

# Tehnični postopki preverjanja pri eSSKJ

**Andrej Perdih**

Znanstvenoraziskovalni center Slovenske akademije znanosti in umetnosti,  
Inštitut za slovenski jezik Frana Ramovša, Ljubljana, Slovenija, andrej.perdih@zrc-sazu.si

DOI: <https://doi.org/10.18690/um.ff.3.2024.6>

ISBN: 978-961-286-878-9

---

Za zagotavljanje čim večje usklajenosti slovarske vsebine in za zmanjšanje števila napak v slovarju je poleg ročnega vsebinskega pregleda mogoče uporabiti tudi različne avtomatizirane postopke preverjanja slovarske vsebine. Na primeru slovarja *eSSKJ: Slovar slovenskega knjižnega jezika* prikazujemo splošno uporabne postopke in specifične postopke, pripravljene namensko za ta slovar. S temi postopki poleg tehničnih napak odkrivamo tudi vsebinske neskladnosti.

**Ključne besede:** eSSKJ, slovar, tehnično preverjanje, leksikografija, slovenščina

In addition to manual content checking, various automated dictionary checking procedures can be used to ensure maximum consistency of dictionary content and to reduce the number of errors in the dictionary. *Dictionary of the Slovene Standard Language* (eSSKJ) is used to illustrate the general and specific procedures developed specifically for this dictionary. These procedures help detect not only technical errors, but also inconsistencies in content.

**Keywords:** eSSKJ, dictionary, technical verification, lexicography, Slovene

## Uvod<sup>1</sup>

Razlagalne slovarje srednjega obsega navadno izdeluje skupina leksikografov skozi daljše časovno obdobje, zaradi česar obstaja potreba po usklajevanju in odpravljanju neuskklajenosti znotraj slovarja ali neskladnosti z dogovorjenimi pravili, po katerih slovar nastaja. Usklajevanje znotraj slovarja med drugim pomeni tudi zagotavljanje, da se podatki, ki se pojavljajo na različnih mestih v slovarju, med seboj ujemajo. Usklajevanje in

---

<sup>1</sup> Prispevek je nastal v okviru raziskovalnega programa P6-0038, ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

odpravljanje napak lahko poteka tako ročno kot z različnimi avtomatiziranimi tehničnimi postopki, ki bodisi neposredno odpravljajo napake ali pa opozarjajo na tiste dele slovarske vsebine, za katere je treba ročno preveriti, ali v resnici predstavljajo napake ali ne. Nekatera pravila namreč veljajo v vseh primerih, druga pa so sicer veljavna v večini okoliščin, vendar ne vedno. Nekaterih odstopanj od pravil tako ni mogoče predvideti vnaprej ali pa jih ni mogoče enostavno algoritmično opredeliti, zato tudi pravila za avtomatizirano preverjanje dosegajo različno stopnjo zanesljivosti.

Možnost, da bi vse z avtomatiziranimi tehničnimi postopki prepoznane napake popravljali samodejno, je, razen v najbolj preprostih primerih (dvojni presledki, presledek pred vejico ipd.), tvegana. Prvič zato, ker lahko obstaja več različnih vzrokov za neko napako, drugič zato, ker napake niso le površinske, ampak so posledica bolj zapletenih notranjih razmerij med različnimi deli slovarja, tretjič zato, ker napaka lahko pokaže na nujnost temeljitejšega posega v vsebino, kar pa je treba reševati pri vsakem primeru posebej.

Priprava avtomatiziranih postopkov preverjanja sicer zahteva nekaj časa, prav tako dodajanje novih in izboljševanje obstoječih postopkov, vendar je to pri večjih slovarjih tudi z vidika prihranka časa upravičeno, saj je manj časa treba nameniti preverjanju tistih medsebojnih odnosov znotraj slovarja, ki jih je mogoče algoritmično opisati. Prav tako se poveča zanesljivost preverjanja, saj gre pri tem tudi za enote, ki jih pregledovalci ne opazijo tudi pri večkratnem pregledu.

