

PEDAGOŠKI LEKSIKOGRAFSKI PORTAL *FRANČEK*

Andrej PERDIH

Inštitut za slovenski jezik Frana Ramovša
(Ljubljana, Slovenija)

<https://orcid.org/0000-0002-2248-9666>
andrej.perdih@zrc-sazu.si

V prispevku je predstavljen leksikografski del prvega slovenskega pedagoškega slovarsko-slovnicega portala *Franček*. Namenjen je osnovnošolcem in srednješolcem, vsebinsko in grafično pa je prilagojen za tri starostne stopnje: 1.–5. razred osnovne šole (6–10 let), 6.–9. razred osnovne šole (11–14 let), srednja šola (15–18 let). Za 100.006 iztočnic so predstavljeni pomen, sinonimija, frazeologija, morfologija, zvočni posnetek izgovora ter narečni, etimološki in zgodovinski podatki.

Ključne besede: pedagoški jezikovni portal, pedagoška leksikografija, šolski slovar, povezovanje slovarjev, slovenščina, knjižni jezik

ПЕДАГОГИЧЕСКИЙ СЛОВАРНО-ГРАММАТИЧЕСКИЙ ПОРТАЛ «*FRANČEK*»

Андрей ПЕРДИХ

Институт словенского языка Франа Рамовша
(Любляна, Словения)

В статье представлена лексикографическая часть первого словенского педагогического словарно-грамматического портала «*Franček*». Портал предназначен для учащихся начальной и средней школы, а его содержание и графика адаптированы для трех возрастных уровней: 1–5 класс начальной школы (6–10 лет), 6–9 класс начальной школы (11–14 лет), средняя школа (15–18 лет). В словарных статьях описаны 100 006 лексем — для каждой приводятся: значение, синонимия, фразеология, морфология, звукозаписи произношения, а также диалектные, этимологические и исторические данные.

Ключевые слова: портал педагогического языка, педагогическая лексикография, школьный словарь, комбинирование содержания словарей, словенский язык, литературный язык

THE EDUCATIONAL LEXICO-GRAMMATICAL LANGUAGE PORTAL *FRANČEK*

Andrej PERDIH

Fran Ramovš Institute of the Slovenian Language
(Ljubljana, Slovenia)

The article presents the lexicographic part of the first educational lexico-grammatical language portal for Slovenian, *Franček*. The portal is aimed at primary- and secondary-school

students, and is specifically adjusted for three age groups: 1st to 5th grade of primary school (6 to 10 years), 6th to 9th grade of primary school (11 to 14 years), and secondary school (15 to 18 years). For 100,006 headwords, semantic, synonymic, morphological, pronunciation, phraseological, dialect, historical, and etymological data is provided.

Keywords: pedagogical language portal, pedagogical lexicography, school dictionary, children's dictionary, dictionary linking, Slovenian language, standard language

I. Uvod

Portal *Franček* (www.francek.si, www.francek.si) predstavlja tip slovarskega portala, kjer so podatki iz različnih slovarjev izbrani in združeni iz različnih slovarjev v novo leksikografsko zbirko, ki temelji na skupnem geslovniku s 100.006 iztočnicami. Leksikografski del je osrednji del širšega, prvega pedagoškega slovarsko-slovničnega portala za slovenščino, saj je namenjen za učence osnovnih in srednjih šol ter prinaša medsebojno povezane slovarske kot tudi slovnične vsebine. S tem se zapolnjuje vrzel pri naboru pedagoških opisov slovenščine, saj so bili dosedanja slovarski opisi slovenskega jezika za pedagoško rabo na voljo le v zelo omejenem obsegu, zlasti pa ne s tako celovitim opisom. Portal je bil objavljen septembra 2021. Učitelji lahko portal uporabljajo kot dodaten pripomoček za delo z učenci pri jezikovnem pouku, obenem pa bo pomagal pri izobraževanju učencev za razumevanje in uporabo slovarjev. Učencem bo v pomoč pri raznovrstnih nalogah v okviru šolskega pouka in v prostem času pri iskanju odgovorov na različna vprašanja: o pomenu, sopomenkah, narečnih in zgodovinskih vidikih, slovnčnih vprašanjih itd.

Poleg slovarskega dela sta v portal integrirani šolski slovnici za osnovno in srednjo šolo (Ahačič 2017a; Ahačič 2017b), na voljo je tudi jezikovna svetovalnica za učitelje slovenščine¹. Celoto dopolnjujejo didaktični materiali, ki so večinoma izdelani po avtomatiziranih postopkih iz obstoječih in novih slovarskih virov. Za potrebe pedagoške leksikografije je nastal *Korpus šolskih besedil slovenskega jezika* (Ahačič idr. 2021), ki ga sestavljajo zlasti besedila učencev, leposlovje za otroke in enakomerno zastopana strokovna področja iz učbeniškega gradiva. Korpus predstavlja gradivsko podlago za *Šolski slovar slovenskega jezika* (Petrič Žižič 2020; Godec Soršak 2015, 2019) v obsegu 2000 iztočnic, ki je namenjen za učence 1.–5. razreda osnovne šole in je prav tako sestavn del portala *Franček*. Registrirani uporabniki so za uporabo portala dodatno motivirani z zbiranjem nagrad za pogosto uporabo portala, s portalom pa je povezana tudi igra za mobilne naprave *Besedoboj*, ki je namenjena zabavnemu ukvarjanju z jezikom.

Za kakovost vsebine in prikaza podatkov na portalu je velikega pomena tesno sodelovanje leksikografov, jezikoslovcev neleksikografov, didaktikov in informatikov ter 72 učiteljev z 22 osnovnih in srednjih šol². Skozi celotno obdobje projekta izdelave portala in

¹ Jezikovno svetovanje je bilo na Inštitutu za slovenski jezik do zdaj namenjeno zlasti splošni javnosti (Dobrovolje, Verovnik, Bizjak Končar 2017) in reševanju terminoloških vprašanj (Žagar Karer 2017), s portalom *Franček* pa se odpira še specializirana svetovalnica kot podpora učiteljem v pedagoškem procesu.

² Število sodelujočih je sodelovala ves čas projekta, nekateri posamezniki so se dodatno pridružili med samim izvajanjem aktivnosti.

pridruženih vsebin³ so potekala preverjanja in testiranja predlaganih rešitev in njihove izboljšave glede na odzive didaktikov, učiteljev in njihovih učencev.

