

Predlagani koncept slovarja bo vseboval:

1. Opis splošnega razmerja med bodočimi partnerji (kdo so subjekti, kakšne so njihove pravice in pravice drugih subjektov do uporabe skupnega koncepta, pogoji za morebitni odstop od sodelovanja pri pisanju skupnega koncepta)

Mnenje ISJFR ZRC SAZU:

Koncept za slovar slovenskega jezika bo izhodišče za morebitno nadaljnje skupinsko delo (javno-zasebnega?) partnerstva pri izdelavi slovarja slovenskega jezika, kjer bi na osnovi dogovora, sklenjenega 3. 7. 2013, sodelovala Inštitut za slovenski jezik Frana Ramovša ZRC SAZU (javni raziskovalni zavod, nosilni partner) in še nedoločeni (zasebni?) partner. Določitev bodočih partnerjev že na začetku pisanja koncepta se zdi nujna, ker želimo, poznavajoč usposobljenost delovnih skupin s tega področja, optimalno uskladiti potrebe s strokovnimi zmožnostmi slovenske akademske krajine, ker je torej vsebina koncepta soodvisna od sodelujočih pri izdelavi slovarske baze in slovarja. Pogoje za morebitni odstop od sodelovanja pri pisanju skupnega koncepta bo treba doreči že pred morebitnim podpisom pogodbe, ki bo opredeljevala to dejavnost, in jih vanjo integrirati.

Mnenje avtorjev Predloga za izdelavo slovarja sodobnega slovenskega jezika:

Menimo, da slovarski koncept ne more vsebovati opredelitev glede bodočih partnerjev in razmerij med njimi, ker je to diskrecijska pravica subjekta, ki financira tako projekt izdelave koncepta kot kasneje projekt izdelave slovarja. Izdelani skupni slovarski koncept bo v lasti financerja, ki na tej podlagi lahko določi razpisne pogoje in izbere najbolj primernega izvajalca za izdelavo slovarja, kar je običajna praksa pri razpisih projektov.

Dogovor z dne 3. 7. se v celoti glasi: "Dne 3. 7. 2013 je potekal sestanek avtorjev Predloga za izdelavo slovarja sodobnega slovenskega jezika in predstavnikov Inštituta za slovenski jezik Frana Ramovša ZRC SAZU. Prisotni na sestanku so se dogovorili, da je mogoče v času enega leta izdelati skupni koncept, vključno z opisom morebitnih konceptualnih razhajanj. Sredstva, ki so potrebna za izdelavo skupnega koncepta, ocenjujemo na 2 FTE. Predvidevamo, da bi bil v projektu, ki bi izhajal iz skupnega koncepta, vodilni partner Inštitut za slovenski jezik Frana Ramovša ZRC SAZU." Iz besedila izhaja, da se nam ISJFR zdi najbolj verjetni vodilni partner pri prijavi na morebitni razpis, ki bi bil izveden na podlagi izdelave skupnega slovarskega koncepta, vendar je v zadnji konsekvenci to stvar morebitnega financerja, samega razpisa in konkretne prijave.

2. Vprašanje licenčnih razmerij partnerjev in drugih subjektov do slovarske baze in slovarja ter vizijo za čas po izdelavi slovarske baze in slovarja v predvidenem obsegu

Mnenje ISJFR ZRC SAZU:

Slovarski koncept bo natančno opredeljeval pravice posameznih partnerjev do objave skupnih rezultatov in njihove nadaljnje uporabe, tudi nadgrajevanja in dopolnjevanja.

Mnenje avtorjev Predloga za izdelavo slovarja sodobnega slovenskega jezika:

Menimo, da dostopnost in avtorske pravice glede slovarske baze oz. slovarja določa financer. Priporočena licenca, ki bi jo zagovarjali avtorji Predloga, je Creative Commons. Licenca CC omogoča najširšo možno rabo slovarja in slovarske baze za različne, tudi vnaprej nepredvidljive namene, to licenco pa za jezikovne vire priporočajo tudi relevantne institucije EU.

3. Finančni in kadrovski načrt izdelave slovarske baze in slovarja

Izvedba opredeljenih vsebin slovarske baze in slovarja bo v konceptu kadrovsko definirana ter časovno in finančno ovrednotena.

