

Polona Gantar in Mija Bon

Dati skozi ali prestat? Napake in jezikovne variante v rabi večbesednih enot pri samostojnem tvorjenju besedil v osnovni in srednji šoli

Povzetek: V prispevku opišemo težave pri rabi večbesednih enot (VBE), ki jih imajo učenci in dijaki pri samostojnem tvorjenju besedil. Pri tem izhajamo iz učiteljskih popravkov v korpusu Šolar 3.0 in jezikovnih odločitev pri izdelavi učne množice Šolar-Eval. V raziskavi smo nabor problemov kategorizirali glede na tip alternativne jezikovne ubeseditve (izrazne, leksikalne in kontekstualne), tip VBE (stalne zveze, frazeološke enote, kolokacije, skladišne zveze) ter glede na to, ali gre za variantno ali napačno ubeseditve. V jeziku je stvari mogoče povedati na različne načine, zato tudi ni enoznačnih popravkov, hkrati pa se spekter besednozvezne problematike razteza od napak na eni strani, ki so razmeroma redke in imajo za posledico nerazumevanje, napačno razumevanje oz. z vidika rojenega govorca nepričakovano izbiro, do variant na drugi strani, pri katerih obstaja (še) druga ubeseditve istega sporočila. Prepoznavanje omenjenih ubeseditvev omogoča oblikovanje povratne informacije v smislu pravilnejše, slogovno in kolokacijsko ustrežnejše ubeseditve, kot tudi prepoznavanje alternativnih ubeseditvev, ki v jeziku delujejo bolj ali manj enakovredno, kar je mogoče ugotavljati s pomočjo referenčnega korpusa.

Ključne besede: popravljanje šolskih besedil, jezikovna varianta, jezikovna napaka, povratna učiteljska informacija, večbesedne enote, frazeološke enote, kolokacije, skladišne zveze

UDK: 373.5

https://doi.org/10.63384/sptB5_z789s

Znanstveni prispevek



Dr. Polona Gantar, znanstvena sodelavka, Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva 2, SI-1000 Ljubljana in Fakulteta za računalništvo in informatiko, Večna pot 113, SI-1000 Ljubljana, Slovenija; elektronski naslov: apoloniya.gantar@ff.uni-lj.si; ©

Mija Bon, asistentka, Univerza v Ljubljani, Filozofska fakulteta, SI-1000 Ljubljana, Slovenija; elektronski naslov: mija.bon@ff.uni-lj.si

Uvod

Ko govorimo o šolskem pisanju in učenju prvega jezika, je pojem jezikovne napake, še zlasti v odnosu do jezikovne variantnosti, kompleksen. Nekatere napake, kot so nepravilno črkovanje, stava ločil, zapis z veliko ali malo začetnico ipd., je mogoče relativno preprosto prepoznati, saj za njihovo ustrezno rabo obstajajo pravila, zapisana v pravopisnih, slovarskih in drugih jezikovnih priročnikih. Pri jezikovnem izražanju, ki zajema najrazličnejše besednozvezne in stavčne ubeseditve, pa napačnosti ni vedno mogoče jasno opredeliti, ustreznosti alternativne ubeseditve, ki jo npr. predlaga učitelj, pa zaradi jezikovne variantnosti v najširšem pomenu besede ne vedno nedvoumno upravičiti.¹ Tvorec besedila namreč med številnimi možnostmi ubeseditve sporočila namerno izbira ne le, kaj bo povedal, temveč tudi, kako bo to povedal, da bi dosegel želen odziv pri naslovniku (Prince v Cameron in Schwenter 2013, str. 2), pri čemer je (stilni) izbor jezikovnih sredstev pogojen tudi z drugimi dejavniki, mdr. z govornim položajem oz. vrsto besedila (Krajnc Ivič 2015), kar diapazon jezikovnih napak precej razširi in hkrati relativizira.

V prispevku predstavljamo raziskavo, v kateri smo želeli ugotoviti, katere težave imajo šolarji in dijaki pri rabi večbesednih enot v šolskem pisanju.² Njihove težave smo opredelili glede na različne pomenske in skladenjske lastnosti z namenom raziskati možnosti variantnih ubeseditvev, z njihovim vrednotenjem in preverjanjem v splošni rabi pa ponuditi tudi možna izhodišča za podajanje učiteljeve povratne informacije. Drugi cilj raziskave je prepoznane vzorce uporabiti pri izdelavi učnih množic, namenjenih računalniški obdelavi, npr. izdelavi nevronskega črkovalnika in pregledovalnika.

Pri rabi besednih zvez, za razliko od enobesedne leksike, kjer se jezikovne napake pojavljajo na ravni zapisa, pregibanja in rabe pomena, (lahko) prihaja še

1 Kot je pokazala raziskava o podajanju povratnih informacij na pisna študentska besedila (Rozman in Stabej 2024, str. 289), nekateri učitelji dvomijo o lastni usposobljenosti za jezikovno popravljanje oz. popravljajo samo tiste jezikovne izbire, ki otežujejo ali onemogočajo razumljivost besedila.

2 Gre za besedila, ki so jih učenci slovenskih osnovnih in srednjih šol samostojno tvorili pri pouku, v največji meri so to šolski spisi oz. eseji, napisani pri pouku slovenščine (Arhar Holdt idr. 2022).

do neustreznega zapisa na ravni skladnje (*pripadništvo do koga – komu*), napačne ali slogovno manj ustrezne izbire katerega od leksikalnih elementov zveze (*se nenormalno zaljubi – neizmerno zaljubi*), izbire ne tipične ali manj tipične besedne kombinacije (*nepridiprava družba – slaba družba; odločen sam vase – prepričan sam vase*). Pri rabi frazeoloških enot in drugih večbesednih idiomatičnih izrazov je pomembno izpostaviti še njihovo metaforičnost in ekspresivnost kot definicijsko lastnost (Gantar idr. 2021), zaradi česar je mogoče predvidevati, da je njihova raba povezana z željo govorca povedati kaj drugače, bolj ekspresivno ter tako morda ustvariti slogovno opaznejše in zahtevnejše besedilo. Izbira besednozvezne problematike v pričujoči raziskavi je motivirana še iz dveh razlogov. Prvi so rezultati analize šolskega pisanja pri pripravi evalvacijske množice Šolar-Eval (Gantar idr. 2023), ki je pokazala, da so težave pri rabi večbesednih izrazov v šolskem pisanju razmeroma pogoste in da hkrati z vidika učiteljskih povratnih informacij velikokrat ostajajo neobravnane.³ Drugi razlog je, da je ustreznost rabe večbesednih izrazov, ki pogosto temelji na ustaljenosti in tipičnosti, mogoče s pomočjo širše jezikovne rabe, kot se odraža v referenčnem korpusu standardnega jezika, tudi preverjati. Vzroke, zakaj ostaja večbesedna problematika zunaj učiteljeve pozornosti, v tem prispevku puščamo ob strani, je pa enega od vzrokov mogoče videti v tem, da gre v večini primerov za poseg v slogovno sfero učenčevega pisanja, ki pogosto ne odpira le ene možnosti za (drugačen) zapis. Namen našega prispevka je zato ne samo prepoznati težave v šolskem pisanju na ravni večbesedne leksike, pač pa raziskati tudi možnosti ponujanja variantnih ubeseditiv kot možne povratne informacije povsod, kjer te variante obstajajo in jih je mogoče upravičiti z rabo v standardni slovenščini.