Portal je nastajal v ugodnih okoliščinah, zaradi katerih je mogoče pričakovati, da bo dobro sprejet med učitelji in učenci. Leta 2014 je bil namreč vzpostavljen slovarski portal Fran, na katerem so za splošno javnost brezplačno ponujeni številni slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU (Ahačič, Ledinek, Perdih 2015a; Ahačič, Ledinek, Perdih 2015b; Perdih 2020). Pomen portala, njegova prepoznavnost in uporabnost se kaže med drugim s povprečno 280.000 iskanji na dan v letu 2021, pri čemer slovenščino govori nekaj več kot dva milijona govorcev. Učbeniki in delovni zvezki za pouk slovenskega jezika vključujejo tudi uporabo slovarjev na portalu Fran. Če bo portal *Franček* dobro sprejet pri učiteljih slovenskega jezika v osnovnih in srednjih šolah, lahko predvidevamo, da bo postopoma prav tako postal vključen v šolske učbenike in delovne zvezke.

II. Osnovne značilnosti

Tako vsebina kot prikaz podatkov sta prilagojena trem starostnim skupinam glede na njihove zmožnosti in šolski sistem: 1.–5. razred osnovne šole (6–10 let, razredna stopnja), 6.–9. razred osnovne šole (11–14 let, predmetna stopnja), srednja šola (15–18 let).

Portal *Franček* z vidika iskanja informacija predstavlja enotno začetno točko, kjer so različni slovarski podatki organizirani z vidika iskane vsebine. Slovarski podatki so na portalu namreč razvrščeni po vsebinskih razdelkih glede na različna vprašanja, ki si jih učenci lahko zastavijo, npr. «*Kaj pomeni ta beseda?*», in prinašajo slovarske podatke o osmih vsebinskih vprašanjih: o pomenu, sinonimiji, frazeologiji, morfoloiji, na voljo je zvočni posnetek izgovora, pa tudi narečni, etimološki in zgodovinski podatki. Bistvena razlika od portalov, na katerih nastopajo slovarji vsak kot samostojna enota, je ta, da so uredniki portala iz slovarjev že vnaprej izbrali tiste podatke, ki so za posamezno jezikovno vprašanje relevantni za šolskega uporabnika, in optimizirali podatke, ki so redundantni zaradi ponavljanja v različnih slovarjih. Ker je *Franček* pedagoški portal, so bile nekatere podrobnosti iz slovarjev ob tem poenostavljene ali izločene, saj presegajo pedagoške potrebe ali pa šolskemu uporabniku otežijo osredotočanje na bistvene jezikovne prvine na račun prevelikega števila podrobnosti. Ob tem so bili pripravljene novi pedagoški leksikografski opisi, ki zapolnjujejo največje vrzeli pri dosedanjih opisih.

Integracija slovarjev v eno leksikografsko celoto obenem povzroča tudi težave, s katerimi se soočajo uredniki portala. Slovarji, ki predstavljajo vsebinski vir za portal, zaradi svoje nepopolnosti niso namreč nekonsistentni le znotraj sebe, temveč zlasti nekompatibilni med seboj. Izhajajo iz različnih teoretičnih postavk, izdelani so bili po različnih metodoloških pristopih, njihove besedilne gradivske podlage se med seboj razlikujejo, slovarji pa so prav tako nastajali z večdesetletno časovno razliko. Njihovi vsebinski

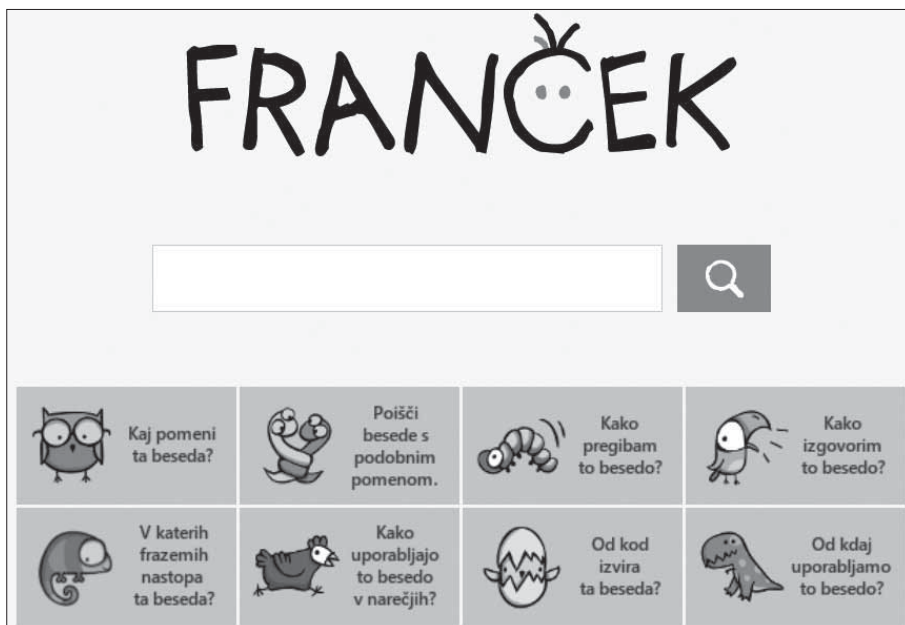
³ Projekt *Portal Franček, Jezikovna svetovalnica za učitelje slovenščine in Šolski slovar slovenskega jezika*. Vodja projekta: Kozma Ahačič. Trajanje projekta: julij 2017–september 2021.

poudarki so različni, zato so nekateri jezikovni podatki v ospredju in natančneje obdelani, drugi le dodani in zato manj natančni (vendar ne napačni). S poenostavljanjem za pedagoški namen se sicer nekatere nekonsistentnosti zabrisujejo, vendar popolne skladnosti med različnimi slovarskimi podatki vseeno ni mogoče pričakovati.

