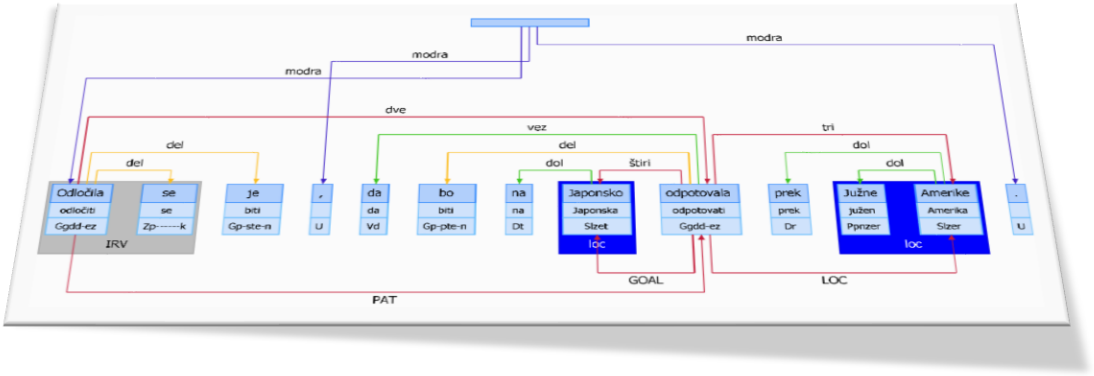


Q-CAT

Orodje za ročno označevanje in analizo besedilnih korpusov

Priročnik za uporabo

Različica 1.0, 25. 10. 2019

Avtorica priročnika: Kaja Dobrovoljc

Avtor programa: Janez Brank

KAZALO

1 O PROGRAMU

Orodje Q-CAT (*Querying Supported Corpus Annotation Tool*) je računalniški program za ročno označevanje besedilnih korpusov, ki poleg pripisovanja in pregledovanja oznak različnih tipov podpira tudi naprednejša iskanja po označenih povedih.

Prva različica programa, takrat še pod imenom SentenceMarkup, je nastala v okviru projekta Sporazumevanja v slovenskem jeziku (2008–2013) za ročno skladenjsko razčlenjevanje in označevanje imenskih entitet v pilotnih učnih korpusih za slovenščino. V okviru projekta Nova slovnica sodobne standardne slovenščine: viri in metode (2017–2020) je bil program izdatno nadgrajen tako, da omogoča dodajanje poljubnega števila označevalnih ravni različnih tipov, dinamično prilagajanje nastavitev in kompleksnejša iskanja, nadgrajene ali dodane pa so bile tudi številne druge funkcije, ki omogočajo prijaznejšo uporabniško izkušnjo.

Orodje Q-CAT, čigar funkcije so podrobneje predstavljene v tem priročniku, tako danes poleg jezikoslovnih analiz referenčnega ročno označenega korpusa za slovenščino ssj500k in njegovih nadaljnjih izboljšav podpira označevanje in/ali analizo kateregakoli drugega besedilnega korpusa, ne glede na jezik in izbrani nabor jezikoslovnih ali drugih oznak.

2 NAMESTITEV

2.1 PRENOS

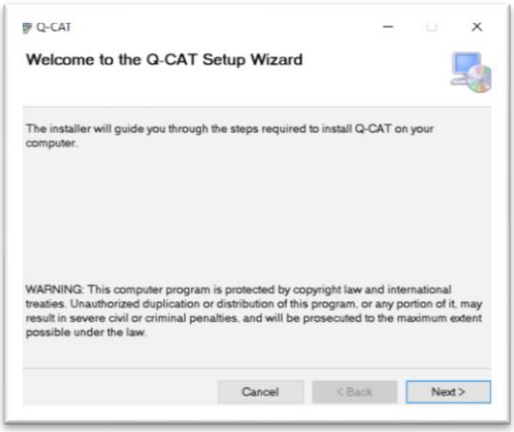
Q-CAT je prosto dostopen program, ki si ga lahko za uporabo na osebnem računalniku prenesemo z repozitorija CLARIN.SI.

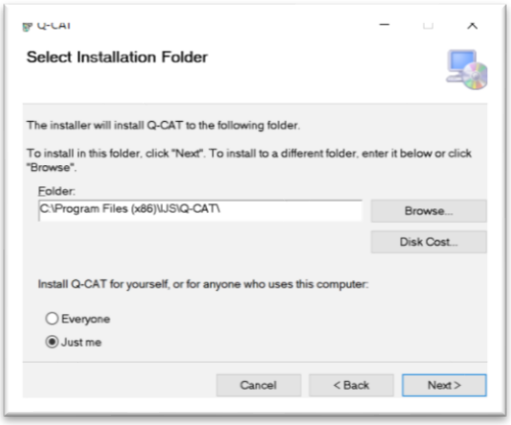
2.2 TEHNIČNE ZAHTEVE

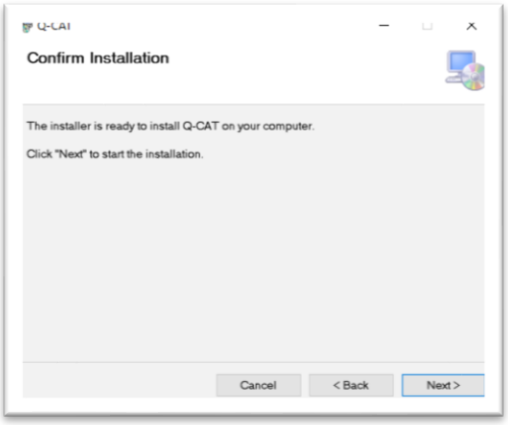
Orodje Q-CAT deluje na operacijskem sistemu **Windows** in za svoje delovanje potrebuje ogrodje .NET Framework (4.5 ali novejša različica). To ogrodje je že vključeno v najpogosteje uporabljene različice sistema Windows (npr. Windows 10), zato ga običajno ni treba nameščati posebej.

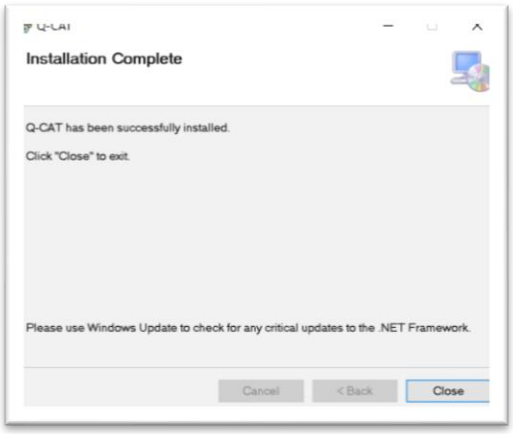
2.3 NAMESTITEV PROGRAMA

Datoteko **Q-CAT.msi** prenesemo na svoj računalnik in jo zaženemo. Sledimo navodilom na ekranu, kjer s klikom na *Next* potrdimo priporočene privzete nastavitve. Namestitev zaključimo s klikom na gumb *Close*.







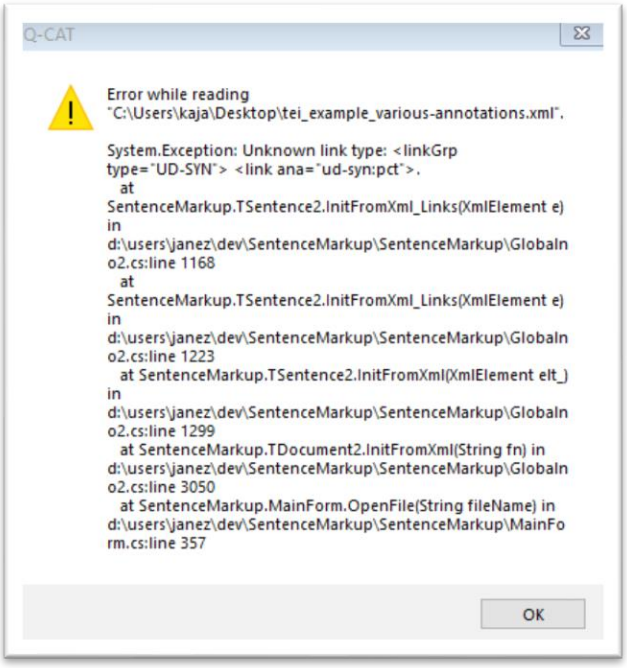


2.4 OPCIJSKA PRILAGODITEV NASTAVITEV OZNAČEVANJA PRED ODPIRANJEM KORPUSA

Ob namestitvi programa se v mapo *Uporabniki > Ime uporabnika > AppData > Roaming* samodejno namesti tudi datoteka **Q-CAT-Settings.xml**, v kateri so specificirane podrobnosti o posameznih ravneh označenosti korpusa, kot je nabor oznak in njihovih lastnosti.

Datoteka Q-CAT-Settings.xml privzeto vsebuje nastavitve, ki ustrezajo oznakam korpusa ssj500k v različici 2.2 (leme in oblikoskladenjske oznake. skladenjske oznake po sistemu JOS, oblikoslovne in skladenjske oznake po sistemu UD, imenske entitete, udeleženske vloge in glagolske stalne besedne zveze). To pomeni, da te datoteke oz. nastavitev označevanja pred začetkom uporabe orodja Q-CAT **ni treba prilagajati**, če nameravamo vanj uvoziti ta korpus ali katerikoli drug korpus z enakimi oznakami na eni ali več ravneh. Enako velja tudi za delo s korpusom, ki vsebuje samo tokenizirano in segmentirano besedilo brez kakršnihkoli drugih oznak.

Če pa nameravamo v orodje uvoziti **korpus z nepoznanimi oznakami**, na kar nas program opozori z obvestilom o napaki, kakršno je prikazano na spodnji sliki, moramo pred začetkom dela ustrezno prilagoditi nastavitve označevanja. To storimo tako, da v orodju kliknemu na gumb ***Settings…*** desno zgoraj in uredimo, tj. izbrišemo ali dodamo, ravni označevanja in oznake, kot je opisano v poglavju *7. Nastavitve označevanja*.