4. Namen izdelave slovarske baze in slovarja ter opis ciljnega slovarskega uporabnika oz. uporabnikov.

V konceptu bo opisan namen izdelave slovarske baze z vidika njene čim bolj optimalne izkoriščenosti, in sicer tako glede (a) izdelave na njej temelječih možnih različnih slovarskih opisov (zlasti z vidika uporabnika, medija ipd.) kot glede (b) možnosti izrabe za namene računalniškega procesiranja in izdelavo jezikovnotehnoloških aplikacij za slovenščino. Posebej bo določen ciljni uporabnik, in sicer glede na stopnjo zahtevnosti prikazanih podatkov, glede na specializiranost (šolski uporabnik, jezikoslovec, strokovnjak določenega področja) in glede na identificirane uporabniške potrebe. Dolgoročni namen izdelave slovarske baze in slovarja je tudi zagotoviti kontinuirano spremljanje in opis besedja slovenskega jezika in v konceptualno zasnovi zajeti čim širše jezikovne potrebe različnih ciljnih slovarskih uporabnikov.

5. Opis splošnih lastnosti slovarja (katero časovno obdobje vključuje, katero občno- in lastnoimensko besedje je zajeto in katero ne)

Koncept bo omejil časovno obdobje slovenskega jezika, ki ga bosta opisovala slovarska baza in slovar, in opredelil, katero občno- in lastnoimensko besedje bo vključeno ter katero ne. Opredeljeval bo tudi gradivne vire za izdelavo slovarske baze in slovarja ter leksikografske metode, ki se bodo pri tem uporabljale.

6. Opis makro- in mikrostrukture slovarja (obseg, katere informacije bo uporabnik dobil, kako bodo v bazi strukturirane in kako prikazane v slovarju)

V slovarskem konceptu bosta na makrostrukturni ravni opredeljena in pojasnjena predvsem obseg in vrsta elementov. Na mikrostrukturni ravni bosta orisana načrt strukturiranja jezikovnih podatkov v slovarski bazi in način podajanja jezikovnih podatkov (slovničnih, pomenskih in pragmatičnih) v slovarju kot končnem izdelku. Predvidena je izdelava testnega vzorca slovarskih enot.

7. Opis normativnosti v slovarju (kaj jo določa in kako bo v slovarju prikazana)

Slovarski koncept se bo opredelil do vprašanja normativnosti v sodobni slovenščini in pojasnil, kako bodo normativni podatki prikazani v slovarju.

8. Opis gradivnih virov in metod dela

Slovarski koncept bo predstavil prednosti in slabosti obstoječih gradivnih virov in med sodobnimi leksikografskimi metodami, upoštevajoč usposobljenost partnerjev, izbral optimalno kombinacijo za izdelavo slovarja slovenskega jezika.

9. Opis slovarskega metajezika

Z načrtovanim slovarskim konceptom uvajamo na Slovenskem sodobnejše leksikografske prakse, zato nekatere klasične leksikografske pojme (re)definiramo in oblikujemo sodobnejši slovarski metajezik, potreben za izdelavo slovarske baze in slovarja.

10. Vprašanje delne ali sprotne spletne objave še nedokončanega slovarja

Koncept se bo opredelil do vprašanja, kdaj in kako, če sploh, v digitalnih medijih objavljati še nedokončane slovarske sestavke oziroma dokončno izdelane slovarske sestavke, preden bo slovar izdelan v predvidenem obsegu.

11. Opis procesa pridobivanja leksikalnih podatkov iz besedilnih korpusov in drugih virov ter njihove analize oz. priprave za ožje leksikografsko delo (lahko vključuje tudi množičenje).

Koncept bo vseboval opis referenčnega korpusa ter proces njegove nadgradnje, postopke avtomatskega in polavtomatskega pridobivanja leksikalno-gramatičnih podatkov iz korpusa, opis jezikovnih orodij za njegovo analizo ter postopkov korpusne analize za posamezne dele leksikografskega opisa. Če bo koncept vključeval tudi množičenje (crowdsourcing), bodo opisana tudi spletna orodja in postopki, ki jih bo predvidel ta del procesa.

12. Opis sprotnega spremljanja sprememb v besedišču (vključuje sistem posodabljanja in nadgradnje besedilnih korpusov).

V tem delu bodo opredeljeni postopki in orodja, s katerimi bo mogoče spremljati in zaznavati spremembe v besedišču, kot so novi pomeni, podpomeni, kolokacije ipd. Kot predpogoj realizacije tovrstnih postopkov je redno posodabljanje korpusnega gradiva, zato bodo opisane tudi metode, časovno obdobje in ostali elementi te aktivnosti.

