

# OBLIKOVANJE VAJ IN NALOG ZA VSEBINSKA SKLOPA SLOVNICA IN PRAVOPIS S POMOČJO KORPUSA IN DRUGIH (JEZIKOVNIH) VIROV

ŠPELA ANTLOGA

Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,  
Maribor, Slovenija.  
E-pošta: s.antloga@um.si

**Povzetek** Temeljni vodili pri oblikovanju vaj in nalog za vsebinska sklopa slovnica in pravopis sta poleg avtomatizacije predvsem izbira avtentičnih primerov jezikovne rabe in dodajanje številnih raznovrstnih primerov k posameznim nalogam, zato je bilo v proces oblikovanja in generiranja nalog poleg samodejnega priklicevanja primerov nujno vključiti tudi pripravo seznamov ustreznih besednih kandidatov, tj. besed, besednih oblik ali besednih zvez, na podlagi katerih so priklicani primeri k posameznim nalogam. V prispevku predstavljam postopek oblikovanja tovrstnih seznamov s pomočjo korpusov MAKS in Gigafida ter drugih (jezikovnih) virov.

## **Ključne besede:**

besedni  
seznam,  
korpusni  
pristop,  
učno  
e-okolje,  
pravopis,  
slovnica.

## 1 Uvod

V projektu *Slovenščina na dlani*, katerega cilj je izdelava interaktivnega učnega e-okolja za učenje slovenščine, sta nastala dva specializirana korpusa besedil slovenskega jezika: korpus besedil, namenjen tvorjenju vaj iz pravopisa in slovnice (MAKS),<sup>1</sup> ter korpus oz. zbirka besedil praktičnega sporazumevanja (BERTA). Besedila korpusa MAKS so torej glavni vir primerov za slovnično-pravopisne vaje. Primeri so bili glede na obravnavano problematiko priklicani samodejno z računalniškim orodjem, ki deluje v ozadju (prim. Majninger 2018), če pa to ni bilo mogoče, je bilo treba za njihov priklic pripraviti dodatne sezname ustreznih besednih kandidatov.<sup>2</sup> Za namen samodejnega priklica primerov in za pripravo seznamov je bil korpus MAKS označen na več ravneh. Poleg podatka o lemi, besedni vrsti in drugih oblikoslovnih lastnostih pojavnic so besedila razčlenjena še skladenjsko, v njih pa so eksplicitno označena tudi lastna imena in druge kategorije, vezane na posamezne tipe vaj (Majninger 2018: 74). Ročna izdelava seznamov besednih kandidatov s pomočjo korpusa je smiselna predvsem zato, ker primeri k nalogam, priklicani s pomočjo seznamov, učencem (in učiteljem) ponujajo številne zanimive primere in odražajo avtentično jezikovno rabo.

## 2 Oblikovanje seznamov besednih kandidatov

Za naloge pri vsebinskih sklopih slovnica in pravopis je bilo pripravljenih 99 seznamov (13 za naloge, ki preverjajo poznavanje slovnice, in 86 za naloge, ki preverjajo pravopis). Pri slovničnih nalogah prevladujejo sezname iz vsebinskega sklopa Samostalnik in sklanjanje (npr. tuja imena, ki se končajo z *-y* ali *-e*, pri katerih preverjamo sklanjanje s podaljšavo, izpad nemega končnega glasu, preglas; domača moška imena na *-a* in *-o*, moška imena, pri katerih pride v orodniku do preglasa, itd.). Pri pravopisnih nalogah je bilo največ seznamov oblikovanih pri vsebinskih sklopih: a) Lastna imena (sezname izlastnoimenskih pridevnikov, eno- in večbesednih naselbinskih in nenaselbinskih imen, stvarnih lastnih imen, različnih kategorij lastnih imen itd.); b) Besedotvorje (samostalniške, glagolske in pridevniške podredne in priredne zloženke, sestavljenke, samostalniški, vezniški in prislovni sklopi, stalne prislovne zveze itd.); c) Prevzete besede (besede z različnimi samoglasniškimi sklopi,

---

<sup>1</sup> Korpus MAKS obsega približno deset milijonov besed lektoriranih besedil iz publicistike, leposlovja in spletnih portalov, ki so zanimiva za mlade med 12. in 19. letom starosti.