Pri tem bomo izhajali iz analiz šolskega pisanja v korpusu Šolar 3.0 (Arhar Holdt idr. 2022, 2024) in problematiko ubeseditiv analizirali z vidika konvencionalnih vzorcev, ki jih lahko preverjamo v korpusu splošnega jezika Gigafida 2.0 (Krek idr. 2019).⁴ Na ta način bomo na eni strani ponudili možnosti za avtomatsko prepoznavanje problematičnih vzorcev v korpusu in za razvoj jezikovnih črkovalnikov za slovenščino, na drugi strani pa vidimo prepoznavanje in vrednotenje jezikovne variantnosti na podlagi korpusnih virov kot koristno izhodišče tudi za učitelje, bodisi pri pripravi učnih gradiv ali pri oblikovanju povratne informacije.

Naša raziskava ima nekaj omejitev, ki jih moramo za ustrezno vrednotenje predstavljenih rezultatov izpostaviti na začetku.

Sprva smo ob predlaganih alternativnih ubeseditvah beležili tudi, ali je popravek predlagal učitelj ali ga je dodal jezikoslovec oz. jezikoslovka pri ročnem

3 Kot ugotavlja Pižorn (2014, str. 78), učiteljem podajanje povratne informacije na pisne izdelke povzroča časovno obremenitev, zato se pogosto osredotočajo le na nekatere, npr. ponavljajoče se napake oz. napake v območju učenčevega bližnjega razvoja. Podobno navajajo učitelji v anketi, izvedeni v okviru projekta Empirična podlaga za digitalno podprt razvoj pisne jezikovne zmožnosti (Rot Vrhovec idr. 2024), ki hkrati poudarjajo zavedanje o pomenu povratne informacije za razvoj pisne zmožnosti.

4 Korpus Gigafida 2.0 je obsežen referenčni korpus, ki vsebuje več kot milijardo besed iz besedil, nastalih v letih 1990–2018, in predstavlja temeljni podatkovni vir sodobne slovenščine, ki se uporablja za jezikoslovne raziskave, jezikovni opis ter za razvoj jezikovnih tehnologij in postopkov. Korpus je za jezikoslovno rabo poleg konkordančnika CJVT prosto dostopen tudi v konkordančnikih NoSketch Engine, Kontext in v orodju SketchEngine.

označevanju učne množice Šolar-Eval. Ker pa se odločitve pri nekaterih (lahko ponavljajočih se) problemih med učitelji samimi (ter med učitelji in jezikoslovci) razlikujejo tako glede (ne)posega v besedilo kot tudi glede predlagane rešitve (npr. učenec: *v zapahe* – učitelj: *v zapor*, jezikoslovec: *za zapahe*; učenec: *brez kaj reči* – učitelj: *ne da bi kaj rekel*, jezikoslovec: *brez besed*) in ker v evalvacijski množici Šolar-Eval za vse zabeležene popravke ni podatka o odločitvi jezikoslovca, te informacije pri nadaljnji analizi nismo upoštevali. Učiteljev odziv na določeno učenčevo ubeseditev je v raziskavi uporabljen v prvi vrsti kot možno izhodišče za analizo jezikovnih težav.

Izbira kategorij za opredelitev jezikovnih problemov učencev na ravni večbesednih enot v raziskavi ni naključna. Izhaja namreč iz načina organizacije podatkov v Digitalni slovarski bazi za slovenščino (DSBS; Gantar 2020) in je primarno namenjena avtomatskemu prepoznavanju podobnih skladenjskih struktur v korpusih ter učenju črkovalnika. Pri kategorizaciji posameznih problemov, izbranih na podlagi šolskih spisov, pa se je pokazalo, da kategorizacija, ki v DSBS omogoča funkcionalno prehajanje med kategorijami, šolski situaciji manj ustreza. Najprej zato, ker lahko posamezni problemi na večbesedni ravni vključujejo tudi spremembe na skladenjski in morfološki ravni, segajo čez okvir stavka ipd., zaradi česar ne predstavljajo klasičnih leksikalnih enot, ki jih v DSBS obravnavamo kot iztočnice v osnovni obliki. In na drugi strani zato, ker možna prehajanja med kategorijami ne ponujajo trdne podlage za odločanje učiteljev. Kljub vsemu smo osnovno razdelitev v raziskavi ohranili, pri čemer menimo, da sama kategorizacija za prepoznavanje problema z vidika učitelja ni ključna. Ne zdi se namreč tako pomembno, ali učitelj opredeli določen jezikovni problem kot kolokacijo, stalno zvezo, frazeološko enoto ali skladenjsko zvezo, ključno je, da pri določeni besednozvezni kombinaciji prepozna možnost alternativne ubeseditve in jo s pomočjo korpusnih virov tudi ustrezno ovrednoti.

Bolj kot prepoznavanje posameznih tipov večbesednih enot je za podajanje povratne informacije pomembno vrednotenje učenčeve ubeseditve v odnosu do predlaganega popravka. Pri tem predlagana delitev na napako in varianto temelji na pragmatičnih izhodiščih, kjer nam napako predstavlja odklon od norme/standarda, otežena razumljivost ali napačno razumevanje, vse druge možnosti pa obravnavamo kot variante. Na podlagi slovarskih oznak v DSBS in korpusnih informacij predlagamo tudi opredelitev variante kot redke, slogovno ali registrsko zaznamovane ali nevtralne. Taka razdelitev vsekakor pušča veliko prostora za dopolnitve, zlasti v smislu načina podajanja povratne informacije in preverjanja različnih možnosti v pedagoški praksi.

Jezikovna napaka

Jezikovne napake, ki jih delamo rojeni govorniki v procesu usvajanja jezika, se razlikujejo od jezikovnih napak tistih, ki se učijo tujega jezika. Hkrati pa definicija napak pri učenju jezika navaja nekatere značilnosti, ki jih je mogoče upoštevati tudi pri razumevanju jezikovne napake pri rojenih govornikih. Pri usvajanju tujega

jezika je jezikovna napaka definirana glede na pričakovano jezikovno produkcijo rojenih govorcev (James v Pirih Svetina 2003), tj. kot »jezikovna oblika ali kombinacija oblik, ki je v identičnem sobesedilu in pod podobnimi pogoji po vsej verjetnosti rojeni govorci jezika ne bi tvoril« (Lennon v Pirih Svetina 2003, str. 20). Kot ugotavlja Pirih Svetina, so jezikovne napake rojenih govorcev drugačne po svojem izvoru in obliki ter predvsem v tem, da bi »rojeni govorniki skoraj večino svojih napak znali popraviti, če bi jih nanje opozorili oz. jim dali možnost, da jih popravijo« (prav tam, str. 18). Mogoče je reči, da gre tako pri učenju tujega jezika kot pri rojenih govornikih pri jezikovnih napakah za »odstopanja od pričakovanega, ustreznega in pravilnega«, kar pomeni določeno vrsto »neuspešnosti v jezikovni produkciji« (prav tam). Napako je zato nujno definirati v odnosu do pričakovanj v ciljnem jeziku (prav tam). Lahko je opredeljena različno, npr. kot odklon od norme ali od katere od jezikovnih različic, nenazadnje kot odklon od »nedosegljivega ideala« rojenega govorca.⁵

Poleg tega lahko ločimo t. i. odkrite napake, ki so kot take prepoznane že na ravni zapisa ali kot oblike, ki ne ustrezajo slovnici in pomenskim zakonitostim jezika. Na drugi strani so težje prepoznavne t. i. prikrita napake, »pri katerih na videz oblika ne odstopa od norme oziroma je besedilo na površini dobro oblikovano, vendar ni v skladu z želeno sporočevalčevo namero« (James v Pirih Svetina 2003, str. 21).