13. Izdelava geslovnika in opis morebitne prednostne obravnave aktualnega besedišča.

Predstavljena bo metodologija, ki bo uporabljena pri izbiri geslovnika, kar vključuje tako vire, na katerih bo temeljil geslovník, kakor tudi tipe leksike, ki bodo v slovarju, podana pa bodo tudi pojasnila za odločitve v primeru izpusta določenega dela besedišča. Opis bo vključeval tudi razmislek o razdelitvi geslovnika v prioritete skupine z vidika leksikografske obravnave, upoštevajoč potrebe uporabnikov in (ne)pokritost v obstoječih jezikovnih priročnikih.

14. Načela vključevanja enobesednih ali večbesednih slovarskih iztočnic v slovarsko bazo.

Koncept bo opredelil načela za oblikovanje izrazne podobe iztočnic oz. njihove osnovne slovarske oblike ter merila za določanje večbesednih leksikalnih enot, njihovo tipologijo ter razmerja glede na vsebovano iztočnico in pomensko odvisnost oz. neodvisnost od registriranih pomenov. Natančneje bodo opredeljene vrste leksikalnih in slovničnih podatkov, vezanih na posamezni tip večbesedne leksikalne enote ter način njihove organizacije in integracije v slovarsko bazo in slovar.

15. Opis formalne strukture slovarske baze (vključno s shemo/DTD).

Predstavljena bo struktura slovarske baze glede na svojo večnamenskost: opisani bodo tipi vključenih leksikalnih in gramatičnih podatkov, njihova notranja organizacija in medsebojna razmerja. Opisan bo tudi računalniški format slovarske baze s podrobno predstavitvijo sheme DTD in glede na možno povezljivost z drugimi bazami podatkov in integracijo že obstoječih jezikovnih podatkovnih baz.

16. Opis leksikografskega sistema oz. postopkov, med njimi:

- a. sistem členjenja pomenov in podpomenov: podana bodo merila za prepoznavanje besedne večpomenskosti ter določena načela za prikazovanje hierarhičnih razmerij v odnosu podpomenov do osnovnih pomenov in med osnovnimi pomeni
- b. sistem slovarskih razlag oz. definicij: določeni bodo tipi slovarskih razlag glede na (i) uporabnika, (ii) skladišne in pomenske lastnosti besed, (iii) način vključevanja slovničnih podatkov ter pragmatičnih lastnosti pomena v pomensko

razlago. Pripravljene bodo študije posameznih primerov in predlagane optimalne rešitve glede na postavljene parametre.

- c. kvalifikatorski sistem: podrobno bodo predstavljeni potencialni kvalifikatorski sistemi za označevanje slovničnih, stilnih, žanrskih in drugih posebnosti v rabi besed in zvez ter izdelana morebitna primerjava med njimi z vidika uporabniške naravnosti, obvestilnosti, razumljivosti, načina prikaza v bazi in slovarju ipd.
- d. vprašanje stalnih besednih zvez (in njihovega razmerja do morebitnih večbesednih slovarskih iztočnic): postavljena bodo merila za uvrstitev stalnih besednih zvez in njihovo umestitev v slovarsko bazo na mesto iztočnice oz. na ustrezno mikrostrukturno mesto
- e. vprašanje vključevanja podatkov o frazeologiji: postavljena bodo merila za uvrstitev frazemov in način njihovega prikaza, vključno z variantnostjo
- f. sistem vključevanja slovničnih in leksikalno-gramatičnih podatkov (npr. kolokacije in stavčni vzorci): določene bodo vrste (i) slovničnih podatkov, (ii) podatkov o tipičnem besedilnem okolju besed ali zvez, (iii) podatkov o vezljivostnih lastnostih besed ter način njihove integracije v slovarsko bazo in slovar.
- g. načela vključevanja zgledov rabe (kolokacije, iztržki, stavčni zgledi): opisana bodo merila za vključitev vseh treh vrst zgledov v slovarsko bazo in način njihovega prikaza v slovarju
- h. sistem vključevanja normativnih podatkov: opisan bo način interpretiranja podatkov o sodobnem jeziku v odnosu do pravopisne norme, ki je vsebovana v pravopisnih pravilih in slovarskem delu Slovenskega pravopisa (2001) ter način beleženja in prikaza tovrstnih podatkov v slovarski bazi oz. slovarju, tako v spletni kot v tiskani različici
- i. sistem vključevanja podatkov o etimologiji: opredeljena bodo temeljna načela za prikaz zgodovinskih in etimoloških podatkov
- j. sistem vključevanja podatkov o sinonimiji: opisan bo način integracije sinonimnih podatkov v slovar in morda v slovarsko bazo, opredeljeni bodo parametri same baze in morebitna integracija podatkov, kot jih je možno izluščiti s pomočjo orodja Sketch Engine (*sketch difference*)
- k. sistem vključevanja podatkov o izgovoru: opisane bodo vrste podatkov o izgovoru, ki bodo vključene v slovar in slovarsko bazo, in način njihovega beleženja, pri čemer bodo upoštevani tako specifika digitalnih oblik slovarja kot potrebe ostalih potencialnih uporabnikov oz. tehnologij (npr. strojna prepoznavna govora)
- l. sistem vključevanja in preverjanja terminološkega gradiva: opisana bodo načela za sprejetje terminološkega besedišča v slovarsko bazo in slovar (npr. izdelava terminoloških podkorpusev, zajem terminov iz šolskih učbenikov ipd.), določen bo nabor področij in način kvalifikatorskega opredeljevanja strokovno specifične leksike in opisan postopek vzpostavitve terminološkega redakcijskega dela.
- m. opis načel sklicevanja: opisan bo način vključevanja podatkov o sklicnih elementih v slovarsko bazo in opredeljeni bodo možni prikazi sklicevanja, pri čemer bodo upoštevani spletni medij in drugi digitalni mediji ter tiskana različica slovarja