Pomemben vidik preverjanja na podlagi tehničnih kriterijev je tudi, da nekatere napake izvirajo iz vsebinskih nedoslednosti. Namreč, preverjanje slovarja z vsebinskega vidika je še vedno prepuščeno človeški presoji in ga z običajnimi algoritmičnimi postopki ni mogoče neposredno preverjati. Z razvojem umetne inteligence bo v prihodnje verjetno mogoče preverjati tudi vsebinske neustreznosti, trenutno pa to poteka zlasti posredno prek preverjanja tehnične ustreznosti in medsebojne povezanosti slovarskih vsebin.

## **eSSKJ**

*eSSKJ: Slovar slovenskega knjižnega jezika* nastaja na Inštitutu za slovenski jezik Frana Ramovša ZRC SAZU na osnovi koncepta iz leta 2015 (Gliha Komac idr. 2015; Bizjak Končar idr. 2016; Gliha Komac idr. 2016). Gre za splošni razlagalni slovar z načrtovanim obsegom 100.000 slovarskih sestavkov sodobnega knjižnega jezika. Opisuje sodobno jezikovno rabo, ki

se kaže v besedilnih korpusih, zlasti korpusu Gigafida (Logar idr. 2012; Krek idr. 2020), dodatno pa so upoštevani tudi drugi besedilni viri. Slovar opisuje jezikovne enote, ki izkazujejo dovolj pogosto in uveljavljeno rabo v knjižnih besedilih. Pri tem prinaša tako glasovne (Mirtič 2018; Mirtič 2020), pisne in slovnične lastnosti jezikovnih enot (Ledinek in Perdih 2022; Mirtič in Snój 2022; Snój 2022; Snój in Žele 2022) kot tudi njihov pomen (Krvina in Petric 2022; Ledinek in Michelizza 2023), skladijsko okolico, ki je ponazorjena s tipičnimi kolokacijami in rabo v stavkih (Michelizza 2021), mesto v jezikovnem sistemu (Gliha Komac in Divjak Race 2022; Divjak Race 2023), besednodružinsko povezanost z drugimi enotami, nefrazeološke in frazeološke stalne zveze, v katerih te jezikovne enote nastopajo (Meterc 2019; Perdih in Ledinek 2019; Meterc 2022; Petric 2022), oblike, normativna obvestila, etimološke osvetlitve, slovar pa prinaša tudi povezave do Slovarja slovenskega znakovnega jezika (Klemenčič in Kern 2022). Slovar je objavljen na portalu Fran<sup>2</sup> (Ahačič idr. 2015a; Ahačič idr. 2015b; Perdih 2018; Perdih 2020).

## **Splošne tehnične možnosti preverjanja**

Slovar eSSKJ nastaja v programu iLex (Erlandsen 2010), kjer so podatki zapisani v formatu XML (Ledinek in Perdih 2012). Formalna struktura slovarja je določena v XSD shemi, ki poleg hierarhije in položaja različnih tipov podatkov po tehnični strani omejuje tudi njihovo vsebino, npr. zgolj na omejen seznam možnosti, kot so spustni seznamei poimenovanj besednih vrst ipd., ali pa dovoljuje povsem prost zapis. Program iLex preverja skladnost vnosov s splošnimi pravili XML, hkrati pa stalno preverja tudi skladnost strukture in vsebine z XSD shemo. Pri tem gre zlasti za preverjanje ustreznosti hierarhije in umestitve elementov v slovarskem sestavku. V nekaterih primerih je dovoljeno izbirati med vnaprej določenimi možnostmi (na spustnem seznamu), program pa opozarja na morebitna odstopanja. Na primer, če je v etimološkem razdelku vneseno poimenovanje jezika, ki v shemi ni predviden, sta na voljo dve možnosti – ali se napako popravi ali pa se po uredniškem dogovoru ta jezik doda na seznam veljavnih možnosti. Program omogoča tudi preverjanje unikatnosti identifikacijskih števil, npr. za označevanje pomenov in podpomenov.