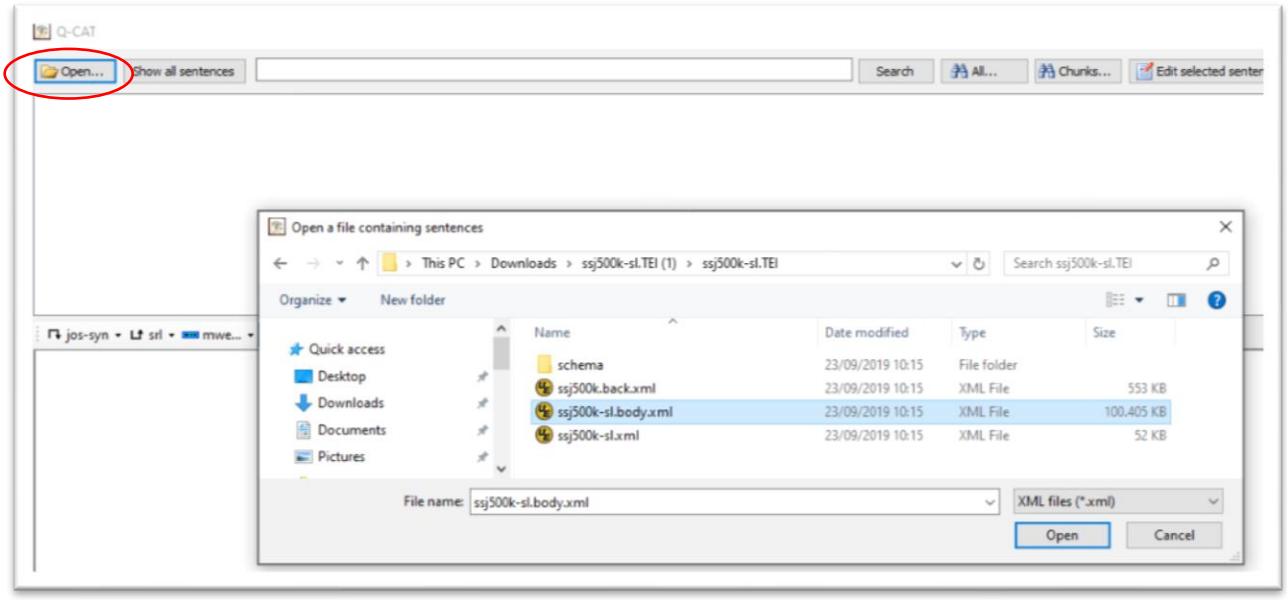


3 ZAGON PROGRAMA IN ODPIRANJE KORPUSA

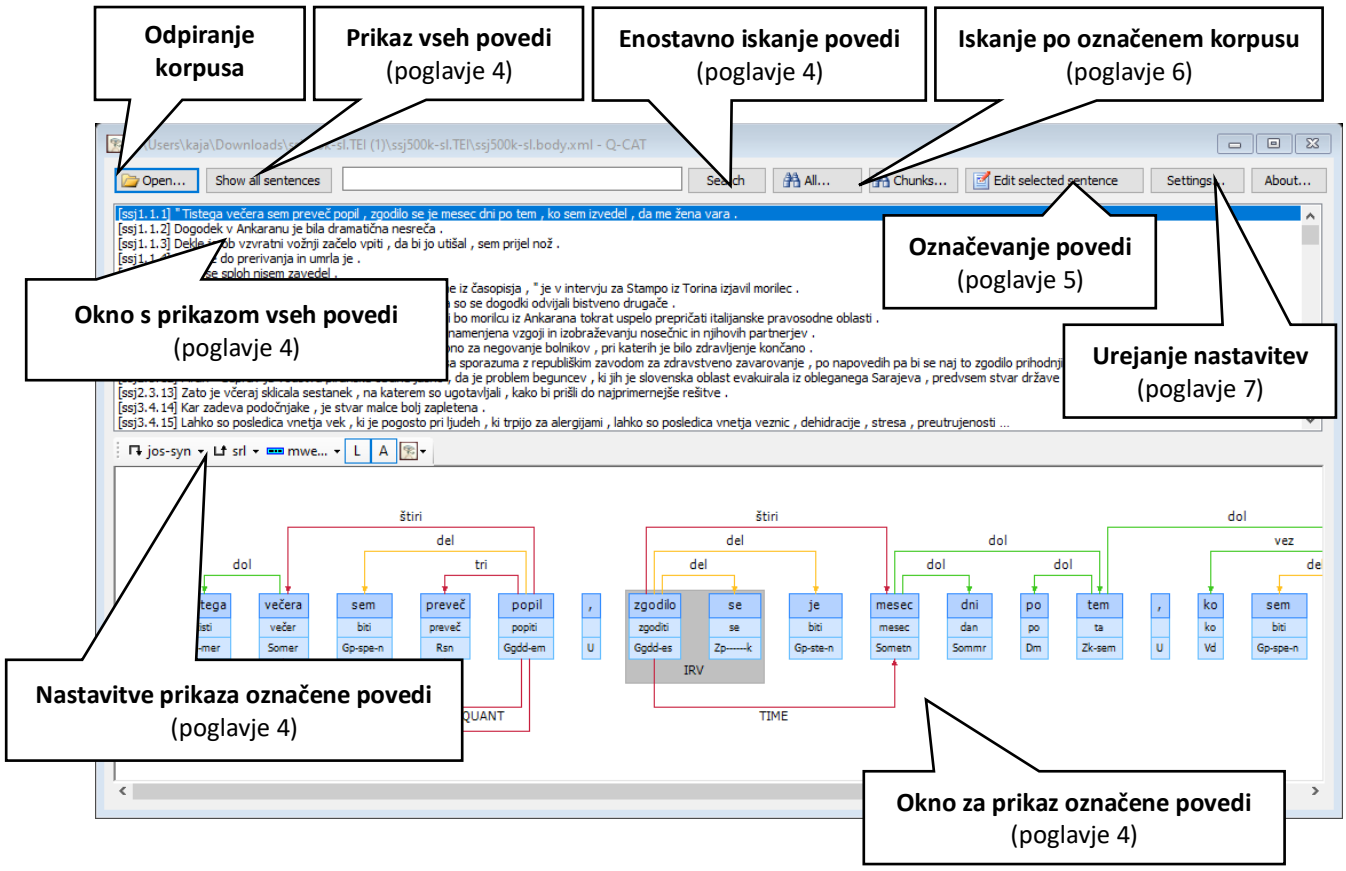
Program zaženemo tako, da v mapi, kjer je program nameščen (privzeto Programi > IJS > Q-CAT), zaženemo datoteko **QCat.exe** ali pa to datoteko poiščemo v iskalnem okencu opravilne vrstice sistema Windows (ikona oken levo spodaj).

Ob zagonu programa se nam prikaže vmesnik, v katerega moramo pred začetkom dela uvoziti korpus. To storimo s klikom na gumb ***Open…*** in na svojem računalniku poiščemo želeni korpus (datoteko vrste XML) v ustreznem formatu (glej poglavje *8. Format*), kot prikazuje spodnja slika.

V orodje lahko uvozimo samo eno datoteko, zato je v primeru dela z večjimi korpusi (npr. več kot milijon besed) priporočljivo, da te pred odpiranjem razdelimo na več manjših datotek. Delitev korpusa na več manjših datotek je priporočljiva tudi v primeru počasnega delovanja programa, ki je sicer odvisno od tehničnih zmogljivosti uporabnikovega računalnika.



Po kliku na gumb *Open* se v vmesnik naložijo vse povedi uvoženega korpusa, kot prikazuje spodnja slika. Posamezne funkcije programa so podrobneje opisane v poglavjih, ki sledijo.

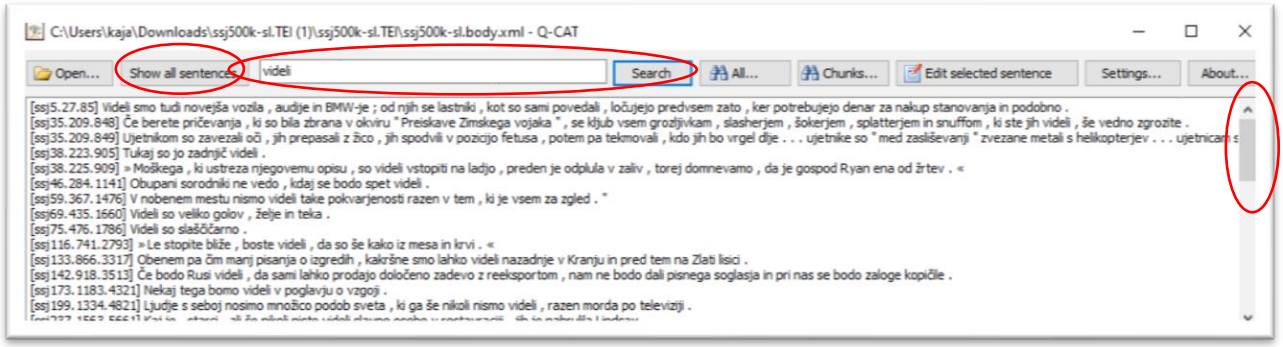

Odpiranje korpusa
Prikaz vseh povedi (poglavje 4)
Enostavno iskanje povedi (poglavje 4)
Iskanje po označenem korpusu (poglavje 6)
Označevanje povedi (poglavje 5)
Okno s prikazom vseh povedi (poglavje 4)
Urejanje nastavitev (poglavje 7)
Nastavitve prikaza označene povedi (poglavje 4)
Okno za prikaz označene povedi (poglavje 4)
Open...
Show all sentences
Search
All...
Chunks...
Edit selected sentence
Settings...
About...
jos-syn
srl
mwe...
L
A
štiri
del
dol
tri
IRV
TIME
QUANT

4 BRSKANJE PO KORPUSU

4.1 BRSKANJE PO POVEDIH

Po uvozu korpusa se v zgornji polovici vmesnika izpiše seznam vseh povedi korpusa in njihovih identifikatorjev. Po seznamu povedi se lahko premikamo s puščicami na tipkovnici ali drsnikom na desni strani.

Iskalno okno za enostavno iskanje povedi nad seznamom omogoča iskanje po **besednih oblikah** z malimi črkami. Kot prikazuje primer spodaj, iskanje po besedni obliki *videli*, ki ga potrdimo s tipko Enter ali s klikom na gumb *Search*, vrne vse povedi s to besedno obliko ne glede na malo ali veliko začetnico vsebovanih črk. Možno je tudi iskanje kombinacije **več besednih oblik** (npr. *so videli*), pri čemer program poišče vse povedi, ki vsebujejo vse navedene oblike, ne glede na njihov dejanski vrstni red v povedi.



Če želimo seznam povrniti na prvotno stanje s prikazom vseh povedi korpusa, kliknemo na gumb *Show all sentences* na levi strani iskalnega okna.

Velikost prikazanih črk lahko spremenimo s hkratnim držanjem tipke Ctrl in premikanjem kolesca miške navzgor (povečanje črk) ali navzdol (pomanjšanje črk) oz. z ustreznim ukazom na sledilni ploščici prenosnih računalnikov.

4.2 PRIKAZ OZNAČENE POVEDI

Ob izbiri povedi s seznama se nam v spodnji polovici vmesnika prikaže poved z vsemi pripadajočimi oznakami, ki jih lahko poljubno vklapljamo in izklapljamo z gumbi na levi zgornji strani okna za prikaz označene povedi.

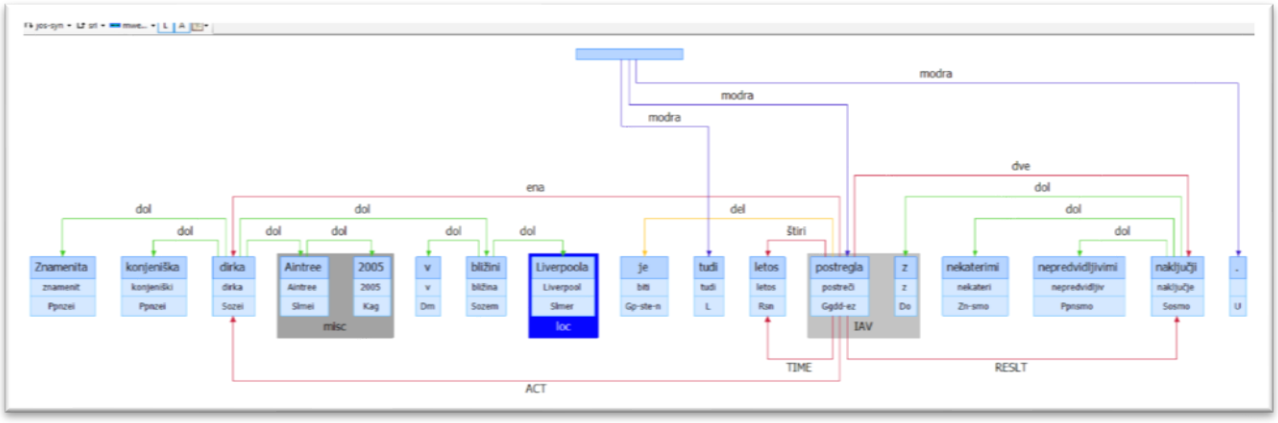
Od leve proti desni si sledijo:

- gumb s puščico navzdol: izbira označevalne ravni s povezavami (npr. skladnja ali udeleženske vloge), ki naj se prikazuje nad povedjo
- gumb s puščico navzgor: izbira označevalne ravni s povezavami (npr. skladnja ali udeleženske vloge), ki naj se prikazuje pod povedjo

(Hkrati sta lahko prikazani največ dve označevalni ravni s povezavami, ena zgoraj in ena spodaj).

- gumb z modro ploščico: izbira ene ali več označevalnih ravni z nizi (npr. imenske entitete ali stalne zveze)
- gumb s črko L: vklop ali izklop prikazovanja osnovnih oblik besed (lem)
- gumb s črko A: vklop ali izklop prikazovanja oblikoskladenjskih oznak besed (msd-jev)
- gumb z očesom: vklop ali izklop prikazovanja korenskega elementa (tipično na ravni skladnje) in njegove oblike (velikost ikone, slika ali grafika).

Spodnja slika prikazuje primer povedi iz korpusa ssj500k z več prikazanimi ravnmi označevanja, pri kateri so pod oblikami (npr. *Znamenita*) izpisane tudi leme (npr. *znamenit*) in oblikoskladenjske oznake (Ppnzei), z barvnimi ploščicami pa so označene imenske entitete (npr. *Liverpoola*) in stalne zveze (npr. *postreči z*). Nad povedjo so izrisane povezave, ki označujejo površinskoskladenjska razmerja med besedami (npr. povezava *ena* med jedrom povedka in osebkom), vključno s tistimi, ki izhajajo iz korenskega elementa (modri pravokotnik nad povedjo). Pod povedjo so izrisane povezave, ki označujejo udeleženske vloge (npr. *letos* kot označevalec časovne okoliščine z oznako *TIME*).



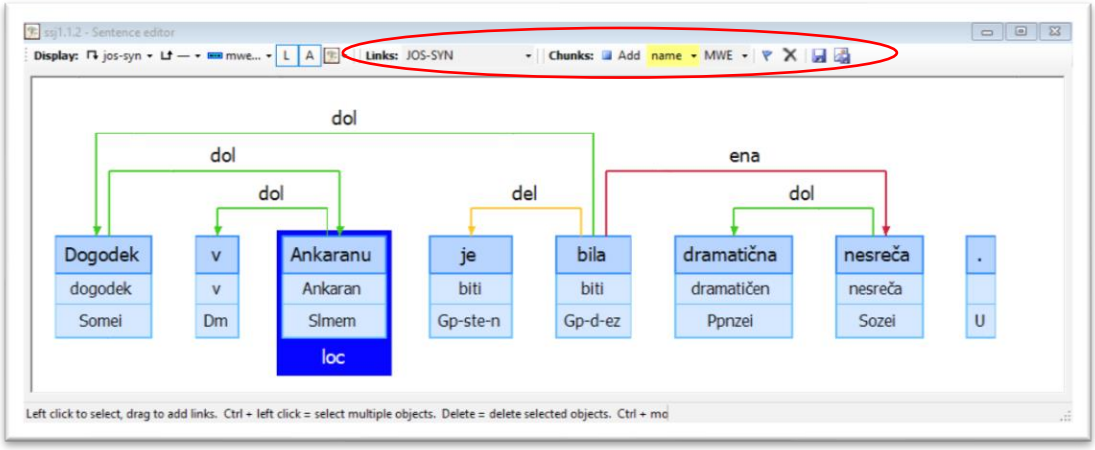
Velikost prikazanih oznak lahko spremenimo s hkratnim držanjem tipke Ctrl in premikanjem kolesca miške navzgor (povečanje črk) ali navzdol (pomanjšanje črk) oz. z ustreznim ukazom na sledilni ploščici prenosnih računalnikov.

5 OZNAČEVANJE KORPUSA

V tem poglavju so opisane funkcije, povezane s tipično rabo orodja Q-CAT, tj. **popravljanje ali dodajanje že opredeljenih oznak** korpusa. Za naprednejše uporabnike, ki jih zanimajo specifike posameznih ravni označevanja, možnosti njihovega spreminjanja ali dodajanje novih oznak, so podrobnejša navodila glede samih nastavitev označevalnih ravni in posamičnih oznak opisana v poglavju *7. Nastavitve označevanja*.

5.1 OKNO ZA UREJANJE POVEDI

Če želimo označiti določeno poved, jo izberemo na seznamu povedi in okno za njeno urejanje odpremo bodisi s klikom na gumb *Edit selected sentence* v zgornji vrstici vmesnika bodisi z dvoklikom na izbrano poved. Prikaže se **okno za urejanje**, kakršno je prikazano na spodnji sliki.