III. Struktura in vsebina

Portal združuje podatke različnih slovarjev, vendar pri tem niso več v ospredju izhodiščni slovarji, ampak so leksikografski podatki organizirani po vsebinskih modulih, kjer so podatki lahko pridobljeni iz različnih izhodiščnih slovarjev:

1. Pomen (*Kaj pomeni ta beseda?*)
2. Sopomenke (*Poišči besede s podobnim pomenom.*)
3. Pregibanje (*Kako pregibam to besedo?*)
4. Izgovor (*Kako izgovorim to besedo?*)
5. Frazemi (*V katerih frazemih nastopa ta beseda?*)
6. Narečja (*Kako uporabljajo to besedo v narečjih?*)
7. Izvor (*Od kod izvira ta beseda?*)
8. Zgodovina (*Od kdaj uporabljamo to besedo?*)



Slika 1: Iskalnik in osem vsebinskih razdelkov slovarskega dela portala *Franček*

Geslovnik kot makrostrukturalni segment portala predstavlja nabor 100.006 iztočnic, ki temelji na geslovnih dveh splošnih razlagalnih slovarjev: druge izdaje Slovarja sloven-

skega knjižnega jezika (SSKJ2) iz leta 2014⁴ in eSSKJ, ki nastaja od leta 2016 dalje (Gliha Komac idr. 2015). Za izdelavo geslovnika nismo uporabili vseh iztočnic iz slovarjev, saj smo izločili tiste, ki so v SSKJ2 v vseh (ali edinem) pomenu opredeljene kot zastarele (kvalifikator zastar.) in nekatere vulgarne (kvalifikator vulg.), dodatno pa so v geslovník portala *Franček* dodane prislovne podiztočnice iz SSKJ2, pa tudi večje število glagolskih podiztočnic, ki vsebujejo pomenske razlage (izmed teh so vključene le tiste podiztočnice, ki imajo v odnosu do iztočnice dodan ali izločen morfem *se/si*, npr. *delati se pri delati*, *sprehajati pri sprehajati se*, ne pa tudi različni deležniki in deležja). Geslovník portala *Franček* tako ne vsebuje podiztočnic — vsi obravnavani (enobesedni) leksemi nastopajo kot iztočnice. Stalne besedne zveze (terminološke in neterminološke) in frazeološke enote niso obravnavane kot iztočnice.

Homografom, tj. iztočnicam, ki so (brez upoštevanja naglasnih znamenj) pisane enako, so dodani **indikatorji** različnih tipov, ki so namenjeni lažji identifikaciji iskane iztočnice (Perdih 2021a). V slovarjih, iz katerih črpamo podatke, so namreč homografom⁵ za razločevanje med njimi pripisani številski indeksi. Ti indeksi ne prinašajo nobene možnosti za vsebinsko razločevanje, za pedagoški portal pa smo želeli učencem omogočiti smiselno razločevanje v procesu izbire s seznama zadetkov iskanja. Za razliko od pomenskih indikatorjev, ki so v slovarjih uporabljeni za razlikovanje med pomeni iste iztočnice (prim. Gantar 2015: 166 za *Leksikalno bazo za slovenščino*; Žmigrodzki 2018: 32–33 za *Wielki słownik języka polskiego*), so indikatorji na portalu *Franček* uporabljeni za razločevanje med homografi. Kadar je mogoče, so uporabljeni **besednovrstni indikatorji**, ki razločujejo iztočnice različnih besednih vrst (*vrh*¹ m : *vrh*² predlog) ali besednovrstnih podkategorij, npr. spol pri samostalniku (*prst*¹ m : *prst*² ž). Poleg 1165 iztočnic s pripisanimi besednovrstnimi indikatorji je 3268 iztočnicam pripisan **pomenski indikator**, kadar iztočnice pripadajo isti besedni vrsti (*atlas*¹ ‘knjiga’ : *atlas*² ‘tkanina’, *atlas*³ ‘vretence’). V 58 primerih so uporabljeni **drugi indikatorji**, kadar se homografi ne razlikujejo niti po besedni vrsti niti po pomenu. Tako je v redkih primerih uporabljena razlika v naglasu ali pregibanju (*brezimen*¹ brezímen : *brezimen*² brezimén), čeprav v obravnavanem primeru obstaja tudi stilistična razlika, vendar je stilistični podatek za šolskega uporabnika slabše razumljiv kot podatek o naglasu. V izjemnih primerih, kadar opisujemo razmerje med več kot dvema homografoma, smo uporabili kombinirane indikatorje (*vaten*¹ ‘vatiran (vátén)’ : *vaten*² ‘v fiziki’ : *vaten*³ ‘vatiran (vatén)’).

Pri oblikovanju pomenskih indikatorjev so bila kot glavna načela upoštevana:

- a) razumljivost za šolskega uporabnika,
- b) kratkost,
- c) pomenska razločevalnost.

⁴ SSKJ2 je nastal kot posodobljena izdaja Slovarja slovenskega knjižnega jezika, ki je v 20. stoletju tako po obsegu kot po zasnovi najambicioznejši slovarski podvig pri Slovencih (Vidovič Muha 2021: 13). V petih zvezkih je izhajal v obdobju 1970–1991 in je bil kasneje večkrat ponatisnjen in objavljen v različnih elektronskih izdajah. Druga izdaja slovarja zapolnjuje vsebinsko in časovno vrzel v leksikalnem opisu slovenskega jezika od 1991 do 2013 (Perdih, Snoj 2015; Perdih 2016: 30–31).

⁵ V nekaterih slovarjih zamejeno ožje, na homonime, pri čemer je potrebno tudi ujemanje v naglasnem znamenju in besedni vrsti.



Slika 2: Indikatorji za razločevanje med homografi na portalu *Franček*

Dodatno je bilo v nekaterih primerih mogoče upoštevati načelo sistematičnosti, tj. da se v homografskem nizu uporabijo istovrstni tipi indikatorja⁶.

V nadaljevanju je opisanih **osem vsebinskih razdelkov**, ki prinašajo mikrostrukturno vsebino. Prvi štirje razdelki (pomenski, sinonimni, morfološki, izgovorni) so na voljo vsem starostnim skupinam, pri čemer je za vsako starostno skupino prilagojena vsebina in vizualizacija. Drugi štirje razdelki (frazološki, narečni, etimološki, zgodovinski) so na voljo le učencem iz druge (6.–9. razred) in tretje starostne skupine (srednja šola). V splošnem so vsi razdelki strukturirani tako, da je najprej podano vsebinsko vprašanje, npr. «*Kaj pomeni beseda **kamen**?*», temu sledi vsebina razdelka, na dnu strani pa je podana povezava na portal Fran neposredno na originalni slovar, iz katerega izhaja vsebina tega razdelka.