---

<sup>2</sup> <https://fran.si/201/esskj-slovar-slovenskega-knjiznega-jezika/>

## Specifični tehnični postopki preverjanja pri eSSKJ

### Preverjanja, nevezana na vsebino

Poleg splošnih preverjanj se pri eSSKJ uporabljajo tudi dodatna preverjanja, za katere se odloči uredniška ekipa in jih postopoma dopolnjuje. Nekatera med njimi so enaka tudi pri drugih nastajajočih slovarjih, druga pa so prilagojena ali pa se drugje ne uporabljajo. Preverjanja, ki načeloma niso vezana na vsebino, zajemajo preverjanje znakov, uporabo črkovalnika, podvojenih presledkov, urejanje presledkov pred vejico in podobno.

### Preverjanje znakov

Pri preverjanju znakov gre lahko za primere, ko v naboru znakov Unicode na videz enaki ali podobni znaki niso identični, saj je njihova koda različna. Tak primer je polglasnik in njemu podobni znaki:  $\Theta$ ,  $\vartheta$  in  $\vartheta$  s šestnajstiškimi kodami 018F, 0259 in 04D9. Prvi znak označuje veliki latinični polglasnik, drugi mali latinični polglasnik, tretji znak pa polglasnik iz ciriličnega podnabora znakov. Pri zapisu v slovarski bazi je predvidena le uporaba malega latiničnega polglasnika, zato je uporaba drugih dveh znakov načeloma napačna – le v izjemnih primerih bi bila raba drugih znakov lahko upravičena (do zdaj še nimamo takega primera, vendar to možnost dopuščamo).

Zadrega se ravno tako pojavlja pri uporabi različnih vodoravnih »črtic« (vezaji, pomišljaji, minus ipd.) in narekovajev, kjer lahko uredniška ekipa natančneje opredeljuje dovoljene rabe posameznih znakov in s tem omogoči strojno preverjanje nepredvidenih znakov.

Zaradi uporabe pisave ZRCola<sup>3</sup> (Weiss 2004) s specifičnimi kodami znakov, ki vsebujejo diakritična znamenja (te kode nastopajo v nestandardizirani podmnožici Unicode: *Območje zasebne uporabe*, ang. *Private use area*), se preverja tudi uporaba načeloma nedovoljenih posamičnih diakritičnih znamenj iz podmnožice »sestavljiva ločevalna znamenja«.

### Odsotnost povezave

Nekatera preverjanja, ki niso vezana na vsebino, potekajo z orodji, vgrajenimi v program iLex, druga preverjanja pa se izvajajo s pomočjo XSL transformacije, ki potencialne napake prikaže kot spletno stran (HTML).

---

<sup>3</sup> <https://zrcola.zrc-sazu.si/>

Med preverjanja, ki so bolj ali manj tehnične narave, spadajo tudi preverjanje **odsotnosti povezave** pri razvezavah simbolov/kratic, povezav v etimološkem razdelku ipd., saj je v dinamičnem delovnem procesu mogoče, da vzpostavitev povezave v določenem trenutku še ni mogoča, ker ciljni slovarski sestavek še ne obstaja, zato je treba povezave dodati naknadno.

### **Skladenjski vzorci**

Vzorci, ki opredeljujejo tipično skladenjsko okolje, npr. pridevnik + samostalnik, samostalnik + samostalnik v rodilniku itd., vsebujejo tudi podatek o tem, katera izmed enot v vzorcu ustreza iztočnici. V samostalniški zvezi je torej opredeljeno, ali je iztočnica jedro ali prilastek. Strojno lahko preverjamo neskladja, kjer je v skladenjskem vzorcu opredeljeno, da je iztočnica ene besedne vrste (npr. pridevnik), čeprav je iztočnica dejansko samostalnik.

### **Preverjanja, vezana na vsebino**

Pri opredeljevanju možnosti, kaj vse je mogoče znotraj slovarja preverjati, uredniki slovarja pri pregledovanju ugotavljajo, da se nekateri tipi napak pojavljajo večkrat, temu pa sledi razmislek o tem, ali in kako bi bilo mogoče tak tip napake algoritmično opredeliti. Temu sledi priprava namenske XSL transformacije in preverjanje, ali je rezultat transformacije skladen s pričakovanji. Zlasti je treba preveriti, ali je algoritem res našel vse podtipе iskanega problema, pa tudi, ali je kot napake identificiral tudi pravilne primere. Postopke preverjanja zato po potrebi izboljšamo in ponovno preizkusimo.