Prvi del zgornje vrstice okna za urejanje (do ikone z očesom) vsebuje enake gumbe kot okno za prikazovanje označene povedi (poglavje *4.2 Prikaz označenih povedi*), s katerimi opredeljujemo, katere izmed obstoječih ravni označevanja naj bodo v urejevalniku **prikazane** in na kakšen način. Ne glede na to, ali želimo oznake dodajati ali samo popravljati, pred začetkom urejanja za prikaz izberemo (vsaj) tiste ravni označevanja, ki jih nameravamo urejati. S tem omogočimo ustrezen prikaz **gumbov za označevanje**, ki so na zgornji sliki označeni z rdečo in jih podrobneje predstavljamo v nadaljevanju.

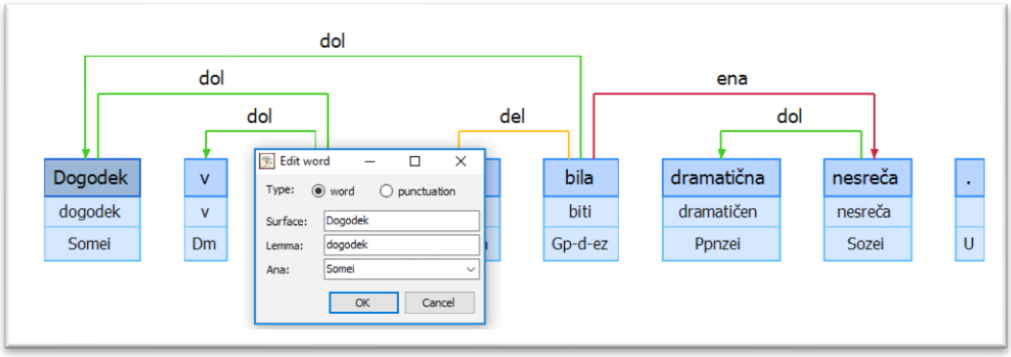
Velikost prikazanih oznak tudi v tem oknu spremenimo s hkratnim držanjem tipke Ctrl in premikanjem kolesca miške navzgor (povečanje črk) ali navzdol (pomanjšanje črk) oz. z ustreznim ukazom na sledilni ploščici prenosnih računalnikov.

5.2 UREJANJE OZNAK

V tem poglavju možnosti urejanja, tj. spreminjanja ali dodajanja oznak, predstavljamo ločeno za različne vrste oznak. Če vas zanimajo podrobnosti glede formalnega razlikovanja med oznakami *oblik* (npr. leme), oznakami *nizov* (npr. stalne besedne zveze) in oznakami *povezav* (npr. skladenjske povezave), si oglejte poglavje *7.1 Vrste označevanja*.

5.2.1 Urejanje oblik, lem in oblikoskladenjskih oznak

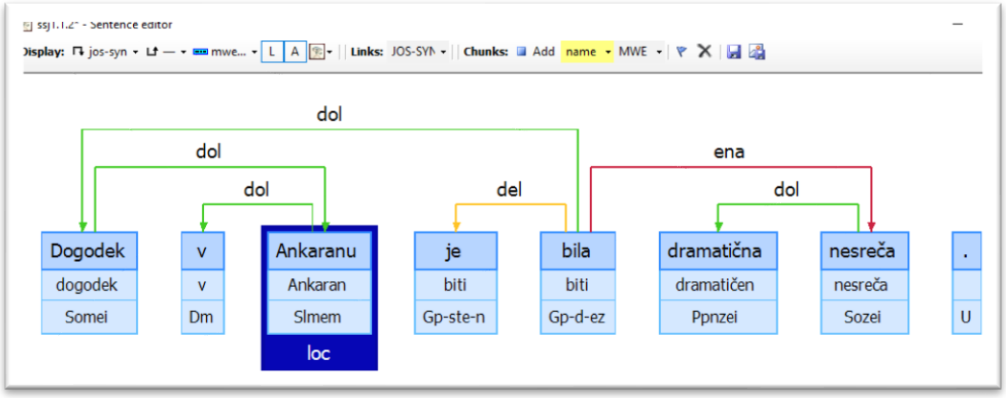
Če želimo besedam v povedi spremeniti obliko ali dodati podatek o osnovni obliki (lemi) besede ali njenih oblikoslovnih lastnostih, z dvoklikom na izbrano besedo odpremo okno za urejanje besednih oblik, kot prikazuje spodnja slika. Vpišemo želeno obliko besede (polje *Surface*), njeno osnovno obliko (polje *Lemma*) ali oblikoskladenjsko oznako (polje *Ana*). Medtem ko prvi dve polji omogočata vnos poljubnega niza znakov, je izbira oblikoskladenjskih oznak lahko omejena z vnaprej določenim naborom dovoljenih oznak (glej poglavje *7.3.3. Nastavitve seznama oblikoskladenjskih oznak*). V tem primeru lahko uporabnik ustrezno oznako izbere tudi s klikom na eno izmed možnosti v spustnem meniju.



Pri urejanju pojavnic, ki so bile v uvoženem korpusu označene kot ločila (Type: punctuation), je mogoče spreminjati samo obliko (polje *Surface*), ne pa leme ali oblikoskladenjske oznake.

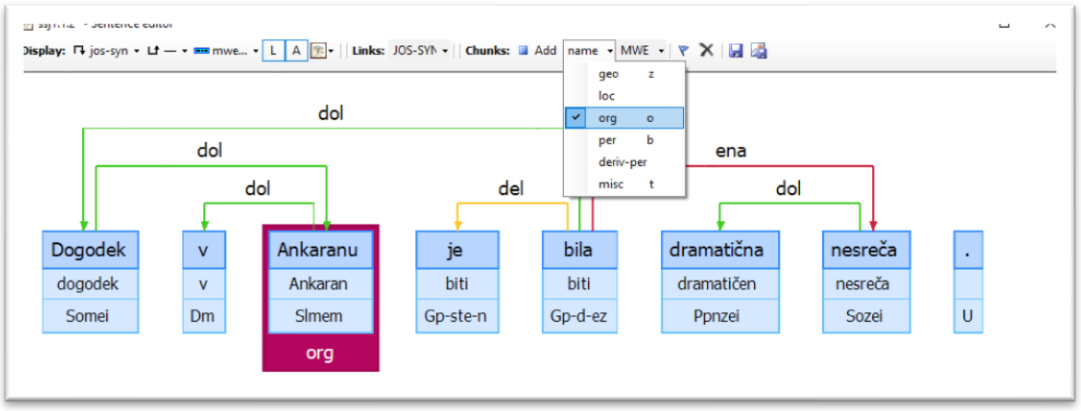
5.2.2 Urejanje oznak nizov

Če želimo urejati oznako, ki je bila pripisana nizu ene ali več besed, nanjo kliknemo. Oznaka nekoliko potemni, v zgornji vrstici okna za urejanje pa se aktivira in rumeno obarva raven označevanja, ki ji oznaka pripada. Slika spodaj denimo prikazuje primer izbire oznake 'loc', ki je bila pripisana besedi *Ankaran* in je ena izmed oznak ravni za označevanje imenskih entitet ('name').



Če želimo oznako spremeniti (npr. iz oznake za lokacijo 'loc' v oznako za organizacijo 'org'), kliknemo na puščico ob rumeno obarvanem imenu ustrezne ravni in izberemo drugo oznako s **spustnega seznama**, kot prikazuje slika spodaj. Za hitrejše spreminjanje oznak lahko uporabimo tudi **bližnjice** na tipkovnici, ki so za lažji priklic tudi izpisane ob posameznih oznakah v spustnem meniju. Po kliku na ploščico z oznako

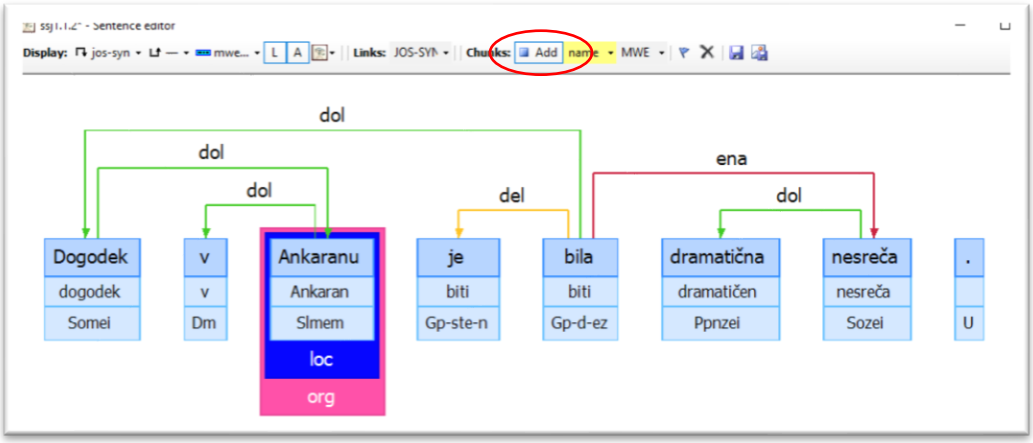
'loc' (zgoraj) lahko torej že s pritiskom na črko *o* oznako spremenimo v 'org' (spodaj), brez zamudnejšega iskanja ustrezne kategorije na spustnem seznamu.



Če želimo oznako **izbrisati**, kliknemo na gumb z ikono križca ali pritisnemo tipko Delete na tipkovnici.

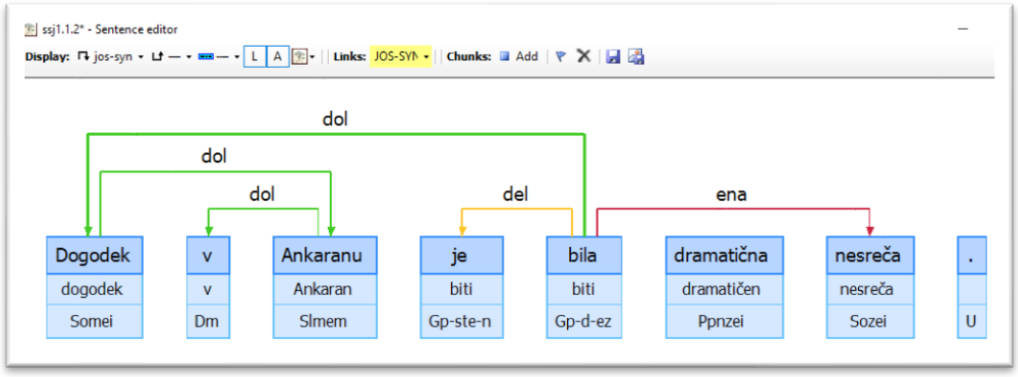
Če želimo oznako pripisati še neoznačenim besedam, kliknemo na želeno besedo in na spustnem meniju izberemo ustrezno kategorijo. Če želimo kot enoto označiti **dve ali več pojavnic**, držimo tipko Ctrl in izberemo vse relevantne pojavnice, nato pa jim oznako pripišemo na enak način, z izbiro kategorije na ustreznem spustnem seznamu.

V redkih primerih, ko želimo identičnemu naboru ene ali več pojavnic pripisati **dve ali več različnih oznak** znotraj iste ravni označevanja (npr. tako oznako za lokacijo 'loc' kot organizacijo 'org'), po izbiri prve oznake (npr. 'loc') kliknemo na gumb *Add* in izberemo še drugo oznako (npr. 'org'). Kot prikazuje spodnja slika, se v urejevalniku nato prikazujeta obe oznaki.



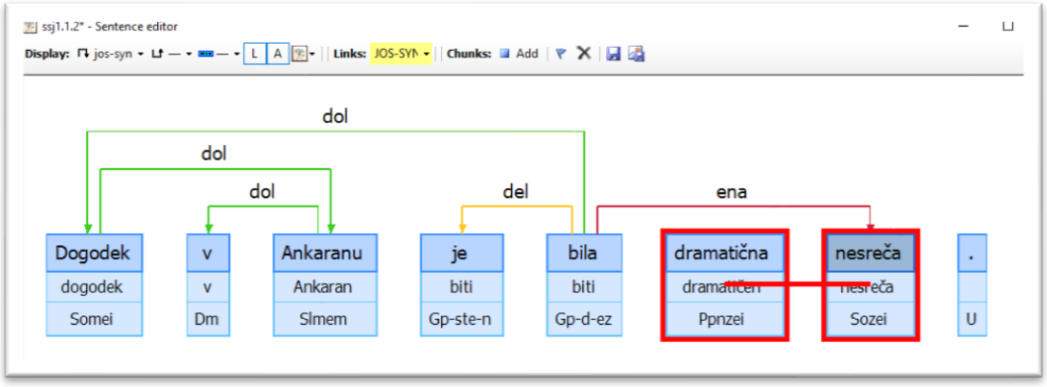
5.2.3 Urejanje povezav

Za urejanje povezav med dvema besedama kliknemo na puščico. Oznaka nekoliko potemni, v zgornji vrstici okna za urejanje pa se aktivira in rumeno obarva raven označevanja, ki ji oznaka pripada. Slika spodaj denimo prikazuje primer izbire povezave 'dol' med besedama *bila* in *Dogodek*, ki je ena izmed oznak ravni za skladenjsko razčlenjevanje po sistemu JOS ('JOS-SYN').

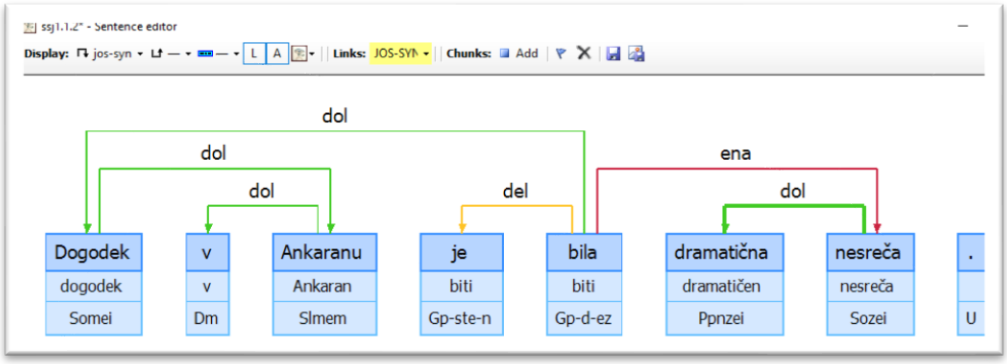


Tako kot pri nizih oznako spremenimo bodisi z izbiro druge oznake na **spustnem seznamu** vseh možnih oznak te ravni bodisi s pritiskom na ustrezno **bližnjico** na tipkovnici, izbrišemo pa jo s klikom na križec ali tipko **Delete**.