1. Pomenski razdelek

V prvem razdelku je predstavljen pomenski opis z razlagami, primeri rabe in slikovnimi materiali. Učencem 1.–5. razreda osnovne šole je primarno namenjenih 2000 slovarskih sestavkov iz novega *Šolskega slovarja slovenskega jezika*, ki nastaja v okviru portala, po zaključku projekta pa naj bi se slovar še dopolnjeval. Ker je slovar namenjen jasno določeni ciljni starostni skupini, so razlage v njem »prilagojene otrokovim izkušnjam tako, da je poudarjen z otrokovega stališča najbolj tipičen vidik pomena, npr. *Zvezek je knjižica iz listov in platnic, v katero pišemo, rišemo pri pouku* (v nasprotju z razlago iz SSKJ: *skupek trdno sešitih, spetih nepotiskanih listov v platnicah iz papirja ali kartona*).« (Petric Žižić 2020: 393) Podatki za druge besede za najmlajšo starostno skupino enako kot za starejše uporabnike izhajajo iz 2. izdaje *Slovarja slovenskega knjižnega jezika* (SSKJ2, pribl. 96.000 iztočnic) in dodatnih 2000 iz slovarja *eSSKJ – Slovar slovenskega knjižnega jezika* (3. izdaja). Ugotovljamo namreč, da je tudi za najnižjo starostno skupine bolj prikazati podatke za vse besede na portalu (približno 100.000), čeprav jim niso v celoti prilagojeni, kot pa da bi bili podatki zamejeni na le 2000 iztočnic.

Za starejše učence so vsi podatki v pomenskem razdelku pridobljeni iz 2. izdaje *Slovarja slovenskega knjižnega jezika* in iz slovarja *eSSKJ – Slovar slovenskega knjižnega jezika* (3. izdaja). Ker sta ta dva slovarja primarno namenjena odraslim govorcem slovenskega

⁶ Kot pomenski indikatorji nastopajo zlasti: nadpomenke, sopomenke, področne opredelitve, opisni tip *o + čem* ipd.

jezika, sta bila za pedagoško rabo prilagojena, kolikor so to dopuščale tehnične in časovne možnosti, zlasti z izpuščanjem podatkov (zmanjšana je količina ilustrativnega gradiva, med njimi tudi večina stilno okvalificiranih zgledov; izločeni so manjši pomenski prenosi in prenesene rabe; izločene so vsebine, ki so prikazane v drugih razdelkih, npr. sinonimija, frazeologija in oblikoslovni podatki; izločeni so zastareli in del vulgarnih pomenov in besednih zvez) in s prilagajanjem prikaza podatkov, zlasti kvalifikatorjev, ki so bili razvezani v polne opise (anat. → anatomija, geol. → geologija). Dodatno so kvalifikatorjem dodana kratka pojasnila, ki se prikažejo v oblaku s pomočjo, tam pa je na voljo tudi hiperpovezava do ustreznega poglavja v integrirani šolski slovnici, kjer lahko uporabniki dobijo podrobnejše informacije o tej vrsti kvalificiranja. Na ta način je na portalu poudarjena tudi zavest o tem, da sta slovar in slovnica med seboj v jeziku neločljivo povezana, in s tem ponuja celovitejši pogled na jezik namesto parcialnega.



Kaj pomeni beseda **otrok**?

1. POMEN

Otrok je deček ali deklica v prvih letih življenja, preden odraste.

Otroci se igrajo na šolskem dvorišču.

2. POMEN

Otrok je hči ali sin.

Starši so s svojimi **otroki** brali knjige in otroške revije.

Slika 3: Pomenski razdelek iztočnice *otrok*

Omejitev pri prilagajanju *Slovarja slovenskega knjižnega jezika* predstavlja njegov obseg, saj bi za celovito prilagoditev razlag in ilustrativnega gradiva pedagoškemu namenu pravzaprav morali postaviti in realizirati povsem nov razlagalni sistem, česar realno ni bilo mogoče opraviti. V danih okoliščinah druge realne možnosti kot uporabiti ta slovar niti ni, saj dosedanji obstoječi slovarji slovenščine za pedagoško rabo obsegajo premajhno število iztočnic.

Slikovni materiali so pridobljeni iz dveh virov. Večina slik je licenčno prostih slik, pridobljenih prek iskalnika Bing. Manjši del slik za predmete iz nedavne preteklosti smo prejeli v uporabo od *Slovenskega etnografskega muzeja*. Sledil je ročni pregled vseh slik za končni izbor 64.063 slik pri 19.760 iztočnicah.

2. Sinonimni razdelek

Sinonimni razdelek prinaša t. i. prave sinonime (na portalu jih razumemo in opredeljujemo kot «besede s podobnim pomenom») za posamezne pomene iztočnice. Temelji na podatkih iz *Sinonimnega slovarja slovenskega jezika*, pri čemer je bil najprej opravljen izbor relevantnih podatkov, nato pa je bil način ubeseditve njihove zvrstno-stilne vrednosti prilagojen šolskim uporabnikom. Tako so bile najprej izločene enote, ki so v izhodiščnem slovarju opredeljene kot delni sinonimi, protipomenke, vzporedne podpomenke in nekatere druge enote, saj je za pedagoški namen najustreznejši glavni tip, tj. «prava» sinonimija. Vsebinska odločitev je v tem primeru sovpadla z grafično, saj bi veliko število povezanih enot izgubilo svoj smisel, kadar njihova preobsežnost zmanjšuje učinkovitost prebranega za to starostno skupino uporabnikov in njihovo stopnjo znanja.

Od «pravih» sinonimov prav tako niso uvrščeni vsi, ampak je izbor omejen na tiste, ki so glede na okvalificiranost uvrščeni v katero od naslednjih skupin:

- nevtralni (v izhodiščnem slovarju: nevtralni, prakt.sp.)
- ekspresivni (v izhodiščnem slovarju: ekspr., ljubk., otr., šalj., vznes.)
- negativni (v izhodiščnem slovarju: slabš., nizk.)
- časovni (v izhodiščnem slovarju: star., knj.izroč., redk.)
- kazalka na (nevtralni) dominantni sinonim v izhodiščnem slovarju, če je iztočnica nevtralna ali označena kot ekspresivna, negativna ali časovno zaznamovana.