- n. sistem vključevanja statističnih podatkov o iztočnicah: določeni bodo tipi statističnih podatkov o iztočnicah, ki bodo vključeni v slovarsko bazo, npr. pogostost, žanrska distribucija, raba po letih, in način(i) njihovega beleženja, pri čemer bodo med drugim upoštevane predvidene vizualizacijske rešitve.

17. Demo in opis vizualizacije slovarskih podatkov v spletnem in drugih digitalnih medijih.

Opisana in prikazana bo okvirna vizualizacija podatkov zlasti v spletnem mediju, pri čemer bo vsak del oz. spletna stran s slovarskimi informacijami komentirana z vidika funkcionalnosti in predvidene uporabniške izkušnje, predstavljena pa bo tudi povezanost vizualizacije s strukturiranjem podatkov v slovarski bazi.

18. Opis transformacije podatkov v slovarski bazi za tiskani medij.

Prikazan in opisan bo način prilagoditve podatkov v slovarski bazi za tiskani medij, po možnosti za različne tipe uporabnikov (npr. šolski, jezikoslovec), in sicer tako za tisk iz spletne vizualizacije kot za pretvorbo v knjižno obliko.

19. Opis (morebitne) integracije zunanjih virov v spletno predstavitev slovarskih informacij.

Predstavljeni bodo zunanji viri, ki bi se vključili v spletno predstavitev slovarskih informacij kot dodatni vir informacij, npr. o govornem jeziku, multimedijskih vsebinah ipd. Za vsak vir bo podan njegov opis, način izrabe, relevantnost oz. povezanost s slovarsko informacijo in predlog vizualizacije podatkov v spletni predstavitvi.

20. Opis načina beleženja in vrste podatkov, namenjenih računalniškemu procesiranju slovenščine in uporabi v jezikovnotehnoloških aplikacijah za slovenščino:

semantični podatki iz slovarske baze so nujni predpogoj za možnost avtomatskega razdvoumljanja pomenov, skladenjskega razčlenjevanja, oblikoskladenjskega označevanja, prepoznavanja imenskih entitet in drugih postopkov računalniškega procesiranja naravnih jezikov. Opisana bo struktura baze, predvsem v povezavi z zahtevami, ki izhajajo iz potreb omenjenih aplikacij.

21. Po potrebi evalvacijo leksikografskih rešitev pri predvidenih uporabnikih.

V predstavitvi koncepta bodo vključene ankete, beleženje dnevnikov in podobne sodobne metode preverjanja odzivov uporabnikov na leksikografske rešitve v novem slovarskem konceptu.

Mnenje ISJFR ZRC SAZU:

Te sugestije bodo lahko vplivale na naknadno spremembo koncepta le ob soglasju partnerjev.

Mnenje avtorjev Predloga za izdelavo slovarja sodobnega slovenskega jezika:

Menimo, da iz istega razloga kot v prvi točki, kjer vnaprejšnje določanje partnerjev negira diskrecijske pravice subjekta, ki financira projekt izdelave koncepta, tudi v tej točki potrebe po soglasju partnerjev ni mogoče prejudicirati.