Če želimo **ustvariti novo povezavo** med dvema besedama (npr. med samostalnikom *nesreča* in njegovim prilastkom *dramatična*), kliknemo na besedo, ki bo izvor povezave, in se med držanjem levega miškinega gumba premaknemo do besede, ki bo cilj povezave, tako da se obe besedi obarvata rdeče, kot prikazuje slika spodaj.



Ko miškin gumb sprostimo, se med besedama izriše puščica, ki poteka od prve označene besede do druge. Pripisano oznako lahko takoj spremenimo s pritiskom na ustrezno bližnjico na tipkovnici (npr. črko *o* za *dol*) ali z izbiro ustrezne oznake na spustnem seznamu rumeno obarvane ravni.



Med dvema pojavnicama je na eni označevalni ravni mogoče vzpostaviti samo eno povezavo določene smeri.

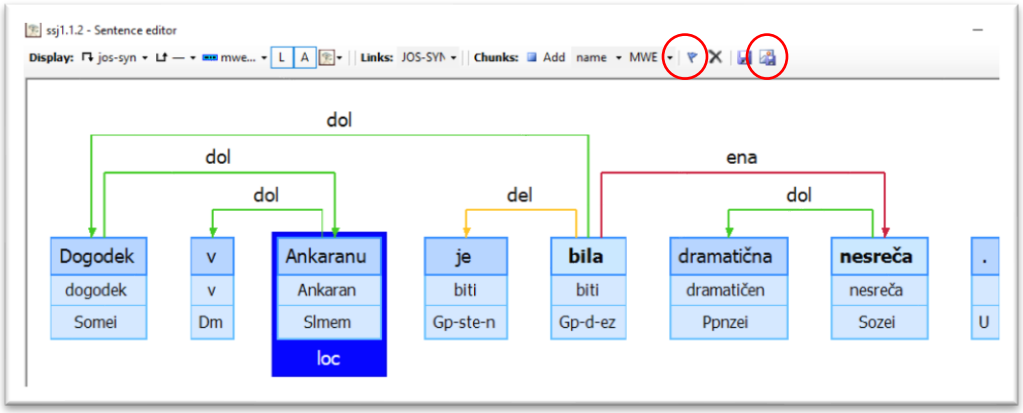
5.3 SHRANJEVANJE OZNAČENE POVEDI

Spremembe v označenosti povedi, ki smo jih izvedli v oknu za urejanje, se shranijo **šele s klikom na ikono za shranjevanje** v zgornji vrstici okna. Če sprememb ne želimo shraniti, okno zapremo s klikom na križec v desnem zgornjem kotu in ob opozorilu programa, da spremembe še niso bile shranjene, izberemo možnost Ne ('No').

Privzeto se spremembe vedno shranijo v datoteko, ki smo jo izbrali ob uvozu korpusa, s čimer nove oznake nadomestijo prejšnje. Če želimo ohraniti izvorni korpus nespremenjen, je priporočljivo, da že pred uvozom korpusa ustvarimo **kopijo izhodiščne datoteke**.

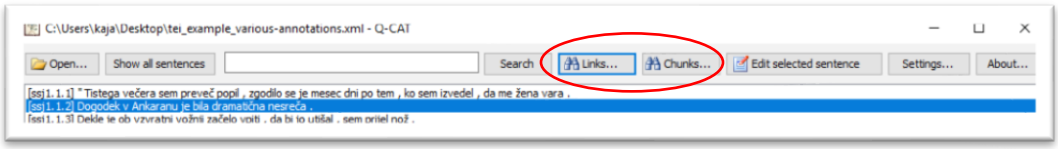
5.4 IZVOZ SLIKE OZNAČENE POVEDI

S klikom na zadnji gumb v zgornji vrstici okna za urejanje lahko označeno poved z izbranimi prikazanimi ravnmi označevanja tudi **shranimo kot sliko v visoki ločljivosti** (.png), denimo za potrebe prikazovanja oznak v strokovnih objavah ali drugih predstavitvah. Če želimo pred izvozom slike posamezne besede v povedi še dodatno izpostaviti, jih izberemo s klikom in uporabimo gumb z zastavico. Izbrane besede se izpišejo z **odebeljenim tiskom**, kot na primer besedi *bila* in *nesreča* na sliki spodaj.



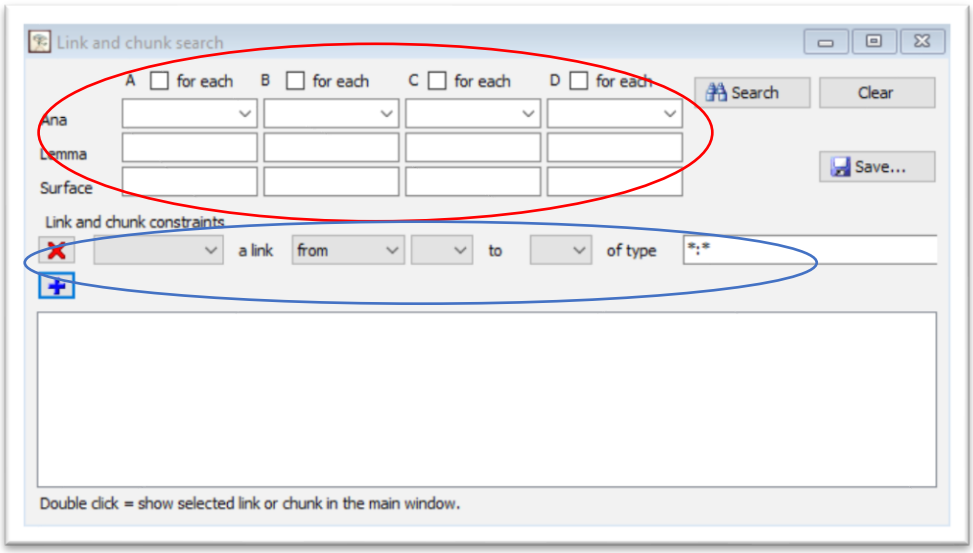
6 ISKANJE PO OZNAČENEM KORPUSU

Orodje Q-CAT poleg označevanja korpusa omogoča tudi **naprednejša iskanja** po označenem korpusu. Uporabnik lahko s posebnim vmesnikom oblikuje različne iskalne pogoje, na podlagi katerih program poišče, prešteje in izpiše vse povedi v korpusu, ki temu iskalnemu pogoju ustrezajo. Ker iskanje po nizih, povezavah ali njihovih kombinacijah predvideva različne potrebe uporabnikov, sta omogočeni dve vrsti iskanj: **splošno iskanje**, ki omogoča iskanje po vseh tipih oznak (ikona z daljnogledom in pripisom All) in **iskanje zgolj po nizih** (ikona z daljnogledom in pripisom Chunks).



6.1 SPLOŠNO ISKANJE

Ob kliku na ikono z daljnogledom in pripisom ***All…*** v vstopnem oknu orodja se nam odpre okno za splošno iskanje, kot prikazuje slika spodaj. Okno je sestavljeno iz dveh delov: območja za opredelitev lastnosti posameznih besed (rdeče) in območja za opredelitev lastnosti oznak nizov ali povezav (modro), ki jih lahko poljubno dodajamo s klikom na modri znak plus.



6.1.1 Iskanje po oblikah ali oblikoskladenjskih oznakah besed

V območju za opredelitev lastnosti besed lahko opredelimo lastnosti največ štirih besed oz. pojavnic (besed A, B, C ali D), ki naj se pojavijo v dani povedi. Pri tem lahko opredelimo:

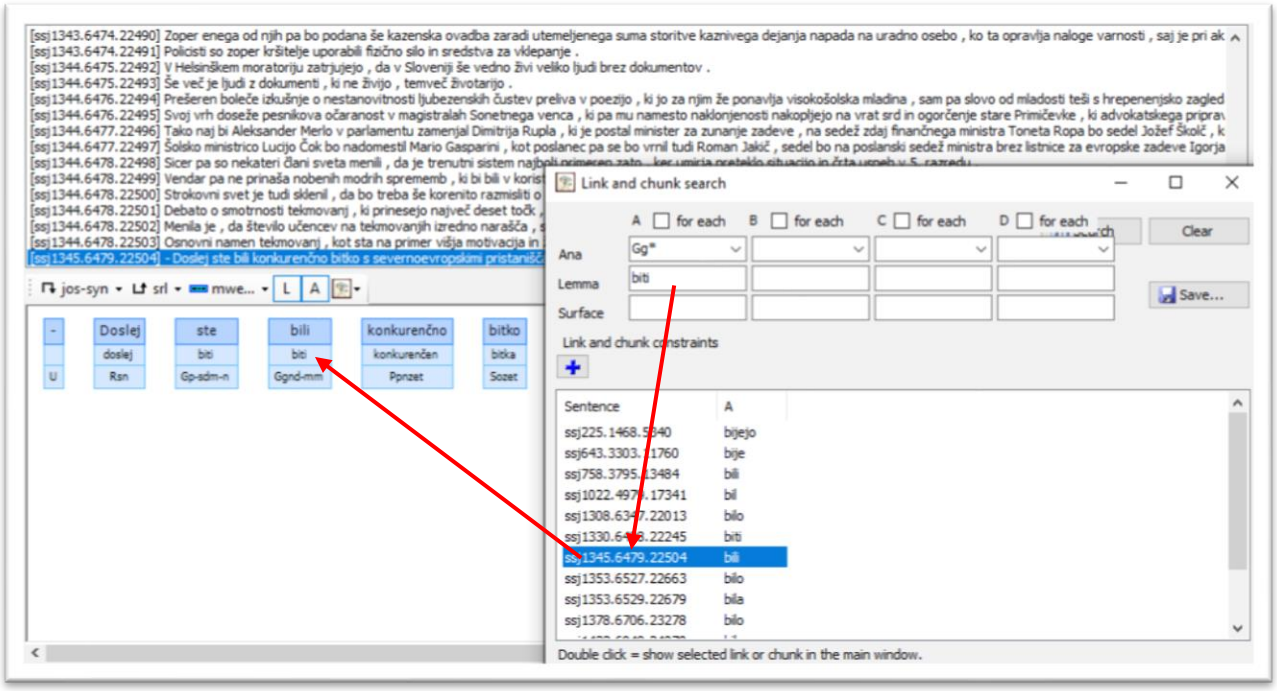
- oblikoskladenjsko oznako besede (polje *Ana*)
- osnovno obliko besede (polje *Lemma*)
- površinsko obliko besede (polje *Surface*)

Naenkrat lahko opredelimo eno ali več lastnosti, pri čemer iskanje podpira tudi uporabo **nadomestnih znakov** – zvezdica (*) za poljuben niz znakov in vprašaj (?) za en sam znak.

Po oblikovanju iskalnega pogoja iskanje izvedemo s klikom na gumb ***Search***.

Primer na spodnji sliki prikazuje iskanje po vseh povedih korpusa, v katerih se pojavlja glagol *biti* z oznako glavnega glagola (Gg*). Okno z rezultati vrne seznam vseh takih povedi z njihovim ID-jem in izpisom konkretne oblike vseh opredeljenih pojavnic (v našem primeru ene, pojavnice A), desno zgoraj pa je izpisano tudi skupno število zadetkov.

Primere najdenih povedi si ogledamo **z dvojnim klikom na posamezen zadetek**, s čimer se v glavnem oknu v ozadju prikaže ustrezna poved.



6.1.2 Iskanje po oznakah povezav ali nizov

S klikom na **modri znak plus** se odpre območje za omejevanje oznak povezav ali nizov, v katerem lahko opredelimo (polja od leve proti desni):

- obstoj (*exists*) ali neobstoj (*doesn't exist)* določene povezave med dvema pojavnicama
- smer povezave med dvema pojavnicama, ki je lahko usmerjena od prve k drugi (*from*) bodisi je smer povezave med pojavnicama poljubna (*between*)
- označevalno raven in oznako povezave (polje *:*). Pri tem lahko v prvi del polja (pred dvopičjem) vpišemo raven označevanja (npr. SRL), v drugi del polja (za dvopičjem) pa konkretno oznako (npr. ACT, PAT …). Podrobnosti zapisa tega dela iskalnega pogoja so opisane v poglavju *6.3 Vmesnik za opredelitev iskane oznake*.

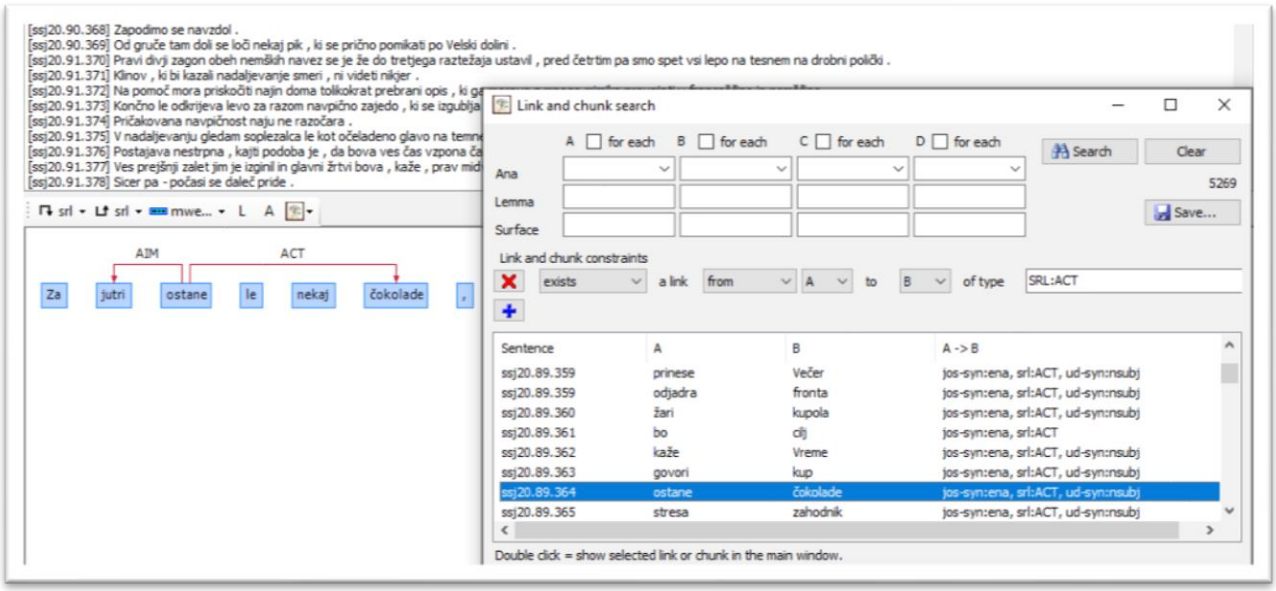
Iskanje izvedemo s klikom na gumb ***Search***.