<sup>2</sup> Za namen tega prispevka sem se odločila za poimenovanje *seznam besednih kandidatov*, čeprav lahko sezname vključujejo tudi besedne zveze ali daljše kolokacijske nize.

npr. *-ia-*, *-io-*, *-eo-* itd., lastne in občne prevzete besede); č) Težji primeri (izpeljanke s priponskim obrazilom, pri katerih se neobstojni samoglasniki vrivajo po istih načelih kot v imenovalniku ednine, npr. *bolezenski*, *kiselkast*, ali roditelju množine, npr. *ladijski*, *sabelj*, izpeljanke s predponskim obrazilom iz podstav, ki se končujejo na *-r*, npr. *metrski*, *sestrski*, itd.).

Načeloma so besedni kandidati na seznamih v lematizirani obliki (torej v imenovalniku ednine oziroma v nedoločniku), razen kadar je za priklic ustreznega primera potrebna točno določena skladenjska oblika, npr. *tekem*, *dekel*, *oken* pri preverjanju zapisa neobstojnega samoglasnika.

Oblikovanje seznamov besednih kandidatov za priklic primerov je potekalo na štiri načine:

1. oblikovanje seznamov na podlagi nekorpusnih virov,
2. oblikovanje seznamov na podlagi korpusa MAKS in drugih jezikovnih korpusov,
3. oblikovanje seznamov z združevanjem nekorpusnih virov in korpusa MAKS,
4. oblikovanje seznamov s paberkovanjem.

## 2.1 Oblikovanje seznamov na podlagi nekorpusnih virov

Pri določenih nalogah so pripadajoči primeri zahtevni z vidika samodejnega priklica, z jezikoslovnega vidika pa gre za zaključen nabor besed ali besednih zvez, ki so relevantne za preverjanje določene vsebine pri nalogi in so že popisane v jezikovnih priročnikih. V takih primerih je bila za oblikovanje seznamov v največji meri uporabljena *Slovenska slovnica* (Toporišič 2000), po potrebi pa tudi drugi relevantni pisni jeziko(slo)vni priročniki in razprave. Tako je bilo na primer pri nalogi, s katero preverjamo sklanjanje zaimkov *kaj*, *malokaj*, *marsikaj*, *mnogokaj* in *nekaj* oziroma rabo njihove tožilniške oblike, za priklic ustreznih primerov najbolj smiselno pripraviti nabor parov vseh iskanih zaimkov v imenovalniški in tožilniški obliki, ki so kot napaka vstavljeni v priklicane primere z zaimkom v tožilniški obliki s seznama (*česa*, *maločesa* ...).

Na podlagi drugih virov so bili narejeni sezname tudi, kadar bi bila korpusna analiza zahtevnejša oziroma bi do enakih rezultatov z njo prišli težje kot z iskanjem primernih kandidatov za priklic v drugih, že obstoječih pisnih ali spletnih virih. Besedne kandidate za priklic primerov pri preverjanju začetnice imen praznikov bi najverjetneje s korpusno analizo iskali veliko dlje, kot če preprosto uporabimo relevantne spletne strani z informacijami javnega značaja, Wikipedijo ipd.

## 2.2 Oblikovanje seznamov na podlagi korpusa MAKS in drugih jezikovnih korpusov

Besedilni korpusi zaradi svoje raznovrstnosti in aktualnosti v jezikoslovje vnašajo številne nove možnosti empiričnih raziskav in aplikacij. Če so zasnovani premišljeno, tj. če zajemajo čim širši in ustrezno uravnotežen nabor relevantnih besedil, jih lahko kot reprezentativne vzorce dejanske jezikovne rabe uporabljamo za različne namene (Krek, Arhar 2010), tudi pri pouku. Ker so bila besedila za korpus MAKS izbrana z mislijo na ciljno publiko, je ta služil kot primarni vir tudi za izdelavo seznamov za priklic primerov. Zgolj v primeru, da v tem korpusu ni bilo mogoče dobiti dovolj relevantnih kandidatov ali z dobljenimi zadetki nismo pokrili celotne obravnavane tematike, smo uporabili tudi korpus Gigafida. Pri oblikovanju seznamov smo si namreč prizadevali, da bi poiskali kandidate, s katerimi bi čim bolj nazorno in realno predstavili dejansko jezikovno rabo.