V trenutni različici portala so podatki o sinonimiji z geslovnikom povezani na nivoju iztočnice. To pomeni, da niso ustvarjene povezave med pomeni iz pomenskega razdelka in pomeni iz sinonimnega razdelka. Kot posledica tega so možna neujemanja med pomenskimi opredelitvami v obeh razdelkih. Velik izziv pri povezovanju podatkov *Sinonimnega slovarja slovenskega jezika* z geslovnikom portala *Franček* je predstavljalo različna obravnava besednih vrst na makrostrukturnem nivoju. *Sinonimni slovar slovenskega jezika* med drugim na makrostrukturni ravni osamosvaja t. i. *povedkovnike* (raba besede v povedkovem določilu ob glagolu *biti*, npr. *On je lisjak*), prav tako slovar za razliko od geslovnika na makrostrukturnem nivoju razlikuje določne (vrstne) in nedoločne (lastnostne, stanjske ...) pridevnike. V slovenščini se ti dve skupini oblikovno razlikujeta le v imenovalniku (1. sklon) ednine za moški spol in oblike za neživo v tožilniku (4. sklon) ednine za moški spol, kjer oblika za vrstni pridevnik dobi končni *-i* (*zelen* : *zeleni*, *pralen* : *pralni*) (Perdih 2021b).

Vizualizacija podatkov in metajezikovni opis sta prilagojena trem različnim starostnim skupinam. Za vse tri je podana razlaga pomena in navedeni sinonimi. Okvalificiranost sinonimov za prvo starostno skupino je prikazana le grafično, tj. z ikonami. Pri prikazu za drugo starostno skupino je okvalificiranje pojasnjeno tako z ikonami kot z besedilnimi opisi, kar je prikazano na Sliki 4. Za tretjo starostno skupino ikone niso prikazane, saj je bralno razumevanje pri srednješolcih že na visokem nivoju, zato so uporabljeni besedilni opisi.

Največje pomanjkljivosti se v tem trenutku kažejo pri oblikah za pregibanje zaimkov in števnikov, pa tudi pri stopnjevanju prislovov in pridevnikov, zlasti zaradi dveh načinov stopnjevanja⁷ in pomanjkanja raziskav.

4. Izgovor

Za vse iztočnice so bili s strojnim sintetizatorjem izgovora *eBralec* (Žganec Gros idr. 2016) izdelani posnetki z ženskim in moškim glasom. Za nekaj več kot dva tisoč iztočnic *Šolskega slovarja slovenskega jezika* je bil posnet izgovor naravnega govorca, da imajo šolarji najnižje starostne skupine (1.–5. razred) na voljo celovito zaokrožen šolski slovar, obenem pa so isti posnetki na voljo tudi starejšim učencem.

5. Frazeološki razdelek

Zgoraj predstavljeni štirje razdelki so prikazani pri vseh starostnih stopnjah. Od tu dalje se razdelki zaradi svoje zahtevnosti prikazujejo le uporabnikom iz druge in tretje starostne skupine.

Slovenski frazemi so v največjem obsegu slovarsko obdelani v SSKJ in *Slovarju slovenskih frazemov*. Ker so frazemi v *Slovarju slovenskih frazemov* predstavljeni v slovarski obliki (v SSKJ pa so večinoma ne) in ker je tudi po tehnični strani ta slovar bistveno bolj primeren za prilagajanje za pedagoški portal, so uporabljeni podatki iz tega slovarja. Izjema so iztočnice, pri katerih so za pomenski razdelek uporabljeni slovarski sestavki iz slovarja eSSKJ. Tudi za frazeološki razdelek imajo prednost podatki iz eSSKJ, saj je po mnenju urednikov portala primernejši, vsekakor pa se naslanja na bolj aktualno korpusno gradivo. Takšna kombinacija nujno pomeni notranjo neskladnost portala, na kar je bilo opozorjeno v uvodnem delu prispevka, vendar ocenjujemo, da je po vsebinski strani takšna rešitev vendarle ustrežnejša.

Pri pripravi frazeološkega razdelka sta bila oba slovarja (*Slovar slovenskih frazemov* in eSSKJ) strukturno poenotena v smislu, da so bili izločeni podatki, za katere menimo, da presegajo potrebe pedagoškega portala, tako da so ostali samo temeljni podatki, potrebni za prikaz in razumevanje frazemov v pedagoškem procesu: slovarska oblika frazema, relevantnost glede na starostno stopnjo⁸, variante frazema⁹, na pomenski ravni pa kvalifikatorji in kvalifikatorska pojasnila¹⁰, razlaga ali razlagalni sinonim in stavčni zgledi iz korpusnega gradiva. Prikaz za drugo in tretjo starostno stopnjo se razlikuje glede na število prikazanih variant frazemov in število stavčnih zgledov.

⁷ Opisno stopnjevanje: *zelen — bolj zelen — najbolj zelen*; priponsko stopnjevanje: *visok — višji — najvišji*. Nekateri pridevniki in prislovi se stopnjujejo na oba načina, pri nekaterih pa je uveljavljen le en način. Na žalost manjkajo teoretične in korpusne raziskave, s katerimi bi lahko ustrezno opredelili veliko število pridevnikov na portalu.

⁸ Kadar iztočnica ne nastopa v slovarski obliki frazema, ampak le v variantah, je tak frazem prikazan le za srednješolsko starostno skupino.

⁹ Zaradi preglednosti je število variant omejeno na štiri, skupaj s slovarsko obliko frazema je torej podanih do največ pet različnih izraznih podob frazema.

¹⁰ Tako kot v pomenskem razdelku so tudi tu kvalifikatorji izpisani neokrajšano, dodani so oblački s pomočjo in hiperpovezava na slovnico, kjer je podrobneje pojasnjen pomen kvalifikatorja.