S klikom na modri znak plus odpiramo nova polja, v katerih lahko opredelimo dodatne restrikcije glede istih povezav oz. nizov ali specificiramo nove. Z **rdečim križcem** polje za določanje lastnosti povezav zapremo in s tem iskalni pogoj izbrišemo.

Za lažjo predstavo spodaj prikazujemo nekaj bolj ali manj kompleksnih primerov uporabe splošnega iskanja po nizih ali povezavah.

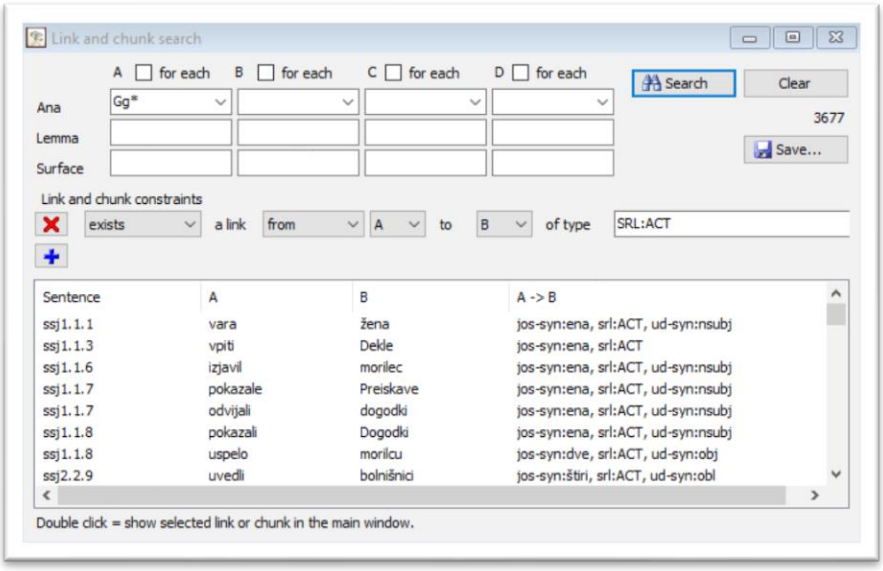
6.1.2.1 Primer enostavnega iskanja po povezavah med pojavnicami

Spodnji primer prikazuje iskanje po ravni označevanja udeleženskih vlog v korpusu ssj500k (raven SRL), v katerem smo poiskali vse pare pojavnic A in B, pri katerih je pojavnica B označena kot vršilec dejanja oz. aktant (oznaka 'ACT') pojavnice A. Prikazana je poved, v kateri je beseda *čokolada* označena kot vršilec glagola *ostati*.



6.1.2.2 Primer kombiniranega iskanja po povezavah in lastnostih pojavnic

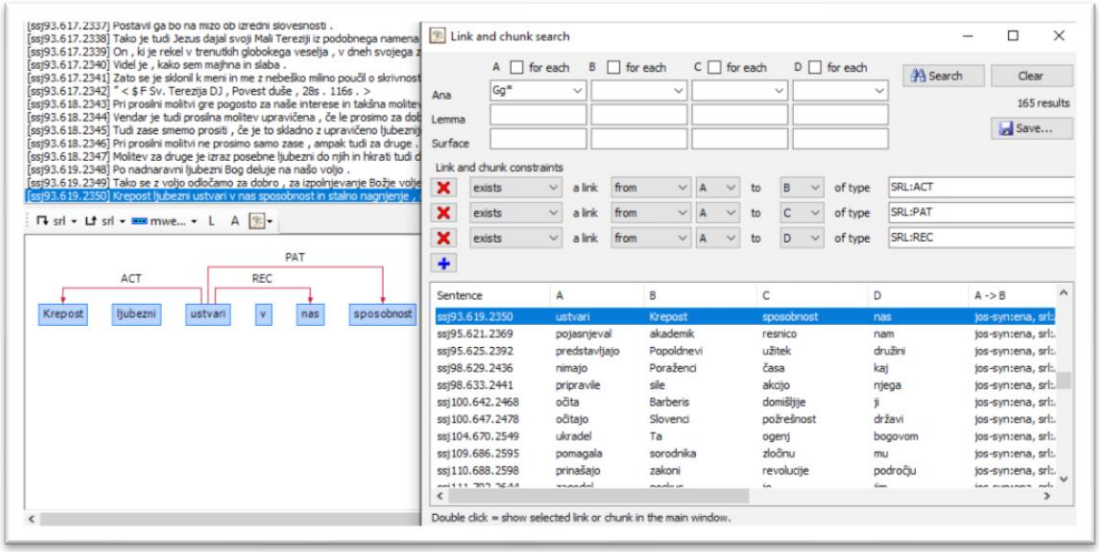
Izvedemo lahko tudi kombinirano iskanje, v katerem poleg vrste povezave med pojavnicama določimo tudi njihove podrobnejše lastnosti (poglavje 6.1.1), kot prikazuje spodnji primer iskanja vršilcev dejanj, ki so povezani na poljubni glavni glagol (oznaka Gg*), s čimer denimo iz rezultatov zgornjega iskanja (6.1.2.1) izločimo pojavitve s pomožnim glagolom *biti*.



6.1.2.3 Primer iskanja po več povezavah hkrati

S klikom na modri znak plus odpiramo nova polja za vnos dodatnih iskalnih pogojev glede nizov ali povezav, ki se lahko nanašajo na iste ali druge pojavnice.

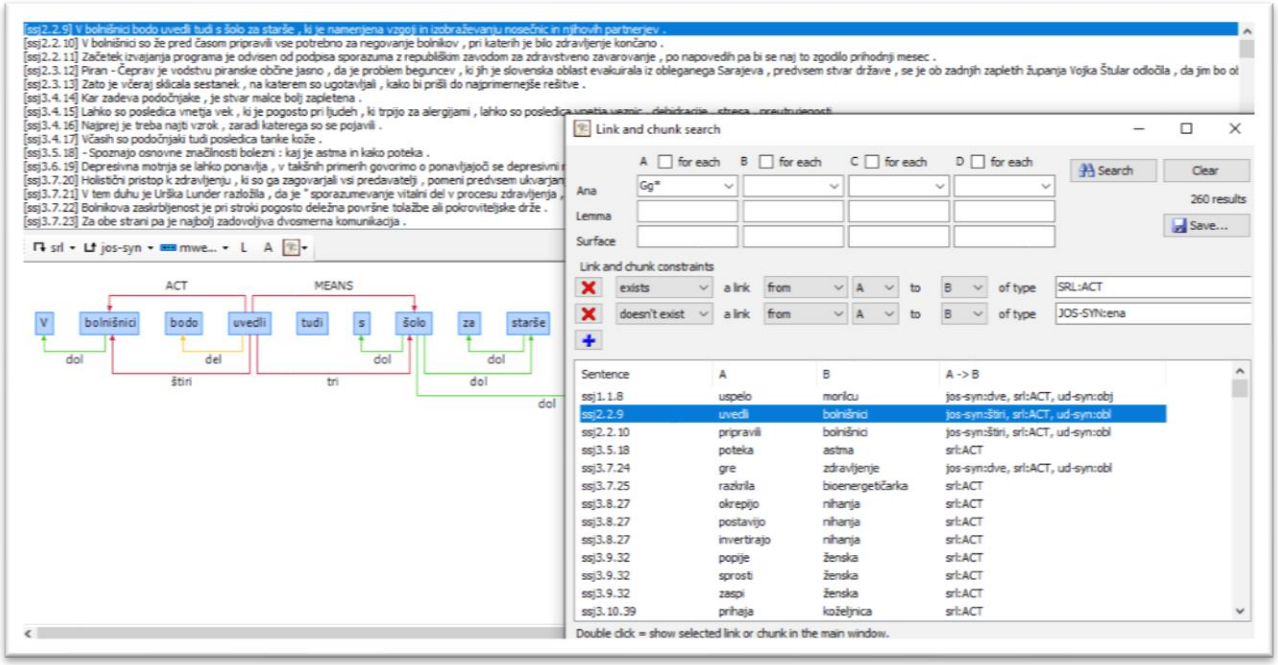
Spodnja slika prikazuje razširjenega zgornjega primera iskanja vršilcev dejanja, izraženega z glavnim glagolom (6.1.2.2), v katerem želimo nabor primerov zožiti zgolj na tiste, kjer sta poleg vršilca izražena tudi prizadeti predmet (oznaka 'PAT') in prejemnik dejanja (oznaka 'REC'), torej primere glagolov z vezljivostjo *kdo komu stori kaj*.



6.1.2.4 Primer iskanja po različnih ravneh označevanja hkrati

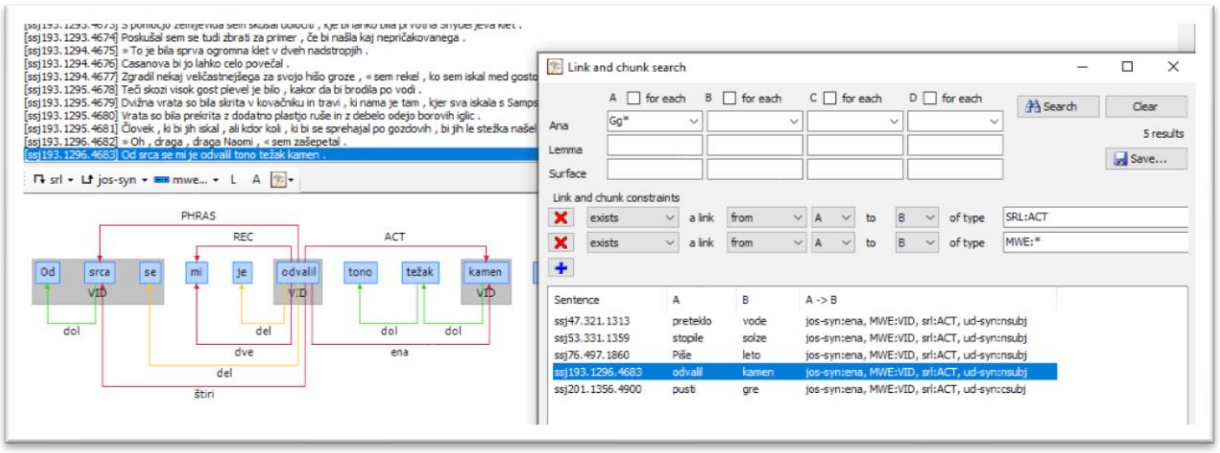
Hkratno iskanje po več povezavah lahko izvedemo tudi s povezavami na več različnih ravneh označevanja, denimo s hkratno opredelitvijo udeleženske vloge in njenih skladenjskih lastnosti. Tak

primer prikazuje spodnja slika, na kateri so prikazani razširjenega iskanja vršilcev dejanj, izraženih z glavnim glagolom (6.1.2.2), v katerem iščemo samo tiste vršilce dejanja (SRL:ACT), ki na skladenjski ravni niso bili označeni kot osebki (JOS-SYN:ena). Med zadetki tako najdemo primere vršilcev dejanja v neimenovalniških sklonih (npr. *morilcu je uspelo*), prostorsko izražene vršilce (npr. *v bolnišnici bodo uvedli*) in podobne primere.



Na podoben način lahko pogoje glede povezav kombiniramo tudi z iskanji po povezavah nizov. Če med istima pojavnicama opredelimo tako določeno oznako povezave kot določeno oznako niza, orodje poišče pare pojavnic, ki hkrati izpolnjujejo oba pogoja.

Spodnja slika prikazuje primer iskanja vršilcev dejanja (SRL:ACT), ki se pojavljajo v stalnih besednih zvezah poljubnega tipa (MWE:*). Med zadetki najdemo frazeme tipa *preteči veliko vode, piše se leto, kamen se odvali od srca* itd.

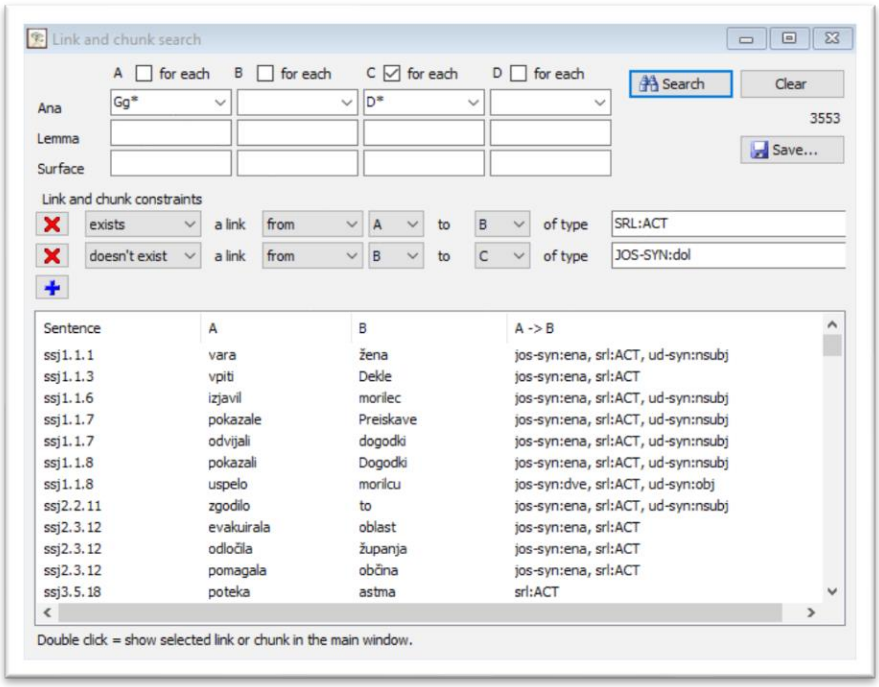


6.1.2.5 Primer iskanja z negacijo

Enostaven primer uporabe funkcije za negacijo povezave (*doesn't exist*) smo že prikazali v prvem primeru poglavja 6.1.4, v katerem smo iskali primere povezav med pari pojavnic A in B – glagoli in vršilci dejanja na semantični ravni, med katerimi obenem ni povezave za osebek na skladenjski ravni. Kadar funkcijo negacije uporabimo za povezave med več različnimi pari pojavnic (npr. A in B ter A in C), pa moramo opredeliti tudi, ali negacija velja za vse možne kombinacije pojavnic ali ne. Temu je namenjeno polje ***for each*** pri vsaki pojavnici.