Analiza jezikovnih napak v šolski produkciji zahteva poleg omenjenega še upoštevanje šolske situacije, znotraj katere je ciljni jezik opredeljen z jezikovnim standardom, preverljivim v jezikovnih priročnikih na eni strani, in z rabo v korpusu standardnega jezika Gigafida 2.0 na drugi.

Jezikovna varianta

V teoriji, ki se ukvarja s to problematiko z leksikalnega in skladenjskega vidika,⁶ je mogoče ločiti dva tipa variant:⁷ gramatične in leksikalne. Gramatične variante vključujejo prepoznavanje semantično povezanih gramatičnih vzorcev v besedilu (Kolbe-Hanna in Szmrecsanyi 2015), največkrat na podlagi specializiranih označenih korpusov in anket, npr. med trpnimi in tvornimi zgradbami (Weiner in Labov 1983) ter med večbesednimi in enobesednimi zvezami (npr. v rabi fraznih glagolov; Hundt in Mair 1999). Gramatične variante so znotraj tega analizirane glede na različne spremenljivke, kot so časovna obdobja (dihrone variante; Gries in Hilpert 2010), besedilni žanri in registri (žanrske in registrske variante; Biber 1988; Hundt in Mair 1999), dialekti (regionalne variante; Szmrecsanyi 2013) ter spol in socialni položaj (sociolingvistične variante; Hundt in Mair 1999). Na drugi

5 Prim. Dodatek k SEJO (2011), kjer se namesto približevanju jezikovni zmožnosti rojenega govorca govori o merilu razumljivosti.

6 Koncept jezikovne izbire je širši in zahteva tako razumevanje ideologij norme kot tudi razmerja med pisno standardnostjo in govorno variabilnostjo, zato puščamo meje v smislu kategorizacije ohlapne in se osredotočamo na prepoznavanje variantnosti na leksikalnogramatični ravni.

7 Izraz jezikovna varianta (oz. varianta) razumemo kot regionalne, družbene ali kontekstualne razlike v načinih, kako ljudje uporabljamo določen jezik.

strani Geeraerts (1994) deli leksikalne variante na konceptualne in formalne. Konceptualne variante zajemajo tako (i) pomenska razmerja med alternativnimi ubeseditvami v smislu so-, nad- in podpomenskosti (*spodnjice* – *spodnje hlače*; *hlače* – *spodnjice*; *džins* – *kavbojke*),⁸ (ii) variante, ki pripadajo različnim leksikalnim oz. pomenskimi kategorijam (npr. *džins*: ‘blago’ – *hlače*: ‘oblačilo’), kot tudi (iii) izrazne variante, pri katerih je isti referent poimenovan z različnimi besedami, ki se s semantičnega vidika med seboj lahko ali pa ne razlikujejo, npr. *spodnje hlače* – *spodnjice*; *gate* – *spodnjice*). Na drugi strani so kot formalne obravnavane variante, ki so sociolingvistične narave. Zanje velja, da je naklonjenost izražanju določenega pomena z enim izrazom namesto z drugim pogojena s kontekstualnimi dejavniki, kot je npr. govorna situacija, ki predvideva določen register (npr. *hlače* – *gate*), odločitev govornika za določen slog (*hlače* – *hlamudračje*), ali pa gre za z geografskimi dejavniki povezane izbire, ki so opredeljene kot narečne variante (npr. *hlače* – *bregše*;⁹ Szmrecsanyi 2013).

Nabor jezikovnih problemov na ravni večbesednih enot

Nabor jezikovnih problemov, ki se kažejo v rabi večbesednih enot pri šolskem pisanju in smo jih vključili v analizo, vsebuje 310 večbesednih enot in je sestavljen iz dveh delov: del jezikovnih problemov izhaja iz učne množice Šolar-Eval (Arhar Holdt idr. 2023; Gantar idr. 2023), ki je bila izdelana za razvoj in evalvacijo strojnega popravljanja. Ta del vsebuje jezikovne probleme, ki so jih v korpusu Šolar 3.0 evidentirali učitelji na ravni besedišča, v nekaterih primerih pa so bili dodani še t. i. jezikoslovčevi popravki,¹⁰ ki smo jih vključili v učno množico za evalvacijske namene. Drugi del nabora problemov je nastal na podlagi analize korpusa Šolar 3.0 z iskanjem po kategoriji označenih napak, ki zadevajo skladnjo in besedišče.¹¹ V teh primerih so predlagani popravki samo učiteljski.

Kategorizacija jezikovnih problemov

Izhodiščne, tj. problematične ubeseditve učencev, ki smo jim dodali predlagane popravke oz. alternativno ubeseditve,¹² smo prenesli v Excelovo datoteko ter posamezni jezikovni problem in zanj predlagani popravek kategorizirali glede na

⁸ Primeri so v slovenskem prevodu povzeti po Geeraerts (1994, str. 78–82).

⁹ Slovar sopomenk slovenskega jezika (Sopomenke 2.1): geslo hlače – uporabniško dodana sopomenka: bregše.

¹⁰ Vsi označevalci evalvacijske množice so jezikoslovke in jezikoslovci, ki se ukvarjajo z izdelavo jezikovnih virov, ne poučujejo pa slovenščine na katerikoli šolski ravni.

¹¹ Iskalni parametri: besedišče: vse besedne vrste razen veznika, izključili smo dodatne oznake in ohranili »menjava prek meje besedne vrste«. Pri skladnji smo upoštevali vse razen besednega reda in dodatnih oznak. Konkordančnik je dostopen na spletni strani Centra za jezikovne vire in tehnologije.

¹² V prispevku uporabljamo izraz alternativna ubeseditve za učiteljeve in/ali jezikoslovčeve popravke ne glede na njihovo normativno ali slogovno vrednost. Izraz varianta v prispevku definiramo v odnosu do napake, zaradi česar smo vpeljali tudi terminološko razlikovanje.

različne strukturne in pomenske vidike. Pri kategorizaciji jezikovnih problemov smo izhajali iz tipov večbesednih enot, kot smo jih definirali v DSBS (Gantar 2020), pri njihovi pomensko-skladenjski klasifikaciji pa iz teorije leksikalne variantnosti (Geeraerts 1994), kot smo jo opredelili zgoraj, pri čemer smo podrobnosti znotraj posamezne kategorije prilagodili specifičnim lastnostim jezikovnih problemov. Na koncu smo izhodiščne jezikovne probleme glede na predlagani popravek ovrednotili še z vidika (ne)zaznamovanosti.

Oprelitev jezikovnih problemov glede na tip večbesedne enote

Prva spremenljivka pri kategorizaciji jezikovnih problemov je določitev tipov večbesednih enot glede na pomenske in strukturne lastnosti. Na podlagi teh kriterijev smo nabor problemov opredelili kot stalne zveze, frazeološke enote, kolokacije in skladenjske zveze, ki jih v DSBS (Gantar idr. 2021) obravnavamo kot podtipke večbesednih enot. V tem naboru predstavljajo *stalne zveze* večbesedna poimenovanja za predmete, nazive, pojave, ki jih govorci ohranjamo v mentalnem leksikonu kot pomensko celoto in imajo relativno ustaljeno zgradbo, vključno z variantami. Tipično gre za polterminološke in terminološke zveze, kar velja tudi za naš nabor problemov, npr. *osebnostna izpoved* – *osebna izpoved*, *materinski* – *materni jezik*.

