



www.slovenščina.eu
sporazumevanje

**Iztok Kosem, Mojca Stritar, Sara Može, Ana Zwitter Vitez,
Špela Arhar Holdt, Tadeja Rozman**

Analiza jezikovnih težav učencev: korpusni pristop

Zbirka *Sporazumevanje*

Urednik zbirke *Simon Krek*

Recenzenta *Helena Dobrovoljc, Nataša Logar Berginc*

Prevod povzetka *Iztok Kosem*

Urednik *Iztok Kosem*

Oblikovna zasnova zbirke *Tomato Košir*

Prelom *Roman Ražman*

Naslovnica *Tomato Košir*

Avtor črkovne vrste »BadNews« *Samo Ačko*

Založila *Znanstvena založba Filozofske fakultete Univerze v Ljubljani*

Izdal *Center za jezikovne vire in tehnologije Univerze v Ljubljani*

Za založnika *Roman Kuhar, dekan Filozofske fakultete Univerze v Ljubljani*

Ljubljana, 2020

Prva e-izdaja.

Publikacija je v digitalni obliki prosto dostopna na

<https://e-knjige.ff.uni-lj.si/>

DOI: 10.4312/9789610603535



To delo je ponujeno pod licenco Creative Commons Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna licenca. / This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Izid knjige je podprla Javna agencija za knjigo

Katalogni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani

E-knjiga

COBISS.SI-ID=21024771

ISBN 978-961-06-0353-5 (pdf)

*Analiza jezikovnih
težav učencev:
korpusni pristop*

avtorji:

Iztok Kosem

Mojca Stritar

Sara Može

Ana Zwitter Vitez

Špela Arhar Holdt

Tadeja Rozman

Kazalo vsebine

9	Kazalo slik
10	Kazalo tabel
13	Spremna beseda
16	1 Korpusi usvajanja maternega jezika in analize napak – neizkoriščen potencial
16	1.1 Uvod
17	1.2 Korpusi usvajanja jezika
18	1.2.1 Korpusi usvajanja maternega jezika
20	1.2.2 Korpusi usvajanja tujega jezika
21	1.3 Lastnosti korpusov usvajanja maternega jezika
21	1.3.1 Velikost
21	1.3.2 Uravnoveženost
22	1.3.3 Prenosnik
22	1.3.4 Tvorci
22	1.3.5 Vrsta besedil in tema
23	1.3.6 Okoliščine zbiranja
24	1.3.7 Označenost
24	1.4 Korpusi usvajanja maternega jezika kot vir novih spoznanj o jeziku
26	1.5 Sklep
28	2 Korpus Šolar in njegova priprava za potrebe analize
28	2.1 Uvod
28	2.2 Korpus Šolar
29	2.2.1 Besedila po regijah in območjih
30	2.2.2 Besedila po šolah in razredih
31	2.2.3 Besedila po predmetih in tipih besedil
31	2.2.4 Karakteristike dela korpusa Šolar z označenimi napakami
33	2.3 Učiteljski popravki
35	2.3.1 Tipi napak
35	2.3.1.1 Napake zapisa
36	2.3.1.2 Napake besedišča
36	2.3.1.3 Napake oblike
36	2.3.1.4 Napake skladnje
37	2.3.2 Odprta vprašanja označevanja napak v Šolarju
38	2.4 Izbira orodja za analizo
42	2.5 Napake in jezikovne težave
43	2.6 Prikazovanje napak in popravkov v zgledih pri rezultatih analize
44	2.7. Sklep
45	3 Napake zapisa
45	3.1 Uvod

45	3.2	Napake ločil
48	3.3	Napake male in velike začetnice
49	3.4	Napake pisanja skupaj in narazen
51	3.5	Napake črkovanja
52	3.5.1	Črkovanje in besedotvorje
54	3.5.2	Črkovanje in oblikoslovje
55	3.5.3	Črkovanje in besedoslovje
55	3.6	Sklep
57	4	Napake besedišča
57	4.1	Uvod
58	4.2	Rezultati
59	4.2.1	Neustrezno izbrane polnopomenske besede in vsebinske napake
60	4.2.2	Slogovni popravki
61	4.2.3	Napake zaimkov
62	4.2.4	Napake predlogov
64	4.2.5	Napake veznikov
65	4.2.6	Napake glagolov
65	4.2.7	Napake pridevnikov
66	4.2.8	Napake prislovov in členov
66	4.2.9	Napačne oznake in napake pravopisa
67	4.3	Sklep
68	5	Napake oblike
68	5.1	Uvod
68	5.2	Obravnavane napake
71	5.2.1	Nedoločnik ali namenilnik
71	5.2.2	Ujemanje samostalnika
72	5.2.3	Ujemanje povedka z osebkom
73	5.2.4	Glagol v dvojini
73	5.2.5	Samostalniški zaimki v rodilniku
73	5.2.6	Samostalniški zaimki v mestniku ali dajalniku
74	5.2.7	Sklanjanje samostalnika <i>mati</i>
74	5.2.8	Zahtevnejši primeri sklanjanja
75	5.2.9	Osebni zaimek v dvojini
76	5.2.10	Glagol v sedanjiku
76	5.2.11	Sklanjanje samostalnika <i>otrok</i>
77	5.2.12	Oblika pridevnika
77	5.3	Sklep
78	6	Napake skladnje
78	6.1	Uvod
79	6.2	Rezultati
80	6.2.1	Napake besednega reda
80	6.2.1.1	Zaporedje stavčnih členov
83	6.2.1.2	Napake znotraj stavčnih členov
83	6.2.1.3	Naslonski niz

84	6.2.1.4	Zaporedje stavkov
84	6.2.1.5	Povezane napake in manj relevantni popravki
85	6.2.2	Izpuščeni podatki
86	6.2.2.1	Vsebinski in slogovni izpusti
87	6.2.2.2	Izpuščeni stavčni členi v večstavčnih povedih
88	6.2.2.3	Napačni izpusti in lapsusi
89	6.2.2.4	Večja razumljivost
90	6.2.2.5	Poudarjanje
90	6.2.2.6	Oziralni in drugi odvisniki
91	6.2.2.7	Neustrezni popravki in napačne oznake
91	6.2.3	Odvečni podatki
92	6.2.3.1	Slogovni popravki
95	6.2.3.2	Odvečni podatki zaradi pomenskega podvajanja
96	6.2.3.3	Napačni odvečni podatki in lapsusi
97	6.2.3.4	Odvečni podatki v večstavčnih povedih
97	6.2.3.5	Odvečni poudarki
98	6.2.3.6	Nerelevantni popravki in napačne oznake
98	6.2.4	Napake strukture
98	6.2.4.1	Najpogostejše napake strukture
101	6.3.	Sklep
103	7	Povezane napake
103	7.1	Uvod
103	7.2	Analiza povezanih napak oblike
104	7.2.1	Iskanje novih problemov
106	7.2.2	Končni nabor
107	7.3	Sklep
108	8	Uporaba rezultatov pri pripravi pedagoške korpusne slovnice
108	8.1	Uvod
108	8.2	Pedagoška korpusna slovnica
109	8.2.1	Nekaj izbranih problemov poučevanja slovenščine
110	8.2.2	Cilj pedagoške korpusne slovnice
111	8.2.3	Struktura slovničnega poglavja
112	8.3	Pilotna slovnična poglavja
112	8.3.1	Seznam slovničnih vsebin
115	8.3.2	Seznam jezikovnih virov
115	8.4	<i>Morati ali moči?</i>
115	8.4.1	Analiza napak
117	8.4.2	Razlaga problema
117	8.4.2.1	Prvi korak
118	8.4.2.2	Drugi korak
119	8.4.2.3	Tretji korak
120	8.4.3	Za navdušene
121	8.4.4	Povzetek, Vaje in <i>Goriška, pozor</i>
122	8.5	Sklep

123	Povzetek
126	Summary
129	Literatura
131	Spletni viri
132	Imensko kazalo

Kazalo slik

- 29 Slika 1: Delež besedil in besed po regijah
- 39 Slika 2: Vtipkanje večznakovnega poimenovanja kategorije v orodju WordSmith Tools
- 39 Slika 3: Brisanje nerelevantnih konkordanc v orodju WordSmith Tools
- 69 Slika 4: Napake paradigme in konteksta
- 80 Slika 5: Število primerov in kategorij napak skladnje
- 116 Slika 6: Število različnih vrst jezikovnih napak, povezanih z glagoloma *moči* in *morati*
- 117 Slika 7: Primerjava števila napak pri sedanjskih in deležniških glagolskih oblikah glagolov *moči* in *morati*
- 118 Slika 8: Razlaga problema *moči* – *morati*, prvi korak
- 119 Slika 9: Razlaga problema *moči* – *morati*, govorjeni jezik
- 120 Slika 10: Razlaga problema *moči* – *morati*, mnemotehniki
- 121 Slika 11: Za navdušene, nezanikani glagol *moči*

Kazalo tabel

- 23 Tabela 1: Prvih 20 ključnih besed v korpusu Šolar (primerjava s korpusom FidaPLUS)
- 29 Tabela 2: Število in delež besedil ter besed po območjih JZ regije
- 30 Tabela 3: Število in delež besedil ter besed po območjih SV regije
- 30 Tabela 4: Število in delež besedil ter besed po šolskih programih
- 30 Tabela 5: Število in delež besedil ter besed po razredih/letnikih
- 31 Tabela 6: Število in delež besedil ter besed po različnih tipih besedil
- 32 Tabela 7: Število in delež besedil ter besed po območjih v podkorpusu
- 33 Tabela 8: Število in delež besedil ter besed po šolskih programih v podkorpusu
- 33 Tabela 9: Število in delež besedil ter besed po razredih/letnikih v podkorpusu
- 33 Tabela 10: Število in delež besedil ter besed glede na tip besedila v podkorpusu
- 34 Tabela 11: Klasifikacija napak v korpusu Šolar
- 35 Tabela 12: Pogostost napak v korpusu Šolar
- 40 Tabela 13: Izvleček iz seznama lem v korpusu Šolar za orodje WordSmith Tools
- 41 Tabela 14: Pregled ohranjenih in odstranjenih oznak v korpusu Šolar (razen učiteljskih popravkov napak)
- 41 Tabela 15: Skrajšane oznake napak zapisa in skladnje
- 45 Tabela 16: Napake zapisa s podtipi
- 47 Tabela 17: Jezikovni problemi, ki so bili določeni pri analizi napak ločil
- 49 Tabela 18: Jezikovni problemi, vezani na rabo velike in male začetnice
- 49 Tabela 19: Jezikovni problemi, vezani na zapis besed skupaj, narazen in z vezajem
- 51 Tabela 20: Osnovne kategorije napak črkovanja
- 53 Tabela 21: Jezikovni problemi, vezani na zapis izpeljank
- 54 Tabela 22: Jezikovni problemi, vezani na zapis sestavljenk
- 55 Tabela 23: Jezikovni problemi črkovanja, vezani na oblikoslovne teme
- 58 Tabela 24: Napake besedišča glede na besedno vrsto
- 59 Tabela 25: Najpogostejše napake besedišča
- 63 Tabela 26: Pogoste napake predlogov
- 69 Tabela 27: Napake oblike
- 80 Tabela 28: Napake besednega reda
- 85 Tabela 29: Napake izpusta

- 91 Tabela 30: Napake odvečnih podatkov
- 99 Tabela 31: Najpogostejši jezikovni problemi, vezani na rabo neustreznih skladijskih struktur
- 104 Tabela 32: Novi jezikovni problemi, ki so bili določeni pri analizi povezanih napak oblike
- 113 Tabela 33: Jezikovni problemi, izbrani za pilotska slovnična gesla

Spremna beseda

Naše izhodišče je bilo jasno – v okviru projekta *Sporazumevanje v slovenskem jeziku*¹ smo želeli analizirati jezikovne napake učencev na podlagi empiričnih podatkov in rezultate uporabiti pri izdelavi pedagoškega vira za njihovo reševanje. Analiza, ki jo predstavljamo v tej monografiji, je tako neke vrste vezni člen med dvema aktivnostma projekta, in sicer med aktivnostjo *Nova didaktika poučevanja slovenskega jezika* (Rozman idr. 2010), v okviru katere je bil izdelan tudi korpus Šolar, ki je bil osnova za našo analizo, in aktivnostjo *Pedagoška korpusna slovnica* (Arhar Holdt idr. 2011).

Pomembna odločitev, ki smo jo sprejeli že med gradnjo korpusa Šolar, je bila, da bomo kot osnovo za analizo vzeli napake učencev ter njihove popravke, kot so jih zapisali učitelji. Napake učencev in učiteljski popravki so bili tako označeni med transkripcijo oz. digitalizacijo besedil, obenem pa smo vsako napako že klasificirali v splošne kategorije glede na jezikovne ravnine. To je za našo analizo pomenilo precejšen prihranek časa, saj so bile napake že združene v osnovne skupine, hkrati pa je vnaprejšnja delitev napak v kategorije olajšala delitev dela. Po drugi strani pa smo si delo načrtno otežili z odločitvijo, da znotraj obstoječih kategorij napak nismo vnaprej določili podtipov oz. podkategorij, ampak smo jih določali sproti. Pri poimenovanju podtipov smo upoštevali tako obstoječe slovnične kategorije kot specifične posamezne napake oz. jezikovne težave. To je bilo ključno, saj smo tako k posamezni napaki pristopali od spodaj navzgor oz. je bila vsaka napaka sprva specifična jezikovna težava, ki je kasneje postala splošna, če je bila pogosta, ali bila uvrščena v širšo kategorijo (skupaj s sorodnimi napakami), če je bila redka. Posledica tega je tudi občasna neizenačenost jezikovnih težav znotraj posameznih jezikovnih ravnin (npr. pri napakah oblike je tip napake glagol v dvojini precej širši od tipa napake sklanjanje samostalnika *mati*).

Analiza napak učencev in popravkov učiteljev je zahtevala dobršno mero kritičnosti, saj smo opazili, da v določenih primerih učiteljski popravki sploh niso bili potrebni ali pa so bili učitelji pri popravljanju nedosledni. Zaskrbljujoče je predvsem to, da lahko takšni popravki po nepotrebem učence zmedejo oz. vanje zasejejo dvom o lastnem jezikovnem znanju, hkrati pa odvrnejo pozornost, tako učencev kot učiteljev, od pravih jezikovnih težav. Vendar pa tu ne gre preveč kritizirati znanja oz. prakse učiteljev, saj podrobnejši pregled pokaže, da tovrstni nepotrebni oz. nedosledni popravki učiteljev velikokrat odražajo slabosti obstoječih referenčnih priročnikov, bodisi pomanjkljivosti ali neaktualnost informacij, ki jih vsebujejo, bodisi neopredeljenost do določene jezikovne rabe.

¹ Operacijo je delno financirala Evropska unija iz Evropskega socialnega sklada ter Ministrstvo za izobraževanje, znanost, kulturo in šport Republike Slovenije. Operacija se je izvajala v okviru Operativnega programa razvoja človeških virov in vseh življenjskega učenja; razvojne prioritete: razvoj človeških virov in prednostne usmeritve: izboljšanje kakovosti in učinkovitosti sistemov izobraževanja in usposabljanja 2007–2013.

Marsikatera jezikovna težava, odkrita pri naši analizi, za učitelje in jezikoslovce najbrž ne bo nič novega. Navsezadnje je opozorila oz. nasvete v zvezi z določenimi jezikovnimi težavami, ki smo jih zaznali v korpusu Šolar, moč najti v že pred desetletji objavljenih jezikovnih koticčkih, lahko pa bi šli po primere še bolj v preteklost – npr. o mestniku besede *otrok* (*pri otrokih* → *otrocih*) je pisal že Ivan Košťal leta 1935 (v Kalin Golob 2005: 53). Seveda pa je šlo v tovrstnih jezikovnih svetovanjih za težavo, ki je bila zaznana v jezikovni praksi odraslih piscev. Prednost podatkov naše analize gre iskati predvsem v tem, da imamo zdaj podatke o pogostosti posamezne jezikovne težave pri šolajoči se populaciji, zlasti v primerjavi z drugimi (sorodnimi) težavami, in vpogled v njeno razširjenost (npr. glede na regijo ali razred/letnik) ter specifično na mikro nivoju (npr. kateri skloni so zlasti problematični pri sklanjanju določene besede). Poleg tega so zdaj vsem dostopni tudi avtentični primeri jezikovne rabe, tako napačne kot pravilne, ki preprosto morajo biti del gradiv in virov, namenjenih reševanju jezikovnih težav učencev.

Ena od pomanjkljivosti analize, ki se je zavedamo, je, da ne zajema vseh napak v korpusu Šolar, saj vsa besedila v korpusu niso vsebovala popravkov. Po drugi strani tudi v besedilih, ki so vsebovala popravke, učitelji pogosto niso popravili čisto vseh učenčevih napak. Kljub temu glede na število analiziranih napak in tudi glede na celotno število ugotovljenih težav (tj. 692) lahko rečemo, da je precej verjetno, da bi analiza še ne označenih napak v korpusu v večini primerov razkrila zgolj ponovitve že odkritih težav in le redke nove težave. Prvi skupini neoznačenih napak (ponovitvam že odkritih težav) se že posvečamo pri pedagoški korpusni slovnici, saj pri izdelavi gesla za posamezen problem analiziramo vse pojavitve jezikovne rabe v Šolarju, tako pravilne kot napačne, in ne zgolj napak, ki so jih označili učitelji.

Upamo, da rezultati naše analize nudijo aktualen in svež vpogled v jezikovno rabo učencev v določenem obdobju ter opozarjajo na vsebine, ki si v okviru jezikovnega pouka zaslužijo več pozornosti, pa tudi na spremembe, ki so potrebne na področju referenčnih virov. Dela je vsekakor še veliko, tako pri nadaljnjih analizah napak učencev in njihovi kategorizaciji kot pri njihovem prenosu v pedagoško prakso. Vendar pa to z vidika potenciala, ki ga ponuja korpus Šolar, še zdaleč ni vse – tu so še primerjave med pisanjem mlajših in odraslih, kjer smo, vsaj kar se tiče napak, odkrili nekaj zanimivih podobnosti, analiza usvajanja besedišča po razredih oz. letnikih, raziskave vpliva govornega jezika na pisanje učencev in še bi lahko naštevali. Izhodiščni cilj smo torej dosegli, deloma pa smo ga tudi že presegle oz. ga še bomo.

Analize, katere rezultate predstavljamo v tej monografiji, ne bi mogli opraviti brez korpusa Šolar, zato hvala vsem, ki so sodelovali pri njegovem nastanku, zlasti učiteljem, ki so posvetili svoj dragoceni čas

za zbiranje in pošiljanje besedil, ter učencem in njihovim staršem, ki so v to privolili. Pomembno vlogo pri analizi je odigral Mihael Arčan, ki je napisal program za ločevanje konkordanc po oznakah in njihovo štetje. Zelo sta nam pomagala Miro Romih in Simon Krek iz podjetja Amebis, ki je sodelovalo pri izdelavi korpusa Šolar, pa tudi pri pripravi korpusa za prostodostopni spletni vmesnik, ki smo ga uporabljali v zaključnih fazah priprave monografije. Hvala pa tudi recenzentkama za neprecenljive komentarje. Kot vidite, nas ni bilo malo – da bi le imeli priložnost še velikokrat sodelovati!

Ljubljana, 10. december 2012

Avtorji

1 Korpusi usvajanja maternega jezika in analize napak – neizkoriščen potencial

IZTOK KOSEM

1.1 Uvod

Korpusi, elektronske zbirke besedil, so že pred časom postali uveljavljen vir za proučevanje (rabe) jezika, njihov razcvet pa sovпада z razvojem sodobnih tehnologij. Prvi korpusi so sicer nastali v 60. letih 20. stoletja, vendar se je intenzivna raba korpusov uveljavila šele v 90. letih, ko so bili mogoči shranjevanje velikih elektronskih zbirk podatkov in hitra iskanja po njih. Sledili sta prava poplava korpusnih raziskav (med vplivnejšimi deli so Sinclair 1991; Hunston in Francis 2000; Sinclair 2004; Hoey 2005) in izdelava na korpusnih podatkih temelječih virov, od slovarjev in slovníc do učnih gradiv in računalniških orodij. Dandanes se izdelujejo vse večji korpusi – če se je pred dvajsetimi leti govorilo o milijonih besed, se danes o milijardah (npr. Pomikálek, Rychlý in Kilgarriř 2009; Kilgarriř in Kosem 2012). To sicer velja bolj za pisne korpuse, medtem ko specializirani korpusi ostajajo manjši predvsem zaradi dejstva, da za raziskovanje in ugotavljanje vzorcev rabe v manjših, specializiranih delih jezika niso potrebne tolikšne količine podatkov kot za raziskovanje rabe splošnega jezika.

Korpusne raziskave so privedle do številnih novih dognanj o jeziku in ovrgle mnoga zakoreninjena prepričanja o jeziku (Leech 1992: 106). Prednost korpusnega pristopa k analizi jezika je, da omogoča opis jezika, kot se dejansko uporablja, ter posledično relativizira na korpusnih podatkih temelječa gradiva in vire za njihove uporabnike. Večina prvotnih korpusnih raziskav je bila izvedenih v okviru slovarskih projektov in se je osredotočila na opis jezika maternih govorcev. Leksikografi so začeli opisovati jezik, kot se je dejansko uporabljal. Izpostavljati so začeli tipične in pogoste pomene, vzorce rabe ipd. oz. jim namenjati več pozornosti kot redkejšim pomenom, vzorcem ipd. Prvi tovrstni projekt je bil slovar COBUILD, izdan leta 1987, ki je bil namenjen tujim govorcem angleščine in je temeljil na sedemmilijonskem korpusu The Bank of English (Sinclair 1987).² V okviru projekta COBUILD so bili izdelani tudi ostali na korpusu temelječi viri, npr. slovnica, slovar fraznih glagolov in korpusna učna gradiva.

V čas izdelave slovarja COBUILD segajo tudi začetki korpusov usvajanja jezika. Ti korpusi spadajo v neke vrste poddisciplino korpusne jezikoslovja, pri čemer je potrebno omeniti, da so tovrstni korpusi in na njih temelječe raziskave še danes povečini osredotočeni na rabo angleščine, bodisi kot maternega ali kot tujega jezika. V zadnjih 25 letih je nastalo predvsem veliko korpusov usvajanja tujega jezika (ang. learner corpora), ki so služili kot osnova za številne raziskave, katerih ugotovitve so bile uspešno prenesene v pedagoško rabo. V primerjavi z jezikom nematernih govorcev je bilo jeziku mlajših maternih govorcev, ki jih tudi lahko uvrstimo med tiste, ki se jezika učijo (Pirih Svetina 2005), z vidika korpusnih raziskav namenjeno manj pozornosti. Sicer obstaja kar nekaj korpusov mlajših maternih govorcev in na njih temelječih raziskav (gl. razdelek 1.2.1), vendar pa rezultati korpusnih študij še niso bili preneseni v pedagoški proces in izdelavo jezikovnih virov do takšne mere kot pri nematernih govorcih.

V Sloveniji so korpusi uveljavljeni že dobro desetletje, vse od izgradnje prvega referenčnega korpusa FIDA. V zadnjih letih se je število korpusov slovenščine precej povečalo, tudi na račun projekta *Sporazumevanje v slovenskem jeziku*, v okviru katerega je bil izdelan tudi korpus Šolar (Rozman idr. 2012), ki je za pričujočo monografijo najbolj relevanten. Korpus Šolar predstavlja pomemben mejnik za slovensko jezikoslovje, ki je dobilo dragocen vir za empirične raziskave pisnega jezika, kot ga uporablja šolska populacija. Namen tega poglavja je pokazati, kam v primerjavi s korpusi usvajanja jezika po svetu se umešča korpus Šolar, nakazati nekaj možnih analiz ter predstaviti potencial enega ključnih segmentov Šolarja, namreč označenih napak učencev.

Pričujoče poglavje najprej ponudi pregled korpusov usvajanja jezika, pri čemer so najprej predstavljeni korpusi usvajanja maternega jezika, ki so za to poglavje in pričujočo monografijo bolj relevantni, nato pa na kratko tudi korpusi usvajanja tujega jezika. Sledi pregled lastnosti korpusov usvajanja maternega jezika ter diskusija o njihovem pomenu za analize. Nato so predstavljene različne analize, ki so bile opravljene na tovrstnih korpusih, vključno z rezultati in ugotovitvami, pri čemer je poudarek predvsem na analizah napak v korpusih oz. njihovem pomanjkanju ter (zaenkrat slabo izkoriščenem) potencialu za izdelavo jezikovnih učnih gradiv in virov. V sklepu so povzete ključne točke o korpusih usvajanja maternega jezika, podanih pa je tudi nekaj razmišljanj o njihovem potencialu.

1.2 Korpusi usvajanja jezika

Korpusi usvajanja jezika imajo dokaj dolgo tradicijo, saj so prvi taki korpusi, tako usvajanja maternega kot tujega jezika, nastali v 80. letih 20. stoletja. Tako kot na splošno v korpusnem jezikoslovju tudi tu prednjačijo korpusi angleščine. Najbolj opazno razliko med korpusi

usvajanja maternega jezika in korpusi usvajanja tujega jezika, poleg očitne razlike med tipom govorcev in proučevanim jezikom, je opaziti na ravni prenosnika – medtem ko med korpusi usvajanja maternega jezika prevladujejo korpusi govornega jezika, večina korpusov usvajanja tujega jezika vsebuje zgolj pisna besedila.

1.2.1 Korpusi usvajanja maternega jezika

Eden najbolj znanih korpusov usvajanja maternega jezika je CHILDES (Child Language Data Exchange System),³ baza korpusov (video)posnetkov otroškega govora, ki se zbirajo že vse od leta 1981. CHILDES vsebuje 20 milijonov besed ali več kot 130 korpusov v 20 različnih maternih jezikih. Polovico korpusov predstavljajo posnetki maternih govorcev angleščine. V bazi je tudi nekaj posnetkov govora otrok z jezikovnimi težavami (npr. disleksijo), tujih govorcev in dvojezičnih otrok.

Po jezikovni raznolikosti je bazi CHILDES podobna zbirka korpusov, ki je nastala v okviru projekta EU SACODEYL, ki je bil izveden med letoma 2005 in 2008. Ta vsebuje transkribirane (video)posnetke najstniških govorcev angleščine, francoščine, nemščine, italijanščine, litvanščine, romunščine in španščine, starih med 13 in 18 let. Za vsak jezik je na voljo 20 do 25 intervjujev, ki so povprečno dolgi po deset minut. Posnetki so na voljo na spletni strani projekta (<http://www.um.es/sacodeyl>), kjer uporabniki lahko tudi iščejo po transkribiranih besedilih.

Korpus COLT (Corpus of Teenage Language),⁴ ki so ga leta 1993 izdelali na Univerzi v Bergnu, vsebuje 100 posnetkov oz. 50 ur govora 31 najstnikov iz londonskih okrožij, starih med 13 in 17 let (več o projektu gl. v Stenström, Andersen in Hasund 2002). Vseh 500.000 besed v korpusu je bilo ortografsko transkribiranih in oblikoskladenjsko označenih. Korpus je del referenčnega korpusa angleščine BNC (British National Corpus).

Govorjeni jezik otrok vsebuje tudi korpus POW (Polytechnic of Wales), ki so ga izdelali med 1978 in 1984 v južnem Walesu. Korpus vsebuje 65.000 besed, posnetih pa je bilo približno 120 otrok, starih med 6 in 12 let.

Od pisnih korpusov usvajanja angleščine kot maternega jezika sta poznana predvsem korpus LUCY in korpus LOCNESS. LUCY je bil izdelan leta 2003 in je sestavljen iz treh podkorpusov – korpusa besedil objavljenih avtorjev (za pričujoči pregled ni relevanten), korpusa besedil „mlajših odraslih“ in korpusa besedil otrok. Korpus mlajših odraslih sestavljajo besedila gradiv za maturo, seminarske naloge in eseji študentov prvega letnika – skupaj 48 besedil oz. 33.000 besed. V korpusu otrok je 150 besedil oz. 30.000 besed, avtorji besedil pa so otroci, stari med 9 in 12 let.

Korpus LOCNESS je bil izdelan na Univerzi v Louvainu in vsebuje pisne izdelke (argumentativne in literarne eseje) dijakov in študentov, maternih govorcev angleščine. Korpus vsebuje 324.304 besede, od tega 60.209 besed predstavljajo eseji britanskih dijakov na maturi, 95.695 besed eseji britanskih študentov, 168.400 besed pa eseji ameriških študentov. Starost avtorjev besedil v korpusu LOCNESS je primerljiva s starostjo avtorjev besedil v korpusu usvajanja angleščine kot tujega jezika ICLE (International Corpus of Learner English) (Granger idr. 2009), kar omogoča primerjave pisanja nematernih govorcev na univerzah s pisanjem maternih govorcev.

Od „neangleških“ korpusov usvajanja maternega jezika velja omeniti korpus jezika tajvanskih otrok (TCLC). Gre za korpus govornice tajvanščine, ki je bil izdelan v obdobju med 1997 in 2000 in vsebuje 300 ur posnetkov oz. 1,6 milijona besed.

Od ostalih korpusov je treba omeniti še korpuse akademskega jezika, kot so MICASE (michiganski korpus govornice akademske ameriške angleščine),⁵ MICUSP (michiganski korpus dobro ocenjenih študentskih pisnih izdelkov),⁶ BASE (korpus govornice akademske britanske angleščine)⁷ in BAWE (korpus pisne akademske britanske angleščine).⁸ Z vidika kategorizacije so ti korpusi problematični, saj vsebujejo pisanje oz. govor tako maternih kot nematernih govorcev (angleščine). Dejansko bi jih lahko uvrstili tudi med korpuse usvajanja tujega jezika, saj raziskave npr. za akademsko angleščino (Römer 2009a) kažejo, da ni bistvenih razlik med akademskim jezikom, kot ga uporabljajo nematerni govorniki in materni govorniki pri študiju.

Za slovenski prostor je od tujih korpusov usvajanja maternega jezika najrelevantnejši korpus Chyby, polmilijonski korpus češčine, ki vsebuje besedila čeških univerzitetnih študentov (Bušta idr. 2009). Besedila v korpusu so dveh tipov: eseji in uvodi diplomskih nalog. Povprečna dolžina besedil je od 600 do 700 besed. Korpus Chyby je eden redkih korpusov usvajanja maternega jezika, ki ima označene napake tvorcev besedil. Te temeljijo na popravkih učiteljev, klasifikacija napak, ki je bila izdelana vnaprej, pa je zaradi predpostavke, da tuji in materni govorniki delajo podobne napake, podobna tistim v korpusih usvajanja češčine kot tujega jezika. Podobnost napak tujih in maternih govorcev bi v tem primeru lahko pojasnili tudi z zgoraj omenjeno specifično akademsko jezikom, vendar lahko predpostavljamo, da vsaka klasifikacija napak, izdelana za jezik tujih govorcev, vsebuje določene kategorije, ki jih je moč prenesti na označevanje jezika domačih govorcev.

Prvi korpus usvajanja slovenščine kot maternega jezika, korpus Šolar (Rozman idr. 2012), je nastal pred nedavnim, leta 2010. Velik je skoraj milijon besed in vsebuje pisne izdelke šolske populacije (osnovnošolcev zadnje triade in srednješolcev) ter je po velikosti in vsebini povsem primerljiv s korpusi, opisanimi v tem razdelku. V korpusu so

5 <http://quod.lib.umich.edu/m/micase/>

6 <http://micusp.elicorpora.info/>

7 <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/>

8 <http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/>

zabeleženi učiteljski popravki napak učencev, katerih klasifikacija, podobno kot pri korpusu Chyby, temelji na klasifikaciji, uporabljeni za označevanje napak v korpusu usvajanja slovenščine kot tujega jezika PiKUST (gl. razdelek 1.2.2 in 2. poglavje). Korpus Šolar je predstavljal osnovo za našo analizo in je podrobneje predstavljen v 2. poglavju.

1.2.2 Korpusi usvajanja tujega jezika

Korpusi usvajanja tujega jezika so bili za namene analize korpusa Šolar pomembni zaradi njihovega vpliva na klasifikacijo napak v Šolarju, jih je pa veliko, zato tu ne navajamo vseh, temveč le najrelevantnejše. Podrobnejši pregled korpusov usvajanja tujega jezika najdemo v Stritar (2012).

Pri korpusih usvajanja tujega jezika še bolj kot pri korpusih usvajanja maternega jezika prednjačijo korpusi angleščine. Med najbolj znanimi in tudi največjimi sta korpusa založb Cambridge (Cambridge Learner Corpus – 35 milijon besed)⁹ in Longman (Longman Learners' Corpus – 10 milijonov besed),¹⁰ ki vsebujeta besedila izpitov iz angleščine, ki so jih pisali materni govorniki različnih jezikov. V korpusu Cambridge so npr. izpiti govorcev 130 različnih maternih jezikov iz več kot 190 držav. Poleg omenjenih korpusov, ki sta zaradi rabe v komercialne namene dostopna le sodelavcem založb, je uveljavljen tudi korpus ICLE (International Corpus of Learner English; Granger idr. 2009), ki vsebuje argumentativne eseje tujih govorcev angleščine. Ta vsebuje 16 različnih podkorpusov glede na materni jezik avtorjev besedil, ki skupaj vsebujejo 3,7 milijona besed.

Pomembna lastnost vseh treh zgoraj omenjenih korpusov so označene jezikovne napake v besedilih. Sistemi kod za označevanje napak v korpusih se resda razlikujejo, a so bili vsi izdelani z namenom detekcije jezikovnih težav avtorjev besedil in izdelave virov, vsebin in učnih gradiv za pomoč pri njihovem odpravljanju.

Večina obstoječih korpusov slovenščine kot tujega jezika je bila, kot ugotavlja Stritar (2012), zbrana priložnostno, za potrebe specifičnih analiz in vsebuje relativno malo besedil oz. posnetkov (npr. Rozman 2006; Briški 2010; Kure 2010). Obsežnejši je korpus usvajanja slovenščine (KUS), ki je bil izdelan v okviru projekta Sporazumevanje v slovenskem jeziku in vsebuje 306 besedil oz. 32.117 besed (Rozman, Stritar 2010). Besedila sestavljajo testi s tečajev slovenščine in uvrstitveni testi. Podobno velik je korpus PiKUST (Poskusni korpus usvajanja slovenščine kot tujega jezika), ki vsebuje 34.873 besed in 128 pisnih besedil 119 govorcev (za več gl. Stritar 2012). Posebnost korpusa PiKUST so označene napake, ki temeljijo na skrbno izdelani klasifikaciji napak. Klasifikacija razvršča napake v dveh stopnjah: na prvi stopnji jih deli v štiri splošnejše kategorije (zapis, beseda, oblika in skladnja), na drugi stopnji pa nadalje deli splošnejše kategorije

v podkategorije. Klasifikacija je bila z manjšimi spremembami, potrebnimi za jezik maternih govorcev slovenščine, uporabljena tudi pri označevanju napak v korpusu Šolar (gl. 2. poglavje).

1.3 Lastnosti korpusov usvajanja maternega jezika

Vsaka korpusna analiza mora pri interpretaciji rezultatov upoštevati lastnosti korpusa, kot so velikost, uravnoteženost, tipi besedil ipd. Pravzaprav lahko lastnosti korpusa že na samem začetku onemogočijo določene vrste analiz, npr. zelo majhni korpusi niso primerni za analize kolokacij. V tem razdelku so predstavljene tiste lastnosti korpusov, ki so relevantne za analize korpusov usvajanja maternega jezika.