6. Narečni razdelek

Zaradi visoke narečne členjenosti slovenskega jezika je narečni razdelek relevanten tudi za šolajočo se populacijo. Vsebinsko odgovarja na dve vprašanji:

- onomaziološko (s katerimi besedami se izrazi koncept X in v katerih narečjih),
- semaziološko (katere koncepte še označuje beseda X in v katerih narečjih),

ob tem pa je lahko dodan še komentar. Podatki temeljijo na dosedanjih zvezkih *Slovenskega lingvističnega atlasa* z onomaziološkim pristopom, za portal *Franček* pa je semaziološki vidik nastal najprej z obrnitvijo onomazioloških podatkov in nato z obsežno dopolnitvijo podatkov. Ob tem je bila zmanjšana kompleksnost podatkov, dodane pa so bile opredelitve razširjenosti (*redko, ponekod, zelo pogosto, v primorski narečni skupini ...*) in dopolnitve z dodatnimi narečnimi podatki, ki v atlasu niso zajeti. Uporabnikom je na voljo funkcija zvočnega snemanja, kjer lahko posnamejo svoj narečni izgovor besede in jo opredelijo z lokacijskim podatkom (Ježovnik, Kenda-Jež, Škofic 2020).

7. Etimološki razdelek

V okviru šolskega pouka slovenščine se etimologiji posveča le malo pozornosti, vendar smo zaradi celovitosti portala in opisa leksike dodali tudi etimološki razdelek. Podatke zanj črpamo iz treh virov:

- novega *Šolskega slovarja slovenskega jezika*,
- nastajajočega eSSKJ,
- *Slovenskega etimološkega slovarja*.

Etimološke osvetlitve za približno 2000 iztočnic, obdelanih v okviru novega *Šolskega slovarja slovenskega jezika* predstavljajo opis, prilagojen za šolskega uporabnika (slika 5). Pri tem je eksplicitno navedeno, ali je beseda podedovana iz prajezika ali prevzeta iz tujega jezika, v katerem časovnem obdobju je nastala oz. za katero obdobje jezikovnega razvoja jo lahko rekonstruiramo ter v katerih sorodnih jezikih jo tudi najdemo. Pri ostalih iztočnicah so uporabljene bodisi etimološke osvetlitve iz eSSKJ bodisi iz *Slovenskega etimološkega slovarja*, ki so za šolskega uporabnika zahtevnejše, saj so primarno namenjene odraslemu govorcju.

8. Zgodovinski razdelek

V zgodovinskem razdelku lahko sledimo rabi besed v slovenščini od 16. stoletja do 19. stoletja. Za slovenščino je tak časovni pregled sploh prvič zbran na enem mestu. Pri tem je uporabljena zbirka *Besedje slovenskega knjižnega jezika 16. stoletja* in geslovnika slovarjev *Slovar jezika Janeza Svetokriškega* in *Slovensko-latinski slovar po Dictionarium Latino-Carniolicum*, ki opisujeta jezik okoli leta 1700, ter geslovník *Slovensko-nemškega slovarja* Maksa Pleteršnika z besedjem do konca 19. stoletja¹¹. Pri povezovanju geslovníkov teh slovarjev in zbirk na geslovník portala *Franček* poleg pravopisnih sprememb (sistematično npr. pripona *-vec* : *-lec* kot v *bravec* : *bralec*, pa tudi posamični primeri, npr. *ambašador* : *ambasador*) večjo težavo predstavljajo tudi različne besednovrstne

¹¹ Gre za slovarje in zbirke, ki so na voljo v digitalni obliki. Za pregled vseh ključnih slovenskih zgodovinskih slovarjev, tudi še nedigitaliziranih, prim. (Merše 2013).

Od kod izvira beseda *miš*?

Beseda je domačega izvora.

Tako kot hrvaška, srbska, bosanska, črnogorska **miš**, ruska **мышь** in češka **myš** se je razvila iz 1200 let stare splošnoslovanske ***myšb**, ta pa iz 5000 let stare praindoevropske ***mūs-**, iz katere so se razvile še staroindijska **mūṣ-**, grška **mῦς**, latinska **mūs**, albanska **mi**, starovisokonemška **mūs** in iz te nemška **Maus**, staroangleška **mūs** in iz te angleška **mouse**, ki vse pomenijo **'miš'**.

Praindoevropska beseda je verjetno tvorjena iz korena ***meus-**, ki je pomenil **'hitro se premikati'** in **'izmikati, krasti'**. Če je domneva pravilna, je beseda prvotno pomenila **'hitro premikajoča se žival (ki krade hrano)'**.



Ta opis je del Šolskega slovarja slovenskega jezika.



Preberi več o preteklosti jezikov.

Slika 5: Etimološka osvetlitev besede *miš*

opredelitve in diahrona homonimija, saj zmanjšujejo uspešnost avtomatskega povezovanja in so zahtevale večji časovni vložek za ročno preverjanje povezanih iztočnic, razvezovanje napačno povezanih in povezovanje tistih, ki jih ni bilo mogoče avtomatsko povezati. Tak primer homonimije je v 16. stoletju *moka* 'moka' : *moka* 'muka', saj je drugi izšel iz rabe in ga je nadomestila slovanska izposojenka *muka*, zato je ni mogoče povezati s sodobno *moka*.

IV. Povezovanje in prilagajanje slovarskih podatkov

Pri pripravi vsebinskih razdelkov je bil prvi korak izdelava geslovnika, kot je bilo že predstavljeno zgoraj. Sledilo je polavtomatsko povezovanje slovarjev na makrostrukturnem nivoju. To pomeni, da smo slovarske iztočnice, ki nastopajo v več slovarjih, povezali na izdelani geslovník¹². Posebno pozornost so zahtevale besednovrstne opredelitve, saj različni slovarji iste besede uvrščajo k različnim besednim vrstam skladno s teorijo in specifičnimi potrebami slovarja (prim. poglavje o sinonimnem razdelku), posebej zahtevno pa je bilo povezovanje iztočnic v zgodovinskih slovarjih (prim. poglavje o sinonimnem razdelku).

Za prilagajanje slovarjev so v veliki meri uporabljeni avtomatizirani postopki izločanja in preurejanja podatkov. Tako smo želeli doseči čim večjo primernost za predvidenega uporabnika portala. Ker struktura nekaterih slovarjev v procesu digitalizacije tiskanih slovarjev

¹² Ob tem ni potekalo povezovanje na pomenskem nivoju, saj presega okvire tega projekta, je pa to ena od možnosti prihodnje nadgradnje slovarskih podatkovnih zbirk po zgledu nekaterih drugih leksikografskih ustanov. O tem prim. npr. (Tavast idr. 2020; Kouvara idr. 2020).