Frazeološke enote so v DSBS opredeljene kot ekspresivni elementi jezika, s katerimi ustvarjamo določeno pripovedno opaznost in nenevtralnost (Gantar 2021), zato morda v šolskem pisanju zahtevajo dodatno pozornost, npr. *zaljubiti se v prvem pogledu* – *na prvi pogled*.

Kolokacije razumemo kot besede, ki se skupaj pojavljajo statistično značilno pogostejše, kot bi bilo pričakovano po naključju. Kot take v jeziku ustvarjajo naravnost, ki je značilna za rojene govorce, ter imajo pomembno vlogo pri učenju (tujega) jezika (Sinclair 1991). Za razliko od skladenjskih zvez zastopajo kolokacije končni seznam struktur v DSBS, ki jih določata vsaj dva leksikalna oz. predmetnopomenska elementa, npr. pridevnik + samostalnik: *tragični konec*, ali glagol + predložna zveza: *naleteti na oviro*. Za kolokacije je značilno, da so pomensko prozorne in v slovarju ne potrebujejo razlage kot celota.

Skladenjske zveze so kombinacije slovničnega in leksikalnega elementa v določeni obliki, ki navadno prek predloga napovedujejo relativno prosto skladenjsko mesto, npr. *pod okriljem koga/česa*, *v zvezi s čim*. Skladenjske zveze zajemajo širok spekter predložnih (npr. *trden do česa* – *v čem*), prislovnih in vezniških zvez (npr. *v četrto* – *četrtič*; *zaradi* – *iz ljubosumnosti*), ki so v jeziku zelo produktivne, niso pa nujno nosilke samostojnega leksikalnega pomena.

Z naslonitvijo na klasifikacijo, ki smo jo uporabili pri organizaciji leksikalnih podatkov v DSBS (Gantar idr. 2021), bo mogoče jezikovne probleme šolarjev obravnavati tudi v povezavi z drugimi podatki, ki jih v zvezi z večbesednimi enotami vključuje DSBS, ter s tem povečati učno množico za učenje avtomatskih črkovalnikov in pregledovalnikov.

Opredelevanje jezikovnih problemov glede na tip alternativne ubeseditve

Glede na tip alternativne jezikovne ubeseditve smo jezikovne probleme opredelili glede na to, ali gre za izrazno, leksikalno ali konceptualno alternativno ubeseditve.

Kot *izrazne* alternativne ubeseditve smo obravnavali zamenjave na ravni slovničnih elementov zveze, kot so denimo predlogi, kjer gre predvsem za problem vezljivosti, npr. *v zapahe – za zapahe, stremeti k idealom – stremeti za ideali*, pogosto v kombinaciji z morfološkimi spremembami leksikalnih elementov, npr. *po mojem okusu – za moj okus*. Kot izrazne alternativne ubeseditve smo obravnavali tudi zamenjavo besedne zveze z enobesedno znotraj besedne družine, npr. *po nagonu – nagonsko, postavljati vprašanja – spraševati*, in obratno, npr. *fizično – na fizični ravni*.

Kot *leksikalne* alternativne ubeseditve smo opredeljevali zamenjave leksikalnega elementa z drugim znotraj pomenskega razmerja, npr. nad- in sopomenskosti: *po dolgem letu – po dolgem času; pripeljati do smrti – privedi do smrti, prikopati si krivdo – nakopati si krivdo*.

Tretji tip alternativnih ubeseditve smo po Geeraertsu (1994) poimenovali *kontekstualne*. Te dejansko opredeljujejo govorcevo izbiro leksikalnih elementov, specifičnih za določeno jezikovno situacijo, predvidoma na podlagi interference iz drugih jezikov: *najti se v situaciji – znajti se v situaciji, odkloniti se komu – odpovedati se komu*; narečne interference: *upati se učiti – lahko se učiti, odpreti luč – prižgati luč*, sloga: *ljubiti nazaj – vračati ljubezen, ne imeti nobene komunikacije – se ne pogovarjati, super – kot po maslu*, ali registra: *niti pod razno – nikoli, za brez veze – po nepotrebnem*.

Opredelevanje jezikovnih problemov z vidika vrste zaznamovanosti

Tretji tip kategorij opredeljuje jezikovne probleme kot napake ali variante. Razumevanje *napake* je v naši raziskavi vezano na primere, kjer izhodiščna učenčeva ubeseditve bodisi otežuje uspešno razumevanje bodisi sugerira drugačno razumevanje sporočila, kot je bil (predpostavljen) učenčev namen, npr. *srečati na oviro – naleteti na oviro; obrati pravilno pot – ubrati pravo pot*. Glede na to, da so predmet naše obravnave večbesedne enote, smo kot napako obravnavali neustrezno rabo le-teh tudi v primerih, kjer je učenčeva ubeseditve možna, vendar se v rabi ne kaže, npr. *pritekel je čas – prišel je čas; kdaj in kdaj – kdaj pa kdaj*. V takih primerih smo stanje preverjali v korpusu standardne slovenščine Gigafida 2.0. Kot napake smo obravnavali tudi neustrezno rabo povratnih zaimkov *se/si*, npr. *biti si podobnih misli – biti podobnih misli*, kjer gre za sovpadanje zveze s povratnosvojljnim zaimkom *biti si podoben* in zveze *biti podobnih misli*, ter za slovenščino nepričakovane vezljivostne vzorce, npr. *bolehati s hudo boleznijo – za hudo boleznijo*. Na ravni frazeoloških enot kot napak nismo obravnavali alternativnih ubeseditve, ki jih izkazuje raba v referenčnem korpusu Gigafida 2.0, npr. *prevzeti v svoje roke –zeti v svoje roke*.

Kot *variante* smo na drugi strani obravnavali vse tiste jezikovne probleme v odnosu do alternativnih ubeseditiv, ki na leksikalni ali skladenjski ravni ohranjajo sporočilo, smisel izrečenega, slovnično pa ne odstopajo od pričakovanega vzorca – ob predpostavki, da ob tem prihaja do žanrskih, registrskih in drugih pomenskih nians, npr. *pogrešati iz dna srca – na smrt pogrešati; imeti trmo – biti trmast*. S tega vidika smo jezikovne probleme ovrednotili glede na stilno ali registrsko zaznamovanost, pri čemer smo izhajali iz vrednosti slovarskih oznak, kot smo jih opredelili v okviru DSBS (Gantar idr. 2021). Z oznako »register« smo označevali besede, ki so značilne za neformalno komunikacijo, ki tipično poteka v nejavnih govornih položajih in je pretežno vezana na govor v domačem, znanem okolju. Tu obravnavamo tudi jezikovne probleme, za katere predvidevamo, da so nastali na podlagi jezikovne interference ali vpliva slenga, npr. *brez kaj reči – brez besed; za brez veze – po nepotrebnem*. Z oznako »slog« pa variante, katerih namen je vzbujati asociacije z metaforičnimi izrazi ali izrazi, vezanimi na določen žanr, npr. leposlovje, ali stil. Kot redke smo označili jezikovne probleme, ki so glede na predlagano ubeseditiv v splošnem referenčnem korpusu redko zastopani.

Analiza jezikovnih problemov na ravni večbesednih enot

Za analizo smo uporabili Excelovo datoteko, v kateri smo izhodiščnim jezikovnim problemom dodali predlagane alternativne ubeseditve ter jih kategorizirali glede na tip večbesedne enote in glede na tip leksikalne variantnosti. V datoteki smo beležili tudi pojavnost izhodiščnega problema in alternativne ubeseditve v referenčnem korpusu Gigafida 2.0.