1.3.1 Velikost

Velikost korpusa je z vidika analize zelo pomembna, saj narekuje izbiro metode in morebitno posploševanje izsledkov. Izdelovalci korpusov težijo k čim večji velikosti, vendar to velja bolj za referenčne kot specializirane korpuse. Korpusi usvajanja maternega jezika so praviloma manjši (obsegajo do milijon besed), saj so besedila in njihovi tvorca težje dostopni tako zaradi medija (pisna besedila so redko na voljo v elektronski obliki) kot pridobivanja (npr. pri govorcih, mlajših od 18 let, je potrebno privoljenje staršev).

Zaradi relativne majhnosti so korpusi usvajanja maternega jezika posledično primernejši za analize zapisa, oblike in besedišča kot za analize kolokacij, frazeologije in skladnje. Slednje so seveda vseeno izvedljive in zanimive, vendar so njihovi rezultati bolj relevantni na ravni posameznikove jezikovne zmožnosti ali na ravni pogostejših jezikovnih pojavov.

1.3.2 Uravnoteženost

Za kakršne koli splošnejše analize korpusnih podatkov je pomembna uravnoteženost korpusa. O uravnoteženosti korpusa govorimo, če vsebuje približno enako besed ali besedil, glede na lastnosti besedil oz. njihovih tvorcev. Pri korpusih usvajanja maternega jezika so relevantne lastnosti tvorcev npr. starost, spol, raven izobrazbe in regija. Uravnoteženost se ponavadi doseže s skrbnim izbiranjem besedil za vključitev v korpus (Sinclair 2005: 8-9), kar pogosto pomeni, da morajo imeti njegovi izdelovalci na razpolago več besedil, kot jih na koncu vključijo. Zaradi težje dostopnosti besedil izdelovalci korpusov usvajanja maternega jezika takšne izbire ponavadi nimajo in morajo vključiti vsa zbrana besedila, tako da je uravnoteženost v takšnih korpusih redko dosežena.

Pri korpusih mlajših govorcev, zlasti pisnih, je pomembno tudi vprašanje, ali sploh težiti k uravnoveženosti po besedah, saj moramo upoštevati, da se količina produkcije jezika pri mlajših govorcih tako zaradi spontanega usvajanja kot formalnega učenja jezika v šoli iz leta v leto povečuje. To navsezadnje do neke mere potrjujejo tudi povprečne dolžine besedil po razredih oz. letnikih v korpusu Šolar: 6. razred: 310 besed na besedilo; 7. razred: 246; 8. razred: 279; 9. razred: 220; 1. letnik: 343; 2. letnik: 320; 3. letnik: 370; 4. letnik: 452; in 5. letnik: 478. Glede na to lahko rečemo, da je dejansko bolje težiti k uravnoveženosti po besedilih.

1.3.3 Prenosnik

Na velikost in uravnoveženost korpusa v veliki meri vpliva prenosnik, saj je gradnja pisnih korpusov precej manj zamudna od gradnje govornih korpusov. Kljub temu to za korpusne usvajanja materne jezika, pa tudi tujega jezika velja v manjši meri, saj je pri izdelavi pisnih korpusov usvajanja jezika pogosto potrebna transkripcija (prepis gradiva v elektronski zapis), ki je resda manj zamudna od transkripcije govora, vendar zaradi specifičnosti pisanja mlajše populacije (napake, težje čitljiva ročna pisava) vseeno vzame precej časa (Granger 2004: 124).

1.3.4 Tvorci

O tvorcih jezika, katerih besedila ali posnetki so zbrani v korpusu, je potrebno pri analizi imeti na voljo čim več podatkov, saj ti kasneje omogočajo odkrivanje vzorcev rabe jezika na ravni posameznih lastnosti tvorcev, kot so spol, starost, regija ipd. Pomemben je tudi podatek o številu vseh tvorcev besedil v korpusu, ki je relevanten predvsem z vidika reprezentativnosti za sklepanje o celotni populaciji.

1.3.5 Vrsta besedil in tema

Od vrste besedil v korpusu je odvisno, kako obsežna bo lahko jezikovna analiza. Nekateri tipi besedil, npr. eseji, so ponavadi daljši oz. vsebujejo tako stavke oz. povedi kot odstavke. Analize takšnih besedil se lahko osredotočijo tako na oblikoslovje in besedišče kot tudi na skladnjo. Po drugi strani besedila, kot so npr. testi, pogosto vsebujejo zgolj kratke stavke ali večbesedne odgovore, kar precej omeji možnosti analize.

Pri analizi je treba upoštevati tudi kategorije besedil znotraj istega tipa, saj ima vsaka kategorija svoje specifične značilnosti glede rabe jezika. Tako se zdi, da so argumentativni eseji verjetno bolj primerni za analizo rabe trpnika kot doživljajski eseji.

Tema je povezana z vrsto oz. kategorijo besedila, v največji meri pa vpliva na rabo besedišča. Za predstavo o temah, ki so tipične za korpus oz. njegove podkorpuse, lahko uporabimo metodo prepoznavne ključnih besed, tj. besed, ki imajo v korpusu veliko višjo (relativno) frekvenco (npr. na milijon besed) kot v primerljivem drugem korpusu, npr. referenčnem korpusu istega jezika. Tabela 1 prikazuje prvih 20 ključnih besed korpusa Šolar, pri čemer je bil za primerjavo kot referenčni korpus uporabljen korpus FidaPLUS (Arhar, Gorjanc 2007).¹¹ Vidimo lahko, da na seznamu prevladujejo lastna imena, in sicer imena iz literarnih del, o katerih so pisali učenci, kar je treba upoštevati pri analizi (besedišča).

11 Seznam ključnih besed je bil izdelan v orodju Sketch Engine (Kilgarriff idr. 2004) s funkcijo *Keywords* (Ključne besede) v zavihku *Seznami*.

12 *XlmeX* in *XPriimekX* sta kodi, ki sta bili uporabljeni za anonimizacijo tistih imen in priimkov v besedilih, s pomočjo katerih bi lahko prepoznali avtorja besedila (npr. ime in priimek avtorja, imena in priimki sorodnikov, prijateljev ipd.).

Tabela 1: Prvih 20 ključnih besed v korpusu Šolar (primerjava s korpusom FidaPLUS)

XlmeX ¹²	Črtomir
Kreon	Ismena
Antigona	Matiček
preveč	Kihot
Polikarp	Nežka
don	Ožbej
Bogomila	Hajmon
manj	Ofelija
XPriimekX	ekstraspekcija
Polinejk	Kandor
Volodja	prepozno
ena	buba
Lojzka	Bronja
Hamlet	Kačur
edini	Pečorin

1.3.6 Okoliščine zbiranja

Okoliščine zbiranja besedil ločimo na realne ali umetne. Realne okoliščine so tiste, pri katerih je bil zajem besedil ali posnetkov izveden v dejanskih situacijah produkcije jezika. To so lahko pisni izdelki, nastali med jezikovnim poukom, npr. eseji, testi, ali posnetki pristnih govornih aktivnosti, npr. govor med poukom, pogovarjanje s prijatelji. Pri govoru je sicer pristnost težje dosegljiva, saj so govorniki ponavadi vnaprej obveščeni o snemanju, kar lahko vpliva na spontanost njihovega govora.

Umetne okoliščine so tiste, pri katerih je situacija za zajem besedila umetno izvana in pogosto ne odraža realne situacije. Primer takšne prakse je metoda izdelave posameznih podkorpusev v korpusu ICLE, pri katerem se eseji pišejo zgolj z namenom vključitve v korpus, poleg tega pa imajo tvorniki na voljo omejen nabor tem in predpisano oz. priporočeno dolžino eseja, kar se sicer lahko pojavlja tudi v realnih okoliščinah.

Seveda je z vidika kakršnih koli jezikovnih analiz boljše, če so besedila zbrana v realnih okoliščinah, saj je tako relevantnost kakršnih

koli gradiv oz. virov, nastalih na podlagi rezultatov analiz, veliko večja. Navsezadnje to lahko rečemo tudi za pričujočo monografijo in njeno relevantnost za poučevanje slovenščine v šolah (za več gl. 8. poglavje). Po drugi strani pa tudi rezultati analiz besedil, zbranih v umetnih okoliščinah, niso nič manj relevantni, o čemer priča npr. izraba rezultatov analiz korpusa ICLE za leksikografske namene (npr. opozorila o pogostih napakah pri določenih geslih v slovarju Macmillan za tuje govorce angleščine¹³).

1.3.7 Označenost

Pri označenosti moramo ločiti dva tipa oznak: oblikoskladenjske oznake in oznake napak. Oblikoskladenjske oznake se korpusnim podatkom dodajo avtomatsko, po končani digitalizaciji vseh besedil ali posnetkov. Korpusi usvajanja maternega jezika so z vidika avtomatskega označevanja problematični zato, ker vsebujejo precej nestandardnega jezika in jezikovnih napak, ki niso v podatkovni bazi avtomatskih označevalnikov in je natančnost označevanja tega segmenta besedil posledično slabša (Granger 2004). Zaradi tega naj bi se pri analizi raziskovalci manj zanašali na avtomatsko dodeljene oznake, temveč, kot pravi Horvath (1999), bolj na kombinacijo jezikovne intuicije, ki temelji na pedagoški praksi in izkušnjah, ter jezikovnih podatkov.

Po drugi strani označevanje napak lahko poteka že med digitalizacijo besedil, kar je bilo storjeno tudi pri gradnji korpusa Šolar, ali po njej. Označevanje je ročno ali (pol)avtomatsko, pri čemer se avtomatsko lahko poiščejo predvsem napake na ravni zapisa, oblike in besedišča. Pomemben je tudi podatek, kdo identificira in označuje napake – napake v korpusih usvajanja tujega jezika (npr. ICLE, CLC) praviloma identificirajo in označujejo raziskovalci oz. zaposleni pri založbah ali na univerzah, medtem ko so napake v korpusih, kot sta Chyby in Šolar, označene na podlagi popravkov učiteljev. Pri tem je zanimiv podatek, da so v korpus Chyby popravke učiteljev vnašali kar avtorji besedil. Kot ponazarja pričujoča monografija, so označene napake ne glede na to, kdo jih označuje, zelo koristne za analize jezikovnih težav tvorcev besedil.

1.4 Korpusi usvajanja maternega jezika kot vir novih spoznanj o jeziku

Med pogostejšimi analizami, ki vključujejo korpusne usvajanja maternega jezika, so primerjave med jezikom mlajših in odraslih materinih govorcev. Andersen (1997) npr. primerja rabo pragmatičnih označevalcev v angleščini, pri čemer uporablja korpus COLT in govorjeni

podkorpus referenčnega korpusa BNC. Njegova ugotovitev, da mladi v govoru uporabljajo pragmatične označevalce podobno pogosto kot odrasli, je zanimiva, ker nasprotuje hipotezi Erikssona (1991) in U.-B. Kotsinas (1994), da jih uporabljajo veliko pogosteje od odraslih.

Pogoste so tudi analize specifičnih jezikovnih pojavov v jeziku mlajših maternih govorcev. Stenström, Andersen in Hasund (2002) so na korpusu COLT izvedli obsežno študijo različnih jezikovnih pojavov v najstniškem jeziku, od slenga, kletvic, nestandardne rabe slovnice, premega govora do značilnosti rabe jezika v različnih govornih situacijah, in med drugim ugotovili, da je raba členkov odvisna od starosti, da najstniki uporabljajo sleng in kletvice pogosteje od najstnic, da najstniki iz nižjih slojev uporabljajo bolj vulgarne izraze od najstnikov iz srednjega in višjega sloja, niso pa opazili razlik med najstniki različnih spolov ali slojev v rabi nejasnih oz. abstraktnih izrazov.

Številne raziskave otroškega govora v različnih jezikih, zlasti v povezavi z usvajanjem jezika, je spodbudila baza CHILDES, pri čemer raziskave poleg jezikoslovja segajo tudi na področja pedagogike, psihologije in defektologije.¹⁴ Precejšen del tovrstnih raziskav se ukvarja z napakami, saj kot pravi Rowland (2007: 106), je „pojasnitev napak v otroškem govoru ena ključnih komponent uspešnih teorij o usvajanju jezika“.¹⁵ Ena od ugotovitev teh raziskav je npr. ta, da se napake v otroškem govoru pri vprašanih s pomožnim glagolom *do* (slov. *biti*) in modalnimi glagoli pojavljajo pogosteje kot pri ostalih tipih vprašanj (Rowland idr. 2005).

Za razliko od korpusov usvajanja maternega jezika, ki vsebujejo govorjeni jezik, so analize napak v korpusih pisnih besedil redke, med drugim tudi zato, ker takšnih korpusov ni veliko. Izpostaviti velja analizo napak v češkem korpusu Chyby, kjer je bilo ugotovljeno, da študenti delajo največ slogovnih napak (23,25 %), sledijo jim napake ločil (21,32 %), povezane s strogimi pravili češkega pravopisa, leksikalno-semantične napake (14,09 %) in napake črkovanja (13,04 %) (Bušta idr. 2009: 118).

Omeniti velja še analize korpusov akademske angleščine (npr. Hewings, Hewings 2002; Römer 2009b), ki vsebujejo tudi pisne izdelke maternih govorcev, vendar pa tovrstne analize obravnavajo večinoma zgolj slogovne napake (tj. napake v rabi neakademskega jezika) v okviru primerjav med pisanjem študentov in akademikov.

Analize napak so dragocene tudi za izdelavo ali izboljšavo avtomatskih jezikovnih orodij, npr. črkovalnikov in slovnčnih pregledovalnikov. Češki korpus Chyby je bil v okviru širšega projekta izdelan ravno za namene izdelave orodij za avtomatsko odkrivanje in popravljanje jezikovnih napak v besedilih (Bušta idr. 2009). Za podobne jezikovnotehnoške namene je bil pri nas nedavno prvič uporabljen tudi korpus Šolar, in sicer je Holozan (2012) uporabil označene napake vejic v korpusu v primerjalni študiji natančnosti dveh slovnčnih

14 Podatki v bazi CHILDES so bili uporabljeni za namene že več kot 3000 študij (<http://talkbank.org/usage/childesbib.pdf>).

15 Angleški izvirnik: „The ability to explain the occurrence of errors in children's speech is an essential component of successful theories of language acquisition.“ (prevod: I. Kosem)

16 Če upoštevamo številčnost in raznorodnost napak, ugotovljenih pri naši analizi Šolarja, so ti pozivi dejansko utemeljeni.

17 Dober model takšne prakse je izbira z napakami označenih korpusov usvajanja tujega jezika (npr. korpusa založb Cambridge in Longman) pri izdelavi učbenikov, delovnih zvezkov in vaj za učitelje in učence angleščine kot tujega jezika. Ker so v korpusih zbrana besedila iz testov na različnih ravneh znanja jezika, na voljo pa so tudi podatki o maternem jeziku testiranca, lahko izdelovalci pripravijo tudi gradiva za reševanje jezikovnih težav ožje skupine uporabnikov, npr. uporabnikov z istim maternim jezikom.

pregledovalnikov (Besana in LanguageTool) pri odkrivanju manjkajočih in odvečnih vejic. Kot poroča Holozan (prav tam: 105), bodo rezultati uporabljeni tudi za izboljšavo slovnice pregledovalnika Besana, pri čemer bo potrebno posvetiti večjo pozornost predvsem odkrivanju odvečnih vejic.

Korpusi usvajanja maternega jezika ostajajo torej še dokaj neizkoriščeni, zlasti za namene izdelave jezikovnih virov in učnih gradiv. Potencial se skriva predvsem v analizah jezikovnih napak oz. težav mlajših maternih govorcev pri pisanju. Razloge za neuveljavljenost analiz napak v pisnih korpusih usvajanja maternega jezika gre poleg majhnega števila tovrstnih korpusov iskati tudi v predpostavki, da se napake pri maternih govorcih pripisujejo pomanjkljivemu oz. še neuvojenemu znanju, katerega pridobitev je pričakovani del jezikovnega razvoja. Takšna predpostavka predvideva, da ti materni govorniki kasneje kot odrasli ne delajo napak oz. nimajo težav pri rabi jezika, kar pa ne drži. To je bilo potrjeno tudi za mnoge jezikovne težave, ki so bile identificirane na podlagi analize korpusa Šolar in so predstavljene v pričujoči monografiji, saj smo napake, kakršne smo zasledili pri učencih, zasledili tudi pri odraslih, in sicer v referenčnem korpusu Gigafida.

O jeziku mlajših maternih govorcev slovenščine, predvsem tistih, ki obiskujejo osnovno ter srednjo šolo, obstaja bolj malo (empiričnih) raziskav, pa še te se v veliki meri nanašajo na govorni jezik (npr. Marjanovič Umek idr. 2008; Marjanovič Umek idr. 2007). Korpusne analize napak na besedilih večjega vzorca populacije lahko ponudijo boljši vpogled v jezik učečih se in podprejo (ali ovržejo) pozive po nekaterih spremembah v obstoječi praksi poučevanja slovenščine (Rozman idr. 2010; Stabej 2011; Rozman idr. 2012).¹⁶ Odkrite jezikovne težave mlajših maternih govorcev se lahko upošteva tudi pri izdelavi učnih gradiv in virov, namenjenih rabi pri jezikovnem pouku ali samostojnemu reševanju jezikovnih zagat.¹⁷ Na tak način bodo izkoriščene tudi odkrite jezikovne težave v Šolarju, saj na njih temelji jezikovni vir pedagoška korpusna slovnica (gl. 8. poglavje).

1.5 Sklep

Kot je pokazal pregled v tem poglavju, korpusi usvajanja maternega jezika niso novost v jezikoslovju, saj njihovi začetki segajo v 80. leta 20. stoletja. Zanimiv je podatek, da med korpusi usvajanja maternega jezika prevladujejo korpusi govornega jezika, kar kaže na to, da jezikoslovci zaenkrat posvečajo več pozornosti govornemu jeziku otrok in najstnikov kot pa njihovi pisni produkciji. Manjše zanimanje za pisni jezik mlajših maternih govorcev je presenetljivo, saj se usvajanje pisnega oz. standardnega jezika začne in sprva poteka predvsem v formalnem oz. šolskem okolju, kjer so za učinkovito odpravljanje številnih jezikovnih težav, s katerimi se mlajši materni

govorci srečujejo, potrebni kvalitetni, na empiričnih podatkih temelječi viri in jezikovna gradiva.

Analize napak v korpusih maternega jezika in izraba njihovih rezultatov v pedagoške namene še niso uveljavljeni, vendar pa njihov potencial nakazujejo analize napak v korpusih usvajanja tujega jezika. Prvi korak v tej smeri za slovenščino je tako analiza napak v korpusu Šolar, ki jo predstavljamo v nadaljevanju in ki ponuja prvi celosten vpogled v dejanske težave, ki jih imajo mlajši materni govorce slovenščine pri rabi standardnega oz. knjižnega pisnega jezika, hkrati pa predstavlja izhodišče za klasifikacijo oznak napak v korpusih usvajanja slovenščine kot maternega jezika ter osnovo za standardizacijo postopka za analizo napak in njihovo nadaljnjo kategorizacijo.

18 Izraz *učenec* v predstavitvi rezultatov naše analize zajema učence osnovnih šol in dijake srednjih šol.

19 V razdelkih 2.2.1–2.2.3 navajamo le nekaj podatkov o korpusu Šolar, zlasti tiste, ki so bili relevantni za našo analizo. Več o karakteristikah korpusa je opisano v Rozman idr. (2012). Od jeseni 2012 je korpus Šolar prosto dostopen preko vmesnika (www.korpus-solar.net), ki temelji na konkordančniku Sketch Engine (Kilgarriff idr. 2004).

20 Podatek o številu besed, pa tudi nekateri drugi statistični podatki v tem poglavju (npr. o številu besedil po območjih), se nekoliko razlikujejo od podatkov v Rozman idr. (2010) in Rozman idr. (2012), ker je bila v vmesnem času objavljena posodobljena verzija korpusa. V njej so bile odpravljene nekatere napake, nastale pri transkripciji besedil, beleženju metapodatkov in konverziji datotek.

2 Korpus Šolar in njegova priprava za potrebe analize

IZTOK KOSEM, TADEJA ROZMAN, MOJCA STRITAR

2.1 Uvod

Slovensko jezikoslovje je v zadnjih letih s projektom *Sporazumevanje v slovenskem jeziku* pridobilo pomembne nove korpusne vire, med drugim milijardni pisni korpus Gigafida (Logar Berginc idr. 2012), ki je nadgradnja referenčnega korpusa FidaPLUS (Arhar, Gorjanc 2007), ter korpus GOS (Verdonik, Zwitter Vitez 2011), prvi referenčni korpus govorenega jezika pri nas. Tretji pomemben korpusni vir, ki je nastal v okviru istega projekta, je korpus Šolar (za več gl. Rozman idr. 2012), katerega vsebina in označenost učiteljskih popravkov sta predstavljali osnovo za našo analizo jezikovnih težav učencev.¹⁸

Poglavje je razdeljeno na tri dele: v prvem delu predstavimo korpus Šolar, pri čemer med drugim kritično ovrednotimo njegovo vsebino z vidika potreb in morebitne interpretacije naše analize. V drugem delu se osredotočimo na učiteljske popravke jezikovnih napak učencev v Šolarju, ki so bili uporabljeni kot izhodišče za analizo. V tretjem delu predstavimo postopek izbire korpusnega orodja za analizo in pripravo korpusa Šolar za potrebe analize v izbranem korpusnem orodju. V zadnjih razdelkih namenimo nekaj besed rabi terminov “napaka” in “jezikovna težava” v naši analizi ter predstavimo različne načine, na katere so v nadaljevanju podani zgledi za opisane jezikovne težave.

2.2 Korpus Šolar

Korpus šolskih pisnih izdelkov Šolar je korpus besedil, ki so jih učenci osnovnih in srednjih šol v Republiki Sloveniji samostojno tvorili pri pouku in je bil narejen za namene raziskovanja pisne jezikovne zmožnosti šolajoče se populacije.¹⁹ Učni jezik sodelujočih šol je slovenščina, zato so glavnino besedil tvorili govorci, katerih prvi jezik je slovenščina. Šolar je sinhroni pisni korpus; besedila, ki jih vsebuje, so večinoma nastala v šolskem letu 2009/2010. Vanj je vključenih 2703 besedil, ki vsebujejo 71.443 stavkov, 20.259 odstavkov oz. 963.233 besed.²⁰ Podatek o številu besed vključuje tudi popravke učiteljev; v besedilih učencev brez upoštevanja popravkov učiteljev je 939.243 besed. To je velikost korpusa, uporabljena za izračun števila besed in deležev besed v statističnih podatkih v nadaljevanju.

2.2.1 Besedila po regijah in območjih

Šolar je delno regijsko uravnotežen, saj je delež besedil, ki so nastala v šolah na jugozahodnem delu Slovenije, in delež besedil iz šol severozahodnega dela Slovenije približno v razmerju 60 % proti 40 %. Vendar so besedila iz severozahodnega dela v povprečju krajša, zato razmerje med številom besed iz dveh delov Slovenije kaže še nekoliko večji delež jugozahoda kot pri razmerju števila besedil (Slika 1). O popolni regijski uravnoteženosti pa ne moremo govoriti tudi zato, ker deleži besedil (in besed) med območji posamezne regije niso v ustreznem razmerju (Tabeli 2 in 3).

Slika 1: Delež besedil in besed po regijah

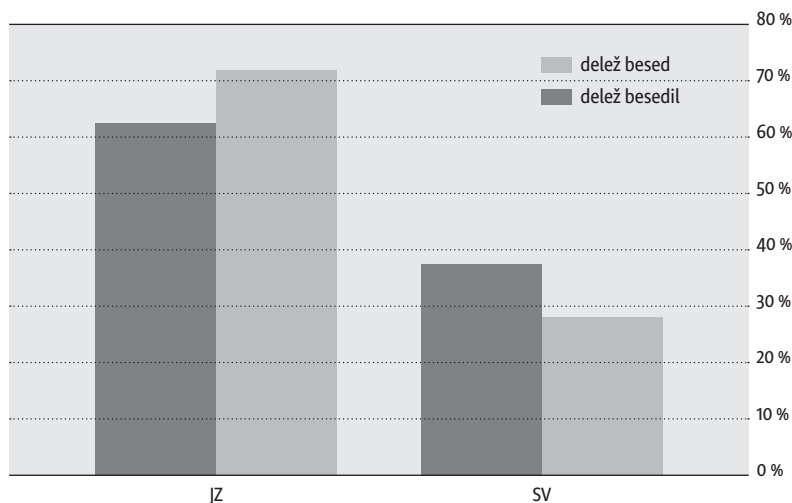


Tabela 2: Število in delež besedil ter besed po območjih JZ regije

Območje	Število besedil	Delež besedil	Število besed	Delež besed
Gorica	422	15,6 %	247.254	26,3 %
Koper	77	2,9 %	22.702	2,4 %
Kranj	12	0,4 %	585	0,1 %
Krško	148	5,5 %	41.012	4,3 %
Ljubljana	667	24,7 %	247.865	26,4 %
Novo mesto	336	12,4 %	98.577	10,5 %
Postojna	28	1 %	17.823	1,9 %
	1690	62,5 %	675.818	71,9 %

Tabela 3: Število in delež besedil ter besed po območjih SV regije

Območje	Število besedil	Delež besedil	Število besed	Delež besed
Celje	580	21,5 %	164.036	17,5 %
Maribor	254	9,4 %	70.274	7,5 %
Murska Sobota	43	1,6 %	4730	0,5 %
Slovenj Gradec	136	5%	24.385	2,6 %
	1013	37,5 %	263.425	28,1 %

2.2.2 Besedila po šolah in razredih

V korpus so vključena besedila učencev tretjega triletja osnovnih šol in srednješolcev, ki obiskujejo različne programe. Deleži besedil po šolah kažejo, da je v korpusu največ gimnazijskih besedil in najmanj besedil, ki so nastala v programih poklicnega izobraževanja (Tabela 4). Deleži besedil, pa tudi deleži besed po posameznih razredih (Tabela 5) kažejo določeno neuravnoteženost, saj je npr. precej več besedil iz srednje kot iz osnovne šole. Vendar je ta neuravnoteženost v tem primeru celo smiselna, saj odraža razliko v količini pisne produkcije učencev na različnih stopnjah šolanja.

Tabela 4: Število in delež besedil ter besed po šolskih programih

Šola	Število besedil	Delež besedil	Število besed	Delež besed
osnovna šola	503	18,6 %	127.048	13,5 %
gimnazija	1170	43,3 %	508.234	54,1 %
strokovno izobraževanje	847	31,3 %	254.396	27,1 %
poklicno izobraževanje	183	6,8 %	49.565	5,3 %

Tabela 5: Število in delež besedil ter besed po razredih/letnikih

Razred/letnik	Število besedil	Delež besedil	Število besed	Delež besed
6. razred	26	1,0 %	8076	0,9 %
7. razred	125	4,6 %	30.837	3,3 %
8. razred	177	6,5 %	49.499	5,3 %
9. razred	175	6,5 %	38.636	4,1 %
1. letnik	671	24,8 %	230.430	24,5 %
2. letnik	433	16,0 %	138.613	14,8 %
3. letnik	521	19,3 %	192.960	20,5 %
4. letnik	513	19,0 %	231.966	24,7 %
5. letnik	46	1,7 %	10.575	1,1 %
maturitetni tečaj	16	0,6 %	7651	0,8 %

2.2.3 Besedila po predmetih in tipih besedil

Šolar vsebuje besedila, ki so nastala pri 11 različnih šolskih predmetih. Njihov daleč največji delež je nastal pri pouku slovenščine (82,3 %), deleži besedil ostalih predmetov so bistveno manjši: psihologija (4,6 %), sociologija (3,7 %), zgodovina (2,7 %), domovinska in državljanska vzgoja ter etika (2,4 %), filozofija (1,6 %), geografija (1,2 %), podjetništvo (0,7 %), umetnostna vzgoja (0,5 %), ekonomija (0,2 %) in kemija (0,1 %).

Teksti so besedilnovrstno različni ter v korpusu klasificirani na eseje, pisne izdelke in teste. Največ je esejev (Tabela 6), ki so bili napisani pri različnih predmetih, večinoma v testni situaciji in so bili zato tudi ocenjeni. Pisni izdelki predstavljajo 18-odstotni delež besedil v korpusu, gre pa za besedila, ki so nastala med poukom slovenščine kot del učne ure, so besedilnovrstno različna (obnove, opisi, prošnje, opravičila ...) in praviloma niso bila napisana za oceno. Testi so razdeljeni v dve skupini: "odgovori na vprašanja" predstavljajo klasične teste z vprašanji in (esejskimi) odgovori, nastali so pri različnih predmetih, v skupino "daljše besedilo" pa sodijo različna praktična besedila (prošnje, zahvale ...), ki so jih učenci morali napisati v okviru daljšega testa iz slovenščine.

Tabela 6: Število in delež besedil ter besed po različnih tipih besedil

Tip besedila	Število besedil	Delež besedil	Število besed	Delež besed
esej/spis	1737	64,2 %	740.002	78,8 %
pisni izdelek (učna ura)	487	18,0 %	66.765	7,1 %
test (odgovori na vprašanja)	434	16,1 %	127.053	13,5 %
test (daljše besedilo)	45	1,7 %	5423	0,6 %

2.2.4 Karakteristike dela korpusa Šolar z označenimi napakami

Nekaj več kot 56 % besedil v korpusu Šolar ima označene učiteljske popravke, ki so bili narejeni za namene ocenjevanja ali pa so služili le kot povratna informacija učencem o jezikovni, besedilni in vsebinski ustreznosti njihovih besedil. Sistem označevanja napak je predstavljen v razdelku 2.3 tega poglavja, na tem mestu pa navajamo karakteristike podkorpusa besedil z označenimi napakami, saj je bila naša analiza narejena prav na tem delu korpusa.

V korpusu Šolar je 1516 besedil z označenimi učiteljskimi popravki, kar je skupaj 581.477 besed oz. 61,9 % vseh besed učencev v korpusu. Skoraj vsa besedila v podkorpusu so nastala pri pouku slovenščine, le štiri pa pri pouku domovinske ter državljanske vzgoje in etike.

Največji delež besedil z učiteljskimi popravki imajo krško (99,3 %), koprsko (96,1 %) in kranjsko območje (83,3 %), sledijo jim goriško (76,1 %), slovenjgraško (75 %), novomeško (74,1 %) in ljubljansko območje (70 %). Precej majhen delež označenih besedil imata mariborsko (36,2 %) in zlasti celjsko območje (5,5 %), medtem ko nobeno besedilo iz postojnskega območja ne vsebuje učiteljskih popravkov. Odstotek označenih besedil s posameznega območja v nekaterih primerih bistveno vpliva na deleže besedil in besed v podkorpusu v primerjavi s celotnim korpusom (Tabela 7). Tako lahko npr. opazimo zelo velik upad deleža besedil in besed s celjskega območja (z 21,5 % oz. 17,5 % v korpusu Šolar na 2,1 % oz. 1,9 % v podkorpusu) ter po drugi strani precejšen porast deležev besedil in besed z goriškega območja (s 15,6 % oz. 26,3 % na 21,2 % oz. 35,3 %). Nizka deleža vključnih besedil s celjskega in mariborskega območja pomembno vplivata tudi na regijsko razporeditev besedil in besed v podkorpusu, saj sta oba še bolj v prid jugozahodnega dela Slovenije (83,6 % vseh besedil in 89 % vseh besed).

Tabela 7: Število in delež besedil ter besed po območjih v podkorpusu

Območje	Število besedil	Delež besedil	Število besed	Delež besed
Celje	32	2,1 %	11.078	1,9 %
Gorica	321	21,2 %	205.342	35,3 %
Koper	74	4,9 %	21.421	3,7 %
Kranj	10	0,7 %	513	0,1 %
Krško	147	9,7 %	40.631	7,0 %
Ljubljana	467	30,8 %	166.292	28,6 %
Maribor	92	6,1 %	27.096	4,7 %
Murska Sobota	22	1,5 %	3222	0,6 %
Novo mesto	249	16,4 %	83.586	14,4 %
Slovenj Gradec	102	6,7 %	22.296	3,8 %

Tudi deleži besedil in besed po šolah v podkorpusu (Tabela 8) kažejo drugačno, nekoliko bolj uravnoteženo sliko od tiste za celotni korpus. Opaziti je precej manjša deleža gimnazijskih besedil in besed, po drugi strani pa porast v deležih osnovnošolskih besedil in besedil, ki so nastala v programih strokovnega in poklicnega izobraževanja. Spremembe v deležih besedil in besed po šolah vplivajo tudi na razmerja med deleži po razredih (Tabela 9): v primerjavi s celotnim korpusom sta v podkorpusu večja deleža besedil in besed v vseh razredih osnovne šole ter prvih dveh letnikih srednje šole, po drugi strani pa so manjši deleži besedil in besed iz tretjega in četrtega letnika srednje šole. Deleži za peti letnik srednje šole se niso skoraj nič spremenili, besedil z maturitetnih tečajev pa v podkorpusu sploh ni.

Tabela 8: Število in delež besedil ter besed po šolskih programih v podkorpusu

Šola	Število besedil	Delež besedil	Število besed	Delež besed
osnovna šola	395	26,1 %	110.828	19,1 %
gimnazija	404	26,6 %	239.146	41,1 %
strokovno izobraževanje	574	37,9 %	186.784	32,1 %
poklicno izobraževanje	143	9,4 %	44.719	7,7 %

Tabela 9: Število in delež besedil ter besed po razredih/letnikih v podkorpusu

Razred/letnik	Število besedil	Delež besedil	Število besed	Delež besed
6. razred	23	1,5 %	7670	1,3 %
7. razred	92	6,1 %	22.928	4,0 %
8. razred	132	8,7 %	43.460	7,5 %
9. razred	148	9,8 %	36.770	6,3 %
1. letnik	427	28,2 %	163.528	28,1 %
2. letnik	236	15,6 %	108.333	18,6 %
3. letnik	252	16,6 %	99.291	17,1 %
4. letnik	181	11,9 %	92.495	15,9 %
5. letnik	25	1,6 %	7002	1,2 %

V podkorpusu z označenimi učiteljskimi popravki je največ esejev (Tabela 10), pri čemer je njihov delež še višji od deleža v celotnem korpusu, kar je med drugim odraz tega, da je večina besedil v podkorpusu nastala pri pouku slovenščine. Precej manjši delež imajo pisni izdelki, nastali med poukom, in testi. Zlasti delež slednjih je v primerjavi z deležem v celotnem korpusu znatno nižji. Podobna slika se kaže tudi pri deležih besed, kjer je razmerje še bolj v prid esejev.

Tabela 10: Število in delež besedil ter besed glede na tip besedila v podkorpusu

Tip besedila	Število besedil	Delež besedil	Število besed	Delež besed
esej/spis	1280	84,4 %	546.779	94,0 %
pisni izdelek (učna ura)	201	13,3 %	32.089	5,5 %
test (odgovori na vprašanja)	35	2,3 %	2609	0,5 %

2.3 Učiteljski popravki

V Šolarju so zabeleženi posegi v besedila učencev, ki so jih naredili učitelji, da bi besedila ocenili ali učencem ponudili povratno informacijo o njihovi ustreznosti. Lahko gre za besedilne ali likovne komentarje: učitelji so k učenčevemu tekstu dopisali komentar, podčrtali del besedila ali narisali grafični znak. Tovrstne oznake se največkrat nanašajo na slogovno ali vsebinsko ustreznost besedila in so za našo jezikoslovno analizo manj relevantne.