Za preverjanja, ki so neposredno povezana z vsebino, npr. preverjanje sinonimov, taksonomskih imen in dvojnic, se uporablja namenska XSL transformacija, ki potencialne napake prikaže kot spletno stran. V resnici gre še vedno za preverjanje na podlagi tehničnih kriterijev (tj. enakost zapisa na različnih mestih v slovarju), vendar te napake pogosto izhajajo iz vsebinske neuskklajenosti. Gre namreč za tiste podatke, ki bi morali biti na različnih mestih v slovarju sistemsko povezani med seboj in usklajeni. Na ta način se torej preverja in zagotavlja notranja konsistentnost nekaterih mikrostrukturnih delov slovarja.

V nadaljevanju prikazujemo različne tipe napak, podajamo njihov opis in jih ponazarjamo s konkretnimi primeri, ki smo jih pri dosedanjem preverjanju našli. Podajamo tudi ugotovitev, ali je najdena težava v resnici napaka ali ne, in kako smo napako odpravili.

## Sinonimija

Sinonimija ima v eSSKJ tezavrsko, razlagalno in normativno vlogo (Gliha Komac idr. 2015: 48), za tehnično preverjanje pa je relevantna zlasti tezavrska in normativna vloga sinonimov, saj so sinonimi v teh primerih navedeni ob razlagi. Sinonimija je urejena z medsebojno povezavo enakih pomenov v različnih slovarskih sestavkih. Razen izjemoma, npr. iz besedotvornih ali pisnodvojničnih razlogov, morajo imeti sinonimi enake razlage. Preverjanje vključuje iskanje razlik med razlagami sinonimov: morda je katera od razlag pomotoma neusklajena z drugimi, morda je sinonim naveden po pomoti ali pa so razlage namerno zapisane drugače. Napake ročno popravimo, namerna odstopanja pa shranimo na seznam izjem, ki jih nato strojno ne pregledujemo več.

### Sinonimi: uskladitev razlag

Pri sinonimih *boksar* : *bokser* sta bili zapisani naslednji razlagi:

- orožje, ki se natakne na prste in poveča učinek udarca s pestjo
- hladno orožje, ki se natakne na prste in poveča učinek udarca s pestjo

Druga razlaga ima glede na prvo razlago na začetku dodano besedo »hladno«. Rešitev je poenotenje obeh razlag.

### Sinonimi: pravilna neskladnost ubeseditve razlag

Pri sinonimnem paru *dateljnov* : *datljev* sta zapisani naslednji razlagi:

- ki je v zvezi z dateljnovimi palmami
- ki je v zvezi z datljevimi palmami

Razlika v razlagah ima besedotvorni izvor (*dateljnovimi* pri *dateljnov* in *datljevimi* pri *datljev*). Tak tip izrazne neenakosti razlag pri sinonimih je v eSSKJ ustrezen, zato je bil ta sinonimni par dodan na seznam izjem in se pri naslednjih strojnih preverjanjih ne preverja več.

### Sinonimi: napačno povezani pomeni

Pri sinonimnem paru *top* : *otopel* sta bili navedeni naslednji razlagi:

- ki kaže, izraža tako neobčutljivost, brezvoljnost
- ki ne reže, ne seka več dobro

Ta pomena nista v medsebojnem sinonimnem razmerju, zato je bila navedba sinonimov izločena, ob tem pa je bilo preverjeno, ali so sinonimna razmerja pri vseh ostalih pomenih obeh iztočnic ustrezno urejena.

### **Sinonimi: navedeni, a nepovezani**

Preverja se tudi, ali so v slovarju sinonimi navedeni, a niso tehnično ustrezno povezani. Pri najdenih primerih se ročno vzpostavijo povezave na ustrezen pomen sinonima. V zaključnem slovarju bi se to sicer lahko preverjalo že v postopku validacije, ker bi bile povezave obvezne. Pri nastajajočem slovarju pa to ni mogoče, saj ciljni slovarskih sestavek pogosto še ni izdelan.