Analiza jezikovnih napak

V odnosu do popravkov so učenčeve ubeseditve prepoznane kot napake v zgoraj definiranem pomenu v približno slabi tretjini primerov, vendar različno znotraj tipa večbesedne enote. Najpogostejše se napake pojavljajo v rabi stalnih besednih zvez, čeprav je na drugi strani v naš nabor vključenih relativno malo primerov, med skupno 310 le 10. Relativno veliko napak delajo učenci pri rabi frazeoloških enot, približno v eni tretjini vseh obravnavanih primerov, pri čemer je iz podatkov mogoče sklepati, da dijaki in učenci tudi sicer v šolskem pisanju bodisi redkeje posegajo po metaforičnih izraznih sredstvih bodisi z njihovo rabo nimajo težav. V približno četrtini vseh primerov so učenci delali napake na ravni kolokacij in skladenjskih zvez. Predlagani popravki, ki jih je mogoče obravnavati z vidika različne slogovne vrednosti in žanrske pripadnosti, so na drugi strani v več kot 70 odstotkih vezani na kolokacije in skladenjske zveze, kot prikazuje Preglednica 1.

Tip VBE	Število primerov	Število napak	Število variant
Frazeološka enota	54	19 (35 %)	35 (64 %)
Stalna zveza	10	7 (70 %)	3 (30 %)
Kolokacija	136	36 (26 %)	100 (73,5 %)
Skladenjska zveza	110	27 (24,5 %)	83 (75 %)
Skupaj	310	89 (28,7 %)	221 (71 %)

Preglednica 1: Delež napak in variant glede na število obravnavanih primerov znotraj posameznega tipa večbesedne enote

Stalne besedne zveze

Kot je razvidno iz Preglednice 1, je največji delež napak v našem naboru problemov vezan na rabo stalnih besednih zvez, kar je pričakovano predvsem pri terminoloških zvezah. V besedilih, ki jih vključuje korpus Šolar, je mogoče prepoznati zlasti napake v rabi literarnih pojmov (Preglednica 2).

Jezikovni problem	Popravek
dramatsko besedilo	dramsko besedilo
osebna izpoved	osebna izpoved
vložena pesem	vložna pesem
materinski jezik	materni jezik
privzeta beseda	prevzeta beseda

Preglednica 2: Izbrani nabor primerov napačne rabe terminoloških stalnih zvez

Frazeološke enote

Napake, ki jih učenci in dijaki delajo pri rabi frazeoloških enot, so največkrat povezane s t. i. kontaminacijo dveh večbesednih enot z ustaljenim pomenom in obliko, npr. *dobil svoje, kar si je zaslužil* – *dobiti svoje + dobiti, kar si kdo zasluži; končati s tragičnim koncem* – *tragični konec + tragično končati*. Pogoste so tudi zamenjave leksikalnega elementa z drugim znotraj določenega pomenskega polja, npr. *priskočiti na delo* – *priskočiti na pomoč; teči okoli koga* – *tekati okoli koga*, pri čemer je tak element znotraj frazeološke enote nepričakovan in v splošnem jeziku ni izkazan. Poleg napačne izbire leksikalnega elementa v frazeološki enoti prihaja tudi do zamenjave slovničnih elementov, npr. predlogov, posledično pa tudi do nepričakovane vezljivosti: *zaljubiti se v prvem pogledu* – *zaljubiti se na prvi pogled; nad koga kazati s prstom* – *na koga kazati s prstom*, ali celo do napačnega razumevanja: *stati komu pri strani* (dobesedni pomen) – *stati komu ob strani* (frazeološki pomen 'podpirati koga'). Nekaj izbranih primerov napačne rabe frazeoloških enot prikazuje Preglednica 3.

Jezikovni problem	Popravek
dobil svoje, kar si je zaslužil	dobiti svoje/dobiti, kar si kdo zasluži
priskočiti na delo	priskočiti na pomoč
tak kot oče, tak kot sin	kakršen oče, takšen sin
zaljubiti se v prvem pogledu	zaljubiti se na prvi pogled
nastati pod roko koga	nastati izpod peresa koga
pustiti se v toku	prepustiti se toku
ljubiti nazaj	vračati ljubezen

Preglednica 3: Izbrani nabor primerov napačne rabe frazeoloških enot

Kolokacije

Največji delež težav, ki jih imajo učenci pri tvorjenju večbesednih enot, je kolokacijske narave.¹³ Kolokacijske napake so, kot lahko razberemo iz pisne produkcije šolarjev in dijakov, vezane predvsem na izbiro napačnega ali netipičnega leksikalnega elementa v kolokaciji: *volja za denarjem – gonja za denarjem; obrati – ubrati pravilno pot; neizhoden – brezizhoden položaj*. Tak element je sicer lahko izbran tudi znotraj enotnega semantičnega okvira, npr. *propasti na tragičen konec – doživeti tragični konec, vzkipljive jeze – vzkipljive narave*, kot prikazujejo izbrani primeri v Preglednici 4. Obstoj alternativne ubeseditve v odnosu do učenčeve kolokacije smo preverjali v korpusu Gigafida 2.0.

Jezikovni problem	Popravek
propasti na tragični konec	doživeti tragičen konec
srečati se v vlogi	znajti se v vlogi
odkloniti se komu	odpovedati se komu
prikopati si krivdo	nakopati si krivdo
prvo morati	najprej morati
srečati na oviro	naleteti na oviro
gojiti obnašanje	vesti se
biti navezan s kom	biti povezan s kom
odločen sam vase	prepričan sam vase
obupano si želeli	obupno si želeli
ravnati obratno svoji vesti	ravnati nasprotno svoji vesti ¹⁴

¹³ Raziskava o procesu usvajanja kolokacij na šolski ravni (Rozman idr. 1918, str. 125) je pri analizi napak izpostavila predvsem slogovno (*cel čas – ves čas*) ali pomensko neustreznost (*posvetno življenje – posmrtno življenje*), manj običajno kombinacijo besed (*rutinski sprehod – rutinski obhod*) in mešanje (*nadmorska gladina – nadmorska višina*).

¹⁴ V korpusu Gigafida 2.0 najdemo predvsem *ravnati ipd. proti svoji vesti*.

povpraševati z uprašanjem	postavljati vprašanja
nazaj se ogreti za	ponovno se ogreti za

Preglednica 4: Izbrani primeri napačne rabe kolokacij z učiteljskimi popravki

Skladenjske zveze

Znotraj tega strukturnega tipa obravnavamo kot napake zveze, ki odstopajo od pričakovanega leksikalnogramatičnega vzorca, tipično v izbiri predloga, npr. *iz otroških oči – skozi otroške oči, čez časoma – čez čas/sčasoma, v kontroli – pod kontrolo*. Do napak prihaja v izbiri vezljivostnega vzorca, tako pri glagolih kot tudi pri samostalnikih in pridevniki, npr. *dodati na zadnjo točko – k zadnji točki, razlika do česa – razlika v čem, trden do česa – trden v čem*, ter pri rabi povratnosvojilnih zaimkov se/si, npr. *biti si podobnih misli – biti podobnih misli*, kjer gre za sovpadanje povratnosvojilnega zaimka v zvezi *biti si podoben* in zveze *biti podobnih misli*. Nekaj izbranih primerov napačne rabe prikazuje Preglednica 5. Prisotnost oz. pojavnost zveze smo preverjali v referenčnem korpusu Gigafida 2.0.