Bistveno pomembnejši je drugi tip učiteljskih posegov, to so oznake napak. Načeloma je vsaka učenčeva napaka v besedilu, ki jo je prepoznal učitelj, v korpusu označena z oznako za napako. Poleg popravka, ki ga je učitelj predlagal in je v korpusu zabeležen, ima vsaka napaka dodano informacijo o tipu napake. Klasifikacija napak, opisana v nadaljevanju, temelji na jezikovnih ravninah (Tabela 11), pri čemer imajo napake zapisa in skladnje še bolj specifične podtipe. V korpusu se pojavljajo tudi dvojne napake, pri katerih je isti del besedila klasificiran z dvema tipoma napak. To je pogosto pri napakah besednega reda ali pri kombinacijah z likovnimi komentarji učiteljev.

Tabela 11: Klasifikacija napak v korpusu Šolar

Tip napake	Podtip	Primer oznake
Napaka zapisa	Črkovanje	<code><u tip="Z" podtip="CRK">bljžnjih<p>bližnjih</p></u></code>
	Pisanje skupaj/narazen	<code><u tip="Z" podtip="SN">neglede<p>ne glede</p></u></code>
	Mala/velika začetnica	<code>napadel mesto <u tip="Z" podtip="MV">tebe<p>Tebe</p></u></code>
	Ločilo	<code><u tip="Z" podtip="LOC">"<p>,</p></u>Ne da sovražim</code>
	Krajšava	<code><u tip="Z" podtip="KR">oz.<p>oziroma</p></u></code>
Številka	<code>je dolgo trajala, <u tip="Z" podtip="STEV">30<p>trideset</p></u> let,</code>	
Napaka besedišča		<code><u tip="B">vzrok<p>razlog</p></u></code>
Napaka oblike		<code>Njeni bližnji so se <u tip="O">jo<p>je</p></u> sramovali</code>
Napaka skladnje	Besedni red	<code><u tip="S" podtip="BR">pokopati simbolično<p>simbolično pokopati</p></u></code>
	Izpuščeni podatki (izpust)	<code>Oba sta bila zelo delavna, saj je Kajn delal na polju, Abel pa <u tip="S" podtip="IZP"><p>je</p></u> pasel drobnico.</code>
	Odvečni podatki (odveč)	<code>Julija mu izpove ljubezen <u tip="S" podtip="ODV">do njega<p></p></u>.</code>
	Neustrezna struktura	<code><u tip="S" podtip="STR">rek Jezusa<p>Jezusov rek</p></u></code>

Prav tako je pomembno povezovanje napak. Povezane napake so denimo zamenjava končnega ločila, ki povzroči popravek velike/male začetnice; napaka sklonске oblike, ki se nanaša na vse besede v samostalniški besedni zvezi; odvečni predlog, ki povzroči tudi spremembo sklonске oblike samostalniške besedne zveze. Povezanost napak je v Šolarju označena, uporabnik korpusa pa se mora zavedati, da nekatere izmed tovrstnih napak dejansko niso prave napake. Marsikdaj je popravek označen le zato, ker je učitelj popravil drugi del besedila.

Pogostost posameznih tipov napak, vključno s povezanimi napakami, je prikazana v Tabeli 12. Vseh označenih napak je 35.035, od tega je 61,1 % napak zapisa, 10,9 % napak besedišča, 10,3 % napak oblike in 17,7 % napak skladnje.

Tabela 12: Pogostost napak v korpusu Šolar

21 Podatki vključujejo povezane napake.

Tip napake	Podtip	Skupno število napak ²¹	Število povezanih napak
Napaka zapisa	Črkovanje	2672	18
	Pisanje skupaj/narazen	1179	
	Mala/velika začetnica	2125	872
	Ločilo	15371	775
	Krajšava	23	
	Številka	50	
	Skupaj	21420	1665
Napaka besedišča		3807	434
Napaka oblike		3618	1241
Napaka skladnje	Besedni red	1265	212
	Izpuščeni podatki (izpust)	1607	240
	Odvečni podatki (odveč)	2665	588
	Neustrezna struktura	653	61
	Skupaj	6190	1101

2.3.1 Tipi napak

2.3.1.1 Napake zapisa

Napake zapisa se nanašajo na neustrezno črkovanje, pa tudi na tipično pravopisne težave: pisanje skupaj oz. narazen, mala oz. velika začetnica, ločila, krajšave in zapis številke.

Napake črkovanja so povezane z neustreznim zapisom glasov. Med napake v pisanju skupaj oz. narazen sta nekoliko poenostavljeno vključena napačna raba oz. izpust vezaja in neustrezen zapis pik v datumu. Oba bi lahko umestili k napakam ločil, a bi bilo takšno označevanje bolj zapleteno in manj pregledno. Pri neustreznosti rabi male oz. velike začetnice je učenec besedo, ki bi morala biti pisana z malo, napisal z veliko začetnico ali obratno. Pri napakah številke gre za menjavo zapisa s številko oz. z besedo pri glavnih, vrstilnih ali drugih števnikih. Napake krajšav pa pomenijo neustrezno rabo okrajšav ali kratic namesto razvezane oblike in obratno.

Po številu napak v Šolarju so najpogostejši tip napak zapisa napačne ločila, ki se nanašajo na manjkajoča, odvečna ali neustrezna ločila. Pri njihovem označevanju je bilo treba sprejeti vrsto pragmatičnih odločitev. Mednje spada (ne)stičnost zapisa ločil, nakazana z (ne)rabo presledka. Zaradi pozornosti, ki jo temu namenjajo učitelji, pa je bila natančno zabeležena tudi oblika narekovajev, ki jih je uporabil učenec. Neustrezni, odvečni ali izpuščeni narekovaji so označeni vsak posebej in ne celotni par skupaj. Prav tako je posebna napaka ločila, ne pa del skladske napake, če je ločilo dodano ali izpuščeno skupaj

z besedo pred njim. Ker je v transkripcijah učenčevo deljenje besed označeno z zavitim oklepajem, so označeni tudi učiteljski popravki napak deljenja.

2.3.1.2 Napake besedišča

Leksikalne napake zajemajo uporabo neustrezne besede ali besedne zveze. Lahko gre za slovnično neustreznost, za besedo z napačnim pomenom, za neustrezno tvorbo besed, uporabo besede neustrezne besedne vrste, pogosti pa so tudi slogovni popravki. Napake besedišča niso členjene na podtipe, pri omejevanju rezultatov v analizi pa uporabniku korpusa, ki se želi omejiti na posamezne besedne vrste, lahko pomagajo oblikoskladenske oznake.

2.3.1.3 Napake oblike

Pri napakah oblike normi ne ustreza oblika pregibnih besed. Podtipov nimajo, saj je večkrat težko odločiti, ali je npr. neustrezna končnica pri samostalniku posledica neustreznega sklona ali števila. Tudi neobstoječe, neparadigemske oblike besed bi bilo težko umestiti med podtipe. Oznako za napako je dobila vsaka beseda posebej, tudi če je šlo za besedno zvezo – v tem primeru so bile to povezane napake.

2.3.1.4 Napake skladnje

Napake skladnje se nanašajo na večbesedne enote oz. neustrezne sintagmatske povezave med besedami. Delijo se na sledeče podtipe: besedni red, izpuščeni podatki, odvečni podatki in neustrezna struktura.

Napake besednega reda zajemajo zaporedje naslonk v naslonskem nizu, mesto naslonskega niza v povedi, zaporedje besed v samostalniški besedni zvezi ali zaporedje sestavin stavčnih členov, pa tudi napake členitve po aktualnosti. Začetek oznake napake besednega reda je pred besedo, ki mora prva zamenjati položaj, konec pa za zadnjo tako besedo, četudi je učitelj originalno označil še večji del besedila.

Kadar v besedilu manjka beseda ali besedna zveza, je to skladnjska napaka s podtipom izpust. Izpustu nasprotna težava, kadar je beseda ali besedna zveza v besedilu odveč, je označena s podtipom odvečnih podatkov. Zaradi enotnosti sistema so vključene tudi pojavitve, pri katerih je učitelj prečrtal vsebinsko neustrezen del besedila (npr. napačna dejstva, imena). To načelo velja kljub temu, da gre za drugačen tip neustreznosti, saj bi se bilo med označevanjem besedil preveč subjektivno odločati, ali gre za vsebinsko ali jezikovno neustreznost. Ročna analiza korpusa pa pokaže, v katerih primerih gre za skladnjo in v katerih za vsebino.

Podtip neustrezna struktura se nanaša na neustrezno sintagmatsko povezavo v večbesedni enoti, besedni zvezi ali celem stavku, ki presega raven leksikalne izbire. V primerih popolnoma spremenjene-ga dela besedila, ko sta prečrtani in dodani del čisto drugačna, pa ne gre za napako strukture, ampak za kombinacijo odvečnih in izpuščenih podatkov.

2.3.2 Odprta vprašanja označevanja napak v Šolarju

Označevanje napak v Šolarju seveda ni brez mest, ki jih je mogoče problematizirati. Najočitnejša med njimi je izbira tipa napak, kadar se prekrivajo. Ker mora označevanje biti enotno, da je sploh smiselno, smo morali v vseh primerih že vnaprej izbrati eno samo možnost in paziti, da so navodila upoštevali vsi označevalci.

V določenih primerih so nejasne meje med napakami črkovanja, besedišča ali oblike. Med napake besedišča smo uvrstili primere, pri katerih neustrezna črka, posebej v korenu besede, spremeni pomen besede, med napake oblike tiste, pri katerih gre za neustrezen glas v končnici besede, med napake črkovanja pa vse ostale, a bi to seveda lahko interpretirali tudi drugače. Problematizirati je mogoče tudi nedoslednost pri rabi predlogov. Če je učenec uporabil neustrezen predlog, je to napaka besedišča, kadar pa se je zmotil pri izbiri *z/s* oz. *k/h*, je to pravopisni problem neupoštevanja premene po zvonečnosti in s tem napaka črkovanja. Zamenjava dveh polnopomenskih besed različne besedne vrste je napaka besedišča, a bi lahko bila tudi napaka oblike ali celo strukture. Med napake besedišča sodijo tudi napake glagolskega vida, toda ker je glagolski vid oblikoslovna kategorija, bi jih lahko umestili tudi k napakam oblike. Za to smo se odločili predvsem zato, ker ti primeri pogosto zajemajo neustrezne glagolske pripone, pri katerih gre tako za neustrezni vid kot za druge pomenske lastnosti. Med napake besedišča smo uvrstili tudi neustrezne besedne zveze, kadar gre za napačno izbiro na leksikalni ravni, čeprav bi jih lahko kategorizirali tudi med napake strukture. Med skladenjske napake izpusta oz. odvečnih podatkov pa spada raba ločila namesto besednega veznika ali obratno, čeprav bi te primere lahko uvrstili tudi k napakam ločil.

Vprašljiva bi lahko bila tudi objektivnost označevanja, ki naj bi zgolj beležilo stanje. Temu smo se namerno odpovedali v primerih, kjer učitelj likovni ali besedilni komentar dovolj verjetno lahko interpretiramo kot popravek napake. Oznake napak namreč zaradi vsebinske klasificiranosti nudijo več enostavno dostopnih in koristnih informacij o besedilih kot likovni ali besedilni komentarji. Tako so denimo pogosto uporabljena korekturna znamenja za manjkajočo ali

22 <http://www.antlab.sci.waseda.ac.jp/software.html>

23 Možnost prikaza označenih napak in popravkov, pa tudi označevanja konkordanc, se je kasneje pojavila v orodju Sketch Engine, katerega razvijalci so izdelali poseben vmesnik za potrebe analiz korpusov usvajanja jezika. Toda ker smo takrat, ko se je ta možnost pojavila, analizirali že več kot polovico konkordanc z učiteljskimi popravki in vmesnik še ni bil testiran, se za zamenjavo orodja nismo odločili.

odvečno vejico interpretirana kot popravek napake ločila. Popravki besednega reda so v izvornih besedilih največkrat označeni s številkami, napisanimi nad besedami, ali s korekturnimi znamenji, toda v korpusu so vedno označeni kot napake besednega reda in ne kot komentarji. Podobno velja, kadar so učitelji napake, posebej pri črkovanju ali drugih pravopisnih težavah, zgolj podčrtali, niso pa navedli popravkov. Če je kljub temu nesporno, kaj so želeli popraviti, je to označeno kot napaka, ki ji je dodana popravljen oblika. Tudi v primerih, ko je učitelj popravil samo besedni red, ni pa označil spremembe velikih oz. malih začetnic, je v popravku navedena pravilna raba, saj je zelo verjetna. Če pa popravljen oblika ni enoznačna, torej če sta enako verjetna dva popravka, je označen samo likovni komentar.

Marsikatero dilemo bi lahko utemeljeno rešili drugače, vendar je treba upoštevati, da je označevanje konkretno delo, za katerega so nujne pragmatične odločitve, ki omogočajo učinkovito izvedbo. Mnoge od morebitnih pomanjkljivosti lahko uporabnik preklasificira pri sprotnem pregledovanju korpusa. Obenem je že ob preletu gradiva mogoče dokaj hitro prepoznati, katere napake učencev so nastale kot lapsusi, pri katerih pa gre za posledico učenčeve pomanjkljive jezikovne zmožnosti (James 1998) – ker bi bil tak poseg že interpretacija podatkov, namreč to v Šolarju ni ločeno. Uporabnik mora sam prepoznati tudi kontekst napake, sklepati o morebitnih razlogih za napako v sobesedilu ali oceniti ustreznost navedenega učiteljevega popravka. Še več omejitev izbranega sistema označevanja pa pomaga odpraviti sistematično ročno pregledovanje določenega tipa napak, ki je nujno za vsako njihovo analizo.

2.4 Izbira orodja za analizo

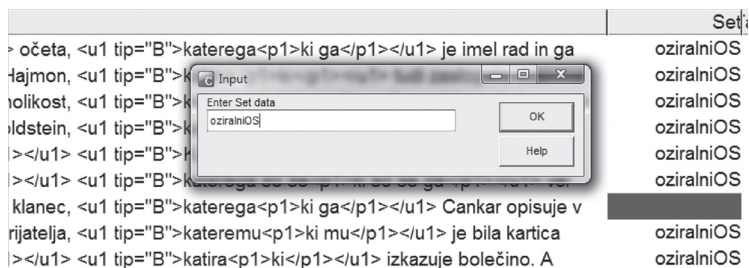
Za analizo Šolarja smo iskali korpusno orodje, ki je moralo po eni strani omogočati ogled učiteljskih popravkov, po drugi pa označevanje konkordanc s kategorijo problema. Prvi pogoj je izpolnjevalo veliko orodij, npr. AntConc,²² Sketch Engine (Kilgarriff idr. 2004), WordSmith Tools (Scott 2008), ni pa nam uspelo najti orodja, ki bi prikazovalo označene napake in popravke brez oznak XML. Drugi pogoj, namreč možnost označevanja konkordanc, se je izkazal za bolj selektivnega od prvega. Eno od redkih orodij, ki je bilo dejansko izdelano za analizo in označevanje konkordanc, je WordSmith Tools, zato smo ga izbrali tudi za svojo analizo.²³ Sestavljajo ga konkordančnik (*Concordance*), orodje za izdelovanje besednih seznamov (*Wordlist*) in orodje za izdelovanje ključnih besed (*Keywords*).

Označevanje konkordanc v konkordančniku WordSmith Tools je zelo preprosto in hitro. Če konkordanco označimo z enim znakom, zadostuje pritisk na tipko, kar pa terja izdelavo enočrkovnih ali enošteviličnih kod za različne kategorije. Pri označevanju s poimenovanji,

ki imajo več kot en znak, je postopek zamudnejši, saj je treba pri vsaki konkordanci klikniti na okence za označevanje in vtipkati poimenovanje kategorije (Slika 2).

24 Avtomatsko deljenje konkordanc po označeni kategoriji problema in njihovo štetje sta bila izvedena s programom, napisanim v programskem jeziku Perl, ki ga je pripravil Mihael Arčan.

Slika 2: Vtipkanje večznakovnega poimenovanja kategorije v orodju WordSmith Tools



Poleg označevanja WordSmith Tools omogoča naprednejše urejanje konkordanc, od sortiranja po vsebini ali oznaki konkordanc, iskanja po vsebini konkordanc in hitre povezave do ogleda posamične konkordance v korpusu in do brisanja nerelevantnih konkordanc.

Slika 3: Brisanje nerelevantnih konkordanc v orodju WordSmith Tools

N	Concordance
2	kaj spleh je narobe z njimi. Vsak človek je samo <u1 tip="B">en</u1> edinstven</p1></u1>
3	je Simona že v naprej določil, da je alkoholik. <u1 tip="B">brez</u1> ne</p1></u1>
4	tip="Z.MV">Poljski<p1>poljski</p1></u1> oficir. <u1 tip="B">kateri</u1> ki je</p1></u1> je bil brez-
5	zato, da bi ozdravili njegovo duševno bolezen. <u1 tip="B">katere</u1> ki je</p1></u1> pa-
6	<p1></u1> je zdravnik rekel v <u2 tip="B">hecu</p2></u2></u1>
7	bi utrdili svoja prepričanja in lastno samozavest. <u1 tip="B">ki</u1> kar</p1></u1> pa je Simonu-
8	uprli sistemu, ki se le poigrava z nami, nas ne <u1 tip="B">odobrava</u1> obravnava</p1></u1>
9	rekli država ponekod še vedno obnaša na tak <u1 tip="B">diskriminanten</u1> diskriminaten</p1></u1>
10	. Simon je hotel ven iz zavoda, Volodja pa mu ni <u1 tip="B">pustil</u1> dovolil</p1></u1>. <u1
11	1. Šolski spis Vodena interpretacija Odlomek iz <u1 tip="B">dramatike</u1> drame</p1></u1>
12	streznitvi. Volodja je Simona zaničeval in ni mu <u1 tip="B">pustil</u1> dovolil</p1></u1>, da
13	/p1></u1> vsi bi morali imeti <u1 tip="B">iste</u1> enake</p1></u1> pravice.
14	bolj izobražen kot Volodja in je na družbeni <u1 tip="B">ravnini</u1> lestvici</p1></u1> on-
15	V romanu nam pisatelj predstavi zgodbo vojaka. <u1 tip="B">kateri</u1> ki</p1></u1> se je trudil,
16	pa revne jugovzhodne države počnejo stvari. <u1 tip="B">katere</u1> ki</p1></u1> so bile
17	so dodali nalogo, da očisti področje polno min, <u1 tip="B">tam</u1> kjer</p1></u1> so bili tudi

WordSmith Tools ima funkcionalnosti, ki so bile obvezne za analizo, npr. podporo slovenščini oz. prikaz šumnikov. Zelo koristno pa je bilo npr. tudi iskanje z uporabo nadomestnih znakov (znak ? nadomešča en znak, znak * nadomešča kateri koli niz znakov), različni načini izvoza podatkov (v programu lastnih formatih in formatih XLS, XML ali TXT) in možnost združevanja shranjenih podatkov, resda le v formatih programa WordSmith. Za shranjevanje analiziranih podatkov smo uporabljali WordSmithov format CNC, za razdeljevanje konkordanc po posamezni kategoriji problema in njihovo preštevane pa format XML.²⁴ Funkcijo združevanja shranjenih podatkov smo pogosto

uporabili po končani analizi, ko smo združevali različne datoteke istega raziskovalca, ker so vsebovale iste kategorije. Pri napakah oblike so se denimo iste kategorije pojavljale tako v napakah paradigme kot v napakah konteksta (gl. 5. poglavje).

Korpus Šolar je na voljo v obliki XML, ki poleg označenih učiteljskih popravkov, komentarjev ipd. vsebuje tudi oblikoskladenjske oznake. Ker je imel program WordSmith Tools določene težave z branjem teh oznak oz. ni bilo mogoče pripraviti nastavitve, ki bi v konkordancah prikazovale samo nekatere oznake XML, smo za analizo pripravili korpus Šolar v formatu TXT brez oblikoskladenjskih oznak. Da bi vseeno omogočili nekatera naprednejša iskanja, npr. iskanje po lemah, smo na podlagi seznama besed v korpusu izdelali seznam lem in oblik v formatu, kot ga zahteva WordSmith Tools. Kot je razvidno iz Tabele 13, so bile pri posamezni lemi navedene le določene oblike, in sicer tiste, ki so se pojavile v Šolarju.

Tabela 13: Izvleček iz seznama lem v korpusu Šolar za orodje WordSmith Tools

amputirati → amputira
analizirati → analiziral, analiziramo, analiziraš, analizirala, analizirat, analizira, analizirati
angažirati → angažirala
angležiti → angleži
anketirati → anketirala, anketiramo
aretirati → aretirali
argumentirati → argumentiral
asimilirati → asimilirali, asimilira, asimilirati, asimilirajo
asociirati → asociira
asocirati → asocira, asocirajo, asociralo
bahati → bahajo, baha
barvati → barva, barvajo
bati → boji, bojim, boj, bali, bojita, bojiš, bala, bati, bat, bojijo, bojte, bal, bojimo
bdeti → bdi
bedeti → bedela, bedele, bedi, bedeli, bedel
begati → begale, bega, begal
beležiti → beležimo
beračiti → beračil, beračila, beračili, beračiti

Poleg oblikoskladenjskih oznak v korpusu Šolar so se po uvodnih analizah za odvečne izkazale tudi nekatere osnovne metajezikovne oznake, saj so bile za potrebe naše analize večinoma nerelevantne in pogosto celo moteče, ker so zmanjšale količino vidnega konteksta konkordanc. Na podlagi teh opažanj smo odstranili učiteljeve besedilne in likovne komentarje, delno čitljive ter nečitljive popravke oz. komentarje in učiteljske popravke, ki se nanašajo na oblikovne

značilnosti besedil (npr. na poravnavo posameznih delov besedila). Obdržali smo delno čitljive in nečitljive zapise učencev in seveda učiteljske popravke napak (Tabela 14).

Tabela 14: Pregled ohranjenih in odstranjenih oznak v korpusu Šolar (razen učiteljskih popravkov napak)

Opis	Primer oznake	Status za analizo
Učiteljev besedilni komentar	<u k="učiteljev komentar">učenčevo besedilo</u>	ODSTRANJENO
Učiteljev likovni komentar	<u l="likovni komentar">učenčevo besedilo</u>	ODSTRANJENO
Nečitljiv učiteljev popravek/komentar	Antigoni vendarle uspe<u l="ravno podčrtano" kn="">naskrivaj</u>pokopati brata.	ODSTRANJENO
Delno čitljiv učiteljev popravek/komentar (prepisano je, kar se da prebrati)	<u kn="poskusi"/>	ODSTRANJENO
Nečitljiv zapis učenca	<un="">	OHRANJENO
Delno čitljiv zapis učenca (prepisano je, kar se da prebrati)	<un="nasprot"/>	OHRANJENO

Učiteljski popravki napak, zapisani z oznakami XML, so zavzeli precejšnje število znakov in posledično zmanjšali količino prikazanega konteksta, kar je bilo še zlasti problematično pri gnezdenih popravkih ali v primerih, ko je bilo v bližini analiziranega popravka še več drugih popravkov. Zato smo oznake skrajšali, tako da smo pri kategorijah popravkov s podtipi, torej pri napakah zapisa in napakah skladnje, podtipi odpravili in njihovo vsebino vključili v atribut tip (Tabela 15).

Tabela 15: Skrajšane oznake napak zapisa in skladnje

Tip napake	Podtip	Oznaka	Skrajšana oznaka
Napaka zapisa	črkovanje	<u tip="Z" podtip="CRK">	<u tip="Z-CRK">
	pisanje skupaj/narazen	<u tip="Z" podtip="SN">	<u tip="Z-SN">
	mala/velika začetnica	<u tip="Z" podtip="MV">	<u tip="Z-MV">
	ločilo	<u tip="Z" podtip="LOC">	<u tip="Z-LOC">
	krajšava	<u tip="Z" podtip="KR">	<u tip="Z-KR">
	številka	<u tip="Z" podtip="STEV">	<u tip="Z-STEV">
Napaka skladnje	besedni red	<u tip="S" podtip="BR">	<u tip="S-BR">
	izpuščeni podatki	<u tip="S" podtip="IZP">	<u tip="S-IZP">
	odvečni podatki	<u tip="S" podtip="ODV">	<u tip="S-ODV">
	neustrezna struktura	<u tip="S" podtip="STR">	<u tip="S-STR">

Z odstranitvijo nepotrebnih metajezikovnih oznak in krajšanjem nekaterih učiteljskih popravkov smo zmanjšali število znakov v TXT verziji korpusa Šolar za več kot 500.000 (z 8.089.523 na 7.539.086

znakov),²⁵ posledično pa se je zmanjšalo tudi povprečno število znakov na stavek, s 113 na 105. Učinek procesov zmanjševanja znakov ponazarja primer stavka iz korpusa Šolar (verzija brez oblikoskladenskih oznak). Najprej je prikazan v prvotni obliki (z vsemi meta-jezikovnimi oznakami), ki ima 550 znakov (1), in potem še v obliki, ki smo jo uporabili pri analizi (torej brez nekaterih metajezikovnih oznak in s skrajšanimi oznakami učiteljskih popravkov), ki ima 362 znakov (2).

1 `<u1 k="Nikjer nisi obravnaval družbenih razmer.">Kandor<u2 tip="Z" podtip="LOC"><p2></p2></u2> glavna oseba Sencnega plesa<u2 tip="Z" podtip="LOC"><p2></p2></u2> je pravkar pokopal svojo ženo na pokopališču na <u2 tip="Z" podtip="MV"><u3 tip="Z" podtip="CRK">gorenskem</p3></u3></u2>, ki jo je zelo ljubil. <u2 k="sosledje" l="ravno podčrtano">Izgubi</u2> tudi smisel za zivljenje, ostane brez <u2 k="?" l="črtkano podčrtano">nacrto</u2>, nima smisla za zivljenje ter <u2 l="obkroženo">zanga </u2> postane vse „TUROBNI PLES V SENCI“.`

2 `Kandor<u2 tip="Z-LOC"><p2></p2></u2> glavna oseba Sencnega plesa<u2 tip="Z-LOC"><p2></p2></u2> je pravkar pokopal svojo ženo na pokopališču na <u2 tip="Z-MV"><u3 tip="Z-CRK">gorenskem</p3></u3></u2>, ki jo je zelo ljubil. Izgubi tudi smisel za zivljenje, ostane brez nacrto, nima smisla za zivljenje ter zanga postane vse „TUROBNI PLES V SENCI“.`

2.5 Napake in jezikovne težave

Pred predstavitvijo rezultatov analize velja nameniti nekaj besed opredelitvi terminov napaka in jezikovna težava oz. jezikovni problem, kot jih uporabljamo v pričujoči monografiji. Kot pravi Pirih Svetina (2005: 35), gre pri napakah za „odstopanja od pričakovanega, ustreznega in pravnega“. V analiziranem podkorpusu so kot napake označene pojavitve v besedilih učencev, ki so jih kot neustrezne označili in večinoma tudi popravili učitelji. Kot je razvidno iz nadaljnjih rezultatov analize, v korpusu najdemo precej primerov napak, ko oblike, ki so jih napisali učenci, dejansko niso slovnično neustrezne, pač pa so jih učitelji prepoznali kot slogovno nesprejemljive oz. neustrezne glede na pragmatični kontekst (prim. James 1998: 65–76). Slogovne in predvsem vsebinske napake so bile za našo jezikoslovno analizo v večini primerov nerelevantne, tudi pri analizi bolj slovnično utemeljenih napak v Šolarju pa je nujno upoštevati vlogo učiteljevega razumevanja napak in jezikovnih ustreznosti ter večkrat tudi njegovih osebnih preferenc. O jezikovni težavi oz. problemu pa govorimo takrat, ko se določena napaka, označena v korpusu, dovolj pogosto pojavlja, da je upravičeno sklepanje, da je značilna tudi splošneje. Jezikovne težave so lahko individualne ali splošne. Individualne se

pojavnjajo redko, v majhnem številu besedil, in so značilne le za posamezne učence. Po drugi strani so splošne jezikovne težave pogoste in jih zasledimo v več besedilih. Za jezikoslovno analizo so precej zanimivejše, saj zaradi sestave korpusa Šolar oz. analiziranega podkorpusa lahko govorimo o težavah večjega dela slovenske šolske populacije.

Poleg pogostosti je pri odločanju, kdaj je jezikovna težava splošna težava šolske populacije, potrebno upoštevati tudi način kategorizacije težav. Pri analizi so bili namreč tipi napak iz klasifikacije, predstavljene v Tabeli 12, nadalje drobljeni v kategorije. Te so v nekaterih primerih precej široke in pogoste, razdelili pa bi jih lahko v še manjše kategorije (npr. raba vejice pred priredji in odvisniki; gl. razdelek 3.2 v 3. poglavju). V drugih primerih pa so zelo specifične (npr. sklanjanje samostalnika *otrok*; gl. razdelek 5.2.11 v 5. poglavju) in posledično v večini primerov manj pogoste, kar pa ne pomeni, da so manj relevantne. Zlasti pri bolj specifičnih kategorijah je namreč poleg pogostosti treba upoštevati delež napak od vseh pojavitev analizirane jezikovne kategorije; npr. v podkorpusu z označenimi napakami je le 22 pojavitev napačne rabe nikalnega zaimka *noben* namesto *nihče*, vendar pa to predstavlja 4,5 % vseh pojavitev zaimka *nihče* v korpusu Šolar. Obstajajo tudi primeri (specifičnih) kategorij, pri katerih je za analizo pomembnejši delež določenega segmenta jezikovne rabe znotraj kategorije, v katerem prevladuje jezikovna težava; npr. pri sklanjanju samostalnika *otrok* je največ napak (30 pojavitev) zaslediti v rabi orodnika množine (*otroci* namesto *otroki*), kar predstavlja nekaj več kot 41 % primerov rabe samostalnika *otrok* v omenjenem sklonu v celotnem korpusu. Takšne ugotovitve so zlasti relevantne za pripravo rešitev za identificirane jezikovne težave.

2.6 Prikazovanje napak in popravkov v zgledih pri rezultatih analize

Rezultate analize napak učencev smo skušali v čim večjem obsegu ponazoriti z zgledi iz korpusa Šolar. V vsakem zgledu je v krepkem tisku samo tista napaka, ki jo zgled ponazarja, napaki sledi učiteljev popravek, ki je tudi v krepkem tisku, uporabili pa smo še naslednja grafična znamenja:

- a) Za napako je puščica, ki ji sledi popravek. Ta prikaz je uporabljen največkrat, saj je večina napak takšnih, da učenec nekaj napiše narobe, učitelj pa to nadomesti s pravilnim besedilom:

Od → Iz prejšnjega zakona pa mu je ostala samo hčerkica.

- b) Prečrtana beseda ali ločilo. Ta prikaz se uporablja za zglede, v katerih je učitelj zgolj prečrtal učenčevo besedilo. Gre za zglede, ki jih najdemo predvsem v napakah odvečnih podatkov (tip napak skladnje) in v napakah zapisa:

26 Npr. pri analizi rabe besedišča bodo zaradi velikega števila besedil, ki so nastala pri pouku slovenščine, rezultati relevantnejši, če bomo analizirali besedišče pri pouku slovenščine ali splošnejše besedišče, kot če nas bo zanimalo besedišče pri drugih predmetih.

Likarju je bilo dolgčas, zato je od Tončka zahteval **en** gumb.

Današnja mladina, na drugačne vrstnike gleda zelo različno.

- c) Oglati oklepaj. Ta prikaz se uporablja za zglede, v katerih je učitelj dodal manjkajoče besedilo ali ločilo. Te zglede najdemo v napakah izpusta (tip napak skladnje) in v manjši meri tudi v napakah zapisa:

Oče ga je sedaj vzel [**s**] sabo v Ljubljano,

- d) Ločena zgled z napako in zgled s popravkom. Ta prikaz se uporablja v primerih, ko je uporaba enega od zgornjih načinov nepregledna, npr. ko so popravljene različni deli povedi, ki niso v neposredni bližini:

Pri tem pa čez nekaj časa **zboli**.

Popravljen poved: Pri tem pa **je** čez nekaj časa **zbolel**.

V zgledih so pogosto še druge napake, ki so jih učitelji popravili, a jih zaradi preglednosti ne izpostavljamo. Podobno z namenom izboljšanja preglednosti marsikateri zgled ni podan v celoti, predvsem zaradi dolžine in/ali prevelikega števila napak.

2.7 Sklep

Korpus Šolar prinaša v slovenski prostor pomembno novost, in sicer empirične podatke o pisnem jeziku šolske populacije, tako o njegovi rabi kot o jezikovnih težavah, s katerimi se učenci srečujejo. Po velikosti in vsebini je povsem primerljiv s tujimi korpusi usvajanja jezika, kar obenem pomeni, da omogoča različne vrste analiz jezika šolske populacije, od analiz rabe besedišča in pogostih vzorcev do analiz napak oz. jezikovnih težav. Seveda pa je pri vsakršnih analizah treba upoštevati lastnosti besedil v korpusu.²⁶ Kot je opisano v razdelku 2.4 tega poglavja, tip analize narekuje tudi izbiro korpusnega orodja in pripravo podatkov v korpusu.

Analiza jezikovnih težav učencev, ki jo opisujemo v tej monografiji, je bila logična izbira za prvo analizo Šolarja, ne samo zaradi potreb projektne aktivnosti Pedagoška korpusna slovnica (gl. 8. poglavje), temveč tudi zaradi učiteljskih popravkov v korpusu, ki predstavljajo preliminarno, splošnejšo analizo pisanja učencev z vidika učiteljev. Učiteljski popravki so bili nadvse dragoceni tudi zaradi prihranka časa, ki bi ga sicer potrebovali za prepoznavanje in označevanje napak. Poglavja, ki sledijo, pa prikazujejo, kako kompleksen je bil kljub temu postopek identifikacije jezikovnih težav učencev, saj je bilo treba učiteljske popravke najprej močno razdrobiti na konkretne jezikovne težave in jih nato kategorizirati.

3 Napake zapisa

SARA MOŽE

3.1 Uvod

Pri analizi korpusa Šolar se je pokazalo, da imajo slovenski učenci zdaleč največ težav s pravopisnimi vsebinami. Napake zapisa, h katerim prištevamo neustrezno črkovanje, postavljanje ločil, velikih in malih začetnic, rabo krajšav, zapis številkk ter pisanje skupaj in narazen (Rozman idr. 2012: 30–33), je v korpusu precej pogostejše zaslediti kot ostale vrste napak; medtem ko se za ostale tri kategorije delež giblje med 10 in 17 %, se napake zapisa pojavljajo v več kot 61 % vseh identificiranih napak. V Tabeli 16 so predstavljeni podatki za vsakega od šestih podtipov napak zapisa. Med podtipi zlasti izstopajo napake ločil (skoraj 72 % vseh napak zapisa), pri čemer jih je večina vezanih na rabo vejice.

27 Napake, vezane na zapis številkk in krajšav, se nanašajo na izključno slogovne učiteljske popravke, npr. na zapis z arabsko številko namesto z besedo ter rabo okrajšav in kratic v šolskih spisih, in so marginalne, zato v tem poglavju niso posebej predstavljene.