### **Sinonimi: navedeni, a še neobjavljeni**

Ker eSSKJ objavljamo v enoletnih intervalih, ni neobičajno, da slovarski sestavki nekaterih sinonimov še niso pripravljani za objavo. Pri objavljenih slovarskih sestavkih zato teh sinonimov ne prikazujemo ne glede na to, ali so v slovarski bazi že navedeni. Izjemo predstavljajo nevtralni sinonimi pri močno zaznamovanih pomenih, saj pri teh nevtralni sinonim praviloma ni vprašljiv, prav tako predstavlja pomembno slogovno in normativno informacijo za uporabnika slovarja. Pri avtomatskem preverjanju izdelamo seznam vseh sinonimov, ki ne bodo objavljeni, in izbrane ustrezno označimo, da naj se kljub temu prikažejo v slovarju.

Primeri parov iztočnica – še neobjavljeni sinonim, pri katerih je sinonim že naveden v slovarju: *bremza – zavora, britof – pokopališče, lojtra – lestev*.

Prav tako se preverja, ali sinonimi niso navedeni pri tistih pomenih, kjer obstaja enaka razlaga tudi kje drugje v slovarju. V tem primeru skupaj zberemo vse pomene z enako razlago in preverjamo, ali pri katerem od pomenov manjka navedba katerega od sinonimov.

### **Sinonimi: enake razlage, manjkajoča navedba sinonimov**

Identificirali smo iztočnici *ornitolog* in *ptičar*, pri obeh je bila v enem od pomenov zapisana naslednja razlaga:

- kdor zlasti ljubiteljsko opazuje, proučuje ptice

V tem primeru smo dodali manjkajoči podatek, da sta to sinonima.

### **Sinonimi: enake razlage, a sinonimi niso upravičeni**

Pri iztočnicah *kvartet* in *sekstet* smo prav tako našli enaki razlagi brez sinonimne povezave:

- skladba za to zasedbo

Kljub enaki razlagi pa to nista sinonima, saj gre pri kvartetu za zasedbo s štirimi, pri sekstetu pa s šestimi instrumentalisti ali pevci. Dejstvo, da sta razlagi kljub temu zapisani enako, izhaja iz sistematične ubeseditve metonimičnih razlag (v tem primeru gre za tip metonimije ‘zasedba’ → ‘skladba’). Tovrstne razlage shranjujemo na seznam izjem, ki jih pri nadaljnjih strojnih preverjanjih ne upoštevamo. Na ravni podatkovne zbirke bi bila še boljša rešitev ta, da bi bili takšni podpomeni označeni kot metonimični in bi jih zato avtomatično izločali iz postopka preverjanja, obenem pa bi imeli v slovarski bazi označene vse metonimične podpomene. Označevanje teh podpomenov v slovarski bazi pa se lahko enostavno naredi tudi naknadno prav na podlagi filtra, ki nastaja po trenutnem pristopu.

Podobni postopki preverjanja se uporabljajo tudi pri preverjanju sistematičnosti navajanja taksonomskih imen pri sinonimih in pisnih dvojnicah.

### **Taksonomska imena: enake razlage, manjkajoča navedba taksonomskega imena**

V paru *plahitica* : *planinska resa*, kjer imata obe enako razlago, je prva imela navedeno taksonomsko ime, druga pa ne:

- *Alchemilla xanthochlora*

Ker pričakujemo navedbo na obeh mestih, je bilo taksonomsko ime dodano tudi pri *planinska resa*.

### **Pisne dvojnice**

Podobne zadrege kot pri sinonimih in taksonomskih imenih preverjamo tudi pri pisnih dvojnicah, od katerih vsaka tvori svoj slovarski sestavek. Ker je že samih pisnih dvojnic razmeroma malo, le nekatere med njimi pa vstopajo v sinonimno razmerje z drugimi leksemi, v tem trenutku ne razpolagamo s primeri napak.

## Smer prihodnjega razvoja

Nove postopke preverjanja dodajamo, ko zaznamo potrebo po njih in ko ugotovimo, ali je mogoče algoritmično opredeliti postopek, po katerem je mogoče odkriti potencialne napake. Poleg prepoznavanja napak je mogoče razmišljati tudi o možnostih preverjanja, ki bi pri razlagah nakazovala šibke točke v ubeseditvah.