Jezikovni problem	Popravek
pripovedovati iz otroških oči	pripovedovati skozi otroške oči
dodati na zadnjo točko	dodati k zadnji točki
povzročati kaj za koga	povzročati kaj komu
po veliki meri	v veliki meri
predati se v resignacijo	predati se resignaciji
pripadništvo do koga	pripadnost komu
v kontroli	pod kontrolo
čez časoma	sčasoma
ven čez okno	skozi okno
skos	ves čas
razlika do česa	razlika v čem
bolehati z boleznijo	bolehati za boleznijo
ljubimkati se s kom	ljubimkati s kom
stopiti ven iz vere	izstopiti iz vere
želja za boljše življenje	želja po boljšem življenju
sebi blizu	ob sebi

Preglednica 5: Izbrani primeri napačne rabe večbesednih enot z lastnostjo skladenjske zveze in učiteljski popravki

Analiza jezikovnih variant

V nadaljevanju predstavljamo analizo alternativnih ubeseditv glede na učenčev jezikovni problem, ki smo jih opredelili kot variante. Izhodiščne ubeseditve smo znotraj posameznega tipa večbesedne enote analizirali glede na to, ali so na tak ali drugačen način zaznamovane ali ne, tj. ali sta izhodiščna in predlagana ubeseditv v splošnem jeziku razmeroma zamenljivi. Med zaznamovanimi variantami smo posebej obravnavali t. i. slogovne variante, ki se nanašajo na učenčevo izbiro ekspresivnih jezikovnih sredstev, in t. i. registrske variante, ki so zaznamovane predvsem z vidika pogovornih in slengovskih elementov v večbesedni enoti.

Stalne besedne zveze

Stalne besedne zveze, ki smo jih zajeli v nabor jezikovnih problemov, učenci uporabljajo napačno, ko gre za strokovne izraze (gl. Preglednico 2). Kot variantne lahko upoštevamo učiteljske predloge, kot sta npr. *free style* – *prosti slog* in na ravni sloga *postaja za avtobus* – *avtobusna postaja*.

Frazeološke enote

Pri jezikovnih problemih, ki izkazujejo lastnosti frazeološke enote – gre torej za zveze besed z lastnim, od sestavin neodvisnim pomenom, predvidljivo leksikalno zgradbo in metaforičnim ali drugače prenesenim pomenom –, učenci navadno zamenjujejo določen leksikalni element z drugim znotraj istega pomenskega polja, pri čemer je zveza s tem elementom izkazana bodisi kot v korpusu redkejša bodisi kot registrsko ali slogovno zaznamovana varianta. Kot sugerirajo slogovne variante v Preglednici 6, lahko v njihovi rabi prepoznamo namerno učenčevo izrabo jezikovne pestrosti in ustvarjanje literarnega učinka. V takih primerih se učiteljski predlogi nagibajo k rabi enobesednih in bolj nevtralnih izrazov. Primeri, kjer je predlagana varianta ekspresivnejša glede na učenčevo, so redki, npr. *takoj se je zaljubil* – *takoj mu je poskočilo srce*.

Jezikovni problem	Varianta	Oznaka zaznamovanosti
imeti samo eno misel v duši	imeti samo eno misel v glavi	redko
skočiti v oči	pasti v oči	redko
prevzeti v svoje roke	vzeti v svoje roke	redko
iz ene strani	po eni strani	redko
imeti za sabo	doživeti	slog
konec koncev	nenazadnje	slog

kaj se komu plete po glavi	o čem kdo razmišlja	slog
pogrešati iz dna srca	na smrt pogrešati	slog
držati se česa	upoštevati kaj	slog
dati skozi	prestati	slog
ne pustiti se komu	postaviti se zase	slog
od malih nog, od majhnega	dolgo/od malega	slog
ostati brez glave	biti kaznovan	slog
imeti slabo vest	slaba vest teži koga	slog
priti zadevi do konca	rešiti zadevo	slog
za brez veze	po nepotrebnem	register
brigati se zase	zanimati se zase	register
niti pod razno	nikoli	register
priti vmes	poseči	register
iti nazaj	vrniti se	register
ne sekirati se za	ne biti mar	register
priti do	zgoditi se	nevtravno

Preglednica 6: Izbrani primeri izhodiščnih frazeoloških enot in variantnih ubeseditiv glede na vrsto zaznamovanosti (poudarjeni elementi izhodiščne zveze označujejo del frazeološke enote, ki je podvržen spremembi)

Kolokacije

Tako kot pri frazeoloških enotah je tudi pri izhodiščnih ubeseditvah z lastnostjo kolokacije pogosto izkazana zamenjava leksikalnega elementa z drugim znotraj istega pomenskega polja, pri čemer so take variante v korpusu bodisi redkejšje, a vseeno zastopane kot kolokacije, bodisi elementi, ki so v kolokaciji podvrženi zamenjavi, tvorijo slogovne ali registrske variante. Podobno kot pri frazeoloških enotah se tudi v tem primeru popravki nagibajo k enobesednim enotam in ekspresivno nevtralnejšim zvezam, zlasti pri t. i. zvezah z glagoli z oslavljenim pomenom, npr. *dati obljubo* – *obljubiti*, *imeti trmo* – *biti trmast*. Drugi tip zamenjav so elementi pogovornega jezika, kot prikazuje Preglednica 7, ki jih obravnavamo kot registrske variante.

Jezikovni problem	Varianta	Oznaka zaznamovanosti
pripeljati do smrti	privedi do smrti	redko
najti se v situaciji	znajti se v situaciji	redko
zgledati mrtev	izgledati mrtev	redko

volja po življenju	volja do življenja	redko
dandanes v svetu	v današnjem svetu	slog
kapljica sreče	kanček sreče	slog
fizično	na fizični ravni	slog
neznosno se razveseliti	neizmerno se razveseliti	slog
dati obljubo	obljubiti	slog
pasti v spanec	zaspati	slog
ne imeti nobene komunikacije	se ne pogovarjati	slog
nima nobenega pomena	ni pomembno	slog
biti jasno	vedeti	slog
v normalnem tonu	normalno	slog
tragično končati	umreti	slog
biti stoprocenten	biti prepričan	register
se nenormalno zaljubi	se neizmerno zaljubi	register
počutiti se krivca	počutiti se krivega	register
imeti trmo	biti trmast	register
totalno nasprotje	popolno nasprotje	register
v zadnjem momentu	v zadnjem trenutku	register
vzeti denar	ukrasti denar	register
imeti nadutost	biti nadut	register
imeti strah	biti strah	register
biti siguren sam vase	biti prepričan sam vase	register
po mojem okusu	za moj okus	nevtravno
narediti prav	ravnati pravilno	nevtravno
dobre in slabe stvari	dobro in slabo	nevtravno
imeti skupno	povezovati	nevtravno
biti proti	nasprotovati	nevtravno
pada dež	dežuje	nevtravno
dajati nasvete	svetovati	nevtravno
biti v dvomih	dvomiti	nevtravno

Preglednica 7: Izbrani primeri izhodiščnih kolokacij in variantnih ubeseditvev glede na vrsto zaznamovanosti (poudarjeni elementi izhodiščne kolokacije označujejo del, ki je podvržen spremembi)

Skladenjske zveze

Velik del učenčevih problematičnih ubeseditv je vezan na skladenjske zveze, kjer prihaja do zamenjave predloga, pri čemer so take variante v korpusu Gigafida 2.0 redkeje zastopane: *zaradi – iz ljubosumja; preko pogovora – v pogovoru*. Kadar prihaja do sprememb na ravni leksikalnega elementa, so variante tipično slogovno ali registrsko zaznamovane, npr. *za zmeraj – za vedno, na račun česa – zaradi česa*, kjer prihaja tudi do interference iz drugih jezikov, npr. *brez kaj reči – brez besed, brez da bi – ne da bi* ipd. Sem lahko štejemo tudi spremembe v vezljivostnem vzorcu, npr. *znajti se na enakem – v enakem položaju; ljubosumnost do – na koga; jezen nad kom – na koga; uvrstiti pod – med kaj; izvedeti o smrti – za smrt*. Zanimivo je, da lahko za kar velik delež izhodiščnih ubeseditv najdemo enakovredne, tj. jezikovno nevtralne predloge. Izbrani primeri za posamezno skladenjsko zvezo so skupaj z vrsto zaznamovanosti prikazani v Preglednici 8.