Tabela 16: Napake zapisa s podtipi

Št.	Podtip napake	Primer	Število napak
1	postavljanje ločil	<i>Knjiga mi je bila zelo všeč[,] saj je zelo napeta in te pritegne k branju.</i>	15371
2	postavljanje velike in male začetnice	<i>Ožbej in meta → Meta sta se zaljubila</i>	2125
3	pisanje skupaj in narazen	<i>Nisi ji hotel povedati, da je nebi → ne bi prizadel.</i>	1179
4	črkovanje	<i>Kmalu za tem se je stemnilo in odšli smo domol → domov.</i>	2672
5	zapis številkk	<i>Prebrala sem ga, ko mi je bilo 9 → devet let.</i>	50
6	zapis krajšav	<i>Torej gre za bratsko oz. → oziroma sestrsko ljubezen.</i>	23
SKUPAJ			21420

V nadaljevanju je predstavljen postopek identifikacije potencialno relevantnih jezikovnih problemov, ki so bili določeni pri analizi prvih štirih podtipov, tj. napak ločil (razdelek 3.2), male in velike začetnice (razdelek 3.3), pisanja skupaj in narazen (razdelek 3.4) ter črkovanja (razdelek 3.5).²⁷

3.2 Napake ločil

Napake ločil so široko zastavljena podkategorija, ki pokriva vsa „manjkajoča, odvečna ali neustrezna ločila“ (Rozman idr. 2012: 31), torej ne le vejice, ki se je pri analizi zgledov iz korpusa izkazala za posebej problematično, temveč tudi vprašaja, klicaja, pike, dvopičja,

tropičja, pomišljaja, oklepajev in narekovajev. Med napake ločil spada tudi nestičnost zapisa, ki bi jo sicer lahko uvrstili tudi med napake pisanja skupaj in narazen, npr.

3 TAKO SE SKUPINA **odloči**, da bo Ketii pripeljala Jurčka do starega rudnika

Iz korpusa je bilo na podlagi učiteljskih popravkov izločenih kar 15.371 napak ločil. Za podrobno analizo je tovrstna količina podatkov težko obvladljiva, zato se je zdelo smiselno natančneje obdelati le manjši vzorec vseh napak zapisa. V ta namen je bilo iz korpusa Šolar najprej izločenih in analiziranih naključnih 1000 napak zapisa, na podlagi katerih je bilo identificiranih dvanajst potencialno relevantnih problemov. Pri analizi se je pokazalo, da je z vidika jezikovnih zagat učencev relevantna le problematika, vezana na napačno rabo vejice; napake v rabi ostalih ločil so bile uvrščene v kategorijo „nerelevantno“ in jih nadalje nismo analizirali. Zaradi močno asimetričnega razmerja med manjkajočimi in odvečnimi vejicami v vzorcu je bilo nato še dodatno izločenih in analiziranih 500 napak odvečnih vejic. Analiza se je sicer izkazala za koristno pri iskanju boljših reprezentativnih zgledov za že identificirane probleme, vendar pa ni prinesla večjih novosti v smislu detekcije novih jezikovnih problemov.

Tabela 17 prikazuje končne rezultate analize napak ločil. Od skupno 1500 analiziranih zgledov se jih je 1380 (92 %) izkazalo za potencialno relevantne, 120 (8 %) pa za popolnoma nerelevantne. Učenci imajo zdaleč največ težav pri rabi vejice med priredji in odvisniki (nekaj več kot 58 % analiziranih zgledov) ter stavčnimi členi (16 %), precej poglavic jim povzročajo tudi večbesedni kontektorji (8 %) – rezultati namreč kažejo, da pogosto ne vedo, kje stoji vejica. Raba vejice je razmeroma problematična tudi pri primerjavah in prislovnih določilih načina, ki jih uvaja veznik *kot*, pristavkih, pastavkih, vrivkih in priredno zloženih besedah oz. besednih zvezah, pogosto jo pozabljajo za odvisniki, vrinjenimi sredi povedi, in jo po nepotrebnem dodajajo v priredno zloženih odvisnikih. Raba vejice ob stavkih z elipso povedka, v pridevniškem nizu, strukturah z *zvezo še posebej* in pred polstavki je po podatkih iz korpusa manj problematična, vendar je treba v tem primeru upoštevati tudi specifičnost identificiranih problemov – medtem ko se problemi z dna tabele nanašajo na specifične jezikovne strukture, so problemi na vrhu zastavljeni kot širše kategorije, ki bi jih bilo mogoče še naprej drobiti (npr. zglede iz kategorije *vejica pred priredji in odvisniki* bi bilo mogoče dodatno razdeliti glede na razmerje med priredjema oz. med priredjem in odvisnikom – časovno, vzročno, dopustno ipd.).

Tabela 17: Jezikovni problemi, ki so bili določeni pri analizi napak ločil

Št.	Jezikovni problem	Primer	Število napak
1	vejica pred prirednji in odvisniki	<i>Knjiga mi je bila zelo všeč[.] saj je zelo napeta in te pritegne k branju.</i>	803
2	vejica med stavčnimi členi	<i>Današnja mladina; na drugačne vrstnike gleda zelo različno.</i>	229
3	vejica pri večbesednih konektorjih	<i>Žena je očeta čakala[.] ne; da bi izgubila upanje.</i>	110
4	vejicah pri primerjavah in prislovnih določilih načina s kot	<i>Morala in ljubezen do brata sta bili zanjo pomembnejši; kot njeno lastno življenje.</i>	51
5	vejica pri pristikah, pastavkih in vrivkih	<i>Karničnik[.] Ožbejev očē[.] je trd, krut, neizprosen brez čustev.</i>	46
6	vejica pri priredno zloženih besedah in besednih zvezah	<i>Ni bil urejen[.] temveč smrdljiv (večinoma zaradi pijanosti), nemaren, potolčen.</i>	44
7	vejica za odvisnikom, ki je vrinjen sredi povedi	<i>Čez teden dni, ko je Fatima prihajala iz šole[.] jo je že čakal Adam s svojim lepim obrazom in avtom.</i>	42
8	vejica v priredno zloženih odvisnikih	<i>Moje mnenje je, da smo vsi mi enaki in; da nebi smel biti razlog za sovraštvu med ljudstvi.</i>	35
9	vejica v stavkih z izpuščenim povedkom	<i>Na eni strani je imel Meto[.] na drugi strani pa očeta in Karnice.</i>	10
10	vejica pri pridevniškem nizu v levem prilastku	<i>To je po mojem mnenju popolnoma napačno ravnanje, nedopustljiva stvar ter najvišji; smrtni greh.</i>	5
11	vejica pred strukturami s še posebej	<i>In da ne smemo zaupati vsakemu[.] še posebej ne neznanu.</i>	4
12	vejica pred polstavkom	<i>Lucija je bila pridna učenka[.] znana po svoji pridnosti pri učenju in v šoli.</i>	1
SKUPAJ			1380

28 Po Slovenskem pravopisu (Toporišič, 2001) je *glede na* prislovna zveza. V gestu je kot primer rabe naveden stavek *oceniti delo glede na umetniško vrednost*, iz katerega je razvidno, da je raba vejice pred (*ne*) *glede na* odvečna. Po podatkih iz korpusa GigaFida pa situacija še zdaleč ni tako preprosta – govornici slovenščine se namreč za zapisa odločajo približno enako pogosto, kar potrjuje izsledke analize korpusa Šolar, da se v tem konkretnem primeru srečujejo z resnično pravopisno zadrego.

Analiza napak iz korpusa temelji na popravkih, ki so jih v besedilo vnesli učitelji, kar pomeni, da končni rezultati deloma odražajo tudi njihove osebne preference in kompetentnost. Učiteljev poseg je čutiti predvsem pri slogovnih popravkih, ki jih je sicer v okviru napak zapisa zaslediti redkeje kot pri ostalih tipih, je pa bilo mogoče na podlagi analize napak, vezanih na rabo vejice, identificirati specifične jezikovne strukture, ki preglavice povzročajo ne le učencem, temveč tudi učiteljem. Za posebej problematično se je izkazalo postavljanje vejice pred strukturami, ki jih uvaja zveza (*ne*) *glede (na)* – nekateri učitelji jo dodajajo (4) oz. je ne črtajo, drugi pa je ne dodajajo (5).²⁸

4 Konec koncev vsi smo ljudje in si zaslužimo spoštovanje **ne glede na** raso, spol, versko izpoved ali spolno usmerjenost.

Popravljen poved: Konec koncev vsi smo ljudje in si zaslužimo spoštovanje, **ne glede na** raso, spol, versko izpoved ali spolno usmerjenost.

5 Pečorin in Črtomir si nista enaka **glede na** odnos do ljubezni.

Podobne nedoslednosti je bilo zaslediti tudi pri postavljanju vejice pred strukturami s z *imenom*, npr.:

6 Punca z imenom **Fatima** je začela hoditi v srednjo šolo, drugi sošolci jo niso kaj marali saj ni bila čista, temveč umazana, revna dandanes bi jo označili za ciganko.

Popravljena poved: Punca, z imenom **Fatima**, je začela hoditi v srednjo šolo, drugi sošolci jo niso kaj marali saj ni bila čista, temveč umazana, revna dandanes bi jo označili za ciganko.

7 Zaljubi se v punco z imenom **Bisa**, ki ga nato prevara s Smako zato, da nebi več imel problemov z njim.²⁹

Podatki iz korpusa Šolar torej prinašajo koristne informacije o jezikovnih pojavih, na katere bi morali biti učitelji posebej pozorni ne le pri popravljanju šolskih izdelkov, temveč tudi pri pripravah na uro, pri čemer bi bilo poleg informacij iz slovnice in pravopisa smiselno upoštevati še korpusne podatke (npr. podatke iz referenčnega korpusa Gigafida), ki so najboljše pokazatelji aktualnega stanja v jeziku.

3.3 Napake male in velike začetnice

Pri analizi napak zapisa je bilo identificiranih 2125 napak, vezanih na neustrezno rabo velike in male začetnice (Rozman idr. 2012: 31). Večina napak (skoraj 96 %) je tipično pravopisne narave – najpogosteje gre za zapis lastnih osebnih (8), zemljepisnih (9) in stvarnih (10) imen ter za napačno rabo velike in male začetnice za ločili (11).

8 Fatima se je z družino odločila da bo obdržala otroka, oče pa je rekel da se bo maščeval **adamu** → **Adamu**

9 Mi vemo, da sta se v petek zvečer odpelale z avtobusom proti **ankaranu** → **Ankaranu** in iskale neko hišo ampak je bila zaklenjena.

10 Obiskovala je gimnazijo v ajdovščini, nato pa študirala grafično oblikovanje na **akademiji** → **Akademiji** za likovno umetnost v Ljubljani.

11 Ko sem pogledala Rozamundo, sem ugotovila, da je postala čisto rdeča v obraz. **da** → **Da**, seveda ona mi je prej rekla, da ji je Ostrovrhar zelo všeč, a je rekel, da so tudi lepša dekleta od nje.

V proučevanem vzorcu je bilo mogoče najti samo 92 potencialno relevantnih zgledov (4 %), ki se delno povezujejo z besedotvorno problematiko, in sicer gre za zapis pridevnikov na -ski oz. -ški (npr. *Slovenski* namesto *slovenski* in *Grški* namesto *grški*) in -ji (npr. *Božji* namesto *božji*) ter svojilnih zaimkov (npr. *cankarjev* namesto *Cankarjev* in *pandorin* namesto *Pandorin*). Problematika se dotika vprašanja, ali se pri izpeljavi pridevnikov iz lastnih imen velika začetnica ohrani ali izgubi, pri čemer imajo učenci zdaleč največ težav s pridevniki na -ski in -ški, zapis svojilnih pridevnikov in pridevnikov na -ji (znotraj te skupine je problematičen le pridevnik *božji*) pa jim povzroča bistveno manj preglavic, kot prikazuje Tabela 18.

Tabela 18: Jezikovni problemi, vezani na rabo velike in male začetnice

Jezikovni problem	Primer	Število napak
začetnica pri izpeljanih pridevnikih na -ski/-ški	<i>Šteje med bolj priznane Slovenske</i> → <i>slovenske arhitekte.</i>	81
začetnica pri svojilnih pridevnikih	<i>In ravno ta cankarjev</i> → <i>Cankarjev roman govori o "navidez" vzgledni družini.</i>	6
začetnica pri izpeljanih pridevnikih na -ji	<i>Mislim, da si je Kajin želel predvsem božje</i> → <i>Božje ljubezni in pozornosti.</i>	5

30 Glavni kriterij pri razločevanju kategorij Sestavljenka ali predložna zveza? in Prislov ali predložna zveza? je pomen: če napaka v zapisu ne vpliva na pomen stavka in povedi, primer spada v prvo kategorijo, če pa je med zapisom skupaj in narazen razlika v pomenu, je primer uvrščen v drugo kategorijo, npr. *Valjhun ima večglavo vojsko in močnejšo in za to tudi premaga Črtomirja in njegove pogane* (prislov zato in predložna zveza za to ne pomenita istega).

3.4 Napake pisanja skupaj in narazen

Napake pisanja skupaj in narazen se nanašajo na „pravopisni problem zapisovanja besed skupaj in narazen: beseda, ki bi morala biti napisana skupaj, je napisana narazen, ali pa sta skupaj napisani besedi, ki bi morali biti napisani narazen“ (Rozman idr. 2012: 31). Na podlagi učiteljskih popravkov je bilo izločenih in analiziranih 1179 napak pisanja skupaj in narazen, od katerih se jih je 1075 (nekaj več kot 91 %) izkazalo za relevantne, in določenih je bilo 15 potencialno relevantnih jezikovnih problemov, ki so prikazani v Tabeli 19.

Tabela 19: Jezikovni problemi, vezani na zapis besed skupaj, narazen in z vezajem

Št.	Jezikovni problem	Primer	Število napak
1	zanikani glagoli	<i>Vsi bi morali nositi te maske in nebi</i> → <i>ne bi prišlo do smrtnih žrtev.</i>	573
2	Sestavljenka ali navadna predložna zveza? ³⁰	<i>To pa zato, ker nakoncu</i> → <i>na koncu knjige glavna junaka umreta.</i>	252
3	sestavljenske z ne-	<i>Ni pa primerno vztrajati, če vemo, da smo ne močni</i> → <i>nemočni in da ne bomo uspeli.</i>	44
4	Prislov ali predložna zveza?	<i>Valjhun ima večglavo vojsko in močnejšo in za to</i> → <i>zato tudi premaga Črtomirja in njegove pogane.</i>	41
5	zapis sklopov	<i>Meta je že na vse zgodaj</i> → <i>navsezgodaj otroke zbudila in napodila bogi za kruhom</i>	38
6	zapis zloženek: narazen namesto skupaj	<i>Menim, da mlado poročenca</i> → <i>mladoporočenca ne zaslužita takšne usode zato sem konec malo spremenil.</i>	31
7	prislovi kratnosti	<i>Nato je Adam še en krat</i> → <i>enkrat priredil hišno zabavo in jo je znova izkoristil.</i>	21
8	števniki	<i>Pisalo se je leto dvatisočtri</i> → <i>dva tisoč tri.</i>	19
9	zaimki	<i>Vsak kdo</i> → <i>vsakdo ima pravico, da s svojim telesom dela tisto, kar se mu zdi pravilno.</i>	18
10	stopnjevani pridevniki in prislovi	<i>Ta del mi je bil tudi naj bolj</i> → <i>najbolj zanimiv.</i>	13
11	zloženke s števnik	<i>Rodila se je šestnajst letni</i> → <i>šestnajstletni gimnaziki, ki je ponesreči zanosila z tankajšim fantom.</i>	10
12	navezne oblike	<i>Prepirala sta se pred Lojzko in sploh se nista zmenila za njo</i> → <i>zanjo.</i>	7

31 Pri obliki gre za dvojno napako – napako zapisa in rabe glagolov *moči* in *morati* (Hamlet je tudi dosti zaprt in nemora sprejeti veliko prijateljev [...]).

32 Poleg zapisa skupaj je opaziti tudi problem rabe nestandardnih glagolskih oblik za 1. osebo dvojine (npr. *grema*, *greve*).

13	zapis zložen: z vezajem namesto skupaj brez vezaja	Od nežne ljubezenske poezije pa vse do trdo-erotičnih → trdoerotičnih in kar nekoliko vulgarnih del.	4
14	izpeljanka razpade na dva dela	Razlog zato odselse vanje → odselsevanje mladih v mesta, kjer je življenje lažje.	3
15	izpeljanke iz tujk ali prevzetih besed: zapis narazen namesto skupaj	Taja je visoka približno sto petdeset centimetrov, ima štirideset kilogramov, dolge zlatorumene lase in je oblečena v bel pulover in bele hlače, obute ima bele all starke → allstarke .	1
SKUPAJ			1075

Učenci imajo zdaleč največ težav z zapisom zanikanih glagolov (53,3 %), pri čemer je najpogosteje zaslediti napačno zapisane oblike glagola *biti* (npr. *nebi*, *nebo*, *nebom*, *neboš*, *nebosta*, *nebomo*, *nebodo*, *nebojo*), redkeje pa tudi oblike glagolov *bati se* (npr. *nebojte*), *dati* (npr. *neda*), *delati* (npr. *nedelaj*), *dohajati* (npr. *nedohajajo*), *dovoliti* (npr. *nedovoliti*), *ganiti se* (npr. *neganejo*), *iti* (npr. *negre*), *ljubiti* (npr. *neljubi*), *marati* (npr. *nemaram*, *nemara*, *nemarajo*, *nemaramo*), *moči* (*nemorem*, *nemoremo*, *nemorejo*, *nemoreta*, *nemore*, *nemoreš*; *nemora*,³¹ *nemorata*), *motiti* (npr. *nemoti*), *obujati* (npr. *neobuja*), *oditi* (npr. *neoditi*), *odločiti se* (npr. *neodloči*), *opaziti* (npr. *neopazi*), *ozirati se* (npr. *neozira*), *pomagati* (npr. *nepomaga*), *pomeniti* (npr. *nepomeni*), *popustiti* (npr. *nepopustimo*), *poslušati* (npr. *neposlušajš*), *povedati* (npr. *nepove*), *pozabiti* (npr. *nepozabiš*), *poznati* (npr. *nepoznam*, *nepozna*), *predstavljati* (npr. *nepredstavljam*), *priti* (npr. *nepride*), *pustiti* (npr. *nepusti*), *razlikovati* (npr. *nerazlikujejo*), *razvaditi* (npr. *nerazvadijo*), *rojevati* (npr. *nerojevajo*), *smeti* (npr. *nesme*, *nesmeš*, *nesmeve*,³² *nesmemo*, *nesmejo*), *starati* (npr. *nestarata*), *storiti* (npr. *nestorijo*), *strinjati se* (npr. *nestrinja*, *nestrinjamo*), *ugajati* (npr. *neugaja*), *umreti* (npr. *neumre*), *upoštevati* (npr. *neupoštevamo*), *vedeti* (npr. *nevem*, *neve*, *nevesta*, *nevemo*, *nevejo*, *neveš*), *verjeti* (npr. *neverjame*), *zavedati se* (npr. *nezavedajo*), *zavzemati se* (npr. *nezavzemajo*), *zdeti se* (npr. *nezdi*), *zganiti se* (npr. *nezgane*), *znati* (npr. *neznajo*, *neznamo*) in *živeti* (npr. *neživi*).

Veliko težav imajo učenci tudi z zapisom predložnih zvez tipa *na koncu*. Problem je zaslediti v 23,5 % vseh relevantnih zgledov, pri čemer prevladujejo predvsem naslednje oblike: *naglas*, *nakratko*, *vredu*, *vresnici*, *nažalost*, *pomoje oz. pomojem*, *ponavadi*, *ponesreči in po nesreči*, *ob enem*, *v pričo*, *na enkrat* in *v naprej*. Jezikovni problem je zanimiv tudi z vidika nedoslednosti učiteljskih popravkov, saj učitelji analogne zglede popravljajo različno:

12 Med pogovorom z materjo **ponesreči** → **po nesreči** ubije dvornika Polonija, ki se je skrival za zaveso.

13 Hamlet med preprirom z materjo Gertrudo **po nesreči** → **ponesreči** ubije Polonija, ki je bil kraljev zvesti minister.

Relativno manj težav imajo učenci z zapisom sestavljenk z *ne-*, npr. *nemočen*, razlikovanjem med oblikovno podobnimi prislovi in

predložnimi zvezami, npr. *za to – zato* in *na to – nato*, zapisom sklopov, npr. *navsezgodaj – na vse zgodaj* in *kajne – kaj ne*, zapisom zloženk narazen namesto skupaj, npr. *mlado poročenca*, prislovov kratnosti, npr. *en krat* in *dosti krat*, števnikov, npr. *dvatisočtri* in 25.3.2010, zaimkov, npr. *vsak kdo*, *marsi kdo*, *kdor koli*, in stopnjevanih pridevnikov ter prislovov, npr. *naj bolj* in *naj srečnejši*. Vsi omenjeni jezikovni problemi predstavljajo med 1 do 4 % napak v vzorcu. Ostali problemi, tj. zapis zloženk s števnikami, naveznih oblik zaimkov, zloženk z vezajem namesto skupaj in brez vezaja, izpeljank narazen in izpeljank iz tujk oz. prevzetih besed, se v vzorcu pojavljajo marginalno (pogostost pojavljanja je enaka ali manjša od 10), kar pomeni, da učencem povzročajo precej manj preglavic kot ostali problemi pisanja skupaj in narazen.

3.5 Napake črkovanja

Napake črkovanja se nanašajo na napačen zapis glasu ali glasovnega sklopa v besedi. Ko neustrezna črka ali glas spremeni pomen besede, je primer uvrščen med napake besedišča (npr. *zavedali* namesto *zavedli*; gl. 4. poglavje). Napake črkovanja so delno prekrivne tudi z napakami oblike (gl. 5. poglavje); besede, pri katerih je neustrezen glas opaziti v končnici, npr. *umoriste* namesto *umorite*, spadajo pod napake oblike in ne črkovanja (Rozman idr. 2012: 30).

Iz korpusa Šolar je bilo izločenih in analiziranih vseh 2672 napak črkovanja. Od teh se jih je kar 785 oz. cca. 30 % izkazalo za manj relevantne – večinoma gre za lapsuse in ozko pravopisno problematiko. Relevantni zgledi (1869 pojavnic oz. 70 %) se povezujejo predvsem z besedotvornimi vsebinami (npr. izpad črke *e* pri izpeljavi pridevnika *mojstrski* iz samostalnika *mojster*), delno pa tudi z oblikoslovnimi in slogovnimi. Na začetku analize so bili zgledi napak na grobo razdeljeni v pet osnovnih podtipov, kot prikazuje Tabela 20.

Tabela 20: Osnovne kategorije napak črkovanja

Št.	Podtip napake	Primer	Število napak
1	problematičen zapis glasu	<i>S svojo sabljo mu je odrezal glavo in medvet → medved je umrl.</i>	1349
2	problematičen zapis sklopa	<i>S tem dejanjem bo dosegla svoj življenski → življenjski smisel, se bo uresničila.</i>	386
3	raba govornega zapisa v formalnem kontekstu	<i>Usedel sem se na zadnji konec jadrnice in si sezu → snel plavutke. Od tistega dne skos → vseskozi mislim nate.</i>	109
4	nestandardno sklanjanje, spreganje ali stopnjevanje	<i>Kralj Ojdip je po prerokbi in naključju dogodkov ubil svojega očeta ter se poročil s svojo matero → materjo.</i>	20
5	zapis prevzetih besed	<i>Človek brez ljubezen je kot avto brez benzina → bencina ali kot mama brez otroka.</i>	5
SKUPAJ			1869

Najobsežnejša kategorija, tj. problematični zapis glasu (72 %), se nanaša na neustrezen zapis črke ali glasu v besedi (npr. *medvet, izkoristiu, rastrgane, gdo, hmalu, opčani, sovražtvo, nacionalni, razmirah, suočiti, očiteljiv* ipd.), manjkajoče črke (npr. *odahnejo, isekavanjem, rojsni* ipd.) in odvečne črke (npr. *bogatstvo, oddlomek, isplača* ipd.). Večina zgledov iz druge kategorije (21 %) je vezanih na zapis sklopov lj (npr. *bolše, življenje, bližini, škatljico* ipd.) in nj (npr. *bivanske, pomankanje, knjižnjici, lažnjiva, najmajn* ipd.). Obe kategoriji se vsebinsko povezujeta predvsem z besedotvornimi temami. Med napakami črkovanja se relativno pogosto pojavljajo tudi besede, ki jih tipično zasledimo v govornem jeziku (cca. 6 %) in so pogosto značilne le za določene pokrajinske govore (npr. *dones, jst, sezu, skos, stauk, pršla, zmerej, lohko, počas, ubistvi* ipd.). Tovrstnih zgledov je sicer mnogo več, kot je razvidno iz Tabele 20, saj tudi ostale štiri kategorije problemov črkovanja vsebujejo besede v govornem zapisu (npr. *dans* v prvi kategoriji), vendar pa je bila med analizo sprejeta odločitev, da se med govorno izrazje uvrstijo samo tisti zgledi, ki jim ni mogoče pripisati druge kategorije. Zgledi govornega izrazja so s slogovnega vidika zanimivi, saj opozarjajo na problem vdiranja govornega jezika v standardni. Pri analizi sta se pokazali še dve marginalni kategoriji, tj. nestandardno sklanjanje, spreganje in stopnjevanje, ki je podrobneje obravnavano predvsem pri analizi napak oblike (gl. 5. poglavje), in zapis prevzetih besed, ki učencem očitno ne povzroča večjih preglavic.

Na podlagi analize relevantnih zgledov iz opisanih kategorij je bilo določenih skupno 34 potencialno relevantnih jezikovnih problemov. V nadaljevanju so zaradi pomanjkanja informacij o pogostosti pojavljanja posameznih problemov predstavljeni le rezultati kvalitativne analize. Posebej so izpostavljeni najbolj pereči jezikovni problemi, ki jih je mogoče obravnavati v sklopu besedotvorja, oblikoslovja in besedoslovja.

3.5.1 Črkovanje in besedotvorje

Pri analizi napak črkovanja je bilo določenih 24 jezikovnih problemov, ki se vsebinsko povezujejo z besedotvornimi temami; ti problemi so bili identificirani pri analizi prvih dveh osnovnih kategorij, tj. problematičnega zapisa glasov in glasovnih sklopov. Posebej problematičen je zapis samostalniških, pridevniških in glagolskih izpeljank; učenci pogosto izpuščajo in dodajajo črke (npr. *bogatstvo, veterček in rojsni*) ali pa enostavno izberejo napačno (npr. *zadustuje*). Izbira napačne črke ali glasu je pogosto povezana z glasovnimi premenami v slovenščini, tj. preglasom (npr. *Romejov*), palatalizacijo (npr. *Mojcin*) in jotacijo (npr. *ustregljív*), kot prikazuje Tabela 21.

Tabela 21: Jezikovni problemi, vezani na zapis izpeljank

Št.	Jezikovni problem	Primer
1	izpad ali dodajanje črke t pri samostalniških izpeljankah	<i>Odšli smo nazaj v apartma, kjer nas je čakal lasnik → lastnik. In tudi bogastvo → bogastvo ne pomeni vedno sreče, če ni pravega odnosa.</i>
2	izpad črke t pri pridevniških izpeljankah	<i>Knjigo sem ji kupil za rojstni → rojstni dan</i>
3	izpad črke e pri pomanjševalnicah	<i>Pihau je nežen veterček → vetrček samo naju ni nič zebla.</i>
4	ohranjanje črke e pri pridevniških izpeljankah	<i>Ker pa je Klavdij sumil, da Hamlet raziskuje smrt, je postavil Hamletu past, s katere se mojstersko → mojstrsko izvlekel</i>
5	neustrezna črka v podstavi samostalniške izpeljanke	<i>Ni maral izdajalstva in prilizevanja → prilizovanja, kajti ni spoštoval takih ljudi</i>
6	neustrezna črka v podstavi pridevniške izpeljanke	<i>Danilo kot oseba je dober po duši in srcu, bojezljiv → bojazljiv, ravnopraven in potrpežljiv</i>
7	neustrezna črka v podstavi glagolske tvorjenke	<i>Antigona mu je povedala, da ji kazen zadustuje → zadostuje.</i>
8	jotacija v podstavi samostalniške izpeljanke	<i>Ta del je vrh komedije – sodenje → sojenje Matičku.</i>
9	jotacija v podstavi pridevniške izpeljanke	<i>Vedel je, da je Polonij zelo ustregljiv → ustrežljiv kralju, kot tudi njemu.</i>
10	preglas pri svojilnih pridevnikih	<i>V Romeovi in Capuletovi družini pa so nastali spori in to je težilo med Julijino in Romejovo → Romejevo poroko</i>
11	palatalizacija pri svojilnih pridevnikih	<i>Ravno ko sta si Tinko in Mojca izmenjala prstane, je v cerkev pritekel kot blisk Mojcin → Mojčin bivši</i>
12	tvorjenje deležnikov	<i>Menim, da so možje poniženi → ponižani, ker jih je veliko v pomenu "copate".</i>
13	težji primeri pridevniških izpeljank	<i>Epika je literalna → literarna zvrst v kateri, je prisotno pripovedništvo.</i>

Učenci imajo težave tudi z zapisom nekaterih sestavljenk; natančneje gre za samostalnike, pridevnike, prislove in glagole s predponskimi obrazili iz-, od- in o-, pri katerih je zaslediti tako manjkajoče (npr. *iztopi*) kot odvečne črke (npr. *oddigrali*). Posebej problematični so izpusti črk pri sestavljenkah z dvojnimi črkami (npr. *odaja*, *sočiti*, *odolžiti* ipd.) in podvajanje črk pri tistih, ki nanje spominjajo (npr. *oddi*). Izpust črke v podstavi ali predponi lahko povzroči tudi spremembo pomena; do pojava pride takrat, ko sta želena in uporabljena beseda blizuzvočnici (npr. *zapil* namesto *zavpil* in *zbosti* namesto *zabosti*). Problemi zapisa sestavljenk so predstavljeni v Tabeli 22.

Tabela 22: Jezikovni problemi, vezani na zapis sestavljenk

Št.	Jezikovni problem	Primer
1	sestavljenske s predponskim obrazilom iz-	<i>Opazovalec prispe na cilj in iztopi → izstopi iz avtobusa</i>
2	sestavljenske s predponskima obraziloma od- oz. o-	<i>Čisto na koncu so oddigrali → odigrali predstavo o Romeu in Juliji.</i>
3	sestavljenske z dvojnimi črkami	<i>V večerni odaji → oddaji je bila Davidova smrt poročana.</i>
4	izpust črke v podstavi (sprememba pomena)	<i>Znašel sem se potleh in sem zapil → zavpil: „Naponoč, me boli.“</i>
5	izpust črke v predponskem obrazilu (sprememba pomena)	<i>Nato vzame Romeo nož in se hoče zbosti → zabosti.</i>

Poleg zgoraj predstavljenih problemov je bilo mogoče identificirati še nekoliko splošnejše besedotvorne kategorije. Pri analizi napak črkovanja se je med drugim izkazalo tudi, da imajo učenci veliko težav z zapisom sklopov lj in nj – ponekod črko j izpuščajo (14), drugod pa jo po nepotrebnem dodajajo (15).

14 Predstavljata partizana, ki pa imata različni **življenski** → **življenjski** zgodbi.

15 Ko je bilo konec **šoljskega** → **šolskega** leta, je Špela domov prinesla spričevalo.

Med splošnejše probleme spada tudi izbira napačne podstave pri tvorjenju besed (16) in napake pri zapisu sklopov (17), vendar so se ti problemi pri analizi izkazali za marginalne:

16 Zaposlila sta se v **slastičarni** → **slaščičarni**, dobivala sta nizke plače in komaj shajala iz meseca v mesec.

17 No **vseno** → **vseeno** vam, bom opisal današnjo družino

3.5.2 Črkovanje in oblikoslovje

Napake črkovanja, ki se povezujejo z oblikoslovnimi temami, spadajo predvsem v osnovno kategorijo nestandardnega sklanjanja, spreganja in stopnjevanja, deloma pa tudi neustreznega zapisa črke oz. glasu (gl. razdelek 3.5). Pogosto gre za jezikovne probleme, ki so bili identificirani že pri analizi napak oblike (npr. sklanjanje samostalnika *mati*; gl. 5. poglavje), v ostalih primerih pa za slovnično manj relevantne vsebine, ki so sicer kot posebna opozorila lahko vključena v razlage oblikoslovnih težav (npr. pri obravnavanju naveznih oblik zaimkov je smiselno posebej opozoriti na zapis sklopa nj). Rezultati analize so predstavljeni v Tabeli 23; zaradi delne relevantnosti problemi v nadaljevanju niso podrobneje opisani, temveč zgolj naštet.

Tabela 23: Jezikovni problemi črkovanja, vezani na oblikoslovne teme

Št.	Jezikovni problem	Primer
1	težji primeri sklanjanja	<i>Izidor pa kljub temu da je bil njegov oče Polikarp zelo težek človek, strog in da je grdo ravnal z njegovo matero → materjo</i>
2	nestandardno stopnjevanje pridevnikov	<i>mati nagovarja svojega otroka, da ga ima vseeno zelo rada, kljub temu, da ga njeni najbližni → najbližji in okolica zaničujejo</i>
3	sklop nj pri naveznih oblikah zaimkov	<i>Vaščani se niso jezili nan → nanj.</i>
4	nestandardne oblike zaimkov	<i>Ali bo z njim pravičen kot to ravna z vsimi → vsemi in bo šel v podzemlje?</i>
5	težji primeri spreganja	<i>Ljubezen pride v najhujših trenutkih in vzplapoli kot ogenj, ki plapota → plapola in sije svetlobo sreče.</i>
6	zapis deležnikov na -l	<i>Delav → Delal se je norca iz njega in ga hotel na vso moč prepričati, da je poljski oficir. Nagovor v tem delu mi je biu → bil všeč</i>
7	raba predlogov z/s, k/h in u/v	<i>Odnesel me je h → k socialni delavki Svoje telo seveda krepimo s → z veliko gibanja, predvsem na soncu. Janez je u → v mestu zelo veliko razmišljal o njej kako je lepa.</i>

3.5.3 Črkovanje in besedoslovje

Pri analizi sta bila identificirana dva problema črkovanja, ki se pozevujeta z besedoslovnimi temami. Prvi problem se nanaša na blizu-zvočnice – učenec v besedi uporabi neustrezno črko, zaradi česar se spremeni tudi pomen:

18 Doma pa je najraje oblečena kar v **trenerko** → **trenirko**, saj se v njej počuti najbolj udobno.

Drugi problem je povezan z mešanjem glagolskih paradig. Učenec, ki ima težave z razločevanjem med podobnimi glagoli, oblike enega glagola delno prenese v paradigmo drugega, npr. pri paru *prizadeti* – *zadeti*:

19 A prav ta ljubezen lahko tudi boli in **prizadane** → **prizadene**.

Tovrstnih primerov napak črkovanja je v korpusu Šolar relativno malo, kar kaže na to, da učenci njimi nimajo večjih težav.

3.6 Sklep

Napake zapisa se nanašajo na tipično pravopisno problematiko zapisovanja besed in besednih zvez, izstopajo pa napake v rabi ločil, zlatih vejice. Podatke, ki so bili pridobljeni z analizo izbranih podtipov, bi bilo mogoče s pridom uporabiti tudi pri dopolnjevanju razlag jezikovnih problemov pri drugih kategorijah: napake ločil so relevantne predvsem za opise skladenjskih tem (npr. raba vejice ob odvisnikih,

stavčnih členih in v pridevniškem nizu), napake črkovanja, velike in male začetnice ter pisanja skupaj in narazen pa za razlage besedotvornih jezikovnih problemov (npr. ohranjanje in opuščanje velike začetnice pri izpeljavi, pisanje tvorjenk skupaj in z vezajem ter zapis črk in črkovnih sklopov v izpeljankah, sestavljenkah in sklopih). Za prihodnje študije o pravopisnih temah in oblikovanje alternativnih jezikovnih virov za slovenščino bi bilo vredno razmisliti o podrobnejši kvantitativni in kvalitativni analizi napak zapisa, pri čemer bi bilo treba več pozornosti nameniti predvsem celostni obdelavi napak ločil.