Za izboljšanje razlag bi lahko uporabili metodo preverjanja korpusne frekvence vseh besed v razlagah in ročnega preverjanja možnosti zamenjave ali izbrisa redkih besed v razlagi ali preoblikovanja celotne razlage. Ta metoda, uporabljena pri nastajajočem *Akademskem slovarju sodobne češčine*, je preprosta in jo je mogoče nadgraditi z dodatnimi kriteriji – za češčino avtorji predlagajo uporabo t. i. »jedrnega besedišča« na podlagi češkega frekvenčnega slovarja, s čimer bi preverjali prekrivnost razlagalnih besed z jedrnim besediščem (Škrabal idr. 2022).

## Zaključek

Pomembno fazo pri nastajanju slovarju predstavlja preverjanje vsebine in odpravljanje tehničnih in vsebinskih napak. Poleg ročnih pregledov se pri nastajanju slovarja eSSKJ uporabljajo tudi avtomatizirani tehnični postopki, ki pomagajo pri hitrejšem in doslednejšem odpravljanju različnih vrst napak in vsebinskih neuskkljenosti. Razlikujemo lahko med preverjanji, ki jih omogoča že sam program za slovarsko delo in splošne vgrajene tehnologije, ter med specifičnimi pravili, ki smo jih opredelili v programu za XSL transformacijo. Rezultate preverjanja za boljšo preglednost prikazujemo kot spletno stran. Ko na ta način dobimo seznam potencialnih napak, je še vedno mogoče, da nekatere med njimi v resnici niso napake. Take primere dodajamo na seznam izjem: mednje sodijo npr. sinonimi, pri katerih se razlage razlikujejo, npr. iz besedotvornih ali pisnodvojničnih razlogov, izjemo predstavljajo tudi enake razlage pri metonimičnih podpomenih, kjer je enakost razlage le površinska, vsebinsko pa gre za drugačno pomensko vsebino. Postopke preverjanja postopoma širimo glede na ugotovljene potrebe.

Čeprav tehnični postopki temeljijo na algoritmičnih opredelitvah, poleg tehničnih nedoslednosti odkrivajo tudi napake, ki izvirajo iz vsebinskih nedoslednosti, zlasti pri ugotavljanju neskladnih odnosov med vsebinami na različnih mestih v slovarju. Tako se izboljšuje tudi vsebina slovarja, zlasti njegova notranja usklajenost.

## Literatura

Kozma AHAČIČ, Nina LEDINEK in Andrej PERDIH, 2015a: Portal Fran – nastanek in trenutno stanje. *Simpozij Obdobja 34, Slovnica in slovar – aktualni jezikovni opis*. Ur. Mojca Smolej. Ljubljana: Znanstvena založba Filozofske fakultete. 57–66.

Kozma AHAČIČ, Nina LEDINEK in Andrej PERDIH, 2015b: Fran: Next Generation Slovenian Dictionary Portal. *Natural language processing, corpus linguistics, lexicography / Eighth International Conference, Bratislava, Slovakia, 21–22 October 2015: proceedings*. Ur. Katarína Gajdošová, Adriána Žáková. Bratislava: RAM-Verlag. 9–16.

Aleksandra BIZJAK KONČAR, Duša DIVJAK RACE, Manca ČERNIVEC, Dejan GABROVŠEK, Nataša GLIHA KOMAC, Nataša JAKOP, Janoš JEŽOVNIK, idr., 2016: eSSKJ: Slovar slovenskega knjižnega jezika.

Duša DIVJAK RACE, 2023: Teoretična izhodišča za kvalificiranje v eSSKJ. *Naslavljanje raznolikosti v jeziku in književnosti*. Ur. Jožica Jožef-Beg, Mia Hočevar, Neža Kočnik. (Zbornik Slavističnega društva Slovenije, 33). Ljubljana: Zveza društev Slavistično društvo Slovenije. 219–229.

Jens ERLANDSEN, 2010: Ilex, a General System for Traditional Dictionaries on Paper and Adaptive Electronic Lexical Resources. *Proceedings of the 14th EURALEX International Congress*. Ur. Anne Dykstra, Tanneke Schoonheim. Leeuwarden/Ljouwert: Fryske Akademy. 306–306.