še ni bila v celoti vsebinsko razdvoumljena, poleg tega pa nekaterih slovarskih podatkov ni mogoče razdvojniti glede na naše potrebe, strojnih postopkov ni bilo vedno mogoče izvesti optimalno. Tako so pri SSKJ izpadli tudi nekateri zelo relevantni podatki, zlasti določene stalne zveze, vzkliki itd. To bi bilo mogoče preprečiti le z ročnim pregledom približno 100.000 slovarskih sestavkov, kar pa v okviru danih možnosti ni bilo realno izvedljivo.

Po drugi strani smo izboljšali razumljivost kvalificiranja v pomenskem razdelku, saj smo kvalifikatorje, ki v tiskanem slovarju stojijo v zaglavju in veljajo za vse pomene (če niso eksplicitno preklicani), prestavili na pomenski nivo k vsem relevantnim pomenom, prav tako smo okrajšane kvalifikatorje razvezali (*ekspr.* → *ekspresivno*, *pog.* → *pogovorno*). V sinonimnem razdelku so bili stilni kvalifikatorji združeni v skupine in opisani z besedilni opisi in grafičnimi ikonami.

V. Novi slovarski podatki

Poleg povezovanja in prilagajanja podatkov so bili pripravljene povsem novi. Tu bomo omenili le najpomembnejše med njimi: *Šolski slovar slovenskega jezika* z etimologijami, indikatorje za vsebinsko razločevanje med homografi, opredelitve frazeoloških stavčnih zgledov glede na varianto frazema, pa tudi veliko količino zbranega slikovnega gradiva.

Velikega pomena za šolske uporabnike je *Šolski slovar slovenskega jezika*, ki je prvi enojezični šolski slovar, ki je nastal z mislijo ne le na razumljivost ciljnemu uporabniku, ampak tudi s sistematičnim pristopom po celotni vsebini slovarja. Zdajšnjim približno 2000 slovarskim sestavkom bodo v prihodnjih letih dodani še novi, tako da se bo slovar še povečal in še naprej zapolnjeval vrzel v slovenski pedagoški leksikografiji. Ob tem so nastajale tudi prve šolske etimologije za slovenščino.

Z indikatorji, ki opredeljujejo razlike med homografi, so pridobljene tudi praktične izkušnje, ki jih je mogoče z določenimi prilagoditvami uporabiti tudi pri pripravi istovrstnih indikatorjev za splošnega uporabnika.

Za potrebe prilagojenega razvrščanja stavčnih zgledov pri frazemih je vsem stavčnim zgledom iz eSSKJ pripisano, h kateri varianti frazema spadajo.

Pridobili smo tudi slikovno gradivo, ki se že uporablja tudi za nekatere zoonime v *Novem etimološkem slovarju slovenskega jezika*, zaradi odprte licence pa bodo na voljo tudi za druge slovarske projekte. V kontekstu digitalnih slovarjev to obenem pomeni, da dodajanje novih slik za potrebe posameznih slovarjev lahko pomeni tudi izboljšavo pri izboru slik za portal *Franček* — slovarji namreč niso več zaprta zaključena enota, ampak se lahko obogatijo tudi kasneje z gradivom, pripravljenim za druge namene. Takšen proces gotovo ne bo samodejen, ampak bo pri tem vedno potrebna kritična presoja, kdaj in katere nove materiale je smiselno uporabiti za druge namene.

VI. Sklep

Portal *Franček* s svojim osrednjim, leksikografskim delom predstavlja prvi slovenski pedagoški slovarsko-slovnčni portal. Namenjen je osnovnošolcem in srednješolcem, po svoji vsebini in vizualizaciji pa je prilagojen za tri starostne stopnje: 1.–5. razred osnovne

šole (6–10 let), 6.–9. razred osnovne šole (11–14 let), srednja šola (15–18 let). Leksikografski del portala sestoji iz novih in obstoječih slovarskih podatkov, pri čemer so slednji v veliki meri prilagojeni potrebam in zmožnostim šolskih uporabnikov. Slovarsko je opredeljenih 100.006 iztočnic z vidika pomena, sinonimije, frazeologije, morfologije, dodani so naravni in sintetični zvočni posnetki izgovora ter narečni, etimološki in zgodovinski podatki. Skupaj s poglavji iz šolske slovnice, jezikovno svetovalnico za učitelje, didaktičnimi materiali in igro za mobilne naprave predstavlja celovit elektronski vir o slovenščini za osnovno- in srednješolce. Na makrostrukturnem nivoju povezani slovarji so pomembni za prihodnjo (pol)avtomatsko izrabo slovarskih podatkov.

ZAHVALE

Prispevek je nastal v okviru projekta *Portal Franček: Jezikovna svetovalnica za učitelje slovenščine in Šolski slovar slovenskega jezika*, ki ga sofinancirata Republika Slovenija in Evropski socialni sklad; del raziskav za projekt je podprl ARRS v okviru programa P6-0038 (*Slovenski jezik v sinhronem in diahronem razvoju*).

LITERATURA

Ahačič 2017a: Ahačič, K. *Kratkoslovnica: slovenska slovnica za osnovno šolo*. Ljubljana, 2017.

Ahačič 2017b: Ahačič, K. *Slovnica na kvadrat. Slovenska slovnica za srednjo šolo*. Ljubljana, 2017.

Ahačič, Ledinek, Perdih 2015a: Ahačič K., Ledinek N., Perdih A. *Fran: the next generation Slovenian dictionary portal* // Gajdošová K., Žáková A. *Natural language processing, corpus linguistics, lexicography: proceedings. Eighth International Conference. Slovakia, 21–22 October 2015. Bratislava, 2015, 9–16.*

Ahačič, Ledinek, Perdih 2015b: Ahačič K., Ledinek N., Perdih A. *Portal Fran — nastanek in trenutno stanje* // Smolej M. *Slovnica in slovar — aktualni jezikovni opis (1. del). Obdobja 34. Ljubljana, 2015, 57–66.*

Ahačič idr. 2021: Ahačič K., Atelšek S., Erjavec T., Holozan P., Jakop N., Jemec Tomazin M., Ježovnik J., Ledinek N., Perdih A., Romih M., Trojar M. *Corpus of Slovenian school texts. Slovenian language resource repository CLARIN.SI. 1. 4. 2021.* <http://hdl.handle.net/11356/1413>.