4 Napake besedišča

MOJCA STRITAR

4.1 Uvod

Napake besedišča oz. leksikalne napake so tipična kategorija napak. Pojavlja se v tako rekoč vsaki klasifikaciji, ki temelji na jezikovnih ravninah (Stritar 2012; Požgaj Hadži, Ferbežar 2001; Rozman 2006; Granger 2001; Nicholls 2003; Lüdeling idr. 2005; Boyd 2008; Hana idr. 2010; Izumi, Uchimoto in Isahara 2004; Bušta idr. 2009; Pala, Rychlý in Smrž 2003). Gre za pojavitve, v katerih je uporabljena beseda ali besedna zveza v danem kontekstu slovnično, semantično ali pragmatično neustrezna. Poleg tega lahko pomenijo tudi neustrezno tvorbo besed, uporabo neustrezne besedne vrste in podobno.

Napake besedišča v Šolarju nimajo podtipa (prim. Stritar 2012: 178), saj njihova nadaljnja členitev med samim označevanjem, ki mora biti hitro in učinkovito, ni smiselna. V navodilih za označevanje korpusa so bile mednje umeščene naslednje napake (Rozman idr. 2012: 33–34):

- zamenjava dveh besed iste besedne vrste;
- zamenjava dveh besed različne besedne vrste;
- zamenjava besedne zveze z besedo ali obratno;
- zamenjava dveh besednih zvez;
- zamenjava moške in ženske oblike samostalnika;
- zamenjava presežne in osnovne oblike pridevnika;
- zamenjava različnih tipov zaimkov;
- zamenjava zaimkov *kateri/ki* (povzroči tudi spremembo daljšega niza);
- zamenjava predlogov (razen *z/s* in *k/h*);
- zamenjava nedoločne in določne oblike pridevnika *mali/majhen*;
- zamenjava glagolskega vida oz. predpon.

Analiza napak je odvisna tudi od določitve kategorij ob začetku označevanja korpusa, zato je treba opozoriti, da je bilo že uvodoma jasno, da se napake besedišča do neke mere prekrivajo z napakami zapisa, predvsem s podkategorijo črkovanja, in napakami oblike. Čeprav se je včasih težko odločiti za eno od kategorij, je to pri označevanju nujno, in to na enoten način, sicer celotno označevanje postane neutemeljeno. Napake, pri katerih je tvorec zamenjal predloga *z* in *s* oz. *k* in *h*, bi lahko bile napake besedišča ali črkovanja, toda ker gre v osnovi za vprašanje pravopisa, smo se odločili, da jih uvrstimo med napake črkovanja. Raba neustrezne oblike pridevnika *mali/majhen* bi bila lahko tudi napaka oblike, enako velja za glagolski vid oz. neustrezne glagolske predpone ali neustrezno obliko svojilnega ali drugega zaimka. Napake, pri katerih je tvorec uporabil besedo neustrezne besedne vrste, bi lahko bile tudi napake oblike (prim. Rozman

2006: 51) ali strukture, pri besednih zvezah pa je težavno ločevanje med napakami besedišča in strukture. Obveljalo je načelno pravilo, da so med napake besedišča uvrščene napačne izbire na leksikalni ravni, pa čeprav se nanašajo na več kot eno besedo. Uporabnik korpusa se mora ob tem zavedati, da gre pri nekaterih napakah besedišča, npr. pri rabi svojilnih zaimkov, bolj za slovnično kot za slovarsko zmožnost. Med napake besedišča sodijo tudi napake pri besedotvorju, čeprav ga nekateri obravnavajo kot poseben del besedoslovja (prim. Toporišič 1992: 8).

4.2 Rezultati

V korpusu Šolar je bilo označenih 3373 napak besedišča,³³ skupaj torej slabih 10 % vseh napak. Vse napake besedišča so bile ročno pregledane in združene v nadaljnje kategorije, ki niso bile določene vnaprej. Oblikovane so bile sproti, kot so se kazale med samo analizo, glede na formalne, predvsem besednovrstne, pa tudi semantične značilnosti. Ob vnovičnem pregledovanju konkordanc po posameznih kategorijah so bile te po potrebi združene ali razdružene. Tako je bilo med napakami, v katerih gre za zamenjavo sinonimnih veznikov, prepoznano pogosto ponavljanje zamenjav protivnih veznikov, zato so ti dobili svojo, bolj specifično kategorijo.

Med napakami besedišča je bilo prepoznanih 191 različnih tipov, ki so v nadaljevanju zaradi preglednosti predstavljeni glede na besedno vrsto. Kot kaže Tabela 24, gre skoraj pri polovici vseh napak za nepolnopomenske besede, pri katerih je lažje prepoznati slovnične značilnosti. Napake pri polnopomenskih besedah, pa tudi besednih zvezah, so bolj semantične, zato je v njih težje identificirati sistemske lastnosti in jih enoumno razvrščati. V Tabeli 24 so zato večinoma združene v kategoriji „razno“.

Tabela 24: Napake besedišča glede na besedno vrsto

Besedna vrsta	Primer	Število napak
zaimek	<i>Tak</i> → <i>Takšen</i> ples bi z veseljem gledala vsak dan.	637
predlog	<i>Od</i> → <i>Iz</i> prejšnjega zakona pa mu je ostala samo hčerka.	423
veznik	<i>Bil</i> je v vojni, <i>kjer</i> je pobijal ženske, otroke, požigal hiše <i>ter</i> → <i>in</i> kradel.	380
glagol	Nato so se stvari začele <i>poboljševati</i> → <i>izboljševati</i>	293
pridevnik	Za <i>Majhne</i> → <i>male</i> ljudi, kamor spada tudi on, pravi, da so le statisti v tej veliki igri.	63
prislov	da se bomo bolj imeli radi <i>ter več</i> → <i>bolj</i> spoštovali.	42
členek	Ko je slišala, da je Hajmon mrtev se je ubila <i>tudi</i> → <i>še</i> ona.	29
samostalnik	Svoje <i>protestiranje</i> → <i>nestrinjanje</i> pokažem z različnimi izrazi.	29
razno	Razlikujejo se od tistih časov po tem, da sedaj ni toliko <i>vojnih težav</i> → <i>vojn</i>	1493
Skupaj		3373

Dvanajst tipov napak besedišča ima v Šolarju več kot 100 pojavitev (Tabela 25). Ker so tako pogosti, upravičeno sklepamo, da so problematični za večji del šolske populacije, zato so v nadaljevanju podrobneje opisani. Kljub temu pa tu niso pojasnjeni tudi vzroki za te napake (prim. Edge 1989: 9–11; James 1998: 83–84; Balažič Bulc 2004: 86), saj bi bila za to potrebnejša podrobnejša analiza, specifična za vsak posamezen primer.

Tabela 25: Najpogostejše napake besedišča

Napaka	Število napak
Napačna polnopomenska beseda	329
Pogovorna/nezaznamovana beseda	248
Vsebinsko neustrezna beseda	192
Splošna/specifična beseda	184
Svojilni oz. povratnosvojilni zaimek	173
Zaimek <i>ki/kateri</i>	169
Samostalnik/zaimek	143
Menjava s sopomenko	141
Pomensko neustrezna beseda	133
Predlog <i>v/na/iz/z/s</i>	119
Neustrezna predpona	114
Specifična/splošna beseda	103
Ostalo	1341

4.2.1 Neustrezno izbrane polnopomenske besede in vsebinske napake

Čeprav je Šolar korpus maternih govorcev slovenščine, so tvorci besedil osnovno- in srednješolci, pri katerih proces usvajanja jezika še ni končan. Zato v skoraj 10 % vseh napak besedišča – verjetno predvsem zaradi nepoznavanja oz. nerazumevanja – uporabijo neustrezno polnopomensko besedo (20). Ta ima lahko neustrezno priponsko obrazilo (*dramatsko besedilo* namesto *dramsko besedilo*), učenec besedo narobe tvori (*poganci* namesto *pogani*) ali pa besedi zamenja, ker podobno zvenita (*etika* namesto *epika*, *vložena pesem* namesto *vložna pesem*, *vsaj* namesto *saj*).

20 Glavni junak Črtomir se je po boju s kristjani udal v **resignacijo**

→ **usodo**

V šolskih besedilih učitelji preverjajo tudi to, kako dobro se učenci naučijo dejstva. Torej ni nenavadno, da so v tretjem najpogostejšem tipu napak popravljene vsebinske napake, povezane z imeni oseb, žanri literarnih del ipd. (21) (prim. razdelek 6.2.3 v 6. poglavju). Čeprav so za jezikoslovno analizo manj zanimive, so bile zaradi objektivnosti označevanja označene kot vse ostale napake besedišča. Zaradi

vsebinske neustreznosti ali sloga učitelji popravljajo tudi različne metonimije oz. posplošitve učencev (22).

21 Pripovedovalec goji veliko sovražstvo do **Inga** → **Lepca**

22 tretji del, pa je Krst in je epika, lirika in dramatika, ker pesnik pripoveduje, se izpoveduje in **pesnitev** → **del pesnitve** je napisana po dialogih.

4.2.2 Slogovni popravki

Ker se učenci prek šolskih besedil učijo pisanja, učitelji zelo pogosto popravljajo njihov slog. Ti popravki so manj zanimivi za jezikoslovne analize težavnih področij mladih govorcev, nudijo pa vpogled v to, kako skušajo učitelji izboljšati slog. V prvi vrsti gre za nadomeščanje slogovno neustreznih besed, saj je drugi najpogostejši tip napak besedišča v Šolarju s 7 % učenčeva raba nestandardne besede ali besedne zveze namesto standardne oz. nezaznamovane (23).

23 kar za sedemnajstletnico ni **gluh** → **ravno** najbolj primerno.

Drugi pogosti poskus izboljšave sloga učencev je izogibanje ponavljanju besed. Učitelji ga odpravljajo s sopomenkami (24) ali z izrazi, ki so sopomenski v danem kontekstu (25), splošnejše besede menjajo z bolj specifičnimi (26), imena ali svojilne pridevnike pa z osebnimi, kazalnimi ali svojilnimi zaimki (27).

24 kako je **možno** → **mogoče**, da se je njen dolgoletni partner tako občutno spremenil.

25 Prav njemu so **določili** → **dali** ukaz, da jo mora ubiti, ker je izdajalka.

26 Besedilo se odvija v mestu Verona v Italiji. **Besedilo** → **Drama** govori o dveh sprtih družinah, ki sta pretrgali vse vezi in nista imeli neposrednega stika.

27 Čeprav za Gregorja Katarina ni izdajalka, izpolni ukaz ter **ustreli Katarino** → **jo ustreli**.

Po drugi strani učitelji popravljajo tudi obratno. Bolj specifične besede nadomeščajo s splošnejšimi (28), parafrazirajo jih s kazalnim zaimkom *ta* in nadpomenko (29), včasih pa zaimke zaradi razumljivosti pri referiranju nadomestijo z bolj nedvoumnimi polnopomenskimi besedami (30).

28 Klavdij se na začetku hoče prikupiti Hamletu in pridobiti njegovo zaupanje, ki bi ga kasneje zlorabil. Na srečo mu to ne uspe, saj ga **Hamlet** → **princ** že od samega začetka na nek način sovraži.

29 Gimnastiko sem trenirala do trinajtega leta. Pri trinajstih sem si poškodovala roko zato sem morala **gimnastiko** → **ta šport** pustiti in se odločiti za nek drug šport.

30 Antigona, trmasto in odločno dekle, si je na vsak način prizadevala pokopati svojega brata, četudi je pri tem kršila vladarjeve zakone. Zaradi svoje ljubezni do **njega** → **brata** bi namreč naredila vse.

Še ena slogovna sprememba besedil učencev je nadomeščanje prevzetih besed s slovenskimi besedami, ki pa se je pojavilo le malo več kot dvajsetkrat (31).

31 **Šofer** → **Voznik** je vzel mikrofon in naznanil: „Pripravite se, da odpotujemo v leto 1825!“

Da bi se učenci izražali bolj jedrnato, učitelji besedne zveze nadomeščajo z eno besedo (32), denimo s prislovom (33) ali z glagolom (34).

32 Razlikujejo se od tistih časov po tem, da sedaj ni toliko **vojnih težav** → **vojn.**

33 Zgodba se konča z **tragičnim koncem** → **tragično,**

34 Bog **mu da kazen** → **ga kaznuje**, da mora hoditi po svetu,

Med manj pogostimi slogovnimi popravki je treba omeniti popravke časovnih referenc (35), pa tudi rabo moške namesto ženske oblike samostalnika ali pridevnika (36).

35 **Včeraj** → **Prejšnji dan** mi je podaril zaročni prstan saj to mi je bilo zelo všeč.

36 V tej zgodbi je Antigona **tragični junak** → **tragična junakinja**

4.2.3 Napake zaimkov

Zaimki, med katere v skladu z besednovrstno razdelitvijo, uporabljeno v korpusih MULTEXT-East, FIDA, FidaPLUS in Gigafida (gl. Erjavec idr. 2001; Erjavec 2004), spadajo tako samostalniški kot pridevniški zaimki, so statistično gledano najbolj problematična besedna vrsta pri napakah besedišča v Šolarju, saj napake zaimkov predstavljajo skoraj 19 % napak besedišča. Med konkretnjšimi primeri izrazito izstopa neustrezna raba svojilnega oz. povratnosvojilnega zaimka (nekajkrat gre tudi za osebne in povratnoosebne zaimke), ki je peta najpogostejša težava v Šolarju (37).

37 To sta predvsem Kocbek z **njegovo** → **svojo** novelo Črna orhideja in Zupan z romanom Menuet za kitaro.

Le malo manj težavna je raba oziralnih zaimkov *ki* in *kateri*, ki je pogosto sicer samo vprašanje sloga, a je šesta najpogostejša napaka besedišča (38). Nekajkrat je nadomeščen zaimenski prislov *kjer* (39). Oziralni zaimki so problematični tudi zaradi zamenjevanja s kazalnimi (40) in vprašalnimi zaimki (41).

38 Navidezna zmaga je pripadla Polikarpu in s tem tudi krivda, **katere** → **ki je** ni pričakoval,

39 Klavdij je dal Hamleta poslati v Anglijo in je zraven poslal dva njegova prijatelja, ki sta imela pismo **kjer** → **v katerem** je pisalo naj ubijejo Hamleta.

40 Oče in mati sta se prikazala v najlepši luči, **tega** → **česar** Lojzka ni razumela.

41 Zato se je tudi on tako spremenil, ker ni povedal **kar** → **kaj** čuti in ni hotel rešiti problema v sebi.

Ponavljajoči se problemi zaimkov, ki so ravno tako stvar slogovne izbire, so še zamenjevanje nikalnih zaimkov *nihče* oz. *noben* (42), raba naslonske namesto navezne oblike osebnega zaimka (43) in raba zaimkov oz. pridevnikov *en*, *eden*, *nek(i)*, *drugi* (44).

42 a se ni zanjo **noben** → **nihče** zanimal.

43 **Za mene** → **Zame** ljubezen pomeni veliko,

44 Spoznala sem, da se je v življenju res potrebno prilagajati, a le do **ene** → **neke** mere

Še vedno tipične, a manj pogoste napake zaimkov so raba nedoločnih zaimkov *nekateri* oz. *neki* (45), raba *nič* oz. *kaj* pri zanikanju (46), menjava oblik *obadva* in *oba* (47), *nobeden* in *noben* (48), *vsak* in *vsakdo* (49), težave pri podobno zvenceh zaimkih *ta*, *tak* in *takšen* (50) in raba osebnega zaimka, ki ga učitelj nadomesti s poudarjenim pridevnikom *sam* (51).

45 Pijanega ga je **nekatera punca** → **neko dekle** peljala v sobi

46 Tudi če nebi, bi lahko vsaj poskusila, saj izgubiti nimata **nič**
→ **kaj**.

47 V medsebojnem boju se **obadva** → **oba** zastрупita z zastrupljenimi konicami mečev in nato podležeta.

48 Pripovedovalec je ugotovil tudi, da se **nobeden** → **noben** drug potnik ni zmenil za to gospo, in niti za to, kdo je iz sebe izpustil neprijeten vonj.

49 Vsak lahko pojma razume drugače, saj je za **vsakega** → **vsakogar** pravičnost nekaj čisto drugega,

50 In vendar je moral postati prav **ta** → **tak**, hladnokrvni morilec, ki brez pomisleka, brez najmanjše misli,

51 Ko je povodni mož to slišal jo je poslal nazaj na kopno, a na drugi konec sveta. **On** → **Sam** pa je od žalosti in živčnosti uničil svoje kraljestvo

4.2.4 Napake predlogov

Napakam zaimkov sledijo napake predlogov, ki jih je med napakami besedišča v Šolarju dobra desetina. Največ, 28 %, se jih nanaša na neustrezne predloge *v* oz. *na* in *iz* oz. *z/s*, kar je deseti najpogostejši problem besedišča (52).

52 **Z** → **Iz** vojne se je vračal skupaj z Joštom

Druge pogoste težave pri rabi predlogov so navedene v Tabeli 26.

Tabela 26: Pogoste napake predlogov

Št.	Tip napake predloga	Primer	Število napak
1	Raba od in z/s	Ko je Klavdij, kar naenkrat odšel od → s predstave	20
2	Raba pri in v	vendar vem da moja prava ljubezen še pride bolj pri → v , „zrelih“ letih.	16
3	Raba za in zaradi	Zevs mu je za to dejanje → zaradi tega dejanja uzel ogenj.	12
4	Raba kot in za	Vzgjajala jih je kot → za pogumne možake.	11
5	Raba na namesto za	Vem, da bi veliko storila na → za preživetje, ampak človeka pa ne bi mogla ubiti.	11
6	Raba do namesto k	Ko sta prišla do mene → k meni sta rekla: „Zdaj si pa nekaj zaželi.“	9
7	Raba od in iz	Od → Iz prejšnjega zakona pa mu je ostala samo hčerka	8
8	Raba zraven namesto poleg	Kadar se vozim v službo, zraven → poleg mene ni nikogar, ki bi spregovoril besedo.	8
9	Raba v namesto o	Pregovorov v → o ljubezni je veliko.	7
10	Raba za namesto v	Vsedla sem se za → v svojo klop.	7
11	Raba čez in skozi	Odlomek govori o veliki članski družini, ki se zelo teško prebijajo čez → skozi življenje.	6
12	Raba na in nad	Ofelija ne bi smela pustiti, da ima njen oče tolikšen vpliv nad → na njo	4
13	Raba za namesto do	za → Do ljudi, katerim je vladar, nima spoštovanja in so zanj manj vredni	4
14	Raba z namesto do	To si razlagam tako, da se želi s tem poudariti odnos posameznika z Bogom → do Boga , ne pa tudi družbeni odnosi.	3
15	Raba iz in zaradi	Antigona mu pove, da je zakon kršila zaradi → iz ljubezni do brata.	3
16	Raba k namesto pri	Kocmur se je Likarju pridružil k → pri igri.	3
17	Raba zraven namesto ob	Danes imajo reklame največkrat zraven večjega besedila → ob večjem besedilu tudi droben tisk	3

Zanimivi so popravki, pri katerih je predlog popravljen zaradi vezave z določenim glagolom ali samostalnikom. Učitelji popravljajo predloge *pod*, *pred*, *med* pri glagolih *uvrščati*, *prištevat*, *spadati*, kjer je najpogostejša popravljena kolokacija *spadati pod* (53), v povezavi s samostalnikom *mnenje* pa so različni predlogi trikrat popravljani v *o* (54).

53 Antigona spada **pod** → **v** dramatiko, vrsta pa je tragedija.

54 Moje mnenje **pri** → **o** tem dogodku je, da je gospa resno jokala

Ostale napake predlogov so bolj sporadične ali omejene na posamezne učence. V nekaj primerih gre za učiteljevo slogovno izbiro, npr. pri menjavi *kot* z *od* (55) ali pri menjavi zaznamovanega *tekem* z *med* (56).

55 On ima večje probleme **kot** → **od** pa tega, da po mestu letajo Italijani in kasneje Nemci.

56 **Tekom** tega pogovora → **Med** pogovorom je spoznal,

4.2.5 Napake veznikov

Napak veznikov je med napakami besedišča dobrih 11 %. Skoraj 20 % napak veznikov je menjav *in* in *ter*, kjer gre verjetno pri učencih tudi za željo po bolj knjižnem izražanju (57).

57 Bil je v vojni, kjer je pobijal ženske, otroke, požigal hiše **ter** → **in** kradel.

Veliko težav imajo učenci pri rabi protivnih veznikov *vendar*, *ampak*, *toda*, *temveč*, *a*, *pa* (58), katerih popravki imajo za posledico pogosto spremenjen besedni red. Mednje sodi tudi nekoliko bolj pogovorno zaznamovani *samo* (59).

58 Tako je bil Kreon kaznovan in je moral živeti dalje, **ampak** → **vendar** sam in žalosten do konca svojega življenja.

59 Kaznuje ga s težkim delom, ponudi mu celo zaščito, **samo** → **a** vedno bo taval zgubljen po svetu.

Pojavljajo se še druge zamenjave sinonimnih veznikov, denimo *saj*, *ker* in *kajti* (60), *čprav* in *četudi* (61).

60 Pri reševanju lastnih težav sem dokaj uspešen, **saj** → **ker** nimam dosti težav, ki bi jih moral reševati.

61 odgovoril ji je, da je zelo trmasta in zato ker je kršila njegov zakon bo obsojena na smrt pa **če prav** → **četudi** je zaročenka njegovega sina.

Priredni vezniki so večkrat nadomeščeni z oziralnim zaimkom (62), s protivnimi (63) ali ločnimi vezniki (64).

62 Po pouku grem k mami, **in** → **ki** me odpelje domov.

63 Tako tudi Antigona naleti na nerazumevanje **in** → **a** na vsak način želi opraviti zadano nalogo.

64 Ob 20:00 si prižgem TV ter pogledam kakšen film ali oddajo. Po filmu **in** → **ali** oddaji se spravim spat.

Učenci imajo manj pogoste težave tudi z rabo *da* in *naj* v poročanem govoru (65), z zamenjevanjem *kadar* in *ko* (66) in z večbesednim veznikom *tako ... kot* (67).

65 Vest mi pravi, **da** → **naj** maščujem smrt očeta, ki ga imam rad.

66 **Kadar** → **Ko** sta se oče in mati slišala, sta bila zelo vesela, ker se je oče spet vrnil domov.

67 Poleg tega, da sta tako Hamlet **in** → **kot** Ofelija krivca za njuno ljubezen, so krivi tudi drugi nastopajoči

Učenci zamenjujejo tudi vzročne in posledične (68) ali druge odvisnike, denimo časovne (69).

68 Sama je že vnaprej dobro vedela kaj hoče, **saj** → **torej** je bil to edini način, da dobi kar je hotela.

69 **Ko** → **Ker** je Polifem oslepel, se je močno zameril Pozejdonu.

Manj pogoste menjave učiteljev so slogovne. Taka je menjava pogovornih struktur z enobesednim veznikom (70), *brez da bi z ne da bi* (71) ali *če z ali* v poročanem govoru (72).

- 70 Svoj drugi jaz, osebni jaz z vsemi čustvi, vrednotami pa v sebi zatira, saj bi **če ne** → **sicer** izpustil ujetnico in bi s tem ogrozil sebe in svoje kolege.
- 71 in ga **brez** → **ne** da bi straža opazila, zasula z zemljo.
- 72 Odisej je hotel poslušati petje božanskih siren, da bi se prepričal na lastna ušesa **če** → **ali** res tako lepo in pogubno pojejo

4.2.6 Napake glagolov

Napake glagolov so delno problem neustreznega pomena ali slogovne izbire (prim. razdelka 4.2.1 in 4.2.2), bolj sistemskih napak, ki jih je lažje združiti v nedvoumne kategorije, pa je okrog 9 %.

Dobra tretjina napak glagolov je povezana z zamenjavo predpon. Najbolj težavni, verjetno tudi zaradi podobnega zvena, sta predponi iz- in z- (73), pojavljajo pa se še drugi primeri, pri katerih gre lahko tudi za drugačen pomen glagola: *napiti* namesto *opiti*, *povprašati* namesto *vprašati*, *pretrgati* namesto *strgati*, *zvezati* namesto *privezati*, *ustrašiti* namesto *prestrašiti* ipd.

73 **zgleдалa** → **izgleдалa** je srečna!

Skoraj tretjina napak glagolov je povezana z naklonskimi izrazi, večinoma z zamenjavami *morati* oz. *moči* (74). Pojavljajo se tudi druge zamenjave, npr. *hoteti* oz. *želeti* (75).

74 da mu je posodila denar in da ji ga **more** → **mora** vrniti ali se z njo poročiti.

75 **Hoče** → **Želi** tudi imeti zvezo s hčerko profesorja De Martinisa, a oba zaradi slabih izkušenj, ki sta jih v življenju doživljala, zveze ne moreta zaživeti.

Četrtno napak glagolov predstavlja raba neustreznega glagolskega vida (76), ki bi jo sicer lahko uvrstili tudi k oblikoslovnim napakam. Le nekajkrat pa se pojavijo napake zaradi zamenjevanja glagolov *vedeti*, *znati* oz. *poznati* (77).

76 Veliko je **ukradla** → **kradla**, znala pa je bit prijazna.

77 V šoli so jo zmerjali, saj ni **vedla** → **znala** slovenskega jezika.

4.2.7 Napake pridevnikov

Z bolj sistemskimi napakami pridevnikov, ki niso opisane v razdelkih 4.2.1 in 4.2.2, sta povezana manj kot 2 % napak besedišča. Najpogostejša tipa sta neustrezna raba količinskih in mernih pridevnikov ali prislovov, npr. *velik*, *visok*, *številen* (78), in neustrezna raba pridevnikov oz. zaimkov *isti*, *enak*, *podoben* (79). Nekoliko redkeje sta bila neustrezno uporabljena pridevnik *cel* oz. zaimek *ves* (80).

78 Plače so zelo **majhne** → **nizke**, kar se mi zdi tudi žalostno

79 **Isto** → **Enako** se je dogajalo Črtomirju.

80 Romeo **cel** → **ves** začuden pogleda Julijo in jo močno objame.

Nekaj napak pridevnikov je uvrščenih k besedišču, čeprav deloma sodijo v oblikoslovje. To velja za primere, pri katerih je učitelj iz vsebinskih razlogov popravil osnovno obliko pridevnika v presežno ali obratno (81). Tudi neustrezna raba določne oz. nedoločne oblike pridevnika načeloma spada med oblikoslovne napake, izjema pa so primeri, pri katerih gre za drugačno osnovo besede (*majhen oz. mali*) ali drugačen pomen (82).

81 Spremembe so lahko dobre lahko pa tudi **slabše** → **slabe**

82 Trudil se je, da bi čim več ljudi govorilo **materin** → **materni** jezik in se pri tem močno namučil.

Enako kot pri glagolih se pri pridevniki pojavljajo napake predponskih obrazil (83), ki so pogoste tudi pri samostalniki (84).

83 Saj v taki družini otrok ne dobi dovolj pozornosti, še nasprotno tukaj je **vidna** → **razvidna** velika ignoranca do Lojzke.

84 Gregor je partizan, ki na rutinskem **sprehodu** → **obhodu** opazi mladenko

4.2.8 Napake prislovov in členkov

Napak prislovov je med napakami besedišča v Šolarju le dober odstotek. V skoraj polovici primerov gre za neustrezno rabo prislovov *veliko oz. več, bolj, zelo* (85). Nekaj jih je povezanih z izbiro krajevnega prislova (86) oz. z izbiro *popoldan* ali *popoldne* (87). Učenci imajo težave z izražanjem trajanja (88) ali z rabo *večino* namesto *večinoma*, nekateri popravki pa so bolj slogovni (89).

85 da se bomo bolj imeli radi ter **več** → **bolj** spoštovali.

86 **ven** → **zunaj** je pa rahlo naletaval sneg.

87 **Dopoldan** → **Dopoldne** je imel šolo,

88 Moje mnenje o televizije je to, da lahko jo gledaš kot sprostitvev prostem času ampak ne **dolgotrajno** → **dolgo časa**

89 Ko je slišala, da je Hajmon mrtev se je ubila **tudi** → **še** ona.

Omeniti je treba tudi sicer redke napake členkov in prostih morfemov. Dvakrat so učenci zamešali *ni* oz. *ne* (90), dvakrat pa *si* oz. *se* (91).

90 Sigurno **ne** → **ni** na naši, temveč je na denarni.

91 Hajmonova mati Evridika pa **si** → **se** zaradi žalosti ob sinovi smrti zabode

4.2.9 Napačne oznake in napake pravopisa

Manj kot odstotek je med napakami besedišča primerov, ki so bili neustrezno označeni in so dejansko napake črkovanja. Med njimi so primeri, v katerih je manjkajoča, odvečna ali neustrezna črka, in so očitno posledica lapsusa (92). Ob tem je treba upoštevati, da so tovrstne napake lahko nastale tudi pri transkribiranju besedil zaradi slabo fotokopiranega gradiva ali drugih, nejezikovnih razlogov (gl. Rozman

idr. 2012: 23). Lahko pa gre za neustrezno zapisane glasove (93), ki so pogost pravopisni problem (gl. razdelek 3.5 v 3. poglavju), in s tem dejansko za napačne oznake.

92 David je **človeška** z masko udaril, vendar se mu ni nič zgodilo.

93 V obeh zgodbah se zrcalijo okoliščine in problemi **socjalizma** pri nas.

4.3 Sklep

Analiza več kot 3300 primerov napak besedišča v korpusu Šolar je pokazala, da so določene težave pri pisanju, kot so raba povratnih oz. svojilnih zaimkov ali nekaterih predlogov, značilne za večino slovenske šolske populacije. Učenci imajo težave tudi z modalnima glagoloma *morati* in *moči* ter z izbiro besed iz iste besedne družine z različnimi predponami.

Na ravni besedišča so za jezikoslovne raziskave zanimivi tudi nekateri slogovni popravki. Izpostaviti velja predvsem popravljanje pogovornih besed, nadomeščanje s sopomenkami, s katerimi skušajo učitelji odpraviti ponavljanje v besedilih učencev, in uporabo bolj specifičnih izrazov, ki naj bi pri manj jasnem referiranju povečala razumljivost besedil. Prav tako so slogovno utemeljeni popravki oziralnih in drugih zaimkov ali zamenjave veznikov.

Napake besedišča v korpusu Šolar so bile dobro opredeljene in zamejene še pred začetkom označevanja, tako da med delom večji popravki niso bili potrebni. Tudi neizbežno prekrivanje določenih kategorij je bilo zaradi jasno postavljenih pravil razmeroma obvladljivo. Treba pa se je zavedati, da je analiza predstavljenega gradiva temeljila zgolj na tistih napakah, ki so jih zaznali učitelji osnovnih in srednjih šol. Ker je sam koncept napake vedno do določene mere subjektiven in odvisen od tega, kaj definiramo kot napako, je tudi učiteljska percepcija napak lahko bolj ali manj stroga od povprečja.

5 Napake oblike

ANA ZWITTER VITEZ

5.1 Uvod

V tem poglavju so predstavljeni rezultati analize napak, ki jih lahko opišemo kot „neustrezne oblike določene besede“. Pri večini napak jezikovnemu standardu ne ustreza sklon, število ali čas določene besedne oblike glede na jezikovni kontekst, v katerega je oblika umeščena.

V korpusu Šolar napake oblike niso razvrščene v podtype, zato se je bilo treba odločiti za večnivojsko kategorizacijo napak, ki je potekala od najbolj grobega razvrščanja do končnega opredeljevanja posameznih jezikovnih težav. Nekatero končne kategorije so homogene, zato ponekod zadošča že en primer, da si bralec ustvari jasno predstavo o tem, kaj učencu povzroča jezikovne preglavice (npr. kategorija *nedoločnik ali namenilnik*). Posamezne kategorije pa so precej raznorodne, zato je potrebnih več primerov, da dobimo ustrezno predstavo o krovni jezikovni težavi (npr. kategorija *ujemanje povedka z osebkom*).

V nadaljevanju je najprej predstavljena statistika najpogostejših oblikovnih napak ter postopek razvrščanja in poimenovanja napak. Sledi predstavitev izbranih kategorij napak, tako homogenih kot raznorodnih. Na koncu poglavja je ponujenih nekaj predlogov za izboljšavo metodologije razvrščanja in združevanja različnih kategorij posameznih oblikovnih napak.

5.2 Obravnavane napake

V korpusu Šolar je zabeleženih 3618 napak oblike, od tega je 1241 povezanih napak, ki so obravnavane ločeno (gl. 7. poglavje). Zabeležene napake nimajo podtipov, saj smo se želeli izogniti prevelikim razlikam v interpretacijah študentov transkriptorjev.

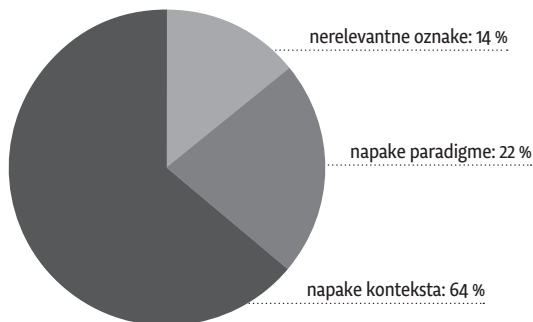
Prvi pregled celotne skupine 2377 obravnavanih napak je pokazal, da delitev na besedne vrste ne bo dovolj. Znotraj posameznih besednih vrst je namreč izstopala še ena lastnost označenih napak, na podlagi katere smo jih ne glede na besedno vrsto razdelili v dve skupini:

- tiste, ki so v standardni slovenščini same na sebi nesprejemljive (*vi meniste*), in
- tiste, ki niso sprejemljive v danem slovničnem kontekstu (*onidve sta šle*).

Zato smo pri prvem razvrščanju napake oblike kategorizirali v napake glede na paradigmo ali glede na kontekst, znotraj teh dveh kategorij pa smo jih razvrstili glede na besedne vrste in vzrok, zaradi

katerega je prišlo do določene jezikovne napake. Slika 4 prikazuje razmerje med napakami paradigme in konteksta ter delež za kategorijo napak oblike nerelevantnih oznak (npr. *ko Klavdij prejme pismo od Hamleta* → *Ko je Klavdij prejel pismo od Hamleta*).

Slika 4: Napake paradigme in konteksta



Končni seznam 29 najpogostejših kategorij napak oblike predstavlja Tabela 27.