Nataša GLIHA KOMAC in Duša DIVJAK RACE, 2022: Tipologija slovarskih oznak za tretjo izdajo Slovarja slovenskega knjižnega jezika (eSSKJ), *Slavistična revija* 70, 611–625. <https://doi.org/10.57589/srl.v70i4.4086>

Nataša GLIHA KOMAC, idr., 2015: *Koncept novega razlagalnega slovarja slovenskega knjižnega jezika*. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.

Nataša GLIHA KOMAC idr., 2016: Novi slovar slovenskega knjižnega jezika – predstavitev temeljnih konceptualnih izhodišč. *Škrabčevi dnevi 9. Zbornik prispevkov s simpozija 2015*. Ur. Franc Marušič, Petra Mišmaš, Rok Žaucer. Nova Gorica: Založba Univerze v Novi Gorici. 19–33.

Simona KLEMENČIČ in Boris KERN, 2022: Povezljivost Slovarja slovenskega znakovnega jezika in Slovarja slovenskega knjižnega jezika, *Slavistična revija* 70, 655–664.

Simon KREK idr., 2020: Gigafida 2.0: the reference corpus of written standard Slovene. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Ur. Nicoletta Calzolari. ELRA – European Language Resources Association. 3340–3345.

Domen KRVINA in Špela PETRIC, 2022: Vrstni pridevniki: teoretična obravnava in razlagalne strategije v eSSKJ glede na SSKJ, predhodne in sodobne slovarje, *Slavistična revija* 70, 525–543. <https://doi.org/10.57589/srl.v70i4.4080>.

Nina LEDINEK in Mija MICHELIZZA, 2023: Issues in treating determinologized lexemes in the eSSKJ general explanatory dictionary of Slovenian, *Collegium antropologicum* 47, 110. <https://doi.org/10.5671/ca.47.2.2>

Nina LEDINEK, Andrej PERDIH, 2012: Izdelava XML-shem za slovarske projekte na primeru nastajajočih tipološko raznovrstnih slovarjev, *Zbornik Osme konference Jezikovne tehnologije, 8. do 12. oktober 2012, [Ljubljana, Slovenia]*, 123–128.

Nina LEDINEK, Andrej PERDIH, 2022: Enobesedne in večbesedne leksikalne enote v slovenščini: opredelitev meje za potrebe izdelave eSSKJ, *Slavistična revija* 70, 475–488. <https://doi.org/10.57589/srl.v70i4.4074>.

Nataša LOGAR idr., 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in ccKRES: gradnja, vsebina, uporaba*. (Zbirka Sporazumevanje). Ljubljana: Trojina, zavod za uporabno slovenistiko; Fakulteta za družbene vede.

Matej METERC, 2019: Analiza frazeološke variantnosti za slovarski prikaz v eSSKJ-ju in SPP-ju, *Jezikoslovni zapiski* 25, 33–45. <https://doi.org/10.3986/JZ.25.2.2>

Matej METERC, 2022: Explanations of meaning for paremiological genres in the eSSKJ and SPP dictionaries: what is expressed with a proverb, saying, superstition, weather proverb, antiproverb, wellerism, slogan and unconventional reply?, *Rasprave Instituta za hrvatski jezik i jezikoslovje* 48, 524. <https://doi.org/10.31724/rihjj.48.2.5>

Mija MICHELIZZA, 2021: Med tipičnim in stereotipnim: družbena občutljivost pri pripravi ilustrativnega gradiva v eSSKJ, *Rasprave Instituta za hrvatski jezik i jezikoslovje* 47, 631–653. <https://doi.org/10.31724/rihjj.47.2.13>

Tanja MIRTIČ, 2018: Pravorečna problematika s splošnoslovaropisnega vidika, *Slavia Centralis* 11, 172–184.

Tanja MIRTIČ, 2020: Odpiranje pravorečnih vprašanj pri pripravi rastočega slovarja eSSKJ. *1. slovenski pravorečni posvet: zbornik povzetkov*. Ljubljana. 17.