Za oblikovanje nekaterih seznamov besednih kandidatov je bilo torej smiselno uporabiti samo korpusni pristop. Sezname smo oblikovali z brskanjem po korpusu preko posebnega spletnega vmesnika (konkordančnika) NoSketch Engine, ki uporabniku omogoča hitro iskanje, razvrščanje, filtriranje, štetje in druge načine obdelave korpusnih podatkov. Najprej je bilo treba definirati ukaz, ki iz korpusa označenih besedil poišče ustrezne dele besedila ali besede, ki bodo potencialno uvrščene na seznam besednih kandidatov. Nekateri iskalni ukazi so bili enostavni in smo z njimi hitro izluščili ustrezne kandidate, ki smo jih po pregledu lahko uvrstili na seznam, npr. analiza pojavnice z ukazom `[tag="P.*" & lemma="*(ski|ški)"]` v korpusu MAKS za priklic izlastnoimenskih pridevnikov na *-ski/-ški* ali z ukazom `[tag="[SPGR].*" & lemma="iz.*"]` za priklic samostalniških, pridevniških, prislovnih in glagolskih sestavljenk s predpono *iz-*. Z nekaterimi iskalnimi ukazi pa je bilo bistveno več dela, saj je bilo ali težje natančno določiti iskalne pogoje za iskanje primernih kandidatov ali pa je bilo zahtevnejše nadaljnje pregledovanje zadetkov.

Zahtevno je bilo npr. iskanje pojavnic z neobstojnimi samoglasnikom s sklopom nezvočnika in zvočnika (tip *pesem – pesmi*), ki smo jih v korpusu poiskali z ukazom [lemma=".\*[bdzžgptsščkfhc][aeiou][mlnrjv]" & word=".\*[bdzžgptsščkfhc][mlnrjv].\*"]. Z vidika obsežnega pregledovanja zadetkov pa so bili zahtevni za pripravo zlasti seznama lastnih in občnih imen, predvsem nekoliko bolj specifična imena, kot so npr. imena hišnih ljubljencev, latinska poimenovanja rastlin in živali, imena vozil, plovil in drugih prevoznih sredstev ipd.

### 2.3 Oblikovanje seznamov z združevanjem nekorpusnih virov in korpusa MAKS

Pri nekaterih nalogah je bilo za pripravo seznamov smiselno združevanje nekorpusnega in korpusnega pristopa. V prvi fazi je bil na podlagi že obstoječih virov in razprav pripravljen model, ki je nato v drugi fazi služil za luščenje relevantnih zadetkov v korpusu MAKS. S pregledom obstoječih jezikovnih virov in razprav je bil na primer narejen model za iskanje samostalniških (po podobnem principu tudi glagolskih, pridevniških) sestavljenk (in tudi nekaterih drugih besedotvornih vrst). Tako smo najprej oblikovali seznam potencialnih predponskih obrazil, npr. *nad-*, *pod-*, *anti-*, *pra-*, *raz-*, *super-*, *eks-*, *ultra-*, *ne-*, *a-*, *proti-*, in v korpusu MAKS poiskati tiste leme, ki so sestavljene iz take predpone in nekega drugega znanega samostalnika (npr. po seznamu iz Sloleksa): *nad + kuhar*, *pod + odbor*, *anti + oksidant* itd. Eden od kompleksnejših vsebinskih sklopov za iskanje so bile besede z neobstojnimi samoglasniki. Zanje smo v korpusu MAKS poiskali vse besede z lemo, ki se konča na priponsko obrazilo, ki se začne s soglasnikom (-*skei/-škei*, -*ček*, -*ni*, -*ka*, -*kast*), niz znakov pred to končnico pa se konča z nizom soglasnika, samoglasnika -*e*- in zvočnika, tj. "[bdzžgptsščkfhcmlnrjv]e(m|r|j|v|lj|nj?)", ki ustreza neki drugi znani lemi (v korpusu MAKS ali v Sloleksu), npr. *vogel*, *kisel*. Vsi zadetki v korpusu so bili pregledani in nato ustrezni uvrščeni na seznam.

### 2.4 Oblikovanje seznamov s paberkovanjem

Nekaterih seznamov nismo mogli oblikovati na podlagi razprav ali s pomočjo korpusov (ponekod je bil sicer v pomoč korpus šolskih besedil Šolar, npr. za iskanje napačno zapisanih glagolov ipd.), zato smo jih oblikovali sami, s paberkovanjem. Takih seznamov sicer ni veliko, so pa naloge, ki temeljijo na teh seznamih, še kako pomembne. Eden takih seznamov je npr. seznam parov glagolov s predponama *u-*

in *v-*, npr. *ustaviti – vstaviti, utreti – vtreti, uliti – vliti, uleči – vleči, uročiti – vročiti, ustaviti – vstaviti, ušteti – všteti* itd., pri katerih preverjamo poznavanje pomena in zapisa.