Tabela 27: Napake oblike

Št.	Kategorija	Primer	Število napak
1	nedoločnik ali namenilnik	„mogoče pa bo prišel to tekmo gledati → gledat kakšen menedžer.“	274
2	ujemanje samostalnika	Po nekaj minut → minutah hoje zaslišimo glasne in lomastenje.	271
3	ujemanje povedka z osebkom	Mnenje okolice se je spremenila → spremenilo , saj so videli, da so otroci pogumni, delovni in samozavestni	178
4	glagol v dvojini	Moja mama in očka so šli → sta šla narazen jast in brat pa smo ostali → sva ostala pri očki.	115
5	ujemanje zaimka	Mislím da prijateljstvo z njem → njim je možno, ker je pošten moški	105
6	samostalniški zaimki v roditeljskem	Antígona pa se ni bila pripravljena prilagoditi, zato jo → je ni čakal dober konec.	75
7	samostalniški zaimki v mestniku ali dajalniku	Moje mišljenje o Meti je, da je ona res junakinja in da je ves čas delala prav in v vsemu → vsem ji dam prav	66
8	sklanjanje samostalnika <i>mati</i>	Prešeren je lepo opisal nezakonsko mati → mater in njene težave s svetom.	61
9	ujemanje pridevnika s samostalnikom (spol, število, sklon)	In če iz tega sklepamo, lahko ugotovimo, da je vsaka sprememba za nekaj dobrega → dobra	61
10	zahtevnejši primeri sklanjanja	Mami je dala gospej → gospe denar in že sva odšla z kužkom.	55

11	osebni zaimек v dvojini	Pomembni sta zato ker je iz njih → njiju zelo lepo razviden motiv odpuščanja in Božjega usmiljenja do svojih sinov.	53
12	glagol v sedanjiku	Antigona se je kljub temu odločila, da Polinejka pokopa → pokoplje .	44
13	sklanjanje samostalnika otrok	Oče z otroci → otroki krene iskati mamo, da še njo seznanj s prihodom	39
14	oblika pridevnika	Bilo je nekega toplega, poletnega popoldneva, ki ga žal nisem izkoriščal zunaj, ampak sem kot zasvojeni → zasvojen buljil v računalnik.	38
15	pogovorni pretekli deležnik	Zato se je župnik odločo → odločil da morata obe družini zvedet resnico	37
16	sklanjanje tujih lastnih imen	Na koncu je izbral še nekaj prostovoljcev da so zaigrali igro o Romeo → Romeu in Juliji.	36
17	ujemanje pridevniškega zaimka s samostalnikom (srednji spol)	Ta izkušnja zaznamuje človeka skozi vso → vse življenje.	29
18	pretekli deležnik glagolov z več osnovami	Jedel je listje, zemljo in vse mehke stvari, ki jih je najdel → našel .	28
19	pogovorna končnica pri sklanjanju	Matiček je živahen, zabaven in zelo predrzen saj večkrat hodi po robi → robu , ki bi mu ob padcu preprečil poroko z Nežko.	27
20	pridevniški zaimек v dvojini	Starša od te sošolke sta se zelo prepirala, edini razlog za obstoj zveze je bila njihova → njuna hčerka	25
21	oblika prislova	Moja starša sta 10. let razlike in se dobar → dobro razumeta.	21
22	svojltni pridevniki iz tujih ali lastnih imen	Ofeljeva → Ofeljina smrt je povzročala bolečine za Hamleta, vendar se je najprej obadal s svojim maščevanjem.	20
23	sedanjik v 2. in 3. osebi dvojine in množine	da ga starša velikokrat pretepesta → pretepeta in, da mu nikoli ne stojista → stojita ob strani.	19
24	preglas	Živeli smo v hiši z babi, dedijem, stricom → stricem in teto.	19
25	oziralni zaimki	Dojela sem, da je ljubezen nekaj, kar → česar se ne da opisati z besedami.	19
26	stopnjevanje pridevnikov	Veliko bolj prepričljivejša → prepričljiva so seveda dejanja Črtomira, ki dela dobro in je zame res pravi junak	18
27	sklanjanje števnikov	Naša družba je prvič v večih → več letih dobila „vodjo“	18
28	redukcija pri sklanjanju	Predvsem zaradi predčasnega ali začetka → začetka zveze že v mladih letih, kar močno vpliva na njuno prihodnost.	17
29	glagol v 1. osebi dvojine	Kadarkoli z bratom naredima → narediva kaj narobe, naju mati obtožuje, da si ni tega zaslužila kar počnema → počneva	16

Med navedenimi kategorijami se jih večina nanaša na konkreten jezikovni problem (npr. dvojina pri glagolu), nekatere kategorije (ujemanje povedka z osebkom) pa zajemajo bolj raznorodne primere in posledično zahtevajo podrobnejšo opredelitev.

V nadaljevanju predstavljamo analizo najpogostejših napak, za katere se je izkazalo naslednje:

- jezikovne težave so pogosto povezane z rabo določene oblike v danem slovničnem kontekstu, pri čemer posebej izstopajo raba dvojine (*sta šle*) in rodilnika (*ne bo jo zapustil*), zamenjevanje mestnika in dajalnika (*pri njemu*) ter nedoločnika in namenilnika (*grem smučati*);
- nekatere težave izvirajo iz dejstva, da učenci tvorijo za jezikovni standard nesprejemljive oblike, kar je razvidno iz napak, vezanih na glasovne premene (*z mečom*), zahtevnejše primere sklanjanja (*gospej je dala denar*) ali pokrajinsko specifične oblike (*vi meniste*).

5.2.1 Nedoločnik ali namenilnik

Učenci pogosto zamenjujejo nedoločniško in namenilniško neosebno glagolsko obliko. Najpogostejša napaka je uporaba t. i. kratkega nedoločnika za naklonskimi glagoli, ki izražajo željo, nujnost, dovoljenje. Zdi se, da gre za vpliv govornih redukcij v neformalnih sporazumevalnih situacijah, ko govorci redko uporabljajo dolgi nedoločnik.

94 Menim, da tudi starši bi se mogli bolj **posvečat** → **posvečati** svojim otrokom

95 ker se začne po tem dejanju se začne vse **razpletat** → **razpletati**.

96 Avtor želi z idejo hlapčevstva **predstavit** → **predstaviti** neumnost in slepo zaupanje ljudi

5.2.2 Ujemanje samostalnika

Gre za zelo obsežno in precej raznorodno kategorijo, ki zajema težave učencev pri rabi samostalnika glede na njegovo stavčno vlogo ali položaj znotraj nje.

97 In s tem mogoče učenec ne razume prav pomena **kazen** → **kazni** in zakaj jo je dobil.

98 Po nekaj **minut** → **minutah** hoje zaslišimo glasne in lomstenje.

99 Tako kot v mnogih besedilih, je tudi v Sofoklejevem delu **Antigoni** → **Antigona** ljubezen ena izmed glavnih tematik.

Zgledi (97), (98) in (99) predstavljajo različne napake oblike, katerih skupna točka je ujemanje samostalnika znotraj hierarhije stavčnega člena. V zgledu (97) gre za predmet v rodilniku (*ne razumeti koga ali česa*), sestavljen iz samostalniške zveze *pomen kazni*, ki zahteva rodilniško obliko desnega prilastka. V zgledu (98) imamo opravka s časovnim prislovnim določilom, sestavljenim iz sintagme *po nekaj*

minutah hoje, ki zahteva mestniško strukturo (*po kom ali čem*). Zgled (99) predstavlja težavo podredne samostalniške zveze v prislovnem določilu, v katerem ime *Antigona* igra vlogo prilastka v imenovalniku, učenec pa ga je postavil v mestniško obliko.

Zelo pogosta je raba samostalnika v tožilniku namesto v roditeljskem ob zanikanih glagolih, zato bi jo bilo v prihodnje smiselno obravnavati kot posebno kategorijo (100). Nekatere napačne oblike samostalnikov so pogosto posledica težav z vezljivostjo glagolov (101). Zgledi 102 do 106 združujejo težave s samostalnikom v raznih predložnih zvezah.

100 Vendar ona ni poslušala **starše** → **staršev** in zato se je odselila k njemu.

101 Nekega dne, ko so **Urško** → **Urški** dvorili snubci, pa se je zgodilo ...

102 dostikrat so jih pretepli z **bičih** → **biči** in še veliko.

103 tam jo je pregledal in cepil proti **boleznimi** → **boleznim**.

104 Za ljubezen se je vredno boriti, če je ljubezen iz obeh **straneh** → **strani** enako močna.

105 Tega bi se morali zavedati vsi, saj smo vsi svobodni pri **večina odločitev** → **večini odločitev**.

106 ko se proda sovražniku – kralju in postavila ljubezen do očeta – Polonija, pred **ljubeznijo** → **ljubezen** do Hamleta.

5.2.3 Ujemanje povedka z osebkom

Problem ujemanja povedka z osebkom je raznorodен zato, ker se lahko nanaša na katero koli lastnost, ki jo osebek lahko določa povedku (torej na osebo, število, spol). V korpusu Šolar se tovrstne napake pojavljajo na vseh od naštetih ravni, najpogostejšo težavo pa predstavlja ujemanje povedka s številom osebka, zlasti če se osebek pojavi za povedkom oz. na koncu stavka:

107 Tako kot Gregor je tudi Berkovo vojak, na čigar mišljenje **vplivata** → **vplivajo** vojna in grozovita dejanja, ki se dogajajo zaradi nje.

108 Ljubezen lahko **ovira** → **ovirajo** družina, spol, narodnost.

109 in vendar na koncu **zmaga** → **zmagajo** dobro, razum in pisateljeve vrednote.

Precej težav učencem povzroča tudi ujemanje glagola s spolom osebka, ko ta zavzema mesto za povedkom. Posebej težavni so primeri s samostalniki srednjega spola (*dekle, obdobje* ipd.).

110 Na odločitev za tako obliko je **imel** → **imelo** velik vpliv obdobje, ker so vsi pisali zelo čustvena dela

111 in tako spoznal njegovo hčerko Noro, ki je **bilo** → **bila** lepo in inteligentno dekle.

5.2.4 Glagol v dvojini

Pri kategoriji glagola v dvojini učenci uporabljajo obstoječe glagolske paradigme, vendar ne upoštevajo ujemanja povedka z osebkom po številu (112). Ta težnja se v številnih primerih ne nanaša na celotno glagolsko strukturo, temveč samo na pretekli deležnik. Ker se to dogaja predvsem pri ženskih oblikah, predvidevamo, da gre za vpliv jezikovne rabe v neformalnih situacijah (113). V nekaterih primerih neustrezne rabe dvojine gre za težavo neujemanja povedka z osebkom, ko osebek stoji na koncu povedi oz. za povedkom (114).

- 112 Moja mama in očka **so šli** → **sta šla** narazen jast in brat pa **smo ostali** → **sva ostala** pri očki.
- 113 Obe sta **dokazale** → **dokazali**, da se da s pravo voljo in željo doseči marsikaj.
- 114 in v njem **je prevladovalo** → **sta prevladovala** samo še sovrašтво in jeza.

5.2.5 Samostalniški zaimki v roditeljski

Velika večina primerov obravnavane kategorije je vezana na uporabo tožilniške namesto roditeljske oblike zaimka v zanikanih povedih. Učenci imajo pri tem težavo z osebnimi in neosebni zaimki ter njihovimi naglasnimi in naslonskimi oblikami.

- 115 Antigona pa se ni bila pripravljena prilagoditi, zato **jo** → **je** ni čakal dober konec.
- 116 Izhaja iz bolj revne družine, zato si **marsikaj** → **marsičesa** nemore privoščiti.
- 117 sprva ne morejo doumeti, da bi lahko bila kriva **kaj** → **česa** takega.
- 118 **To** → **Tega** ga je naučilo prijateljstvo z Antonom

V nekaterih primerih neustrezna raba samostalniških zaimkov ni povezana z zanikanim glagolom v povedi, ampak se nanaša na vezljivost glagola:

- 119 Vsi so jo začeli zavračati in se **jo** → **je** sramovati.
- 120 Pa tudi, da bi bilo več pravice, ki **jo** → **je** v današnjem svetu primanjkuje

5.2.6 Samostalniški zaimki v mestniku ali dajalniku

V veliki večini primerov napak te kategorije so učenci uporabili obliko zaimka v tretjem sklonu namesto v petem sklonu (121, 122, 123). Težava pa je prisotna tudi v obratni smeri: v nekaj primerih so učenci

namesto zaimka v tretjem sklonu uporabili zaimek v petem sklonu. Predvidevamo, da gre za hiperkorektnost (124, 125).

- 121 Ima okrogel obraz, na **kateremu** → **katerem** nima nobene posebnosti.
- 122 in da je ves čas delala prav in v **vsemu** → **vsem** ji dam prav.
- 123 Pri **njemu** → **njem** se je počutila zelo dobro.
- 124 in mi je povedal, kako veliko sem pripomorel k **tem** → **temu**.
- 125 Pri družini Khallan je sin kljub **vsem** → **vsemu** spoštoval starše

5.2.7 Sklanjanje samostalnika *mati*

Samostalnik *mati* ima v spisih šolske mladine izredno visoko pogostost pojavljanja in tudi izredno visoko pogostost zabeležene napačne rabe. Ena najbolj tipičnih težav, povezanih s sklanjanjem tega samostalnika, je raba imenovalniške oblike namesto tožilnika, dajalnika ali orodnika (126, 127, 128). Veliko zabeleženih napak se nanaša na rabo neparadigmatske oblike *matera* (129, 130). V nekaterih primerih pa se učenec odloči za tožilniško obliko namesto imenovalniške (131).

- 126 Prešeren je lepo opisal nezakonsko **mati** → **mater** in njene težave s svetom.
- 127 **Mati** → **Materi** Barbari je bilo zelo hudo
- 128 Domol je odšla z **mati** → **materjo** in pri njej tudi rodila prvega otroka.
- 129 Kralj Ojdip je po prerokbi in naključju dogodkov ubil svojega očeta ter se poročil s svojo **matero** → **materjo**.
- 130 Zelo očitno je da je **matera** → **mati** varala svojega moža
- 131 Oče in **mater** → **mati** sta bila razjokana od veselja

5.2.8 Zahtevnejši primeri sklanjanja

V formalnem položaju pisnega sporočanja, kot je pisanje šolskih spisov, povzročajo učencem največ zagat pogosti samostalniki s specifičnimi sklanjatvenimi vzorci. Eden najtežavnejših je tako samostalnik *hči*, ki ga v neformalnih okoliščinah spontanega govora nadomešča pogovorna oblika *hčerka* (tudi *hčera*), v pisnem sporočanju pa prihaja do zamenjevanja imenovalnika in tožilnika (132). Zgled 133 pa je verjetno posledica hiperkorektnosti, saj je učenec namesto imenovalniške oblike uporabil tožilniško.

- 132 ker *mati* Herodiada uporabi svojo **hči** → **hčer** Salomo kot predmet maščevanja
- 133 in ministrova **hčer** → **hči**, ki je bila vedno ubogljiva, je to tudi naredila.

Zahtevnejše primere sklanjanja predstavljajo določene oblike samostalnika *gospa*. Analiza korpusa je pokazala, da sta za učence problematični predvsem obliki dajalnika ednine (*gospe*) in orodnika

množine (z *gospemi*), pri čemer učenci tvorijo pogovorne, v standardni slovenščini nesprejemljive oblike:

134 Mama je dala **gospaj** → **gospe** denar in že sva odšla s kužkom.

135 na kavi, s tremi ljubljanskimi **gospami** → **gospemi**.

Učencem pogosto povzročajo težave samostalniki, ki spadajo med t. i. posebnosti prve moške sklanjatve in pri katerih v rodilniku prihaja do naglasne in glasovne premene (npr. *grad, mir, strah*).

136 ko se je odločil pobegniti iz **grada** → **gradu** in premagati Valjhuna

137 Preklinjal je v vseh jezikih zato pod domačo streho ni bilo božjega **mira** → **miru**.

Težavni so tudi ostali primeri posebnosti prve moške sklanjatve (npr. *mož, las, zob*) v množini, pri čemer sta problematična predvsem imenovalnik in rodilnik:

138 one samo govorijo o tem koliko imajo njihovi **moži** → **možje** premoženja.

139 Bile so zveste do svojih **možev** → **mož**.

Samostalnik *človek* povzroča učencem največ preglavic pri dvojini, ki v rodilniku in mestniku zahteva osnovo množinske oblike, učenci pa sledijo dvojinškemu vzorcu in tvorijo neparadigmatske oblike:

140 Lojzka je bila primorana nemočno opazovati nečedna dejanja svoje matere in očeta, skratka **človekov** → **ljudi**, ki sta ji podarila življenje

Raba množinske oblike za parna poimenovanja predstavlja posebej trd oreh pri poimenovanju oseb (*starši*) zaradi konkretne predstave o dveh označencih (141). Dodatno težavo predstavlja preglas (142):

141 Včeraj so v Sevnici osnovno in srednjo šolci s svojimi **staršimi** → **starši** očistili opčino Sevnico.

142 Ko sem šla s **staršoma** → **staršema** v nakupovalni center sem srečala Sonjo

5.2.9 Osebni zaimek v dvojini

Pri osebnih zaimkih učenci uporabljajo sicer obstoječe, a kontekstualno neustrezne oblike. Najpogosteje gre za zamenjevanje dvojine in množine osebnega zaimka v tretji osebi, pa naj gre za naslonsko (143) ali naglasno obliko (144). Tudi pri osebnih zaimkih v prvi osebi je najpogostejša težava raba množinske namesto dvojinške oblike (145, 146).

143 Iščemo dve dekleti, ki sta zbežali od domo. Če **jih** → **ju** bo kdo videl, naj to sporoči policiji.

144 Pomembni sta zato ker je iz **njih** → **njiju** zelo lepo razviden motiv odpuščanja

145 Tako sva se že kar nekajkrat sva se spustila, nato **nam** → **nama** je to smučiče postalo že dolgočasno

146 XlmeX **nas** → **naju** je vprašal, če va midva odprla šotorske zadrge.

5.2.10 Glagol v sedanjiku

Sprejanje glagolov v sedanjiku je glede na analizirane napake za učence najbolj težavno pri zelo pogosto rabljenih, a oblikovno kompleksnih pomožnih in naklonskih glagolih (*iti, biti, moči*) (147). Preglavice povzročajo tudi nekateri glagoli z dvema osnovama (*upogibati, kopati, nakopati*), pri katerih učenci pogosto nadaljujejo sprepatveni vzorec glede na nedoločniško osnovo (148, 149). Nekatero glagole na -iti v drugi ali tretji osebi ednine učenci spregajo z obrazilom -eš ali -e namesto -i- (150, 151). V obravnavani sklop je smiselno vključiti tudi težavo, ki se nanaša na ujemanje povedka z osebkom ženskega spola v dvojini (152, 153, 154). Učenci namreč pogosto uporabljajo neparadigmatske oblike *sve, imave, smeve* ipd.

147 ali **bojo** → **bodo** šli z njim v boj ali pa **bojo** → **bodo** odšli domov

148 Antigona se je kljub temu odločila, da Polinejka **pokopa**
→ **pokoplje**.

149 Ta lik predstavlja drevo, ki ne **upogiblja** → **upogiblje** vej.

150 Ali se **spomneš** → **spomniš** tistih dni, ko sva se spoznala?

151 Herod na koncu **izpolne** → **izpolni** svojo obljubo, vendar se zgodba za Salomo tragično konča.

152 Mislili **sve** → **sva**, da **bove** → **bova** zelo hitro prišli do Amerike in na hitro **obogatele** → **obogateli**

153 Žabici odvrneta: **nimave** → **Nimava** ne muhe ne črviča.

154 Starša sta se odločila, da se nikoli v življenju **nesmeve** → **ne smeva** več videti

5.2.11 Sklanjanje samostalnika *otrok*

Samostalnik *otrok* predstavlja težave predvsem v množini: učenci napačno obliko najpogosteje zapisujejo pri orodniku množine s predlogom *z* (155), pa tudi pri manj tipičnih predlogih orodnika (156). Pogost je tudi pojav hiperkorektnosti, ko učenci v imenovalniku množine uporabijo obliko *otroki* (157).

155 Oče z **otroci** → **otroki** krene iskati mamo, da še njo seznanj s prihodom

156 Odnosi med starši in **otroci** → **otroki** bi morali biti pristni.

157 Drugi **otroki** → **otroci** so jih zmerjali in jih ves čas spraševali čigavi so oni

5.2.12 Oblika pridevnika

Pri napakah te kategorije gre za neustrezno določno ali nedoločno obliko kakovostnega pridevnika ali pa za neustrezno rabo vrstnega pridevnika. Večina primerov spada v prvo skupino (*nazadnje je nosila zelenobel črtasti pulover*). Ker so zgledi, kot sta (158) in (159), odprli tudi vprašanje, ali gre za regionalno pogojeno težavo, smo z analizo korpusa govornjene slovenščine (www.korpus-gos.net) ugotovili, da je pojav izrazito prisoten v regijah severovzhodne Slovenije, drugod pa praktično ni zaznan.

158 ampak sem kot **zasvojeni** → **zasvojen** buljil v računalnik.

159 Pod vplivom drugih, predvsem Polonija in kralja ima njuna zgodba **žalostni** → **žalosten** konec.

V nekaterih primerih gre za neupoštevanje pridevniških kategorij:

160 Postal je tudi **časten** → **častni** doktor Ljubljanske univerze.

161 Tam je začel **star ate** → **stari ata** delati malo prej ko je dopolnal 18 let.

162 saj je pred nas postavljen **pogumen** → **pogumni** bojevnik Črtomir, ki se bori za svoj narod.

5.3 Sklep

V poglavju smo predstavili naslednje faze analize 2377 napak oblike v korpusu Šolar:

- metodologijo analize,
- postopek razvrščanja označenih napak v posamezne kategorije,
- ugotovljene težave, razvrščene glede na frekvenco pojavitev v korpusu.

V splošnem lahko ugotovimo, da učencem največ preglavic povzročajo napake konteksta, kar pomeni, da uporabljajo sicer obstoječe oblike standardne slovenščine, ki so v določenem slovničnem kontekstu neustrezne. Pri tem je veliko težav povezanih z rabo dvojine (*sta šle*), roditeljskega (*ne bo jo zapustil*), zamenjevanjem mestnika in dajalnika (*pri njemu*) ter nedoločnika in namenilnika (*grem smučati*). Nekatere napake izvirajo iz dejstva, da učenci tvorijo neparadigmatske oblike, ki so ponavadi vezane na glasovne premene (*z mečom*), zahtevnejše primere sklanjanja (*gospej je dala denar*) ali pokrajinsko specifične oblike (*vi meniste*).

Izkušnja celotnega postopka je pokazala, da je večina na ta način formiranih kategorij homogena in lahko prepoznavna, zaradi česar bodo te kategorije (npr. *nedoločnik ali namenilnik*) neposredno uporabne kot vir za izdelavo posameznih poglavij nastajajoče pedagoške korpusne slovnice. Za posamezne kategorije (npr. *ujemanje samostalnika*) pa je natančen vpogled v končni nabor problemov pokazal, da združujejo več različnih slovničnih težav. Zato bi znotraj kategorije oblikovnih napak predlagali naknadno razčlenjevanje nekaterih problemov.

6 Napake skladnje

MOJCA STRITAR, SARA MOŽE

6.1 Uvod

Napake skladnje se nanašajo na večbesedne enote oz. neustrezne sintagmatske povezave med besedami. Večinoma so omejene na stavčno raven, redkeje pa segajo tudi na nadpovedno raven napak kohezije, diskurzivnih, slogovnih, pragmatičnih napak ipd. Napake skladnje so standardna kategorija napak v klasifikacijah, ki temeljijo na jezikovnih ravninah (Stritar 2012; Požgaj Hadži, Ferbežar 2001; Rozman 2006; Granger 2001; Nicholls 2003; Lüdeling idr. 2005; Boyd 2008; Hana idr. 2010; Izumi, Uchimoto in Isahara 2004; Bušta idr. 2009; Pala, Rychlý in Smrž 2003).

Napake skladnje v korpusu Šolar imajo štiri podtipe: napake besednega reda, izpuščenih podatkov, odvečnih podatkov in strukture. V navodilih za označevanje korpusa napake besednega reda zajemajo (Rozman idr. 2012: 35–36):

- zaporedje naslonk v naslonskem nizu;
- mesto naslonskega niza v povedi;
- zaporedje besed v samostalniški besedni zvezi ali zaporedje sestavin stavčnih členov;
- napake členitve po aktualnosti.

Nekateri besedni red umeščajo k morfosintaktičnim napakam (npr. Požgaj Hadži, Ferbežar 2001: 62), v nekaterih klasifikacijah napak (npr. Balažič Bulc 2004: 85–86) je samostojna kategorija napak, v nekaterih bolj omejenih raziskavah pa je v celoti izenačen s skladnjo (npr. Rozman 2006). Za korpus Šolar, ki naj bi prikazoval bolj celostno jezikovno zmožnost učencev, je izenačevanje skladnje in besednega reda vendarle preozko, zato imajo skladijske napake še tri podtipe.

Podtip izpusta oz. izpuščenih podatkov se nanaša na izpuščeno besedo ali besedno zvezo in na menjavo ločila z besedo, podtip odvečnih podatkov oz. vstavitve pa ravno obratno, na odvečno besedo ali besedno zvezo, odvečen odstavek, menjavo besede z ločilom in tudi vsebinsko neustrezne podatke. Oba podtipa sta osnovna v klasifikacijah, ki napake delijo po površinski strukturi, torej glede na način, kako dejanska, neustrezna oblika v besedilu ne ustreza pričakovani (prim. Dulay, Burt in Krashen 1982: 150–163; James 1998: 106–113). Tako ločevanje pa ne daje zadostnih informacij o tem, kako poteka usvajanje jezika (Ellis 1994: 56), in ker Šolarjeva klasifikacija temelji na jezikovnih ravninah, sta v njej del skladijskih napak.

Podtip neustrezna struktura se nanaša na neustrezno sintagmatsko povezavo v večbesedni enoti, besedni zvezi ali celem stavku, ki

presega raven leksikalne izbire in med drugim zajema (Rozman idr. 2012: 37–38):

- neustrezno oblikovan niz besed;
- zamenjavo samostalnika v rodilniku s svojilnim zaimkom pri izražanju svojine;
- menjavo glagola z naklonskim izrazom;
- menjavo glagolskega načina;
- neustrezno glagolsko vezavo (s predlogom oz. brez predloga);
- popolnoma spremenjen del besedila.

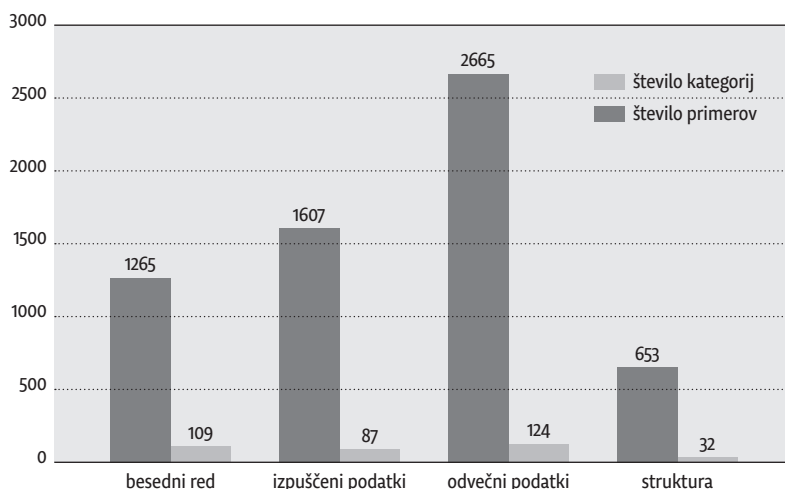
Tako kot pri ostalih tipih napak se tudi napake skladnje ali njihovi podtipi že v zasnovi označevanja prekrivajo z drugimi kategorijami. Neustrezna raba oziralnih zaimkov *ki* oz. *kateri* je uvrščena med napake besedišča, čeprav je povezana tudi s skladenjskimi vprašanji. Podobno velja za nekatere napake pri rabi naklonskih glagolov, ki prav tako spadajo med napake besedišča. Napake, pri katerih učitelj učenčevo ločilo, običajno vejico ali piko, nadomesti s prirednim veznikom ali obratno, bi namesto k skladnji lahko uvrstili k napakam ločil. Primeri, pri katerih učitelj popolnoma spremeni del učenčevega besedila, tako da del prečrta, del pa doda, so sicer del napak skladnje, vendar so označeni kot kombinacija odvečnih in izpuščenih podatkov, lahko pa bi jih umestili tudi k napakam strukture.

Zaradi svoje kompleksnosti so skladenjske napake pogosto del povezanih napak (gl. razdelek 6.2.1.5).

6.2 Rezultati

Napake skladnje v Šolarju predstavljajo več kot 17 % vseh označenih napak, med njimi pa imajo štiri podtipi zelo različne deleže: največ je odvečnih in izpuščenih podatkov, sledijo jim napake besednega reda, najmanj pa je napak strukture (Slika 5). Vsi primeri so bili ročno pregledani in združeni v kategorije glede na različne formalne, predvsem besednovrstne, pa tudi semantične značilnosti. Največ različnih kategorij napak imajo odvečni podatki in napake besednega reda, najmanj pa spet napake strukture. Ker so vsi štiri podtipi tako različni, so v nadaljevanju predstavljeni ločeno.

Slika 5: Število primerov in kategorij napak skladnje



34 Frekvenca je nekoliko višja od podatka v Tabeli 12 v 2. poglavju, saj v ta tip napake sodijo povezane napake besednega reda, ki so bile kot take prepoznane že pri označevanju korpusa, pa tudi tiste, ki so bile kot take določene pri nadaljnji analizi.

35 Popravki v primerih iz Šolarja, v katerih so zaradi preglednosti izpuščene korpusne oznake, so navedeni z odebeljenim tiskom za puščico.

6.2.1 Napake besednega reda

V Šolarju je bilo označenih 1265 napak besednega reda, skupaj torej slabi 4 % vseh označenih napak. Prepoznanih je bilo 109 različnih kategorij, v nadaljevanju predstavljenih glede na del besedila, na katerega se nanašajo (Tabela 28).

Tabela 28: Napake besednega reda

Del besedila, na katerega se nanaša napaka	Število napak
napake v zaporedju stavčnih členov	681
povezane napake besednega reda ³⁴	310
napake v naslonskem nizu	168
napake v zaporedju stavkov	29
napake znotraj stavčnih členov	28
nerelevantni primeri	49
Skupaj	1265

6.2.1.1 Zaporedje stavčnih členov

Klasični tip napake besednega reda je zaporedje stavčnih členov. Takih je slabih 54 % napak besednega reda v Šolarju. Nanašajo se lahko na različne stavčne člene. Opozoriti pa je treba, da gre v teh primerih pogosto za stvar interpretacije oz. slogovne odločitve učitelja, ki je popravljaval besedilo, in ne toliko za dejansko napačno rabo učenca.

Povedek v sedanjiku učitelji v svojih popravkih največkrat prestavljajo za prvi stavčni člen:³⁵

163 Televizija v današnji družbi ima → ima v današnji družbi velik pomen

Učenci imajo težave z mestom nedoločnika (164) ali namenilnika v stavku, posebej pri povratnih glagolih (165).

164 S Hamletom je **biti težko** → **težko biti** prijatelj

165 jaz in XImeX smo **odšle si** → **si odšli** ogledovat okolico.

Včasih učenci začnejo stavek s prislovnim določilom, učitelj pa tja prestavi osebek ali predmet (166). Pretekli deležnik je včasih pred osebkom in predmetom, kar učitelj popravi (167). V nekaj primerih je težaven položaj navezne oblike osebnega zaimka (168).

166 **Vsakič so OI** → **OI so vsakič** v drugem kraju

167 Mene pa je v tistem trenutku, ko je **udaril oče** → **oče udaril** sina

168 Iz odlomka je razvidno da se vsak **zase briga** → **briga zase**,

Prislovno določilo učenci včasih pišejo za povedkom in ga učitelj prestavi predenj (169), kadar se nanaša na naslednji stavčni člen, pa za povedek (170).

169 saj Nežki in ženi **vrjame takoj** → **takoj vrjame**.

170 Danes **bistveno ni** → **ni bistveno** drugače.

Med manj pogostimi, a zanimivimi napakami je besedni red v ozi-
ralnih stavkih s pridevniškimi zaimki:

171 vendar ona je spila strup, katerega **je učinek** → **učinek je da**
deluješ mrtev za nekaj časa

Z zaporedjem stavčnih členov so povezane bolj slogovne napake zaradi neupoštevanja členitve po aktualnosti (172). Te so med napakami besednega reda pogoste in precej raznovrstne, slovnično pa so manj zanimive zaradi subjektivnosti. Od tvorca je „namreč odvisno, katero informacijo bo videl kot pomembnejšo in jo bo zato postavil na konec oz. v remo in katera bo z njegovega stališča manj pomembna in ji bo zato namenjeno mesto v temi, to je na začetku“ (Kranjc 2004: 150; gl. tudi Toporišič 1991: 532-533). Ne zgolj slogovne narave pa so primeri, v katerih učenec besede, na katero se nanaša odvisnik, ne piše na koncu stavka (173).

172 Zaradi tega sem **jaz postal kralj Teb** → **kralj Teb postal jaz**.

173 Ob tem dogodku spoznamo absurdno stanje, ko se **odločata Raymond in Mersault** → **Mersault in Raymond odločata** ali bi se vrnila h Arabcem.

Pri zaporedju stavčnih členov je treba izpostaviti nekatere nestavčnočlenske besedne vrste. Med vezniki³⁶ učencem izrazito največ težav povzroča besedni red za veznikoma *pa* (174) in *vendar* (175). Problematičen je tudi *da* (176), nekoliko manj pa *saj* (177), *ker* (178), *namreč* (179) in *ampak* (180).

174 Vseeno **je pa** → **pa je** po drugi strani lahko

175 Tako Črtomir kot Werther se želita poročiti, vendar **na koncu se** → **se na koncu** izkaže da je to nemogoče.

176 Mislim, da **vsak človek se bi** → **bi se vsak človek** odločil

36 Veliko primerov napak veznikov bi lahko uvrstili tudi med napake naslonskega niza (gl. razdelek 6.2.1.3), vendar pa smo jih uvrstili v to kategorijo, ker se je izkazalo, da nastonski nizi največkrat delajo težave, ko jih začenejo določeni vezniki.

177 Tudi dela je čedalje manj saj **veliko podjetij gre** → **gre veliko podjetij** v stečaj

178 Že začetek mi je bil čuden ker **začelo se je** → **se je začelo** tako:

179 **Namreč Izidorju je** → **Izidorju je Namreč** celo odsekal prst

180 Bila sta tako srečna, da sta se hotela poročiti, ampak **sta ugotovila** → **ugotovila sta**, da prihajata iz družin, ki se sovražita.

Redkejšje težave se pojavljajo v stavkih za vezniki *kajti* (181), *ko* (182) in *medtem ko* (183), v bolj osamljenih primerih pa tudi za nekaterimi drugimi. Zanimiva je napaka, ko učenec veznik postavi za prislovno določilo, učitelj pa ga prestavi na začetek stavka (184).

181 kajti **se je grdo** → **grdo se je** obnašal do nje.

182 Začeli so ob 9.00 ko so **učitelji jim razložili vse** → **jim učitelji vse razložili**.

183 saj so bile prozorne in se jih je dalo zložiti v škatljico, medtem ko **stare so bile** → **so bile stare** večje in jih nisi mogel imeti povsod seboj.

184 **Danes če bi** → **Če bi danes** bil kakšen primer, ki bi

Pri členkih imajo učenci težave z umeščanjem *tudi* (185), *še* (186) in *že* v stavek. Manj problematični so *le*, *samo*, *zgolj* (187), *seveda* (188) in *komaj* (189), pojavijo pa se tudi drugi (190). Zanimivi so predvsem primeri, pri katerih položaj členka, kot ga popravi učitelj, spremeni pomen stavka (191).

185 Lahko **povem tudi** → **tudi povem**, da sem sam

186 A Juliji je bilo žal, ker **še je** → **je še** vedno ljubila Romea.

187 priznal da se je **le z njim** → **z njim le** igral in

188 Ona, **seveda pa je** → **pa je seveda** doštudirala,

189 ker je bila ona stara 24 let in ji je bilo ime Tina, on pa je **komaj imel** → **imel komaj** 18 let.

190 Če dinosavri ne bi izumrli tudi nas **ne bi verjetno** → **verjetno ne bi** bilo.

191 **Predvsem to vrednoto naj bi imel** → **To vrednoto naj bi imel predvsem** njegov prijatelj.

Pri prislovih so nekajkrat na neustreznem mestu *več* ali *bolj* (192) in *lahko* (193), dvakrat pa tudi *hkrati* (194). Nekajkrat je neustrezen tudi položaj *rad* (195).