Tanja MIRTIČ, Marko SNOJ, 2022: Merila za razlikovanje med dvojnicami in sinonimi v eSSKJ, *Slavistična revija* 70, 459–473. <https://doi.org/10.57589/srl.v70i4.4073>

Andrej PERDIH, 2018: Dictionary Portal Fran: Current State and Future Developments. *Slovanská lexikografie počátkem 21. století: sborník příspěvků z mezinárodní konference, Praha 20.–22. 4. 2016*. Ur. B. Niševa, D. Blažek, I. Krejčířová, K. Skwarska, E. Šlaufová, M. Vašíček. Praha: Slovanský ústav AV ČR, v. v. i. 57–65.

Andrej PERDIH, 2020: Portal Fran: od začetkov do danes, *Rasprave Instituta za hrvatski jezik i jezikoslovje* 46, 997–1018. <https://doi.org/10.31724/rihjj.46.2.28>

Andrej PERDIH, Nina LEDINEK, 2019: Multi-word Lexical Units in General Monolingual Explanatory Dictionaries of Slavic languages, *Slovenski jezik – Slovene Linguistic Studies* 12, 113–134.

Špela PETRIC, 2022: Besedne zveze z vidika slovaropisja na primeru sestavkov eSSKJ, *Slavistična revija* 70, 489–502. <https://doi.org/10.57589/srl.v70i4.4077>

Marko SNOJ, 2022: Prispevki k slovnici slovenskih medmetov, *Slavistična revija* 70, 237–253. <https://doi.org/10.57589/srl.v70i3.4047>

Marko SNOJ in Andreja ŽELE, 2022: Merila za razločevanje med medmeti in členki, *Slavistična revija* 70, 557–572.

Michal ŠKRABAL, Michaelá LIŠKOVÁ in Martin ŠEMELÍK, 2022: On defining vocabulary in a monolingual online dictionary: Some remarks from the lexicographical practice on the Academic Dictionary of Contemporary Czech. *Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress*. Ur. Annete Klosa-Kückelhaus, Stefan Engelberg, Christine Möhrs, Petra Storzjohann. Mannheim: IDS-Verlag. 71–74.

Peter WEISS, 2004: Vnašalni sistem ZRCola. *Jezikovne tehnologije: zbornik B 7. mednarodne multi-konference Informacijska družba IS 2004, 9. do 15. oktober 2004*. Ur. Tomaž Erjavec, Jerneja Žganec Gros. Institut »Jožef Stefan«. 124–125.

## TECHNICAL VERIFICATION PROCEDURES IN eSSKJ

### Summary

The paper presents automated procedures for the technical verification of the structure and content of the eSSKJ (*Dictionary of the Slovene Standard Language*, 3rd Edition), which is being compiled at the Fran Ramovš Institute of the Slovene Language at ZRC SAZU. The technical verification procedures are applied during the compiling of entries as well as in the word-processing software and in the final phase of preparing the dictionary for publication. The procedures help to identify potential errors reflecting inconsistencies in the content.

The eSSKJ dictionary is compiled by using the iLex software with the built-in functions to prevent violation of basic XML rules (well-formedness) and by checking entries against an XSD schema designed for each dictionary. The software also allows for checking the uniqueness of identification numbers, e.g. for senses and sub-senses. In addition to the general checking, further verification procedures are determined by the editorial team. The majority of these procedures are performed by means of an XSL transformation, resulting in a web page displaying potential errors.

The verification process includes checking the use of appropriate characters, the number and order of quotation marks, brackets and other characters. For the senses with the same explanation, the indication of synonyms is checked to ensure that they are not accidentally omitted. Allowable exceptions are metonymic meanings that superficially have the same explanation, although the meanings themselves are different. All synonyms are checked to ensure that they have the same explanation. Permitted exceptions are explanations that differ among themselves due to word-formation or spelling reasons. Synonyms are also checked for their collocating with words with potentially inadequate meaning. For the senses that enter into a synonymous relationship and have taxonomic names attached to them, the correspondence of these taxonomic names is checked. In addition, the semantic links in the case

of variants, the consistency of the structural patterns used in relation to the part of speech of the headword, the absence of a link in the case of symbols and abbreviations, the adequacy of the links in the etymological section, etc. are checked. The automated checking of the structure and content of the eSSKJ dictionary helps to ensure high quality and consistency of the data while saving time on the content checking and proofreading.