Jezikovni problem	Varianta	Oznaka zaznamovanosti
v zapahe	za zapahe	redko
zaradi ljubosumja	iz ljubosumja	redko
preko pogovora	v pogovoru	redko
prek zakonov	z zakoni	redko
izvedeti o smrti	za smrt	redko
za zmeraj	za vedno	slog
na zviti način	zvito	slog
tekom noči	ponoči	slog
s hitrostjo	hitro	slog
za povrh vsega	poleg tega	slog
ne dolgo nazaj	pred kratkim	slog
eno leto nazaj	pred enim letom	slog
nedolgo tega	pred nedavnim	slog
na lažji način	lažje	slog
v negativnem smislu	negativno	slog
ravno tako	podobno	slog
za razliko od	v primerjavi s	slog
smatrati kot	imeti za	slog
na račun česa	zaradi	slog
po moje	po mojem mnenju	register
v bistvu	pravzaprav	register
za vikende	med vikendi	register
brez kaj reči	brez besed	register
od koga	čigav	register

s strani koga	kdo	register
sebi blizu	ob sebi	register
napram čemu	v nasprotju s	register
brez da bi	ne da bi	register
jezen nad kom	jezen na koga	register
uvrstiti pod kaj	uvrstiti med kaj	register
na sploh	na splošno	nevtravno
kar se tiče	kar zadeva	nevtravno
poleg tega	ob tem	nevtravno
mnenje glede česa	mnenje o čem	nevtravno

Preglednica 8: Izbrani primeri izhodiščnih skladenjskih zvez in variantnih ubeseditiv glede na vrsto zaznamovanosti (poudarjeni elementi izhodiščne zveze označujejo del, ki je potrjen spremembi)

Zaključek

Raziskava je pokazala, da je številne jezikovne probleme, ki jih imajo učenci pri šolskem pisanju, mogoče prepoznavati pri ubeseditvi večbesednih enot, še zlasti, ko so te v jeziku relativno strukturno in pomensko ustaljene. Odstopanja od pričakovanega jezikovnega standarda, ki smo jih preverjali v korpusu Gigafida 2.0, je mogoče vsaj deloma predpostaviti na podlagi pomenskih in strukturnih lastnosti različnih tipov večbesednih enot. Pri tem se zdi smiselno učenčeve ubeseditve, ki bodisi ne dosegajo sporočanje namena bodisi ne predstavljajo v jeziku pričakovane izbire, opredeliti kot napake, vse druge možnosti pa obravnavati kot variantne in podati tudi njihovo vrednost v smislu jezikovne zaznamovanosti oz. nevtralnosti.

Napačne ubeseditve učencev so vezane zlasti na zamenjavo posameznih elementov večbesedne enote, pri čemer zamenjave odražajo bodisi nepoznavanje večbesedne enote kot take bodisi prihaja do odstopanj na ravni pričakovanega skladenjskega vzorca, zlasti pri vezljivostnih vzorcih (*pripomoči za kaj – k čemu*) in rabi povratnosvojilnega zaimka (*ljubimkati se*). Variantne ubeseditve na drugi strani izkazujejo različne možnosti ubeseditve, ki imajo v različnih kontekstih različne vrednosti in so tipično v rabi v posameznih govornih položajih ali žanrih. Pri tem je mogoče opaziti, da se učenci nagibajo k uporabi večbesednih enot tudi v primeru v jeziku obstoječih enobesednih alternativ, npr. *imeti vpliv – vplivati*. Predvidevamo, da pogosta zamenjava glagola *imeti – biti* kot v zvezah *imeti trmo – biti trmast* poteka pod vplivom tujejezičnih interferenc, ki so sicer opazne tudi v prevzetih vezljivostnih vzorcih. Številne napake so v šolskem pisanju vezane na rabo predložnih in prislovnih zvez, kar je mogoče pripisati dejstvu, da so v jeziku zelo produktivne in hkrati izkazujejo različne variantne možnosti. To potrjuje tudi prisotnost takih zvez v korpusu standardne slovenščine.

Za ustvarjanje širše slike o rabi večbesednih enot v šolskem pisanju bi potrebovali analizo celotnih besedil in ne samo evidentiranih napak. Naša raziskava

tako ne more odgovoriti na vprašanje, v kolikšni meri učenci in dijaki pri svojem pisanju uporabljajo metaforična izrazna sredstva, kot so denimo frazeološke enote, in kako uspešni so pri uporabi tipičnih besednih zvez, kot so kolokacije, nasploh. Je pa naša raziskava pokazala, da v primerih, kjer do napačne ali variantne rabe prihaja, lahko prepoznamo in do določene mere predvidimo tako tip napake kot opredelimo slogovno vrednost variante.

Predstavljeno metodo vidimo kot uporabno za učitelje iz dveh razlogov: najprej z vidika uporabe korpusa pri prepoznavanju večbesednih enot, ki so potencialno problematične, npr. že omenjene zveze glagolov, kot so *imeti*, *biti* ipd., in samostalnika, predložne zveze ali vezljivostni vzorci. Nadalje z vidika prepoznavanja jezikovne variantnosti ob preverjanju v korpusu standardnega jezika, ki lahko učitelju ponudi objektivno sliko, kar je izhodišče za oblikovanje povratne informacije – ob predpostavki, da je šele zavedanje o vrednosti posameznega popravka ključno za njegovo uporabo pri nadaljnjem tvorjenju besedil. Pri tem se zdi pomembno ne samo prepoznavanje napak, ampak tudi opozarjanje na variantnost, in sicer tudi kadar gre z vidika sloga in registra za enakovredne ubeseditve. Hkrati s podrobnejšo opredelitvijo variant povratna informacija lahko sugerira bodisi jezikovno pravilnejše bodisi slogovno ustrežnejše ubeseditve. Prav zato je bil naš namen s pomočjo preverjanja realnega stanja tudi zunaj šolskega pisanja, tj. v korpusu Gigafida 2.0, ponuditi alternativne ubeseditve, s čimer se po našem mnenju krepi zmožnost prepoznavanja sorodnih problemov pri učiteljih, posledično s posredovanjem povratne informacije pa tudi izrazna zmožnost učencev.

Izsledke bomo v prihodnje uporabili tudi pri pripravi učne množice za izdelavo avtomatskega črkovalnika za slovenščino v okviru projekta Veliki jezikovni modeli za digitalno humanistiko (LLM4DH), katerega eden od ciljev je na podlagi evidentiranih tipov napak in variantnih ubeseditv s pomočjo orodij umetne inteligence generirati večje število čim bolj avtentičnih primerov, ki vsebujejo napačne in alternativne ubeseditve.