Za primer vzemimo iskanje iz poglavja 6.1.2, ki nam vrne vse vršilce dejanj, med katerimi najdemo tudi primere vršilcev v obliki predložnih zvez tipa *v bolnišnici so uvedli*. Če želimo iz zadetkov izločiti predložne zveze tipa *v bolnišnici*, poleg glagola (A) in vršilca dejanja (B) opredelimo še pojavnico C z oznako za predlog (oznaka D*) in dodamo pogoj, da B in C ne smeta biti povezana s skladenjsko oznako *dol*, ki se uporablja za označevanje predložnih zvez (npr. *v <--dol-- bolnišnici*). Če želimo s tem priklicati samo primere, kjer opredeljeni pogoji (tako povezava med A in B kot B in C) veljajo za vsak C, odkljukamo polje *for each* pri pojavnici C.

V primeru na sliki to denimo pomeni, da iščemo take A in B, pri katerih za vsak C velja, da če je C predlog, potem ne obstaja povezava *dol* od B do C. S tem dobimo vse vršilce dejanja, ki nimajo povezave na predlog, tj. niso predložne zveze.



Za naprednejše uporabnike dodajmo še podrobnejšo razlago tega mehanizma. Recimo tistim pojavnicam oz. spremenljivkam (A, B, C, D), pri katerih je stikalo "for each" vključeno, da so "vezane", ostalim spremenljivkam pa, da so "proste". Omejitve glede povezav lahko potem v mislih razdelimo v tri skupine glede na to, ali v omejitvi nastopata dve prosti spremenljivki, dve vezani ali pa ena prosta in ena vezana. Tem trem skupinam omejitev recimo "proste", "vezane" in "mešane".

Poizvedba potem poišče vse take nabore vrednosti prostih spremenljivk, pri katerih velja naslednje:

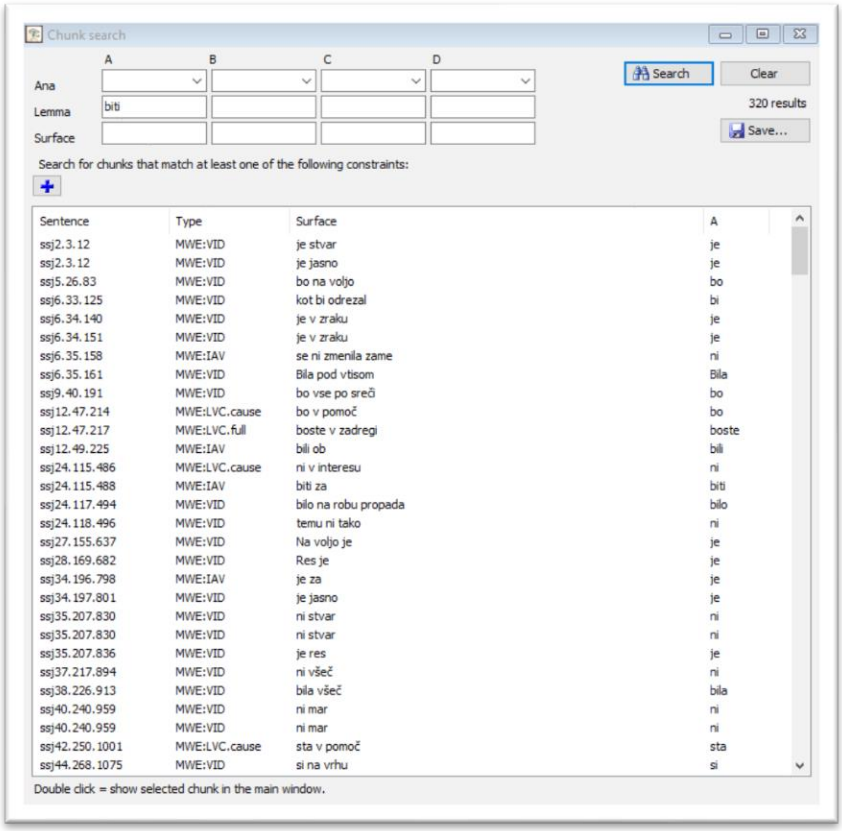
- vse proste spremenljivke so različne med seboj in ustrezajo vsem omejitvam glede lastnosti besed (lema, oblika ali oblikoskladenjska oznaka) in tudi vsem prostim omejitvam povezav; in
- za vsak možen nabor vrednosti vezanih spremenljivk velja: če so te vezane spremenljivke vse različne med seboj in tudi od prostih spremenljivk in če ustrezajo vsem omejitvam glede lastnosti oblik in vsem vezanim omejitvam povezav, potem ta nabor vrednosti vezanih spremenljivk skupaj z našim naborom vrednosti prostih spremenljivk ustreza tudi vsem mešanim omejitvam povezav.

6.2 ISKANJE PO NIZIH

Ob kliku na ikono z daljogledom in pripisom ***Chunks…*** v vstopnem oknu orodja se nam odpre okno za iskanje po nizih. To okno je oblikovano podobno kot okno za splošno iskanje (poglavje *6.1 Splošno iskanje*), vendar so njegove funkcije prilagojene posebej za iskanje po oznakah nizov.

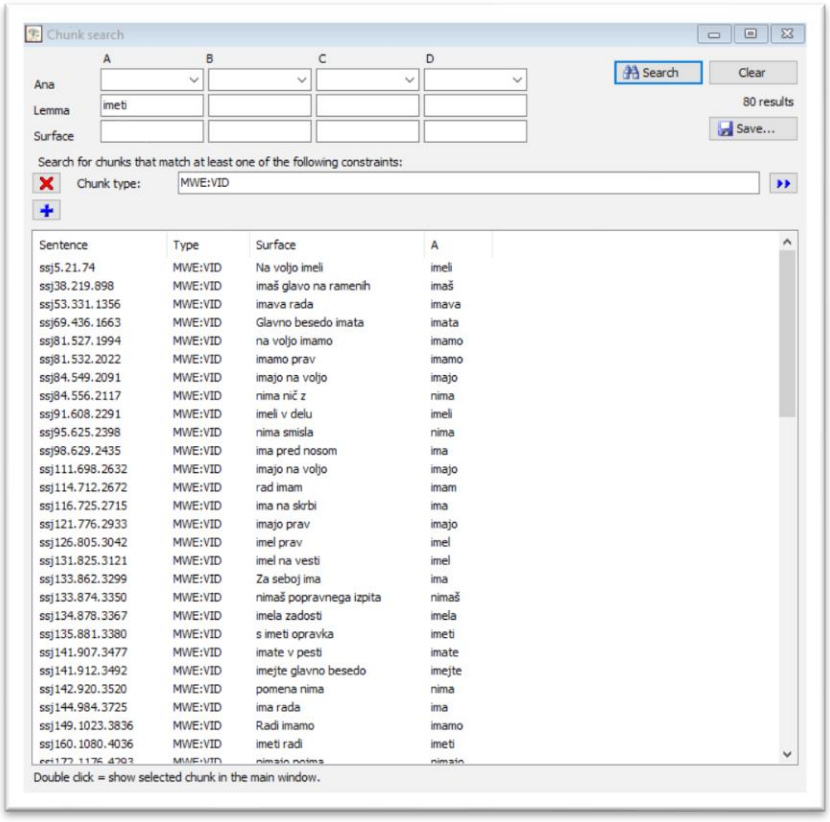
Najpomembnejša razlika je ta, da vsi **iskalni pogoji iščejo izključno po oznakah nizov**, med rezultati pa se v stolpcu *Surface* **izpišejo celotne besedne zveze** in ne zgolj pojavnice, ki smo jih opredelili v iskalnem pogoju.

Če kot iskalni pogoj vnesemo lemo *biti*, se med rezultati pojavijo vse besede ali besedne zveze, ki so bile označene z eno izmed oznak za nize, ne glede na raven označevanja.



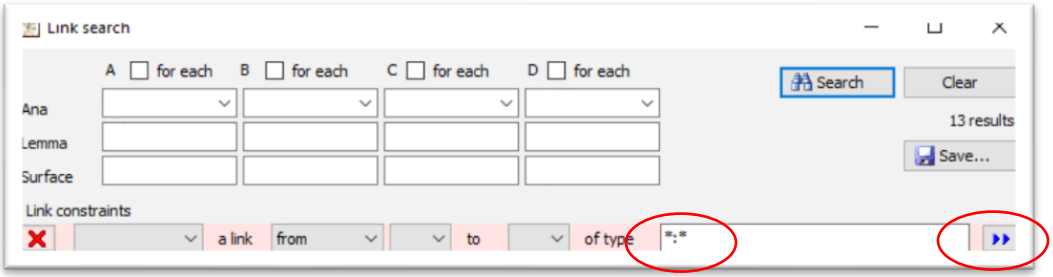
Če želimo dodatno opredeliti tudi raven označevanja, kliknemo na **modri znak plus** in opredelimo podrobnosti glede označevalne ravni in oznake (polje *:*). Pri tem lahko v prvi del polja (pred dvopičjem) vpišemo raven označevanja (npr. MWE), v drugi del polja (za dvopičjem) pa konkretno oznako (npr. VID, IRV …). Podrobnosti zapisa tega dela iskalnega pogoja so opisane v poglavju *6.3 Vmesnik za opredelitev oznake iskanega niza ali povezave*.

Spodnja slika prikazuje primer rezultatov iskanja vseh besednih zvez tipa VID (glagolski frazemi), ki vsebujejo pojavnico z lemo *imeti*.



6.3 VMESNIK ZA OPREDELITEV OZNAKE ISKANEGA NIZA ALI POVEZAVE

Tako pri splošnem iskanju (6.1) kot iskanju po nizih (6.2) polje za opredelitev oznake povezave ali niza (slika spodaj) omogoča dva načina oblikovanja iskalnega pogoja: z **vpisom iskalnega pogoja** v polje, v katerem je privzeto izpisan niz *:*, ali z uporabo **vmesnika za oblikovanje iskalnega pogoja** (klik na modri puščici).



V nadaljevanju predstavimo uporabo vmesnika za oblikovanje iskalnega pogoja in pri vsaki možnosti pripišemo tudi njegovo besedilno različico, ki jo lahko brez klikanja vpišemo neposredno v iskalno polje.

V vmesniku za oblikovanje iskalnega pogoja, ki je prikazan na spodnji sliki, izberemo označevalno raven (npr. SRL) in želeno oznako (npr. ACT).

Nato med danimi grafičnimi prikazi načinov iskanja izberemo ustrezno obliko iskalnega pogoja:

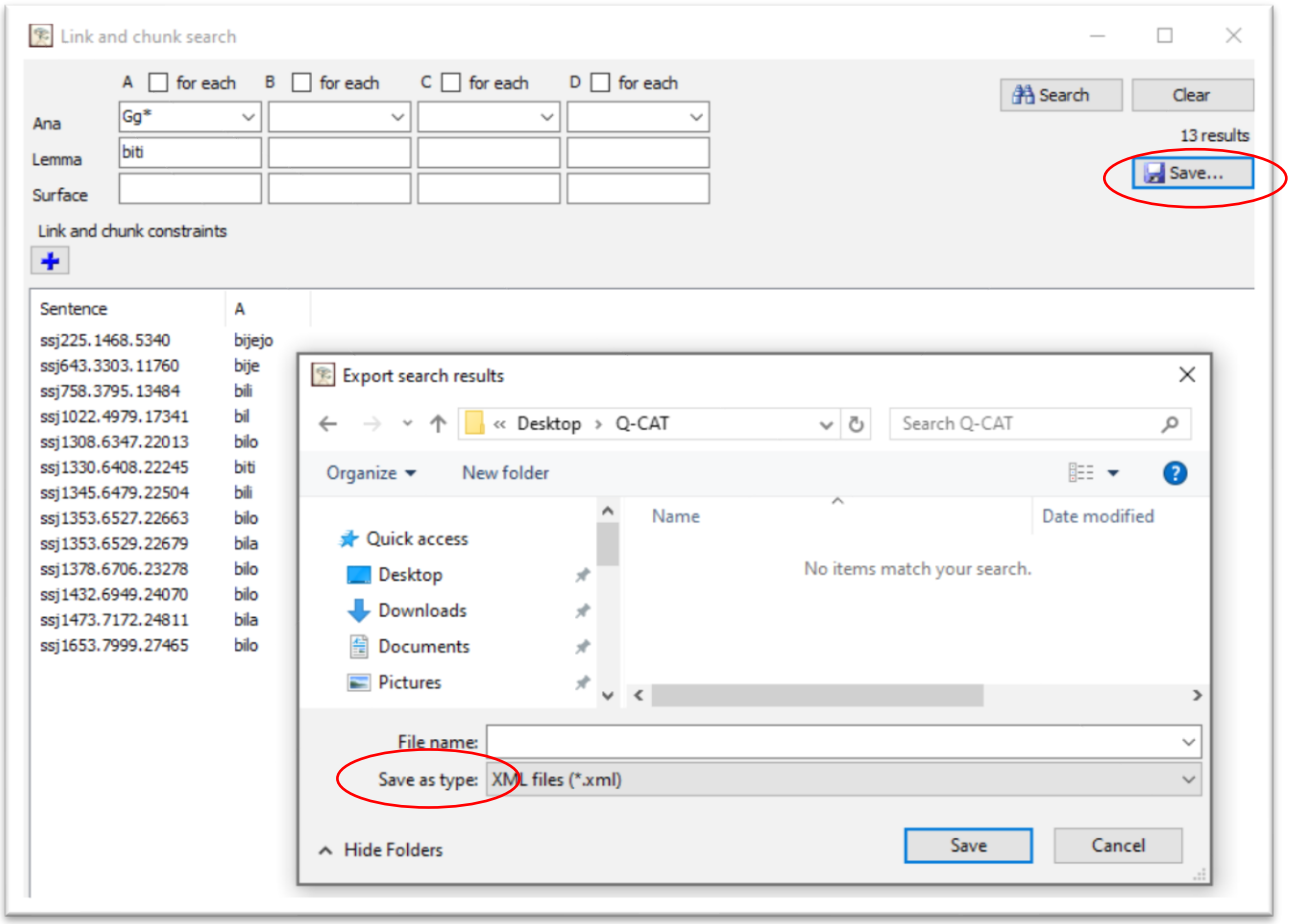
- *At any level, of any type*: pogoj se nanaša na katerokoli oznako med danima pojavnicama na katerikoli ravni označevanja, tj. na vse možne povezave/nize v korpusu. Besedilni opis identičnega iskalnega pogoja: ***:***.
- *At level X, any type*: pogoj se nanaša na katerokoli oznako na izbrani ravni označevanja (npr. SRL). Primer besedilnega opisa identičnega iskalnega pogoja: **SRL:***.
- *At level X, of type Y*: najbolj tipična oblika iskanja, ki se nanaša na izbrano oznako (npr. ACT) na izbrani ravni označevanja (npr. SRL). Primer besedilnega opisa identičnega iskalnega pogoja: **SRL:ACT**.