### 3 Relevantnost jezikovnega korpusa pri preverjanju znanja

Ob pripravi seznamov smo s korpusno analizo med drugim pridobili tudi podatke o pogostnosti rabe posameznih izrazov in prevladujočih kolokacijah. Ker smo težili k temu, da učenci znanje pridobijo in utrjujejo z avtentičnimi primeri besed in besednih zvez, smo na seznam vključili besedne kandidate, ki so splošno znani, pogostejši, aktualni ter pomensko in oblikovno nedvoumni. Na posebne in manj številčne zadetke v korpusu, ki pa so zanimivi zaradi aktualnosti, zapisa ali drugih lastnosti, smo opozorili v spremljajočih sprotnih razlagah k nalogam.

Izbira korpusa kot primarnega vira za priklic primerov in izdelavo seznamov za preverjanje znanja pri nalogah iz slovnice in pravopisa je torej smiselna predvsem zato, ker učencem (in učiteljem) ponuja avtentične primere rabe, ki odražajo aktualno jezikovno stanje, poleg tega pa nas namesto intuitivnega navajanja (včasih že izrabljenih in zaradi ponavljanja nezanimivih) primerov opremi z besedami ali zvezami, na katere morda sploh ne bi pomislili, a so s pomenskega, oblikovnega ali katerega drugega vidika za učence in učitelje še kako zanimivi.

### 4 Sklep

Ročna priprava seznamov je torej zapleten in dolgotrajen proces, ki vključuje: 1) interpretacijo naloge in razmislek o primernih kandidatih za priklic primera, ki bodo najbolje zastopali preverjano problematiko; 2) razmislek, ali seznam pripraviti na podlagi obstoječega pisnega gradiva, kombinacije pisnega gradiva in uporabe korpusa, samo uporabe korpusa ali pa bo (izjemoma) seznam temeljil na jezikovnem znanju skupine, ki pripravlja sezname; 3) načrt, kako najučinkoviteje poiskati čim več relevantnih in čim boljših besednih kandidatov za priklic primerov; 4) proces izdelave seznama z luščenjem primernih kandidatov; 5) preverjanje besednih kandidatov na seznamu in morebitno popravljanje seznama. Šele po tem je seznam uporaben za priklic primerov pri posamezni nalogi. Ob tem je nezanemarljivo dejstvo, da nekateri sezname vključujejo tudi tisoč ali več primernih besednih kandidatov.

## Literatura

- Jeziškova svetovalnica Inštituta za slovenski jezik Frana Ramovša ZRC SAZU. Dostop 6. 11. 2020 na <https://svetovalnica.zrc-sazu.si/>.
- Korpus mladinske književnosti MAKŠ. Dostop 6. 11. 2020 na [https://www.clarin.si/noske/run.cgi/corp\\_info?corpname=maks&struct\\_attr\\_stats=1](https://www.clarin.si/noske/run.cgi/corp_info?corpname=maks&struct_attr_stats=1).
- Korpus šolskih besedil Šolar. Dostop 9. 11. 2020 na <http://viri.trojina.si/v/solar/>.
- Simon KREK, Špela ARHAR, 2010: Slovenski besedilni korpusi: kako v razred? *Sodobna pedagogika* 1, 224–241.
- Sandi MAJNINGER, 2018: Priprava učnega e-okolja – programerski vidik. *Slovensčina na dlani* 1. Ur. Natalija Ulčnik. Maribor: Univerzitetna založba Univerze. Dostop 6. 11. 2020 na <https://press.um.si/index.php/ump/catalog/view/341/302/543-1>.
- Slovenski referenčni korpus Gigafida-*proto*, nededupliciran, v2.0 (1990–2018). Dostop 6. 11. 2020 na [https://www.clarin.si/noske/run.cgi/corp\\_info?corpname=gfida20&struct\\_attr\\_stats=1](https://www.clarin.si/noske/run.cgi/corp_info?corpname=gfida20&struct_attr_stats=1).
- Spletni konkordančnik NoSketch Engine. Dostop 9. 11. 2020 na <https://www.clarin.si/noske/>.
- Jože TOPORIŠIČ, 2000: *Slovenska slovnica*. Maribor: Obzorja.

