Pri tem ne gre samo za dodajanje novih gesel, ampak predvsem za preurejanje starih in dopolnjevanje ter dodajanje novih pomenskih skupin in podskupin. Glavni nameni projekta so pokazati na možnost izdelave celotnega pomenskega slovarja za slovenščino, pokazati uporabnost računalniško podprtega pomenskega slovarja in vzpodbuditi možnost izgradnje »standardizirane« baze pomenskih skupin, podobno kot je to narejeno za angleške glagole (Levin 1993: II del.). Vzporedni namen je tudi preveriti zanimanje za tovrstno delo v slovenski strokovni in laični javnosti in vzpodbuditi odgovore ter dobiti nasvete za nadaljevanje začetega dela in morda tudi za sodelovanje.

Težave pri nastajanju pomenskega slovarja so večplastne. V prvi vrsti gre za pomanjkanje ustrezne programske podpore, saj tak slovar dobi svojo pravo uporabnost šele z uporabniku prijazno odzivnostjo. Druga težava je v samem konceptu in členitvi pomenov. Pri tem ne gre toliko za pomensko členitev na najširšem oziroma na najvišjem nivoju, ampak za členitve oziroma razčlenitve večjih skupin v manjše, ki se prekrivajo preko različnih nivojev in preko različnih pomenskih skupin. Omejitve so v iskanju kompromisa med neko splošno, lahko bi ji rekli »naravno« ali »povprečno sprejemljivo« členitvijo na posamezne pomenske skupine in specifično, individualno, uporabiško usmerjeno členitvijo. Ocenjujemo, da se lahko neka »povprečna« ali »splošno sprejemljiva« členitev pomenov izkaže ali ustali šele po daljšem časovnem usklajevanju, izpopolnjevanju, dopolnjevanju z raznovrstno uprabo in ob izmenjevanju izkušenj. Zgledovanje po podobnih tujih programih je do neke mere lahko uspešno. Opisana struktura nekoliko sledi strukturi *OpenCyc* (<<http://www.cycfoundation.org/>>) in angleškemu slovarju *VisuwordsTM* (<<http://www.visuwords.com>> z bazo *WordNet* (Fellbaum 1998), vendar se primerjave ustavijo prav pri implementaciji povezav med gesli, ki so za vsak jezik tako specifične in težko prevedljive, da je neposreden prenos ali samo prevod v veliki meri neizvedljiv.

Viri in literatura

Amebis, d.o.o., Kamnik, Institut »Jožef Stefan«, Univerza v Ljubljani, Znanstvenoraziskovalni center SAZU, Trojina, zavod za uporabno slovenistiko, 2008–2013: konzorcij projekta *Sporazumevanje v slovenskem jeziku*, ki ga financirata RS MŠŠ in Eu SS.

Fellbaum, Christiane (ur.), 1998: *WordNet. An Electronic Lexical Database*. Princeton: MIT Press.

IMS Corpus Workbench (CWB): <<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/>>. (Dostop 29. 07. 2009.)

Korpus slovenskega jezika FidaPLUS: <<http://www.fidaplus.net.>>. (Dostop 29. 07. 2009.)

Levin, Beth, 1993: *English Verb Classes. A Preliminary Investigation*. Chicago: The University of Chicago Press.

Lexical analysis software for the PC, Oxford University Press, 1996, version 5.0: <<http://www.lexically.net/wordsmith/>>. (Dostop 29. 07. 2009.)

Nova beseda – Besedilni korpus na Inštitutu Frana Ramovša, ZRC SAZU: <http://bos.zrc-sazu.si/s_beseda.html>. (Dostop 29. 07. 2009.)

SAZU, 1961: *Slovenski pravopis*. Ljubljana: DZS.

SAZU, 1995: *Slovar slovenskega knjižnjega jezika*. Ljubljana: DZS.

The Cyc foundation: <<http://www.cycfoundation.org/>>. (Dostop 29. 07. 2009.)

Visuword™, On-line graphical dictionary and thesaurus: <<http://www.visuwords.com/>>. (Dostop 29. 07. 2009.)

Wikivir, prosta zbirka virov v slovenščini: <<http://sl.wikisource.org/wiki/>>. (Dostop 29. 07. 2009.)

Zupan, Jure, in Čeh, Blaž, 2008: *Navodila za uporabo računalniškega programa SLON-13* (Analiza slovenskih besedil). DP-KI 2448. Ljubljana: Kemijski inštitut.

Zahvala

Delo je preko ARRS v sklopu projekta J7-0382 in programa P01-0017 sofinanciralo Ministrstvo za visoko šolstvo, znanost in tehnologijo Republike Slovenije.