192 Po 3 mesecih se mi je kost zacelila in nisem imel **gipsa več** → **več gipsa**.

193 verjame, da **lahko sta** → **sta lahko** s Črtomirom srečna v nebesih.

194 Bog je bil **skrati pravičen nepravičen in usmiljen** → **pravičen, nepravičen in usmiljen hkrati**

195 XlmeX XPriimekX ne **rada nastopa** → **nastopa rada** kot vodja,

6.2.1.2 Napake znotraj stavčnih členov

Znotraj posameznih stavčnih členov je bilo v Šolarju označenih le malo manj kot 30 primerov napak. Učenci imajo težave z zaporedjem pridevnikov oz. zaimkov (196). Nekajkrat pišejo pridevnik za samostalnikom (197), po mnenju učitelja pa je slogovno neustrezen tudi pristavek za samostalnikom (198).

196 Prav tako se tudi Pečorin predaja zabavi, ki je **njegov nekakšen**

→ **nekakšen njegov** ideal,

197 Danes smo lahko srečni, da živimo v **razvitem svetu in strpnem**

→ **razvitem in strpnem svetu.**

198 **Polonij, Ofelijin oče** → **Ofelijin oče Polonij** njej svetuje

Zanimiv je vljudnostni popravek, v katerem učitelj osebni zaimek *jaz* prestavi na drugo mesto v priredno sestavljenem stavčnem členu (199). Napake učenci delajo tudi pri izražanju starosti (200).

199 **Jaz in Sternfeldovka** → **Strenfeldovka in Jaz** sva sklenili,

200 Ko je bila **10 let stara** → **stara 10 let**

6.2.1.3 Naslonski niz

Naslonski niz, njegov položaj v povedi in razporeditev besed znotraj njega so problematični v dobrih 13 % napak besednega reda. Najpogostejša težava glede na popravke učiteljev je deležnik glagola, ki ni takoj za naslonskim nizom (201). Pol manj je primerov, pri katerih je naslonski niz na neustreznem mestu v stavku brez veznika na začetku (202).

201 vendar je **njegov govor bil obarvan rahlo narečno** → **bil njegov govor rahlo narečno obarvan,**

202 V današnjih časih **starši se** → **se starši** ne odločajo

Položaj naslonskega niza je problematičen v glavnem stavku večstavčnih povedi, ki se začnejo z odvisnikom (203), v drugem prirednem stavku (204) ali odvisniku (205).

203 Ko so odšli v kazino **prvič so** → **so prvič** nekaj malega dobili,

204 Ožbej in Meta sta se zaljubila in **ji je naredil** → **naredil ji je** otroka.

205 Tako je doživela svojo pogubo čeprav je mislila in verjela, da jo bo Jazon sprejel za kraljico in **s tem ji** → **ji s tem** dal pravico vladanju ljudstvu.

Včasih učenci osebni zaimek postavijo pred povedek v sedanjiku (206) ali stavek začnejo s pomožnim glagolom *biti* (207).

206 **Vam pripovedujem** → **Pripovedujem vam**, kot je duh mojega očeta pripovedoval o resnici

207 **Sem bila** → **Bila sem** razigrana, včasih pa prav malo tečna.

Mesto naslonskega niza je problematično tudi pri zanikanju, in sicer na začetku (208) ali na sredini stavka (209).

208 **Bili niso bojzljivi** → **Niso bili bojzljivi**, ali ljudi so se izogibali

209 Nikakor pa jim ne lažemo in **ne vrivamo jim** → **jim ne vrivamo** drugo identiteto

Znotraj naslonskega niza imajo učenci nekaj težav z besedo *bi* (210) ali *naj bi* (211) pri pogojniku, z *naj* pri poročanem govoru (212), z *je* pri pretekliku (213) in z *bo* pri prihodnjiku (214). Včasih učenec naslonski niz razbije (215).

210 Bil sem tako jezen, da **ju bi** → **bi ju** kar mahnil

211 Ampak še več nadutosti **bi naj** → **naj bi** imela,

212 Pri izbirnem predmetu Sodobna priprava hrane smo dobili naloga, da **si naj** → **naj si** vsak ogleda eno od restavracij v Piranu.

213 Zgodba se začne, ko Antigoni v boju za domovino umreta oba brata Etekel in Polinej, ki **je ju** → **ju je** imela zelo rada

214 Drugi dan smo se odločili, da **bomo si** → **si bomo** šli pogledat ptujski grad.

215 Oče **je odvrnil mi** → **mi je odvrnil**:

6.2.1.4 Zaporedje stavkov

V šolarju sta le dobra 2 % napak, ki bi se nanašale na zaporedje stavkov, in so pretežno slogovne. Bolj vsebinski oz. pogojeni z žanrsko ustreznostjo so le nekateri primeri v praktičnih besedilih (216).

216 S spoštovanjem. Zahvaljujem se vam za vaše napotke, ki ste mi jih napisali. Moj kivi se je sedaj uspešno zarasel in lepo raste. Lep pozdrav

Popravljeni primer: Zahvaljujem se vam za vaše napotke, ki ste mi jih napisali. Moj kivi se je sedaj uspešno zarasel in lepo raste. Lep pozdrav S spoštovanjem.

Učitelji nekajkrat zamenjajo položaj odvisnika in glavnega stavka (217) in popravijo primere, v katerih učenci odvisnik razbijejo (218).

217 Že naslednji dan sledi sojenje, v katerem obtoženega spoznajo za krivega, saj menijo, da **če ga ne gane niti materina smrt je zagotovo pokvarjen in sovražen družbi** → **je zagotovo pokvarjen in sovražen družbi če ga ne gane niti materina smrt**.

218 **Zanj mislim, da** → **Mislim, da zanj** nova vera ni ravno kot neka odrešitev,

6.2.1.5 Povezane napake in manj relevantni popravki

Slabo polovico vseh napak besednega reda v šolarju predstavljajo povezane napake besednega reda, ki so za analizo manj zanimive in večinoma opisane v drugih poglavjih. Besedni red se namreč spremeni, kadar učitelj:

- črta osebek, prislovno določilo ali veznik ali kadar združuje povedi (gl. razdelek 6.2.3.1),
 - doda glagol (gl. razdelek 6.2.2.2), veznik, osebek ali prislovno določilo ali eno poved ločuje na dve (gl. razdelek 6.2.2.1),
 - nadomesti polnopomensko samostalniško besedno zvezo z zaimkom (gl. 4. poglavje),
 - spremeni glagolsko število, spremeni strukturo (219) ipd.
- Spremembe besednega reda se pojavijo tudi, kadar učitelj zamenja učenčev veznik s *pa* (220), *zato* (221) ali *a*, naglasno obliko osebnega zaimka z naslonsko (222) ali obratno. Podobno se dogaja pri nekaterih drugih menjavah veznikov ali prislovov (223).
- 219 In čeprav **mi nista starša** → **nista moja starša**, me je stisnilo pri srcu.
- 220 Druge znamenitosti so tudi spomenik sredi trga, ki je delo Jožeta Plečnika, tam je tudi etnološki muzej, cerkev sv. Ožbolta, ki slovi po svoji urejenosti, v Mušeniku je graščina in zapuščene fužine, **in v Podpeci** → **v Podpeci pa** so zelo znani snežni gradovi kralja Matjaža,
- 221 **ker pa je bilo mene strah** → **mene pa je bilo strah zato** sem šla nekaj korakov stran.
- 222 Če sedaj pomislim **je mene** → **me je** doletela hujša usoda
- 223 Na svojem obglavljanju želi čim več ljudi, saj bo tako dobil dokaz, da se ni podredil družbi **in to bo** →, **tako bo to** zanj velika notranja zmaga.
- Nekaj popravkov besednega reda v Šolarju je bilo napačnih, torej je učitelj predlagal slovnično neustrezen popravek (224). Slabi 3 % popravkov pa so bili nejasni in nerazvrščeni.
- 224 Njegova prva bitka je bila bitka z mlino na veter ampak seveda **je mislil Don Kihot** → **Don Kihot je mislil**, da so velikani

6.2.2 Izpuščeni podatki

V Šolarju je bilo označenih 1607 napak izpusta, ki predstavljajo 4,6 % vseh napak. Med njimi je bilo prepoznanih 87 različnih tipov (Tabela 29).

Tabela 29: Napake izpusta

Tip napake izpusta	Število napak
napake razumljivosti oz. sloga	487
izpusti v večstavčnih povedih	385
dejanske napake in lapsusi	276
napake zaradi razumljivosti	182
popravki zaradi poudarjanja	88
napake v oziralnih odvisnikih	68
nerelevantne napake	121
Skupaj	1607

6.2.2.1 Vsebinski in slogovni izpusti

Skoraj četrtnina napak izpusta je vsebinska, s popravki, ki so delno tudi slogovni, pa učitelji povečajo razumljivost besedil. Sem sodi sklicevanje na literarno delo, o katerem piše učenec (225), na njegov žanr (226) ali naslov (227), ki ga učitelj doda v uvodu spisa. Podobno velja za natančnejšo družinsko opredelitev literarne osebe (228) ali njen priimek (229).

- 225 Zato tudi Polikarpa **[iz Tavčarjeve Visoške kronike]** ne smemo le grajati, a tudi hvaliti ne.
- 226 V ljubezenski zgodbi **[tragedije]** Romea in Julije gre za dva potomca sprtih družin, ki se zaljubita.
- 227 Besedilo **[Samorastniki]** je napisal P. Voranc
- 228 S tem bogastvom si je kupil kmetijo na Visokem in si ustvaril družino z **[ženo]** Barbaro.
- 229 Mislim, da je za Silvana **[Kandorja]** kriva le usoda,
V 7 % napak izpusta manjkajo časovne (230), krajevne ali druge reference (231) in razne kolokacije (232) ali frazemi (233).
- 230 Zaustavijo jo želijo praktično vsi, razen **[sprva]** sestre Ismene, ki se na koncu zaroti proti njej in postane njena nasprotnica.
- 231 Več bi se morali **[iskreno]** pogovarjati in ceniti, kar nam je dano, iskati tisto pravo, resnično ljubezen,
- 232 V vojni ni vse črno-belo, ni dobrih in slabih ljudi, saj se povsod dogajajo krivice in umirajo nedolžni in tudi naša **[literarna]** lika sta bila konec koncev morilca.
- 233 To dejansko uresniči bolničar Volodja, ki je držal niti odzadaj **[v rokah]** in s tem dejanjem prevzel oblast v zavodu.

Manjkajoči vezniki v Šolarju predstavljajo skoraj 7 % vseh napak izpusta. Lahko gre za slogovne (234), zlasti pri bolj pogovornih oblikah (235), ali za vsebinske spremembe (236). Takšne napake pogosto povzročijo različne povezane napake, denimo besednega reda (gl. razdelek 6.2.1.5).

- 234 Ni minil teden, **[ko]** imel si → **si** imel drugo.
- 235 Sedaj pa pomislimo, **[ali]** je res to to, da zaradi ljudi okoli nas
- 236 Ljubezen te lahko čaka kjerkoli, **[a]** ni nujno, da je tista prava.
Zaradi sloga, pa tudi zaradi razumljivosti, je včasih dodan samostalnik za zaimki oz. pridevniki *ves, ostali, oba* ipd. (237). Lahko gre za ponovljene strukture (238) ali za samostalnik pred prilastkom (239).
- 237 Meni se zdi, da je ta ljubezen žalostna, ker na koncu tudi oba **[ljubimca]** umreta.
- 238 Vse je bilo osvetljeno in še cesta je bila ravna, tako, da sem se lahko v miru rolala. Vse do nekega dne. Predzadnjega **[dne]**, kot sem povedala.
- 239 V Romeovi in Capuletovi družini pa so nastali spori in to je težilo **[vez]** med Julijino in Romejovo poroko,

Učitelj včasih relativizira učenčevo trditev z dodanim prislovom *lahko* (240) ali *npr.* (241). Pri glagolih, ki so lahko povratni, pa doda morfem *se* ali *si* (242).

240 A na koncu je potem **[lahko]** vse še slabše.

241 Razlike, ki se pojavljajo med Črtomirom in Pečorinom, glavnim junakom romana *Junak našega časa*, so **[npr.]** njun odnos do žensk.

242 To bi po mojem mnenju pomenilo, da če **[si]** človek srčno želi, da bi mu Bog odpustil

Vsebinski in slogovni popravki so tudi dodani manjkajoči pridevniki (243) ali prislovi (244), ki so manj zanimivi pri analizi slovničnih težav. Čisto slogoven pa je popravek pogovorne fraze *po mojem* (245).

243 Skoraj so že odnehali, a so prijeli **[nenavaden]** klic.

244 Danes je to **[marsikje]** malo drugače oz. kakor kje.

245 Po mojem **[mnenju]** Meta ravna prav.

Kompleksnejša slogovna izpusta sta, kadar učitelj združi dve povedi v eno, kar povzroči povezane napake izpusta, besednega reda, zapisa ipd. (246), in kadar spremeni način izražanja svojine (247).

246 Napredovale smo predvsem v izobrazbi, imamo drugačne odnose v družini. Smo vsi na svetu enakovredni

Popravljeni primer: Napredovale smo predvsem v izobrazbi, imamo drugačne odnose v družini, **saj** smo vsi na svetu enakovredni

247 bal svojega očeta, ki **mu je bil** → **je bil njegovo** pravo nasprotje

6.2.2.2 Izpuščeni stavčni členi v večstavčnih povedih

Opazno težavo učencem predstavljajo izpusti v večstavčnih povedih. Pri 11 % napak izpusta gre za manjkajoči pomožni glagol v drugem stavku povedi (248), zlasti če je prvi stavek zanikan, drugi pa trdilen (249), ali obratno (250). V drugem, zanikanem stavku lahko manjka nikalni členek *ne* (251).

248 On pa je ostal miren in **[je]** v trenutku našel rešitev.

249 zato ga občudujem ker ni odnehal in se **[je]** upiral do konca.

250 Dandanes bi to bilo kršenje avtorskih pravic, dvesto let nazaj pa **[ni bilo]** nič slabega, celo več!

251 Obljubila sem si, da ne bom nikomur povedala tega tudi najbolšima prijateljicama **[ne]**.

Prav tako v drugem delu večstavčnih povedi (252) ali v povedi, ki se navezuje na prejšnjo (253), učenci izpuščajo zaimenske predmete.

252 Mora opraviti vse izpite, ker **[mu]** podjetje omogoča šolanje,

253 Imel sem brata Antonia. Ta si **[me]** je neskončno želel izriniti iz kraljestva.

V drugem stavku učenci ne ponovijo predlogov (254) ali ne uporabijo drugih predlogov zaradi drugačne vezave (255), večkrat pa učitelj

predlog doda v priredno sestavljeni stavčni člen (256). Podobno slogovno je verjetno utemeljena učiteljjeva, sicer redka ponovitev pridevnika zaradi ujemanja s samostalnikom (257).

254 Pesem govori o Valjhunu in Črtomirju, ki se bojujeta vsak za viro in tudi **[o] ljubezen** → **ljubezni** med Bogomilo in Črtomirjem.

255 Razlikujeta pa se v svojem pogledu na ljubezen in **[po]** obdobju v katerem živita.

256 Pečorin pa je predstavnik egoističnega romantičnega junaka, ki ostaja ravnodušen do vsega razen **[do]** sebe.

257 Ampak menim, da se tega ne da nadomestiti z nobenim dejanjem in **[nobeno]** stvarjo.

V nekaterih primerih v drugem delu učenčeve povedi manjka veznik (258), še pogosteje pa prosti morfem *se/si* pri glagolu (259).

258 Vesel sem, da se mi ni nič naredilo in **[da]** imam tako trdo glavo.

259 in ves večer sta plesala in **[se]** pogovarjala.

6.2.2.3 Napačni izpusti in lapsusi

Za razliko od prejšnjih popravkov, ki so pogosto samo slogovni, gre v nekaterih primerih izpustov dejansko za napake. Učenec izpusti pomožni glagol *je* (260), nedoločnik, zlasti v večstavčnih povedih (261), ali pretekli deležnik glagola *biti* (262). Zaradi enakega zapisa besed je nekajkrat izpuščen pomožni glagol *je* oz. osebni zaimek *je*, kadar bi morala biti pisana skupaj (263), v povezavi z nedoločnikom, deležjem ali deležnikom pa manjkata *se* oz. *si* (264).

260 Kar je razvidno v tragediji, **[je,]** da Shakespeare rad opisuje stvari

261 Tako kot je v orkestru potrebno prilagoditi številu violin število trobent moramo ljudje živeti v sožitju in slogi ter **[biti]** pripravljeni pomagati drugim,

262 Povedal ji je da bi **[bil]** zelo vesel

263 V tretji kritici pove, da je njen ljubljeni in oče otroka odšel, ker se je **[je]** sramoval.

264 da **[se]** je Izidol podvrgel temu in spremenil svoj tok življenja za očeta, ki ga je moral imeti očitno zelo rad.

V skoraj 9 % napak izpusta gre za manjkajoče predloge oz. napačno vezavo (265), vprašanje pa je, ali je to posledica nezadostne slovnice ali slovarske zmožnosti učencev. Nekajkrat manjka predlog *s* pred povratnim zaimkom *sabo* oz. *seboj* (266).

265 in zato je kriv **[za]** vse on, ker jim ne stoji ob strani.

266 Oče ga je sedaj vzel **[s]** sabo v Ljubljano,

Tudi pri nekaterih manjkajočih veznikih gre za lahko prepoznavne napake, npr. za izpuščeni veznik, kadar bi skupaj morala stati dva veznika (267), za manjkajoči *tako* v povezavi s *kot* (268) ali za povezavo veznika *da* in členka *naj* pri poročanem govoru (269).

- 267 Moje mnenje je, **[da]** če se že dve osebi odločita, da bi imela družino potem naj tudi skrbita za njo.
- 268 Antigona je prikazana, kot dobra oseba, ki se bori, da bi Polinejka pokopali **[tako]** kot Eteokla.
- 269 ker se niso držali tega **[da]** naj bi vsak dobil pomoč na sodišču ampak so vsi pomagali baronu.

Med napakami izpusta sta dobra 2 % raznorodnih primerov (270), za katere lahko s precejšnjo gotovostjo trdimo, da so nastali kot lap-susi, torej zaradi nepazljivosti tvorcev, ne pa zaradi njihove pomanjkljive jezikovne zmožnosti, torej so manj zanimivi za jezikoslovno analizo. Pri tem izstopajo manjkajoči predlogi (271) ali vezniki (272).

- 270 Nato pa je bila ura že enajst. In mama **[je]** komaj slišno izjavila:
„Nekam dolgo jih ni! Kako to? Kaj se je zgodilo?“
- 271 zagrozi, da bo **[v]** primeru če ga ne najdejo
- 272 Nekaterim je mar **[in]** se zgrožajo nad takimi novicami, nekaterim pa je to vsakdanje.

6.2.2.4 Večja razumljivost

V precejšnjem deležu napak izpuščenih podatkov učitelji posegajo v besedila učencev zaradi večje razumljivosti in dodajajo npr. osebke:

- 273 Na terasi je bila tudi Herodiadina hči Saloma. Že prej je **[Herod]** velel, da jo pokličejo nazaj notri, vendar se ni odzvala povabilu.

Razmeroma pogost vsebinski ali slogovni popravek je manjkajoči predmet (274). Lahko gre za osebne zaimke, denimo v povezavi z manjkajočim predlogom (275). Učitelji v besedila učencev dodajajo tudi kazalne zaimke (276), lahko v povezavi s predlogi (277).

- 274 Nenačrtno ubije Polonija, ki je prisluškoval **[pogovoru med Hamletom in materjo]** in uboj zelo obžaluje, vendar pa ga kralj Klavdij zaradi svoje varnosti pošlje v Anglijo,
- 275 Drugi, pravi vzrok pa ni nikoli direktno povedan, ampak **[nanj]** lahko sklepamo iz Volodjevega ravnanja in razmer v okolici.
- 276 Prosijo vse, ki karkoli vedo, naj **[to]** sporočijo.
- 277 In se ti **[za to]** tudi zelo zahvaljujem!

Učitelji razumljivost besedil učencev povečajo tudi z dodanim svojilnim zaimkom (278), pridevnikom (279) ali samostalniškimi prilastki (280), ki jasneje opredeli, na koga se nanaša lastnina.

- 278 Tudi dvema zaljubljenca lahko onemogočajo veliko stvari. To, da se **[njuni]** starši sploh ne marajo in nočejo dopustiti, da bi se poročila.
- 279 Bila sta bivša **[Hamletova]** sošolca in Hamlet ju je na začetku spoštoval in prijateljeval
- 280 saj si noče mazati rok z krvjo **[nečakinj]**.

6.2.2.5 Poudarjanje

Pomemben razlog, zaradi katerega učitelji dodajajo dele v besedila učencev, je poudarjanje:

281 Po nekaj minutah sva vstala še **[midva]** z XI meX.

Včasih dodajo zaimke (282), pridevnike (283) ali prislove (284). V nekaj redkih primerih poudarijo, da gre za dvojino, z zaimkom *oba* (285) ali s števnikom *dva* (286).

282 ohranil bo vero svojih staršev in pridobil domovino, v kateri bodo **[vsij]** ljudje svobodni.

283 Ko sem jaz bila **[sama]** v taki situaciji, sem storila kot sem zgoraj opisala.

284 Smisel celotne zgodbe je v tem, da med vojno pozabiš na medsebojne odnose — tako sta se tudi Anton in Berk **[skupaj]** bojevala, kasneje pa sta se spoprijateljila.

285 Antigona in Kreon bi lahko bila **[oba]** tragična junaka.

286 menim, da si ti **[dve]** osebi nista zaslužili biti oče in mati,

Za poudarjanje in izboljšave sloga učitelji dodajajo členke (287), denimo *tudi* (288), *več* (289) ali *že* (290).

287 Zmeraj sem se izobraževal in **[kar]** požiral knjige.

288 Mislil sem **[tudi]**, da lahko zaupam in računam na svoja dva tovariša

289 Vendar, ko je otrok enkrat samostojen, mislim, da mu pa ni **[več]** najbolje, če se starši vmešavajo v njegovo življenje.

290 Njegov ideal je bil, da bi z Bogomilo živela srečno **[že]** na tem svetu,

6.2.2.6 Oziralni in drugi odvisniki

Še en težaven sklop pri napakah izpusta so oziralni in drugi odvisniki. Tu učenci izpuščajo oziralni zaimke na začetku odvisnika (291) ali pri ponovljenih odvisnikih (292), lahko pa ga učitelj doda in spremeni tudi besedni red (293). Za oziralnim zaimkom v neprvem sklonu včasih manjka osebni zaimke (294). Včasih učitelj doda cel oziralni odvisnik (295).

291 Pred njenim sinom pa policija **[, ki]** ga pelje v zapor.

292 Vsak, ki se zna odločati, **[ki]** je pameten, zna obračati vse v svoj prid je na svoj način moder,

293 Nezakonska mati spada v obdobje romantike na Slovenskem, **[ki] trajala je** → **je trajala** od 1830–1849.

294 Upam, da bo zdrava in vesela, zato, ker je skrbna in ima rada svojo družino in druge ljudje, ki **[jih]** je spoznala.

295 Plečnik je bil pomemben živeči Slovenski arhitekt **[, ki je živel]** v 20 stoletju.

Precej pogost problem učencev je manjkajoča referenca pred odvisnikom, ki je lahko splošnejši samostalni (296), še pogosteje kazalni zaimke *to* (297), včasih pa zaimenski prislov (298).

- 296 poudarja prijaznost ter enakost med ljudmi, ter [idejo,] da lahko vsi dosežejo enako.
- 297 Nato so se težave Fatimi začele vrstiti, vendar največja je bila [ta], da je postala noseča,
- 298 to naredi tudi [takrat,] ko pridem jaz na obisk.

6.2.2.7 Neustrezni popravki in napačne oznake

V Šolarju se pojavljajo primeri napak izpusta, ki so bili napačno označeni zaradi napake učitelja pri popravljanju ali označevalca pri transkripciji. Tako jih je bilo 16 označenih, kot da gre za izpuščene oz. odvečne podatke, čeprav bi morali biti označeni kot oblikoslovne napake zaradi zamenjave sedanjika in preteklika (299). Nekaj je bilo napak ločil (300), pa tudi neustrezno zapisana oznaka za uro (301) bi bolj sodila med napake zapisa. Popravki, kot so manjkajoči podatki v naslovu spisa, so za našo analizo manj relevantni.

- 299 Po menjavi mečev [je] tudi Hamlet rani Laerta
- 300 ki jih je doživel v 30[-]letni vojni,
- 301 so se zbrali pred šolo ob 8[h] in

6.2.3 Odvečni podatki

Napake odvečnih podatkov so nasprotne izpuščenim podatkom, vendar so bile pri analizi združene v nekoliko drugačne skupine (Tabela 30). V Šolarju je bilo označenih 2665 tovrstnih napak, skupaj torej slabih 8 % vseh napak. Prepoznanih je bilo 124 različnih tipov.

Tabela 30: Napake odvečnih podatkov

Tip napake odvečnih podatkov	Število napak
slogovne napake	1749
redundantni podatki	171
dejanske napake	166
napake v večstavčnih povedih	128
napake zaradi poudarjanja	51
nerelevantne napake	400
Skupaj	2665

6.2.3.1 Slogovni popravki

Pri odvečnih podatkih izrazito, z več kot 65 %, prevladujejo primeri, pri katerih učitelji črtajo besede ali daljše dele besedil učencev zaradi ustrežnejšega sloga. Eden najpogostejših je odvečni osebek, ki ga učitelj lahko črta tudi zato, da se izogne ponavljanju (302).³⁷ Velikokrat gre za osebni zaimek *jaz* (303), pojavljajo pa se tudi popravki manj običajnih struktur (304).

302 Mati je na mizo prinesla vino. **Mati je bila** → **Bila je** zelo vesela in zaljubljena.

303 Če bi **jaz** imela fanta, ki bi bil druge vere bi to pač zanemarila.

304 Po nekaj letih sva **s-Petkom** rešila še nekaj žrtev ljudožercev.

Učenci napišejo odvečni predmet (305), tudi v povezavi s predlogom (306). Tipično slogovno pogojeni so primeri, kjer učitelj črta ponovljene besede (307) ali besede, ki stojijo skupaj z drugimi iz iste besedne družine (308), s čimer se spet izogne ponavljanju.

305 Čutimo in razumemo **stvari**, toda odločitve so v istih situacijah različne.

306 V pesmi izvemo tudi, da jo je oče tepel in klel **čez njo**

307 Vloga žene danes in **žene** nekoč se je skozi čas zelo spremenila.

308 Tudi sama sem že občutila **občutek**, ko se ti zgodi krivica,

Učitelji črtajo različne pridevnike in druge dele besedila, zlasti, kadar so pomensko prazni in dejansko nepotrebni, recimo *določen* (309), *nekako*, *na nek način* (310), *na primer* (311). Podobno velja za prislovna določila (312), ki so zelo pogost tip napak odvečnih podatkov, in količinske ali merne prislove (313).

309 Optimističen človek v življenju zna postaviti **določene** cilje, ki so zanj primerni,

310 da Pečorin ta svoja čustva **na nekakšen način** zapira vase in na ta način deluje zelo prazen

311 Te ovire so lahko **na primer**, da starši ne pustijo, da sta skupaj

312 To pa ne pomeni, da Polikarpa **sedaj** zagovarjam, ampak dejstvo je, da je res tako.

313 Iz odlomka ugotovimo da so bile nezakonske matere v preteklem času **zelo** šibki člen v družbi.

Pri zaimkih učitelji popravljajo odvečne pridevniške zaimke učencev, kot sta *tisti* (314) in *neki* (315). Pogosto črtajo kazalni zaimek *ta* v samostalniški rabi (316), denimo v frazi *temu ni tako* (317), pred samostalnikom, ki je bil že omenjen (318), ali kadar zaimku sledi pojasnilo (319). Slogovna popravka sta tudi zamenjava zaimka *le-ta s ta* (320) in črtanje oziralnih zaimkov (321).

314 V ušesih sem slišala le **tisti** piskajoč zvok, ki mu ni bilo videti konca.

315 A ljudje ne znamo živeti. Postavljamo **neke** zakone, s katerimi morimo sebe in druge.

316 Takrat sta sistem in družba nasilno in odločno obračunala

z nasprotniki, danes pa se to dogaja prijazno, brez nasilja in se tega tako skoraj ne zavedamo. Posledice **tega** pa se kažejo v občutkih drugačnosti, osamljenosti in odtujenosti.

317 Toda danes **temu** ni več tako, ker če se maščuješ nekomu, ki ti je storil nekaj žalega potem razmišljaš enako kot tisti,

318 Le tako je Prometej naredil prve ljudi včasih. Ker pa je bil zelo pameten in razumen, je **te** ljudi učil vse.

319 O temu govori tudi **ta** odlomek, ki prikazuje

320 Preden je Jošt umrl mu je **le** ta zajecjal, da pride ponj.

321 Novico je izvedela tudi Evridika, **ki je bila** Kreonova žena.

Številnih slogovnih popravkov so deležni vezniki v besedilih učencev. Najpogostejši med odvečnimi podatki s slabimi 9 % primerov je odvečni *pa* (322), pogosto za drugim veznikom (323) ali za drugim *pa* v povedi (324).

322 Z njim niso delali človeško, kar **pa** je osnova vsakega odnosa.

323 Naredil se je vihar, a Urška in njen soplesalec **pa** sta še kar naprej plesala.

324 Ko pa smo prišli domov **pa** sen ju povabil na računalnik.

Slogovno ali pomensko so odvečni tudi drugi vezniki (325), denimo priredni vezniki pri naštevanju (326). Zelo pogosto so na začetku povedi (327), tako da črtanje veznika povzroči povezane napake besednega reda.

325 Meta ni bila kriva, da se je zaljubila v sina bogataša, **ampak** krivo je to zaničevanje revežev.

326 Oblečena je v kratke kavbojke **in** kratko rdečo majico in temne športne čevlje.

327 **Vendar** se mi vseeno zdi → Vseeno se mi zdi

Učitelji črtajo veznike v sklopih dveh zaporednih, posebej *da*, *tako* in *naj* (328). Podobno velja za *zato* in *ker*, kadar stojita skupaj (329) ali vsak v svojem stavku (330).

328 Naročil jim je, **da** naj najdejo storilca, ker če ne bojo oni odgovarjali za prekršek.

329 Lahko je vzor za vse nas, **zato** ker moramo tudi mi kdaj tvegati in premagati strah.

330 Ker je bilo delo na kmetiji pomembnejše **zato** ni imela časa za učenje, zato po osnovni šoli ni nadaljevala šolanja.

Učitelji popravljajo tudi podvojene predloge (331) ali primere, v katerih se jim raba brez predloga zdi ustrežnejša (332).

331 Prva polovica je dobila frutabelo druga pa cd za **na** računalnik.

332 Nekega dopoldneva je kmet Martin Krpan tovoril pistacije **na** skrivaj,

Členki so po mnenju učiteljev v besedilih pogosto odveč: *tudi* (333), *že* (334), *kar* (335), *res* (336) in številni drugi.

333 Berk doživi pravi boj iz oči v oči v Suhi krajini, kjer potekajo **tudi** boji z Nemci.

- 334 Ljudje smo premalo strpni **že** do vseh drugih, ampak se mogoče ne opazi tako zelo, v odnosu med dvema pa lahko to postane zelo velik problem,
- 335 V noveli Samorasniki je bilo **kar** podobno.
- 336 Plače so zelo majhne, kar se mi zdi tudi žalostno da moramo **že-res** gledati na vsak evro.

Med daljšimi frazami učitelje motijo zveze, kot sta *po imenu* (337) ali *z naslovom*, *trpnik s se* (338), uporaba zveze *eden (iz)med* s primernikom ali presežnikom (339) in raba faznih ali naklonskih glagolov (340). Napake izpusta so delno tudi vsebinske, npr. kadar učitelj črta učenčevo sklicevanje na literarno delo (341).

- 337 Včeraj sta brez sledu izginili deklici **po imenu** Taja in Vlasta.
- 338 Človeka **se** ne sme obsojati na prvi pogled.
- 339 saj naj bi veljal za enega **med** boljših nasprotnikov,
- 340 Od takrat se Špela **začne družiti** → **druži** s Smajkom.
- 341 Odlomek spada v Uvod **h Krstu pri Savici** in opisuje boj med kristjani in pogani.

Od bolj pogovornih besed ali struktur učitelji popravljajo *en* v pomenu *nek* (342), *se* v zvezi *gre za* (343) ali ob določenih glagolih, ki se lahko uporabljajo v povratni obliki (344), ali predlog *za* v povezavi z nedoločnikom (345). Kadar učenci izražajo svojino s samostalnikom z rodilniku in predlogom *od*, pa učitelji prečrtajo predlog (346).

- 342 On je bil že enkrat zaljubljen in to v **eno** njegovo prijateljico, ki je precej starejša.
- 343 Gre **se** jim samo za svojo korist,
- 344 zato želi, da ji on govori resnico, da **se** mu ni treba lagati in naj raje ostane pri njej, k
- 345 ne bo nikoli več ker naj bi danes imel edino priložnost **za** dobiti službo.
- 346 Za domače branje sem prebral roman V sedemnajstem **od** avtorja Iva Zormana.

Slogovno ali vsebinsko neustrezni so lahko daljši deli v besedilih učencev: stavki v večstavčnih povedih, kadar se vsebina v njih ponavlja (347), vmesni naslovi (348), metabesedilni stavki (349) ali spremni stavki pri premem govoru (350). Sem bi lahko umestili tudi različne zaimke, s katerimi učenci nagovarjajo bralca (351).

- 347 Antigona je za to izvedela in povedala svoji sestri **,kaj je Kreon storil.**
- 348 **Jedro** Hamlet, hlini blaznost, da bi rešil uganko, kdo je v resnici očetov morilec.
- 349 **Ta mit sem prebrau za domačo branje, zdaj bom to tudi opisal.**
- 350 Ko sta starša odšla mi je učiteljica rekla: „Vlasta zakaj si pobegnila? Zakaj si vzela še Tajo s seboj? Zaradi tega nam je padlo povprečje v razredu, pa še izgubili smo prijazno in odlično učenko. Sram te bodi!!“ **mi še zaupije.**

351 To **nam** pokaže s tem ko želi pokopati svojega brata kljub ukazu Kreona, da ga ne sme.

Nekatere napake izpusta so povezane z učiteljevim združevanjem ali ločevanjem učenčevih povedi, pri čemer črta oziralne zaimke ali veznike:

352 V Uvodu Prešeren napove komu posveča to **pesem, kjer se razbere**
→ **pesem. Razbere se** da je ta pesniti posvečeno umrlemu prijatelju Matiju Čopu

6.2.3.2 Odvečni podatki zaradi pomenskega podvajanja

V dobrih 6 % napak odvečnih podatkov se kaže želja učiteljev po jedrnatosti izražanja, zaradi česar črtajo dele besedila učencev, ki so redundantni: samostalnike (353), svojilne zaimke (354) ali pridevnike (355), kadar je jasno, na koga se nanašajo, števnike *en* (356) ali *dva* (357), nedoločnik *iti* ob naklonskih glagolih (358).

353 Bila je zelo neurejena **ženska**.

354 Laert, jezen zaradi umora **svojega** očeta,

355 Zaradi **Polonijeve** pomoči Klavdiju, je Hamlet sovražil Polonija

356 Likarju je bilo dolgčas, zato je od Tončka zahteval **en** gumb.

357 in tudi moja **dva** bratranca imata podobno zgodbo.