- *At level X, of any type except Y*: pogoj se nanaša na katerokoli oznako razen izbrane oznake (npr. ACT) na izbrani ravni označevanja (npr. SRL). Primer besedilnega opisa identičnega iskalnega pogoja: **SRL:!ACT**.
- *At any level except X*: pogoj se nanaša na katerokoli oznako na katerikoli ravni razen izbrane ravni označevanja (npr. SRL). Primer besedilnega opisa identičnega iskalnega pogoja: **!SRL:***.
- *Anything except X:Y*: pogoj se nanaša na katerokoli oznako na katerikoli ravni razen izbrane oznake (npr. ACT) na izbrani ravni označevanja (npr. SRL). Primer besedilnega opisa identičnega iskalnega pogoja: !**SRL:ACT**.

Ko izberemo želeni iskalni pogoj, se z dvoklikom na grafični prikaz ali klikom na gumb *OK* ta v besedilni obliki (npr. SRL:ACT) izpiše v polju za opis oznake v začetnem iskalnem oknu. Iskanje izvedemo s klikom na gumb ***Search***.

6.4 SHRANJEVANJE REZULTATOV ISKANJA

Rezultate iskanja shranimo s klikom na gumb Save. Odpre se nam pogovorno okno, v katerem določimo mesto datoteke z iskalnimi rezultati in v okencu Save as type izberemo med dvema formatoma zapisa rezultatov: v obliki datoteke XML (.xml) ali v obliki tekstovne datoteke (.txt).



Z izbiro shranjevanja v obliki datoteke **XML** shranimo **vse povedi**, v katerih je bil najden vsaj en rezultat iskalne poizvedbe, in sicer v enaki obliki, kot je poved zapisana v izhodiščnem korpusu, z vsemi oznakami vred. S tem načinom shranjevanja torej ustvarimo **podkorpus** izhodiščnega korpusa, ki ga lahko na enak način uvozimo v orodje Q-CAT za nadaljnje podrobnejše pregledovanje ali označevanje.

Z izbiro shranjevanja v obliki tekstovne datoteke **.txt** pa shranimo vse rezultate v obliki podobnega **seznama**, ločenega s tabulatorji, kot se prikaže pri rezultatih iskanja v orodju. Na shranjenem seznamu sta v zadnjem in prvem stolpcu izpisana poved in njen identifikator, v vmesnih stolpcih pa so izpisane relevantne pojavnice, kot sta izhodišče ali cilj povezave (splošno iskanje) oz. vse pojavnice znotraj niza (iskanje po nizih). Poleg oblike pojavnice so izpisani tudi vsi njeni metapodatki (oblika, lema, oblikoskladenjska oznaka, ID pojavnice v stavku), ločeni s poševnico. Tak seznam je za lažji pregled in nadaljnjo analizo primeren za uvoz v orodja za delo z razpredelnicami, kot je Excel.

Primer dela shranjenih rezultatov za iskalno poizvedbo po glavnem glagolu *biti* (poglavje 6.1.1.).

ssj225.1468.5340	bijejo/biti/Ggnstm/ssj225.1468.5340.t10	Premoč se izraža v bojih , ki jih samci bijejo med seboj .
ssj643.3303.11760	bije/biti/Ggnste/ssj643.3303.11760.t6	Ptujsko mestno gostinstvo že dolgo bije plat zvona .
ssj1022.4979.17341	bil/biti/Ggnd-em/ssj1022.4979.17341.t3	Trud ni bil zaman .

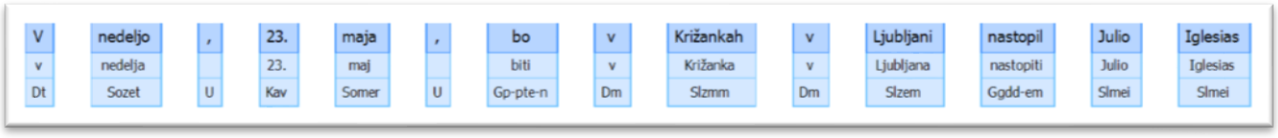
7 NASTAVITVE OZNAČEVANJA

V tem poglavju so predstavljeni načini označevanja, ki jih omogoča program Q-CAT (poglavji 7.1 in 7.2), ter vmesnik, ki omogoča njihovo spreminjanje (poglavje 7.3). Privzete ravni so opredeljene v datoteki Q-CAT-Settings.xml, ki se na računalnik ob namestitvi programa naložijo samodejno in izhajajo iz označevalnih ravni korpusa ssj500k (različica 2.2).

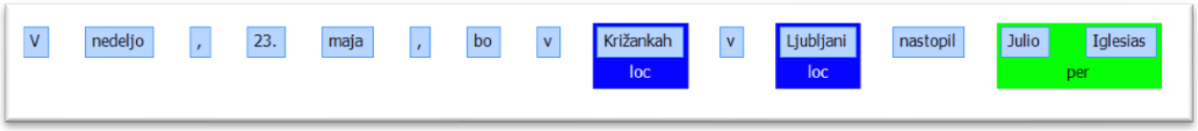
7.1 VRSTE OZNAČEVANJA

Program Q-CAT omogoča označevanje povedi s tremi vrstami oznak: oznakami oblik, nizov in povezav.

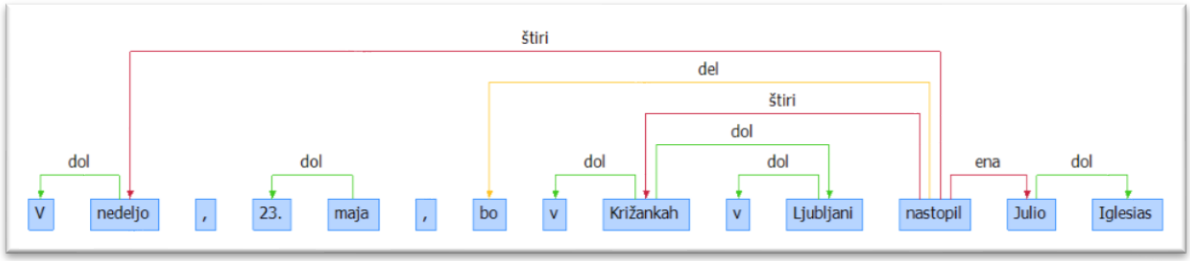
Oznake oblik so oznake, ki jih pripisujemo pregibnim oblikam besed (npr. obliki *nastopil*) in so tipično treh tipov: površinska oblika besede (npr. *nastopil*), osnovna oblika besede oz. lema (npr. *nastopiti*) in oblikoskladenjska oznaka, ki prinaša informacije o slovničnih lastnostih te oblike (npr. oznaka *Ggdd-em*, ki označuje glavne dovršne glagole z deležnikom v ednini množine). V programu so besedne oblike prikazane v modrih ploščicah, leme in oblikoskladenjske oznake pa v svetlejših ploščicah pod njimi.



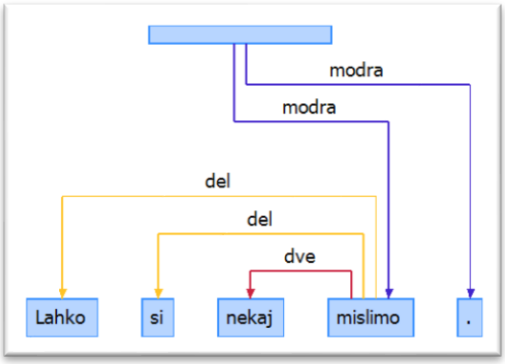
Nizi (angl. *chunk*) so oznake, ki jih pripisujemo povezanim ali nepovezanim nizom ene ali več besed, denimo za potrebe označevanja stalnih besednih zvez ali lastnih imen. V programu so oznake nizov prikazane kot **ploščice**.



Povezave (angl. *link*) so oznake, ki jih pripisujemo parom besed in usmerjeno potekajo od nadrejene k podrejeni besedi, denimo za potrebe označevanje skladenjskih ali semantičnih odvisnosti med besedami. V programu so oznake povezav prikazane kot **puščice**.



Pri nekaterih ravneh označevanja se lahko povezava vzpostavi tudi med besedno obliko in t. i. **korenskim elementom** (angl. *root*), ki označuje izvorno vozlišče nastalega drevesa. Na spodnji sliki je korenski element prikazan v obliki modrega pravokotnika, s katerim sta povezani dve pojavnici v povedi.



7.2 RAVNI OZNAČEVANJA

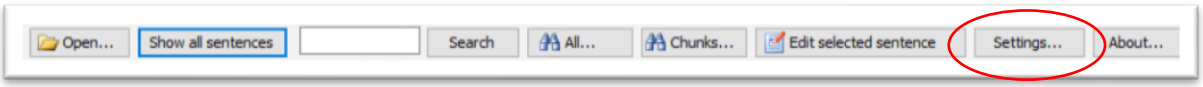
Medtem ko imajo *oznake oblik* vnaprej določene ravni označevanja (pripišemo lahko največ eno lemo in največ eno oblikoskladenjsko oznako na obliko), lahko pri oznakah vrste *nizi* in *povezave* uporabnik ustvari **eno ali več ravni označevanja** (angl. *levels of annotation*), znotraj katerih opredeli tudi nabor **ene ali več možnih oznak** (angl. *annotation types*).

V korpusu ssj500k v2.2 na primer najdemo dve ravni označevanja nizov in tri ravni označevanja povezav, pri čemer ima vsaka raven opredeljen svoj končni nabor oznak:

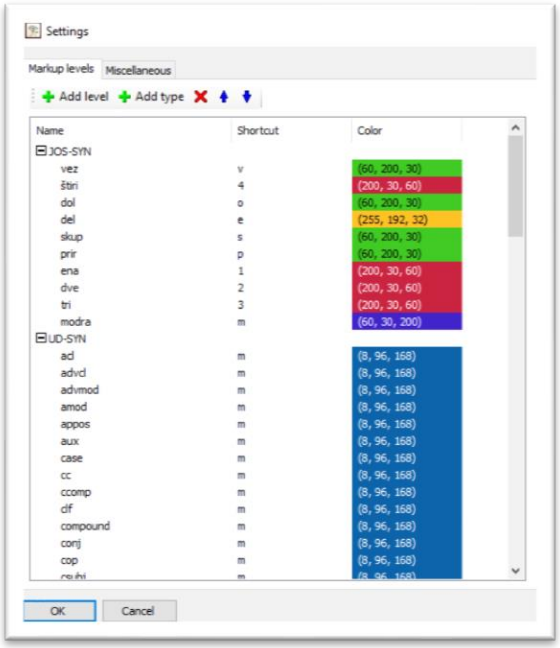
- nizi
  - imenske entitete (raven 'name'): oznake *loc, org, per, deriv-per, misc* …
  - glagolske stalne zveze (raven 'MWE'): oznake *IRV, VID, IAV, LVC.full* …
- povezave
  - skladnja JOS (raven 'JOS-SYN'): oznake *ena, dve, tri, štiri, modra* …
  - skladnja UD (raven 'UD-SYN'): oznake *nsubj, csubj, obj, amod, advmod* …
  - udeleženske vloge (raven 'SRL'): oznake *ACT, PAT, REC, ORIG, RESLT* …

7.3 VMESNIK ZA UREJANJE NASTAVITEV OZNAČEVANJA

Vmesnik za urejanje nastavitev možnosti označevanja odpremo s klikom na gumb **Settings…** v zgornji vrstici osnovnega okna programa, kot prikazuje spodnja slika.