Literatura in viri

- Arhar Holdt, Š. in Kosem, I. (2024). Šolar, the developmental corpus of Slovene. *Language Resources & Evaluation*, 59, str. 1151–1177.
- Arhar Holdt, Š., Gantar, P., Bon, M., Gapsa, M., Lavrič, P. in Klemen, M. (2023). *Dataset for evaluation of Slovene spell- and grammar-checking tools Šolar-Eval 1.0*. Dostopno na repozitoriju Clarin.si: <http://hdl.handle.net/11356/1902> (pridobljeno 29. 8. 2025).
- Arhar Holdt, Š., idr. (2022). *Developmental corpus Šolar 3.0*. Dostopno na repozitoriju Clarin.si: <http://hdl.handle.net/11356/1589> (pridobljeno 29. 8. 2025).
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Cameron, R. in Schwenter, S. (2013). Pragmatics and variationist sociolinguistics. V: R. Bayley, R. Cameron in C. Lucas (ur.). *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press, str. 1–34.

- Gantar, P. (2020). Dictionary of modern Slovene: from Slovene lexical database to digital dictionary database. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 46, št. 2, str. 589–602.
- Gantar, P. (2021). Zapis kanonične oblike frazeoloških enot v Leksikonu večbesednih enot za slovenščino. V: Š. Arhar Holdt (ur.). *Nova slovnica sodobne standardne slovenščine: viri in metode*. 1. izd. Ljubljana: Znanstvena založba Filozofske fakultete, str. 198–230.
- Gantar, P., Krek, S. in Kosem, I. (2021). Opredelitev kolokacij v digitalnih slovarskih virih za slovenščino. V: I. Kosem (ur.). *Kolokacije v slovenščini*. Ljubljana: Znanstvena založba Filozofske fakultete, str. 15–41.
- Gantar, P., Bon, M., Gapsa, M. in Arhar Holdt, Š. (2023). Šolar-Eval: Evalvacijska množica za strojno popravljanje jezikovnih napak v slovenskih besedilih. *Jezik in slovnstvo*, 68, št. 4, str. 89–108.
- Geeraerts, D. (1994). Varieties of lexical variation. V: W. Martin, W. Meijs, M. Moerland, E. Pas, P. van Sterkenburg in P. Vossen (ur.). *Proceedings of the 6th EURALEX International Congress*. Amsterdam, the Netherlands: Euralex, str. 78–83.
- Gigafida 2.0: Korpus pisne standardne slovenščine*. (2019). Dostopno na: viri.cjvt.si/gigafida (pridobljeno 29. 8. 2025).
- Gries, S. T. in Hilpert, M. (2010). Modeling diachronic change in the third person singular: A multifactorial, verb- and author-specific exploratory approach. *English Language and Linguistics*, 14, št. 3, str. 293–320.
- Hundt, M. in Mair, C. (1999). »Agile« and »uptight« genres: The corpus-based approach to language change in progress. *International Journal of Corpus Linguistics*, št. 4, str. 221–242.
- Kolbe-Hanna, D. in Szmrecsanyi, B. (2015). Grammatical variation. V: D. Biber in R. Reppen (ur.). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, str. 161–179.
- Krajnc Ivič, M. (2015). O neposrednem, spontanem, govornem besedilu z jezikovnostilističnega in jezikovnopragmatičnega vidika. V: M. Smolej (ur.). *Simpozij Obdobja 34: Slovnica in slovar – aktualni jezikovni opis*. Ljubljana: Založba Univerze v Ljubljani, str. 393–400.
- Krek, S., Arhar Holdt, Š., Erjavec, T., Čibej, J., Repar, A., Gantar, P., Ljubešić, N., Kosem, I. in Dobrovoljc, K. (2020). Gigafida 2.0: the reference corpus of written standard Slovene. *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, Marseille, France: European Language Resources Association, str. 3340–3345.
- LLM4DH: Large language models for digital humanities. ARIS Gravitacija. Dostopno na: <https://www.cjvt.si/llm4dh/> (pridobljeno 2. 9. 2025).
- Pirih Svetina, N. (2003). Napaka v ogledalu procesa učenja tujega jezika. *Jezik in slovnstvo*, 48, št. 2, str. 17–26.
- Pižorn, K. (2014). Učinki učiteljeve povratne korektivne informacije pri razvijanju pisne zmožnosti v tujem jeziku. V: V. Eržen (ur.). *Izzivi razvijanja in vrednotenja znanja v gimnazijski praksi: Angleščina*. Ljubljana: Zavod RS za šolstvo, str. 65–81.
- Rot Vrhovc, A., Arhar Holdt, Š., Pižorn, K. in Godec Soršak, L. (2024). *Popravljanje pisnih besedil učencev/dijakov: raziskovalni podatki*. Dostopno na: <https://repozitorij.uni-lj.si/IzpisGradiva.php?lang=slv&id=153481> (pridobljeno 2. 9. 2025).
- Rozman, T. in Stabej, M. (2024). Univerzitetno pisanje in popravljanje besedil: Prakse in stališča. V: S. Štumberger (ur.). *Predpis in norma v jeziku: Vol. Obdobja 43*. Ljubljana: Založba Univerze v Ljubljani, str. 285–292.
- Rozman, T., Arhar Holdt, Š., Pollak, S. in Kosem, I. (2018). Kolokacije v korpusu Šolar. *Jezik in slovnstvo*, 63, št. 2/3, str. 117–128.

- SEJO: Dodatek = Skupni evropski jezikovni okvir: učenje poučevanje, ocenjevanje. Dodatek. (2023). Ljubljana: Zavod Republike Slovenije za šolstvo. Dostopno na: <https://www.zrss.si/pdf/SEJO.pdf> (pridobljeno 2. 9. 2025).
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sopomenke 2.1: Slovar sopomenk slovenskega jezika. (2023). Dostopno na: <https://viri.cjvt.si/sopomenke/slv/> (pridobljeno 2. 9. 2025).
- Szmrecsany, B. (2013). *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge, New York: Cambridge University Press.
- Weiner, J. in Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19, str. 29–58.

Polona GANTAR (University of Ljubljana, Faculty of Arts and Faculty of Computer and Information Science, Slovenia)

Mija BON (University of Ljubljana, Faculty of Arts, Slovenia)

GETTING THROUGH OR OVERCOMING IT? ERRORS AND LINGUISTIC VARIANTS IN THE USE OF MULTI-WORD UNITS IN INDEPENDENT TEXT PRODUCTION IN PRIMARY AND SECONDARY SCHOOL

Abstract: This paper describes the difficulties students encounter when using multiword expressions (MWEs) in independent text production. Our analysis is based on teachers' corrections in the Developmental corpus Šolar 3.0 (Arhar Holdt et al. 2022) and linguistic decisions made during the development of the evaluation dataset for automated error correction Šolar-Eval (Gantar et al. 2023). We categorize the identified issues in the use of MWEs according to the type of alternative linguistic expression (formal, lexical, and contextual; Geeraerts 1994), the type of MWE (fixed expressions, phraseological units, collocations, syntactic constructions), and whether the expression is a variant or an error. Language allows multiple ways of expressing the same idea, which means that corrections are not always unambiguous. At the same time, the spectrum of MWE issues ranges from errors – relatively rare occurrences that lead to misunderstanding, misinterpretation, or unexpected choices from a native speaker's perspective, to variants, where an alternative phrasing of the same message exists. Recognizing these different expressions enables the formulation of feedback that offers not only more correct, stylistically appropriate, and collocationally suitable phrasing but also alternative expressions that in the language function more or less equivalently. In school writing, there is a noticeable intentional use of more expressive linguistic choices to enrich the text and enhance narrative vividness, whereas teacher's corrections tend to favor more neutral expressions and simplification (e. g., *go through* – *endure*).

Keywords: correction of school texts, linguistic variation, language error, teacher's feedback, multiword expressions, phraseological units, collocations, syntactic combinations

Email for correspondence: apolonija.gantar@ff.uni-lj.si