Dobrovoljc, Verovnik, Bizjak Končar 2017: Dobrovoljc H., Verovnik T., Bizjak Končar A. *Spletna jezikovna svetovalnica za slovenski jezik ter normativne zadrege uporabnikov in jezikoslovcev* // Škrabčevi dnevi 10: *Zbornik prispevkov s simpozija 2017 / Bizjak Končar A., Dobrovoljc H. Nova Gorica, 2018, 52–69.*

Gantar 2015: Gantar P. *Leksikografski opis slovenščine v digitalnem okolju*. Ljubljana, 2015.

Gliha Komac idr. 2015: Gliha Komac N., Jakop N., Ježovnik J., Klemenčič S., Krvina D., Ledinek N., Michelizza M., Meterc M., Mirtič T., Perdih A., Petric Š., Snoj M., Žele A. *Novi slovar slovenskega knjižnega jezika — predstavitev temeljnih konceptualnih izhodišč* // *Zbornik prispevkov s simpozija 2015 / Marušič F., Mišmaš P., Žaucer R. Nova Gorica, 2016, 16–33.*

Godec Soršak 2015: Godec Soršak L. *Slovenski otroški šolski slovar* // Smolej M. *Slovnica in slovar — aktualni jezikovni opis (1. del). Obdobja 34. Ljubljana, 2015, 243–250.*

Godec Soršak 2019: Godec Soršak L. *Zasnova šolskega slovarja za otroke v 1. in 2. vzgojno-izobraževalnem obdobju. Doktorska disertacija*. Ljubljana, 2019.

Jakopin 1999: Jakopin P. *Zgornja meja entropije pri leposlovnih besedilih v slovenskem jeziku. Doktorska disertacija*. Ljubljana, 1999.

Ježovnik, Kenda-Jež, Škofic 2020: Ježovnik J., Kenda-Jež K., Škofic J. *Reduce, Reuse, Recycle: Adaptation of Scientific Dialect Data for Use in a Language Portal for School children* //

Gavriilidou Z., Mitsiaki M., Fliatouras A. Proceedings of XIX EURALEX Congress: Lexicography for Inclusion. Vol. I, 31–37.

Kouvara idr. 2020: Kouvara E., González M., Grosse J., Sauri R. Determining Differences of Granularity between Cross-Dictionary Linked Senses // Gavriilidou Z., Mitsiaki M., Fliatouras A. Proceedings of XIX EURALEX Congress: Lexicography for Inclusion. Vol. I, 109–118.

Merše 2013: Merše M. Slovenska leksikografija in leksikologija v zgodovinski perspektivi // Славянская лексикография. Международная коллективная монография / Отв. ред. М.И. Чернышева. Москва, 2013, 217–251.

Perdih, Snoj 2015: Perdih A., Snoj M. SSKJ² // Slavia Centralis. 2015. VIII/1, 5–15.

Perdih 2016: Perdih A. Splošni razlagalni slovarji slovanskih jezikov. Ljubljana, 2016.

Perdih 2020: Perdih A. Portal Fran: od začetkov do danes // Rasprave: Časopis Instituta za hrvatski jezik i jezikoslovlje. 2020. 46/2, 997–1018.

Perdih 2021a: Perdih A. Indikatorji pri homografih na portalu *Franček* // Jezikoslovni zapiski. 2021. 27/2, 7–21.

Perdih 2021b: Perdih A. Učenje o slovarjih v šoli: Portal *Franček* kot most med splošno in pedagoško leksikografijo // Лексикографија и лексикологија у светлу актуелних проблема. Зборник научних радова / Стана Ристић, Ивана Лазич Коњик, Ненад Ивановић. Београд, 2021, 1119–1132.

Petric Žižić 2020: Petric Žižić Š. Tipologija razlag v šolskem slovarju slovenskega jezika. // Slavistična revija. 2020. 68/3, 391–409.

Tavast idr. 2020: Tavast A., Koppel K., Langemets M., Kallas J. Towards the Superdictionary: Layers, Tools and Unidirectional Meaning Relations // Gavriilidou Z., Mitsiaki M., Fliatouras A. Proceedings of XIX EURALEX Congress: Lexicography for Inclusion. Vol. I, 215–223.

Vidovič Muha 2021: Vidovič Muha A. Slovensko leksikalno pomenoslovje. Ljubljana, 2021.

Žagar Karer 2017: Žagar Karer M. Terminološko svetovanje: izkušnje in perspektive. // Slovenska terminologija danas: Primljeno na V skupu Odeljenja jezika i književnosti od 30. maja 2017. godine, na osnovu referata akademika Predraga Pipera i prof. dr Rajne Dragičević / Piper P., Jovanović V. Beograd, 2017, 361–370.

Žganec Gros idr. 2016: Žganec Gros J., Vesnicer B., Rozman S., Holozan P., Šef T. Sintenzator govora za slovenščino eBralec. // Zbornik konference Jezikovne tehnologije in digitalna humanistika. Proceedings of the Conference on Language Technologies & Digital Humanities. Ljubljana, 2016, 180–185.

Žmigrodzki 2018: Žmigrodzki P. Makro- i mikrostruktura słownika. Sposób prezentacji materiału. // Wielki słownik języka polskiego PAN: Geneza, koncepcja, zasady opracowania / Žmigrodzki P., Bańko M., Batko-Tokarz B., Bobrowski J., Czelakowska A., Grochowski M., Przybylska R., Waniakowa J., Węgrzynek K. Kraków, 2018, 25–38.

KORPUSI IN SLOVARJI

Besedje slovenskega knjižnega jezika 16. stoletja. 2014 [2011]. [www.fran.si]

eSSKJ: Slovar slovenskega knjižnega jezika. 2016–. [www.fran.si]

Furlan M. NESSJ: Novi etimološki slovar slovenskega jezika. 2017–. [www.fran.si]

Keber J. Slovar slovenskih frazemov. 2015 [2011]. [www.fran.si]

SSKJ2: Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja. 2014. [www.fran.si]

Slovenski lingvistični atlas 1. 2014 [2011]. [www.fran.si]

Slovenski lingvistični atlas 2. 2016. [www.fran.si]

Snoj J., Ahlin M., Lazar B., Praznik Z. Sinonimni slovar slovenskega jezika. 2018 [2016]. [www.fran.si]

Snoj M. Slovenski etimološki slovar³. 2015 [2016]. [www.fran.si]