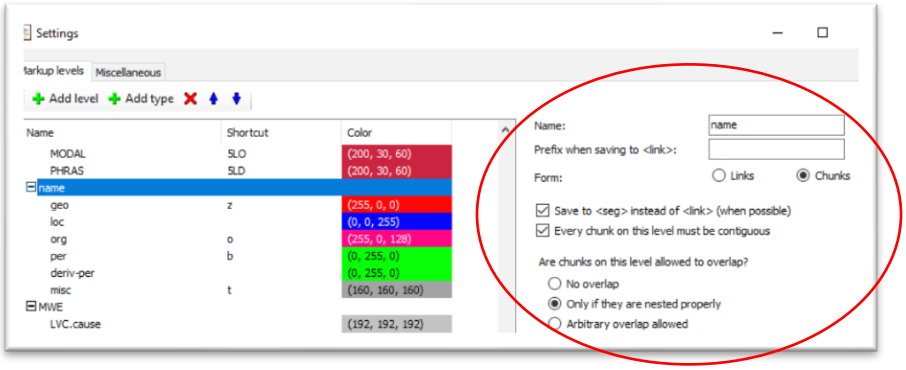


Odpre se nam okno z dvema zavihkoma. V zavihku **Markup levels** (ravni označevanja) so naštete vse obstoječe ravni označevanja za **nize ali povezave** (prvi stolpec, vrhnja kategorija), npr. raven skladenjskega razčlenjevanja JOS-SYN, vse oznake znotraj vsake izmed ravni (prvi stolpec, spodnja kategorija), npr. oznake *vez, štiri, dol* …, njihove bližnjice na tipkovnici (drugi stolpec) in barve (tretji stolpec). Nastavitve oblikoskladenjskih **oznak oblik** urejamo v zavihku **Miscellaneous**.



7.3.1 Nastavitve ravni označevanja

Lastnosti posameznih ravni označevanja lahko spreminjamo tako, da z miško izberemo ustrezno raven, da se na desni strani odpre urejevalnik ravni.



Določimo:

- **ime ravni** (polje *Name*), kakršno se bo prikazovalo v vmesniku
- **kratico** ravni (polje *Prefix when saving to <link>*), kakršna bo zapisana v XML datoteki, če se oznake te ravni shranjujejo znotraj elementa <link>
- **vrsto** označevalne ravni (polje *Form*), tj. ali gre za označevanje povezav (angl. *links*) ali nizov (angl. *chunks*)

Pri izbiri označevanja povezav dodatno določimo še:

- morebitno povezovanje na **korenski element** (*Links to this level may involve the root*).

Pri izbiri označevanja nizov dodatno določimo še:

- **način shranjevanja** oznak v XML datoteki korpusa (*Save to <seg> instead of <link> (when possible)*) – če odkljukamo, bodo oznake nize shranjene v element <seg>, kolikor je to mogoče
- omejitev glede morebitnega **označevanja prekinjenih nizov** (*Every chunk on this level must be contiguous*) – odkljukamo, če lahko oznake pripišemo samo neprekinjenim nizom pojavnic
- omejitev glede morebitnega **prekrivanja oznak** (*Are chunks on this level allowed to overlap?*) – izberemo ustrezno možnost:
    - brez prekrivanja oznak (*No overlap*)
    - samo prekrivanje ene oznake znotraj druge (*Only if they are nested properly*)
    - poljubno prekrivanje oznak (*Arbitrary overlap allowed*)

Novo raven označevanja ustvarimo s klikom na gumb z zelenim znakom plus *Add level* in njene lastnosti nastavimo na enak način, kot je opisano zgoraj.

Vrstni red ravni spreminjamo z modrima puščicama gor/dol. Raven **izbrišemo** s klikom na rdeči križec.

7.3.2 Nastavitve oznak

Lastnosti posameznih oznak spreminjamo tako, da z miško izberemo ustrezno oznako, da se na desni strani odpre urejevalnik oznak. Določimo:

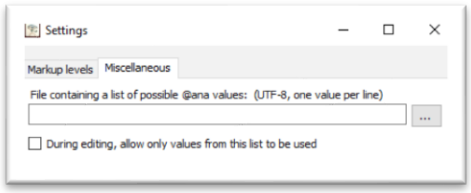
- **ime** oznake (polje *Name*), kakršno se bo prikazovalo in shranjevalo
- morebitno **bližnjico** na tipkovnici (polje *Shortcut*) za hitrejše označevanje
- **barvo** prikazane oznake (polje *Color*) – ploščice v primeru nizov oz. puščice v primeru povezav. S klikom na gumb se odpre podrobnejši urejevalnik barv.

Novo oznako znotraj dane ravni ustvarimo s klikom na gumb z zelenim znakom plus *Add type* in njene lastnosti nastavimo na enak način, kot je opisano zgoraj.

Vrstni red oznak znotraj posamezne ravni označevanja spreminjamo z modrima puščicama gor/dol. Oznako **izbrišemo** s klikom na rdeči križec.

7.3.3 Nastavitve seznama oblikoskladenjskih oznak

Če želimo pri označevanju oblik (poglavje 5.2.1) označevalcem omejiti nabor vseh možnih pripisanih oznak, lahko tak seznam naložimo v zavihku **Miscellaneous**, in sicer v obliki besedilne datoteke (.txt), pri kateri je vsaka oznaka vnesena v svojo vrstico. Po nalaganju seznama lahko dodatno izberemo, ali lahko uporabniki izbirajo samo med oznakami tega seznama (v tem primeru odkljukamo *During editing, allow only values from this list to be used*) ali pa lahko med označevanjem dodajajo tudi nove.



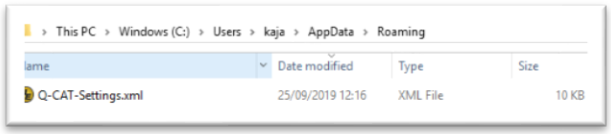
7.3.4 Shranjevanje nastavitev

Morebitne spremembe nastavitev označevanja shranimo tako, da kliknemo na gumb OK na dnu urejevalnika. Če sprememb ne shranimo, nas na to opozori tudi program ob izhodu (gumb *Close* spodaj ali križec desno zgoraj). Spremembe se shranijo v datoteko z nastavitvami **Q-CAT-Settings.xml**.

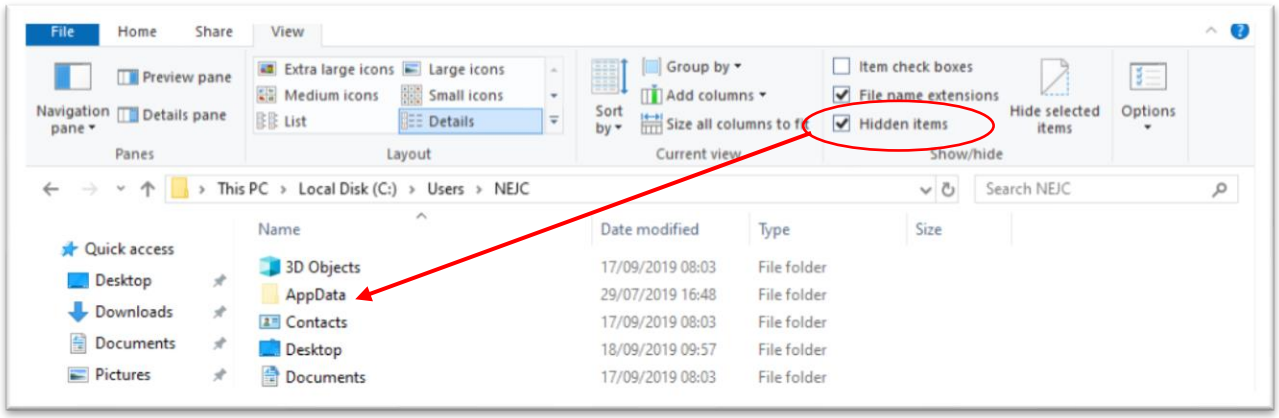
7.4 NEPOSREDNO UREJANJE DATOTEKE Z NASTAVITVAMI OZNAČEVANJA

Naprednejši uporabniki lahko namesto urejanja nastavitev v vmesniku za označevanje (poglavje 7.3) nastavitve spreminjajo tudi neposredno v datoteki z nastavitvami Q-CAT-Settings.xml, če se jim zdi ta način prikladnejši.

Datoteko z nastavitvami Q-CAT-Settings.xml najdemo v mapi Uporabniki > Uporabnik > AppData > Roaming, kot prikazuje spodnja slika.



Če mape AppData ne vidimo, kliknemo na zavihek *Pogled* in izberemo nastavitev *Pokaži skrite datoteke, mape in pogone,* kot prikazuje spodnja slika.



Datoteko odpremo s poljubnim urejevalnikom datotek XML ali Beležnico. Po vzoru obstoječe strukture opredelimo lastnosti poljubnega števila označevalnih ravni za oznake nizov in povezav, vključno z naborom možnih oznak in njihovih lastnosti.

7.4.1 Nastavitve označevanja nizov

Označevanje nizov opredeljujemo znotraj elementa `<chunkTypes>`, kjer podelement `<chunkType>` vsebuje lastnosti označevalne ravni, podelement `<chunkSubType>` pa lastnosti posameznih oznak.

Za posamezno raven označevanja nizov (`<chunkType>`) določimo:

- vrednost atributa ***name****:* ime ravni, kakršno se bo prikazovalo v vmesniku

- vrednost atributa ***anaPrefix*:** kratica ravni, kakršna je zapisana v XML datoteki, če so njene oznake shranjene znotraj elementa `<link>`. Če so oznake shranjene znotraj elementa `<seg>`, pustimo prazno.
- vrednost atributa ***savedAsSeg:*** način shranjevanja oznak v XML datoteki korpusa – vpišemo vrednost *true*, če so oznake zabeležene znotraj elementa `<seg>`, ali vrednost *false*, če so oznake zabeležene znotraj elementa `<link>`
- vrednost atributa ***overlap*:** restrikcijo glede morebitnega prekrivanja oznak – vpišemo vrednost *none,* če prekrivanje ni dovoljeno; vrednost *nested*, če je možno pojavljanje ene oznake znotraj druge; vrednost *any*, če je dovoljeno poljubno prekrivanje oznak.
- vrednost atributa ***contiguous*:** restrikcijo glede morebitnega označevanja prekinjenih nizov – vpišemo vrednost *true*, če se oznake pripisujejo samo neprekinjenim nizom pojavnic, sicer vpišemo vrednost *false*

Za posamezno oznako znotraj označevalne ravni nizov (`<chunkSubType>`) določimo:

- vrednost atributa ***name**:* ime oznake

Lastnosti oznak, kot so bližnjice na tipkovnici (vrednost elementa *shortcutKey*) in barva oznake v modelu RGB (vrednosti elementov *colorR*, *colorG* in *colorB*) za uspešen uvoz korpusa ni treba določiti vnaprej.

7.4.2 Nastavitve označevanja povezav

Označevanje nizov opredeljujemo znotraj elementa `<linkTypes>`, kjer podelement <link> vsebuje lastnosti označevalne ravni, podelement `<linkSubType>` pa lastnosti posameznih oznak.

Za posamezno raven označevanja povezav (`<linkType>`) določimo:

- vrednost atributa ***name**:* ime ravni, kakršno se bo prikazovalo v vmesniku
- vrednost atributa ***anaPrefix*:** kratica ravni, kakršna je zapisana v XML datoteki
- vrednost atributa ***mayLinktoEye:*** morebitno povezovanje na korenski element – vpišemo vrednost *true*, če se povezuje tudi korenski element, sicer vpišemo *false*.

Za posamezno oznako znotraj označevalne ravni povezav (`<linkSubType>`) določimo:

- vrednost atributa ***name**:* ime oznake

Lastnosti oznak, kot so bližnjice na tipkovnici (vrednost elementa *shortcutKey*) in barva oznake v modelu RGB (vrednosti elementov *colorR*, *colorG* in *colorB*) za uspešen uvoz korpusa ni treba določiti vnaprej.

7.4.3 Primer prilagojene datoteke z nastavitvami

Spodnja slika prikazuje izmišljeni primer prilagoditve datoteke Q-CAT-Settings.xml, če bi v program Q-CAT želeli uvoziti korpus, v katerem so poleg morebitnih oznak oblik (lem in oblikoskladenjskih oznak) označene tudi:

- konektorske besede in besedne zveze (nizi z oznakami *dodajanje, specifikacija, nasprotje*)
- odvisniki različnih tipov (povezave z oznakami *osebkov, predmetni, prislovni, prilastkov*).

```
<?xml version="1.0"?>
<QCatSettings fnMsdList="" forceMsdFromList="false">
  <chunkTypes>
    <chunkType name="konektorji" anaPrefix="" savedAsSeg="true" overlap="any"
contiguous="false">
      <chunkSubType name="dodajanje" />
      <chunkSubType name="specifikacija" />
      <chunkSubType name="nasprotje" />
    </chunkType>
  </chunkTypes>
  <linkTypes>
    <linkType name="odvisniki" anaPrefix="odvisniki" mayLinkToEye="false">
      <linkSubType name="osebkov" />
      <linkSubType name="predmetni" />
      <linkSubType name="prislovni" />
      <linkSubType name="prilastkov" />
    </linkType>
  </linkTypes>
</QCatSettings>
```

8 FORMAT

Program kot vhodno datoteko sprejme besedilni korpus v formatu XML TEI. Datoteke z opisom ustrezne različice tega formata in datoteke za njegovo validacijo so dostopne kot del korpusa ssj500k v2.2 (http://hdl.handle.net/11356/1210) in na naslovu https://github.com/clarinsi/TEI-schema.

Program kot vhodno datoteko sprejme tako označene kot neoznačene korpuse, pri čemer morajo biti vsi korpusi **predhodno tokenizirani in segmentirani**, tj. razdeljeni na povedi in pojavnice.

Vzorec korpusa v ustreznem formatu si lahko ogledamo tudi v namestitveni mapi programa (Programi > IJS > Q-CAT > Samples), v kateri so:

- vzorec korpusa brez oznak (**tei_example_no-annotations.xml**),
- vzorec korpusa z lemami in oblikoskladenjskimi oznakami (**tei_example_morphosyntactic-annotations.xml**) in
- vzorec korpusa z več raznolikimi oznakami (**tei_example_various-annotations.xml**).

Za morebitna vprašanja ali pomoč pri pripravi ustreznega zapisa korpusa podporo nudi CLARIN.SI (info@clarin.si).