358 Sin in Nevesta sta hotla **it** v tujino

Zaradi pomenskega podvajanja, pa tudi zaradi večje slogovne ustreznosti učitelji popravljajo različne prislove ali prislovna določila (359), posebej ob glagolih ali predlogih s podobnim pomenom (360), pridevnike (361), odvečna imena (362), oznake oseb ali literarnih del (363). Zanimive so nekakšne dvojne formule, pri kateri učenec uporabi več besed s podobnim pomenom (364). Večkrat tako pojasnilo sledi pojasnjevalnemu vezniku *oziroma* (365). Za učitelje so moteče tudi nekatere primerjave (366).

359 Tudi zdaj, **v tem času** revni ljudje živijo

360 Ravnatelj in Darja sta me grdo pogledala, nato pa odvihrala **ven** iz razreda.

361 mimo pa je prišel čolnar ter ga odpeljal k Bogomili na drug **rečni** breg.

362 Mati **Meta** jim je govorila, da grejo na boj za kruh

363 **Zgodba** Prilika o izgubljenem sinu govori o dveh sinovih.

364 Horacij pomaga Hamletu, ker se z njim pozna že od otroštva, saj je z njim **treniral in** uril v borilnih veščinah.

365 Ker sem odšla v samostan k nunami, nisem smela iti **zabave oz.** plese,

366 Če bi se sama znašla v njihovih položajih, bi ravnala nekoliko drugače **kot oni sami**.

Med napake odvečnih podatkov sodi zanimiv sklop primerov rabe zveze *mi ljudje*. Po eni strani učitelji črtajo samostalnik *ljudje* za osebnim zaimkom (367), včasih pa črtajo zaimek in pustijo samostalnik (368).

367 Mislim, da to nam **ljudem** manjka

368 spoznal, kako pomembna je pravzaprav materni jezik za **nas** ljudi.

6.2.3.3 Napačni odvečni podatki in lapsusi

Večina do sedaj opisanih napak odvečnih podatkov niti ni zares slovnično neustrezna, učitelji jih popravljajo predvsem zaradi želje po boljšem slogu učencev. V dobrih 6 % odvečnih podatkov pa ne gre samo osebni slog, pač pa za dejanske, običajno oblikoslovne napake: dvojno stopnjevanje pridevnika (369) ali rabo deležnika glagola *biti* v prihodnjiku (370). Tudi veznik *kot* je razmeroma pogosto rabljen neustrezno (371).

369 Danes obstajajo **bolj** milejše kazni za otroke,

370 Načrt za prihodnost je, da bo na olimpijskih igrah **bilo** več športnih ponog

371 Takšne družine imenujemo **kot** nezdrave.

Za učence je težavna raba *se* oz. *si* ob glagolih, ki niso povratni (372) ali pa vsaj ne v kontekstu, ko je v stavku tudi predmet (373). Nekateri primeri precej očitno nastanejo kot lapsusi zaradi nepozornosti učencev, npr. ponovljeni predlog (374), ponovljeni samostalnik za osebnim (375) ali drugim zaimkom (376), ponovljeni členek oz. členek s podobnim pomenom (377), neustrezna raba dveh veznikov v večstavčni povedi (378) ali raba *lahko* (379).

372 Vsi smo **se** vstali in jo pozdravili.

373 Njeni otroci pa so **se** doma pazili eden na drugega,

374 potem pa gledam televizijo do približno ~~do~~ 23.00 ko se odpravim v posteljo.

375 ali pa bo zvest vojak in jo brez vesti likvidiral **dekle**.

376 Kava se mu je zdela zelo dobra, ampak boljša se mu zdi tista, kakršno **jo** naredi sam.

377 Nekateri starši so zelo kritični do partnerjev njihovih otrok a to le zato, ker jim želijo **le** najboljše.

378 Ko je Valjhun premagal vojsko Črtomirja, **a** le Črtomir je ustal živ.

379 Prav tako ji je hotel pomagati Ožbej, a ga oče samo grdo pogledal in se **lahko** usedel nazaj.

6.2.3.4 Odvečni podatki v večstavčnih povedih

Podobno kot pri izpuščenih podatkih (gl. razdelek 6.2.2.2) so tudi pri odvečnih težave v večstavčnih povedih. Takih je slabih 5 % napak odvečnih podatkov v Šolarju. Problematični so odvečni pomožni glagoli (380) ali deležniki v drugem stavku (381). V drugem delu priredja je lahko odveč veznik (382), negacija (383) ali predlog (384).

V nekaterih primerih učenec glagol najprej uporabi v glavnem stavku pred odvisnikom, nato pa ga še ponovi (385). Podobno se dogaja z drugimi besedami, denimo *lahko* (386).

380 Hlabci so se prestrašili in hitro **so** rekli

381 Na nek način ga razujem, saj je bil v tistem času tako vzgojen, odraščal je med vojaki in bil iz prve roke deležen agresivnega ravnanja s sočlovekom in **bil** priča, pa tudi izvajalec mnogih nasilj.

382 Tudi otroci so bili visoki, ravni ter močni, kljub temu da so neskončno garali, ter **da** so morali že pri mladih letih začeti delati.

383 Berk se ne pogloblja v svoje misli, vsaj ko se bojuje **ne**.

384 Zgodba pripoveduje pač o dekletu **o** bosanki ki se je preselila.

385 Moje mnenje o nezakonskih materah je, da so tiste ženske, ki jih mož zapusti in niso poročene, **so** lahko tudi več vredne

386 Odlomek govori kako je lahko meta pri svojem trdem in težkem življenju **lahko** ohranila svojo lepoto, kar se je videlo tudi na otrocih.

Pri oziralnih in drugih odvisnikih (gl. razdelek 6.2.2.6) je najpogostejši problem odvečni kazalni zaimek pred odvisnikom (387), pojavlja pa se tudi odvečni osebni zaimek za oziralnim v odvisniku (388).

387 V današnjem svetu se tudi vse prevečkrat dogaja **to**, da od drugih zahtevamo pravico,

388 Problematika o spolnosti in premalo komunikaciji, katero **je** je pisatelj Ivo Zorman izpostavil, je prisotna tudi danes.

6.2.3.5 Odvečni poudarki

Kot je razvidno iz razdelka 6.2.2.5, učitelji dodajajo elemente in z njimi poudarjajo dele besedil učencev. Delajo pa tudi obratno, torej črtajo po njihovem mnenju nepotrebne poudarke učencev, denimo zaimek *oba* (389) ali pridevnika *lasten* (390) ali *sam*.

389 V tragediji Romeo in Julija se dva nesrečno zaljubita, vendar nato **oba** po spletu nesrečnih naključij umreta.

390 Človek se skozi celotno življenje oblikuje in ustvarja svojo **lastno** identiteto.

Pri načinu, na katerega učenci v spisih izražajo svoje mnenje, učitelji črtajo osebni zaimek *jaz* ob glagolih *misliti* ali *meniti* (391). Podobno velja za zveze z *osebno* (392), *po mojem mnenju* (393), glagolom *zdeti se* (394) ipd.

391 **Jaz** mislim, da bi se jaz tudi kot Francka počutila odveč v taki družini in bi poiskala zavetišče

392 Jaz **osebno**, zaradi ljubosumja sem sposoben marsikaj narediti, ubijati pa ne, bog ne daj.

393 oba sta bila mlada, ko sta postala par, a je ta moški hotel čimprej postati finančno neodvisen da bi **po mojem mnenju** zaživela samostojno življenje.

394 S to spremembo **se mi zdi, da** sem postala bolj samostojna.

6.2.3.6 Nerelevantni popravki in napačne oznake

Približno 15 % napak odvečnih podatkov je za slovnične analize manj relevantnih. Besed, stavkov ali odstavkov, ki jih učitelj prečrta predvsem zaradi vsebine (395), je med napakami odvečnih podatkov v Šolarju slaba 2 %. Manj je odvečnih učenčevih komentarjev ali pojasnil v oklepajih (396). Že ob snovanju sistema označevanja korpusa je bilo jasno, da je enačenje takih primerov z jezikovnimi napakami lahko vprašljivo (Rozman idr. 2012: 35), vendar je to omogočalo enotnejše in učinkovitejše označevanje. Sorodne temu so napake besedišča, pri katerih učitelj zamenja besedo zaradi podobnih razlogov (gl. razdelek 4.2.1 v 4. poglavju).

395 Delo Kaplan Martin Čedermac je **dramske** zvrsti.

396 Prvi del odlomka pripoveduje (**saj ne vem čisto če je res**) o njihovi veri in nebožjim pravilom.

Za analizo manj relevantni so primeri, kjer učenec dele besedila oštevilči, učitelj pa to prečrta:

397 1. Knjiga se mi je zdela zelo zanimiva ker je govorila o naravi, prijateljstvu in zaupanju. 2. Ni mi bilo všeč, da je pisatelj pisal tako odkrito.

Nekaj primerov je neustrezno označenih, saj bi zaradi neuporabe različnih oblikoslovnih kategorij – dvojine, ustreznega glagolskega časa (prim. razdelek 6.2.2.7) – morali biti označeni kot oblikoslovne napake. Nekaj primerov pa bi lahko uvrstili med napake zapisa.

6.2.4 Napake strukture

Na podlagi učiteljskih popravkov je bilo iz korpusa Šolar izločenih, analiziranih in kategoriziranih 653 napak, vezanih na rabo neustrezne strukture; od tega se je 447 zgledov (cca. 68 %) izkazalo za potencialno relevantne, ostalih 206 zgledov (32 %) pa za nerelevantne. Nerelevantni zgledi se nanašajo predvsem na slogovno in vsebinsko motivirane popravke, kot so denimo popravki manj formalnih besed in struktur v slogovno ustrežnejše (398), vsebinski popravki (399) in popravki nenavadnih struktur, ki jih ni bilo mogoče kategorizirati:

398 Pobral je ves denar, truplo **zakopal** → **pustil ležati** in odšel.

399 **Znak za ljubezen kupi Janez Meti srce.** → **Kot znamenje ljubezni Janez Meti podari lectovo srce.**

6.2.4.1 Najpogostejše napake strukture

Pri analizi napak neustreznih struktur je bilo skupaj določenih 32 potencialno relevantnih jezikovnih problemov. V Tabeli 31 je navedenih 15 najpogostejših problemov, ki so podrobneje predstavljeni v nadaljevanju.

Tabela 31: Najpogostejši jezikovni problemi, vezani na rabo neustreznih skladenjskih struktur

Št.	Jezikovni problem	Primer	Število napak
1	izražanje svojine	<i>Polonij je bil oče Ofelije</i> → <i>Ofelijin oče.</i>	95
2	besedna zveza namesto stavka in obratno	<i>Ko je Ofelija izvedela za očetovo smrt</i> → <i>, da je oče umrl in da je Hamlet bil morilec, je od žalosti padla v depresijo.</i>	65
3	neustrezen veznik	<i>Prišlo je prvo državno tekmovanje, kjer</i> → <i>na katerem je zmagala.</i>	64
4	strukture s pomensko izpraznjeno besedo	<i>Izidor se ga je bal nista imela nobene komunikacije med seboj</i> → <i>med sabo se nista pogovarjala</i>	35
5	glagolska in pridevniška vezljivost	<i>Obiskal me je tudi slepi starec, ki mi je svetoval naj ponovno premislim svojo odločitev</i> → <i>o svoji odločitvi</i>	27
6	napačna kolokacija	<i>Svit se je odrekel vse nepridprave družbe</i> → <i>slabi družbi ter se je zaprl vase.</i>	24
7	modifikacija: napačni sklon ali predlog v prilastku sam. zveze	<i>Toda dandanes [...] smo ljudje bolj svobodni in lažje izrazimo svoje mišljenje, svoje neodobranje nad nečem</i> → <i>nečesa.</i>	16
8	Zveza z nedoločniki ali polnopomenski glagol?	<i>V primeru vojne so se vsi šli bojevati, saj jim je bilo to v ponos, braniti svojo deželo</i> → <i>da so branili svojo deželo.</i>	14
9	raba trpnika	<i>Oba pojma pa sta upoštevana iz strani Cerkve</i> → <i>upošteva Cerkev, ki se ju trudi vpeljati v življenje sodobnega človeka.</i>	14
10	dajalniške zveze ob glagolu biti	<i>Antigona [...] želi na vsak način pokopat Polinejka, saj ji je brat</i> → <i>ga kot brata ljubi.</i>	10
11	(ra)združevanje povedi	<i>Vendar je imel na prikolici naložene cigarete. Cigarete</i> → <i>, ki jih je peljal od Strojne do Kranja.</i>	9
12	izražanje splošnega osebkca oz. vršilca dejanja	<i>Ljudje kadar smo zaljubljeni počnemo</i> → <i>Ko je človek zaljubljen, počne veliko stvari za svojo ljubezen.</i>	8
13	zveze tipa leto in pol nazaj	<i>S prijatelji smo nekaj let nazaj</i> → <i>pred nekaj leti razpravljali o nekaterih stvareh.</i>	7
14	zveze samostalnika in predložne zveze z od ali izmed	<i>Za razliko od</i> → <i>V primerjavi z originalne Antigone pa se pri tem ne zanaša na božje zakone</i>	7
15	neživo „oživi“	<i>Sam odlomek nam sporoča</i> → <i>Iz odlomka izvemo, da v zavodu ni možnosti, da bi pacient lahko povedal svoje mnenje</i>	6

Učenci imajo precej težav pri izražanju svojine; natančneje gre za zveze z desnim neujemalnih nepredložnim in predložnim prilastkom v rodilniku, npr. *oče Ofelije* (gl. zgled v Tabeli 31) in *sestra od Katarine* (400), ki jih učitelji dosledno popravljajo v svojilne pridevnike v funkciji levega ujemalnega prilastka.

400 *Bianka je bila namreč sestra od Katarine* → *Katarinina sestra.*

Za posebej problematični sta se izkazali tudi izbira ustreznega veznika (401) in besedne zveze namesto stavka (402) oz. stavka namesto

besedne zveze (403). Medtem ko je izbira veznika tipično skladenjski problem, je pretvorba besedne zveze v odvisnik in obratno tipično slogovni popravek, ki se nanaša na (boljšo) berljivost besedila in ne na slovnično pravilnost povedi.

401 Prišlo je prvo državno tekmovanje **kjer** → **na katerem** je zmagala.

402 Ko je Ofelija izvedela **za očetovo smrt** → **da je oče umrl**

403 **Potem pranje cunj** → **Perilo so prali** v potokih, vse na roke ni bilo pralnih strojev.

Tipičen slogovni popravek so tudi pretvorbe struktur s pomensko izpraznjenimi besedami, predvsem glagoli, v strukture s polnopomenskim glagolom. Najpogosteje gre za strukture z glagolom *biti* (404) ali drugimi splošnimi glagoli (405) namesto samostalnikov, ki so pomensko jedro zveze. V tem primeru popravki delno odražajo tudi osebne preference učiteljev, saj nekateri tovrstne strukture popravljajo, drugi pa jih uvajajo tam, kjer učenec uporabi polnopomenski glagol (406).

404 **Sem proti** → **Nasprotujem temu**, da gre več kot polovico hrane ki je v restavraciji namenjena dijakom v smeti

405 **nista imela nobene komunikacije med seboj** → **med sabo se nista pogovarjala**

406 Nato je poklicala eno od učenk in **jo začela spraševati** → **ji začela postavljati vprašanja**.

Učencem povzročajo preglavice tudi vsebine, ki od njih zahtevajo poznavanje leksikalnih in skladenjskih značilnosti uporabljenih besed. Natančneje gre za napake, vezane na vezljivost glagolov in pridevnikov ob povedkovem določilu, tj. napačna izbira sklona in predloga (407), napačne izbire sklona in predloga v desnem prilastku samostalniške zveze (408) in napačne kolokacije (409). Nekoliko manj težav imajo učenci pri odločanju med polnopomenskimi glagoli in strukturami z nedoločniki (410) in z rabo trpnika (411).

407 Obiskal me je tudi slepi starec, ki mi je svetoval naj ponovno premislim **svojo odločitev** → **o svoji odločitvi**, jaz pa sem ga ignoriral.

408 saj smo ljudje bolj svobodni in lažje izrazimo svoje mišljenje, svoje neodobravanje **nad nečem** → **nečesa**.

409 Svit se je odrekel **vse nepridiprave družbe** → **slabi družbi** ter se je zaprl vase.

410 saj jim je bilo to v ponos, **braniti** → **da so branili** svojo deželo.

411 Oba pojma pa **sta upoštevana iz strani Cerkve** → **upošteva Cerkev**

Popravki, ki se nanašajo tudi na rabo dajalniških zvez ob glagolu *biti* (412, 413, 414) in na združevanje več povedi v eno oz. razdruževanje stavkov v samostojne povedi, so tipično slogovne narave.

412 želi na vsak način pokopat Polinejka, saj **ji je brat** → **ga kot brata ljubi**.

413 V vsem tem dogajanju **mi je** → **se mi je zdelo** smešno že Matičkovo jezikanje,

414 **Največja sramota ji je bila** → **Največjo sramoto pa je doživela**, ko so jo druge službnice dvignile krilo,

Med probleme struktur, s katerimi imajo učenci manj težav, spadajo še izražanje splošnega osebkca (npr. *Ljudje kadar smo zaljubljeni počnemo* namesto *Ko je človek zaljubljen, počne*), raba časovnih zvez tipa s prislovom *nazaj* (npr. *nekaj let nazaj* namesto *pred nekaj leti*), slogovno neustrezna raba zvez z *od* in *izmed* (npr. *za razliko od* namesto *v primerjavi z*) in personifikacija (npr. *Sam odlomek nam sporoča* namesto *Iz odlomka izvemo*).

Ostali jezikovni problemi, ki so bili detektirani pri analizi napak neustreznih struktur in v Tabeli 31 niso predstavljeni, se med drugim nanašajo na rabo povedkovega prilastka, namenilnika, pogojnika, deiktčnih sredstev, zastarelih besed in struktur z zvezo s *strani*, neustrezno izbiro zaimka oz. zaimkovne zveze, izražanje modalnosti, povratnosti in recipročnosti ter tvorjenje primerjav z veznikom *kot*. Naštete vsebine so se glede na pogostost pojavljanja v korpusu izkazale za manj problematične, zato v pričujočem poglavju niso bile opisane.

6.3 Sklep

Analiza označenih skladenskih napak v korpusu Šolar in njihovih popravkov je pokazala, da učenci na nadbesedni ravni delajo malo dejanskih napak. Najpogostejše, ki terjajo največ pedagoške in jezikoslovne pozornosti, so nekatere napake besednega reda, npr. v povezavi z naslonskim nizom ali za veznikoma *pa* in *vendar*. Težavna sta tudi položaj členkov *tudi*, *še*, *že* ter zaporedje pridevnikov in zaimkov v samostalniški besedni zvezi. V večstavčnih povedih učenci izpuščajo ali uporabljajo odvečne pomožne glagole, deležnike, nikalne členke, zaimenske predmete ali prosta morfema *se* oz. *si*. Enako velja za referenco pred odvisnikom – včasih učenci uporabijo odvečni zaimek ali prislov, včasih pa ga izpustijo. Med napakami strukture izstopa izražanje svojilnosti, pri kateri učenci namesto svojilnega pridevnika uporabijo zvezo s samostalnikom v drugem sklonu. Pri mnogih težavah, ki so v korpusu Šolar označene kot napake skladnje, pa gre zaradi narave označevalne sheme v resnici za probleme oblikoslovja, npr. pri dvojnem stopnjevanju pridevnikov ali prihodnjiku glagola *biti* z rabo deležnika (*bom bil*), kar še enkrat kaže na že večkrat omenjeno prekrivanje označevalnih kategorij.

Nekoliko presenetljivo se več popravkov, ki jih v besedilih učencev delajo učitelji, nanaša na slogovno neustreznost, kadar ne ustrezajo žanru šolskega pisanja, ali na željo učiteljev po večji razumljivosti besedil (npr. manjkajoči osebek ali predmet). Še večkrat pa popravki odražajo okus učiteljev. Tipičen primer so nekateri popravki besednega reda zaradi upoštevanja členitve po aktualnosti. Tovrstne, pretežno

slogovne napake skladnje so za odpravljanje najpomembnejših napak učencev manj relevantne, njihova podrobna analiza pa bi nudila dragocen vpogled v to, katere skladenjske rešitve se slovenskemu pedagoškemu kadru zdijo slogovno ustreznejše in po njihovem jezikovnem čutu verjetno tudi „lepše“.

7 Povezane napake

SARA MOŽE

7.1 Uvod

V besedilih učencev se pogosto pojavljajo napake, ki so med seboj povezane, zato se je v fazi priprave korpusa Šolar zdelo smiselno razmerja med tovrstnimi napakami posebej zabeležiti.³⁸ Povezane napake so različnih vrst – v korpusu je tako zaslediti povezane napake ločil, oblike, besedišča in skladnje. Tipičen primer povezanih napak so denimo napake ločil in velike oz. male začetnice; gre za to, da učitelj v besedilu popravi končno ločilo v vejico ali obratno, kar sproži popravek velike ali male začetnice (Rozman idr. 2012: 38):

415 Ko sta se obo njegovo otroka poročila je kmetijo zapustil sinu. Saj je vedel da bo zanjo dobro skrbel.

Popravljena poved: Ko sta se obo njegovo otroka poročila, je kmetijo zapustil sinu, saj je vedel, da bo zanjo dobro skrbel.

Čeprav sta popravka v zgledu 415 dva, je učenec v povedi napravil dejansko samo eno napako – piko je uporabil tam, kjer bi bilo ustrezneje uporabiti vejico in povedi združiti v eno. Popravek, ki se navezuje na rabo velike začetnice, je torej rezultat nekakšne „veržne reakcije“, ki jo je sprožil izhodiščni popravek. Na začetku analize povezanih napak je tako postalo jasno dvoje:

- Končno število povezanih napak ne odraža dejanskega števila napak, ki so jih v besedilu napravili učenci.
- Pri analizi povezanih napak ni smiselno upoštevati opazovanega popravka, temveč izhodiščno napako, ki je popravek sprožila.

Večina povezanih napak je bila že analiziranih v okviru analiz napak zapisa, besedišča in skladnje, zato v tem poglavju ne bodo posebej predstavljene.³⁹ Povezane napake oblike pa so se izkazale za nad vse relevantne, zato jih je bilo smiselno iz korpusa posebej izločiti, analizirati in kategorizirati. V nadaljevanju sledi opis postopka obdelave podatkov, ki je nekoliko drugačen kot pri ostalih tipih napak, in identifikacije novih in najrelevantnejših jezikovnih problemov, ki so bili določeni pri analizi povezanih napak oblike.

7.2 Analiza povezanih napak oblike

Iz korpusa Šolar je bilo skupaj izločenih 1241 povezanih napak oblike – od tega se je kar 835 zgledov oz. več kot dve tretjini (67 %) nanašalo na jezikovne probleme, ki so bili že identificirani pri analizi drugih tipov napak, predvsem oblike in skladnje, 322 (26 %) se jih je

38 Za več informacij o označevanju povezanih napak v korpusu Šolar gl. Rozman idr. (2012).

39 Večina povezanih napak zapisa se nanaša na rabo ločil in male oz. velike začetnice, kot je prikazano v obravnavanem zgledu. Pri tovrstnem združevanju (oz. razdruževanju) povedi gre navadno za slogovni popravek in ne za večjo napako.

izkazalo za nerelevantne, zgolj 84 (skoraj 7 % vseh zgledov) pa se jih je nanašalo na še neidentificirane, potencialno relevantne jezikovne probleme.

Pri analizi je bila vsaki pojavitvi povezane napake oblike pripisana unikatna oznaka za jezikovni problem, ki se nanaša na izvirno napako v besedilu; npr. če je učitelj v povedi popravil sklon besede (napaka oblike, npr. *primero* namesto *primera* v zgledu 416), ker je pred tem vnesel popravek besedišča (*najdemo* namesto *se nahajajo* v zgledu 416), je bil zgled uvrščen v kategorijo *sprememba besede*, ki spada med nerelevantne povezane napake oblike, in ne v eno od oblikoslovnih kategorij.

416 V pesmi Studenca **se nahaja primera** → **najdemo primero**:

„in kakor pot ju žene,“.

Večini izločenih zgledov je bilo mogoče pripisati oznako za že identificirani jezikovni problem, ostali pa so bili nadalje razvrščeni glede na relevantnost problema, na katerega se nanašajo. Med nerelevantne povezane napake oblike spadajo vse spremembe oblike, ki so posledica predvsem slogovnih (417) in vsebinskih (418) popravkov:

417 V vsakdanjem življenju se nam velikokrat zgodi, da slišimo ali preberemo kakšne pretesljive novice **v zvezi z dogajanjem**

→ **o dogajanju** po svetu.

418 Ima 203 strani in roman je bil tudi nagrajen leta 1998 **v oddaji**

Večernica → **z nagrado večernica**.

V nadaljevanju so predstavljeni novi, potencialno relevantni jezikovni problemi, ki bo bili določeni pri analizi povezanih napak oblike.

7.2.1 Iskanje novih problemov

Na podlagi 84 relevantnih zgledov, ki jih ni bilo mogoče uvrstiti v eno od obstoječih kategorij, je bilo identificiranih deset novih jezikovnih problemov, ki so predstavljeni v Tabeli 32.

Tabela 32: Novi jezikovni problemi, ki so bili določeni pri analizi povezanih napak oblike

Št.	Jezikovni problem	Primer
1	raba preteklika v pripovedih	<i>Matiček je tudi optimist, ljubi Nežko in se ni menil</i> → <i>ne meni</i> , če mu je šlo kaj narobe.
2	izražanje sredstva (od/preko namesto z/s)	<i>raje se preživljajo od socialne pomoči</i> → <i>s socialno pomočjo</i>
3	raba tožilnika namesto roditelja (slogovni popravek)	<i>Otroke je treba od malih nog učiti red</i> → <i>reda in kultura</i> → <i>kulture</i> .
4	raba edninskih in dvojskih oblik samostalnika <i>starši</i>	<i>Sem doživela tudi sama, ko sta</i> → <i>so mi starša</i> → <i>starši</i> govorila, da ne smem imeti konja
5	Prislovna določila kraja: smer ali statična lokacija?	<i>Če ste videli katero od njiju se oglasite na policijsko postajo</i> → <i>policijski postaji</i> .

6	raba skupnega imena za posamezne člane skupine	<i>Matiček s svojo pametjo, razsodnostjo in razumom pokaže baronu zobe, čeprav je samo grajski vrtnar in tako nižji sloj -> predstavnik nižjega sloja.</i>
7	ločevanje med dajalnikom in orodnikom	<i>Mislím, da sta onadva kljub tragičnim koncem -> tragičnemu koncu še vedno zaljubljena drug v drugega.</i>
8	mernostna modifikacija pridevnikov	<i>Prvo preneseno gibanje je bilo na približno enem decimetru -> en decimeter velikem zaslonu.</i>
9	glagol v prilastkovem odvisniku vpliva na sklon odnosnice	<i>Bil je na podeželju in prvo osebo -> prva oseba, katero je spoznal je bil Albin.</i>
10	glagolski čas v poročanem govoru	<i>Tudi jaz sem ji povedal, kako zelo sem jo ljubil -> jo ljubim.</i>

Analiza povezanih napak oblike je sicer omogočila identifikacijo novih jezikovnih problemov, vendar pa se ti v korpusu pojavljajo precej poredko, kar kaže na to, da učenci z njimi nimajo večjih težav. Problemi z najvišjo pogostostjo pojavljanja v korpusu, tj. izbira med preteklikom in sedanjikom v pripovedih, izražanje sredstva (npr. *preživljati se od socialne pomoči* namesto *preživljati se s socialno pomočjo*) in raba tožilnika, namesto katerega bi bilo morda bolje uporabiti roditelja (npr. *učiti red in kulturo* namesto *učiti reda in kulture*), so pogosto slogovne narave in v večji meri odražajo jezikovne preference posameznih učiteljev. Dober primer nedoslednosti učiteljskih popravkov je raba preteklika v pripovedih; medtem ko so nekateri učitelji naklonjeni rabi preteklika v pripovedih in obnovah (419), se drugi z učenčevim rabo sedanjika strinjajo in nedosledno rabljene glagole v pretekliku popravljajo v sedanjik (420):

419 Pri tem pa čez nekaj časa **zboli**.

Popravljen poved: Pri tem pa **je** čez nekaj časa **zbolel**.

420 Matiček je tudi optimist, ljubi Nežko in se **ni menil** -> **ne meni**, čemu je šlo kaj narobe.

Z jezikoslovnega vidika so zanimivi predvsem jezikovni problemi s sredine in dna tabele, tj. sklanjanje samostalnika *starši*, ločevanje med prislovnimi določili, ki izražajo smer in statično lokacijo, raba skupnih imen, npr. *družina* in *sloj*, ko imamo v mislih le posamezne člane skupine, zamenjevanje dajalniških in orodniških oblik samostalnikov in pridevnikov, mernostna modifikacija pridevnikov (npr. *na približno enem decimetru velikem zaslonu* namesto *na približno en decimeter velikem zaslonu*), vpliv vezljivostnih lastnosti glagola iz prilastkovnega odvisnika na sklon odnosnice v nadrednem stavku (npr. *osebo katero je spoznal* namesto *oseba, ki jo je spoznal*) in izbira glagolskega časa v povedih s poročanim govorom. Naštete vsebine smiselno dopolnjujejo končni nabor problemov, ki vključuje tako nove kot stare jezikovne probleme.

7.2.2 Končni nabor

Končni nabor relevantnih jezikovnih problemov, ki so bili identificirani pri analizi povezanih napak oblike, vključuje 10 novih in 37 starih jezikovnih problemov, na katere se nanaša skupno 919 zgledov iz proučevanega vzorca. Glede na pogostost pojavljanja v korpusu in relevantnost je v nadaljevanju predstavljenih le nekaj pomembnejših jezikovnih problemov, ki učencem povzročajo največ preglavic.

Za posebej problematično se je izkazalo ujemanje med osebkom in povedkom oz. med prilastkom in jedrom zveze v spolu (421), sklonu (422) in številu (423) ter ob zvezah, ki izražajo količino (424). V proučevanem vzorcu je bilo zaslediti skupno 216 zgledov napak ujemanja (več kot 23 % vseh relevantnih zgledov), kar kaže na visoko relevantnost identificiranega jezikovnega problema.

421 **To isto** → **Ta isti** večer je na zabavo prišel nepovabljeni Romeo Monteg.

422 To se najbolj vidi iz **prva dva verza** → **prvih dveh verzov**

423 Bolj **pomembna je postala** → **pomembni sta postali** zvitost in iznajdljivost, pa čeprav lahko v odzadju **skriva** → **skrivata** same laži in bedo.

424 Njen pogreb je bil zelo skromen in **udeležili so** → **udeležilo se je** zelo malo ljudi med njimi tudi Hamlet.

Druga pomembnejša kategorija je raba rodilnika ob zanikanem povedku. V korpusu je bilo mogoče najti kar 103 zglede rabe tožilnika ob zanikanem povedku (425), tj. več kot 11 % vseh relevantnih zgledov, kar kaže na to, da imajo učenci veliko težav pri izbiri ustreznega sklona.

425 Tudi otroci niso imeli **normalno otroštvo in življenje** → **normalnega otroštva in življenja**.

Za posebej težavno se je izkazalo tudi tvorjenje prilastkov samostalniških besednih zvez (88 zgledov oz. skoraj 10 %). Sem spadajo predvsem napake, vezane na izbiro napačnega predloga oz. sklona v predložnih zvezah, ki imajo funkcijo desnega prilastka:

426 Obdal me je srh, moja sovražnost do strica pa je prerasla v željo **do uboja** → **po uboju**.

Pri analizi je bilo zaslediti tudi relativno veliko zgledov, vezanih na neustrezno rabo dvojine (39 zgledov oz. nekaj več kot 4 %); natančneje gre za napake v ujemanju oz. rabi množinske oblike, pri kateri bi bilo treba uporabiti dvojinško:

427 Poniževanje in posmehovanje **so prisotne** → **sta prisotni** povsod po svetu po mojem mnenju.

Raba sklonov je pereč problem, s katerim se pri pisanju soočajo številni učenci. Pri analizi relevantnih povezanih napak oblike je bilo mogoče najti 46 zgledov (5 %) neustrezne izbire predloga, pri čemer je treba poudariti, da ne gre za izbiro napačne oblike, kot je npr. raba *s* namesto *z* – takšne napake so bile namreč obravnavane v okviru

napak črkovanja (gl. 3. poglavje) – temveč predvsem za slogovno in delno tudi pomensko manj ustrezne predloge, kot je razvidno iz zgl. 428:

428 Ko sta prišla **do mene** → **k meni** sta rekla: „Zdaj si pa nekaj zaželi.“

Zadnji jezikovni problem, ki ga velja posebej izpostaviti, je vezan na razločevanje med skloni (skupno 21 zgledov oz. nekaj več kot 2 %) – učenci namreč pogosto zamenjujejo oblike v mestniku in dajalniku (429) oz. v mestniku in orodniku (430), včasih pa celo oblike v dajalniku in orodniku (gl. razdelek 7.2.1).

429 Kljub **vsem tem** → **vsemu temu** je Prometej ustrajal.

430 Knjiga ima 189 strani in govori o današnjih najstnikih in **njihovimi problemi** → **njihovih problemih** v šoli.

Poleg opisanih zgledov je zaslediti tudi precej napak, povezanih z glagolsko in pridevniško vezljivostjo, frazemi in izražanjem splošnega osebk, ter slogovne in vsebinske popravke, ki v pričujočem poglavju bodisi zaradi nerelevantnosti bodisi zaradi tega, ker so bili natančneje obdelani in opisani v okviru analize ostalih tipov napak, niso natančneje predstavljeni.

7.3 Sklep

V pričujočem poglavju je bil na kratko predstavljen koncept povezanih napak s posebnim poudarkom na napakah oblike, ki so bile v okviru analize podatkov iz korpusa Šolar dodatno izločene, analizirane in kategorizirane. Povezane napake so se izkazale za koristen vir podatkov, na podlagi katerih je bilo mogoče obstoječi nabor jezikovnih problemov smiselno dopolniti in s tem pridobiti celostno sliko o jezikovni kompetenci učencev. Eksplicitno označevanje povezanih napak omogoča tudi večjo objektivnost v postopku analize podatkov iz korpusa Šolar, saj tovrstne oznake jezikoslovcem služijo kot opozorilo, da količina učiteljskih popravkov v pisnem izdelku navadno ne odraža dejanskega števila napak, ki jih je napravil učenec. Vsaj del popravkov je rezultat verižne reakcije, ki ga je sprožil izvorni učitelj popravek, kar pomeni, da je pri kvalitativni analizi in statistični obdelavi podatkov treba upoštevati tudi posebnosti povezanih napak.

40 Sestava vzorca je takšna, da omogoča nadaljnje ločevanje jezikovnih napak, ki so tipične za celotno populacijo, od tistih, ki so bolj regionalno vezane.

8 Uporaba rezultatov pri pripravi pedagoške korpusne slovnice

ŠPELA ARHAR HOLDT

8.1 Uvod

Korpusna analiza jezikovnih napak slovenskih učencev pri naši poučevanju slovenščine pomemben vpogled v težavna mesta slovenskega pravopisa in slovnice. Že samo preprost seznam tipov in podtipov aktualiziranih jezikovnih napak je s stališča jezikovne didaktike izrednega pomena, še bolj dragocen in povsem nov pa je podatek o tem, kako pogosta je določena jezikovna napaka v uravnoteženem vzorcu šolajoče se populacije (gl. razdelek 2.2 v 2. poglavju).⁴⁰

Učna gradiva za učenje slovenščine (npr. učbeniki, delovni zvezki, raznovrstna e-gradiva) se trenutno pripravljajo z namenom čim bolj pokriti široko paleto različnih *potencialnih* težav, s katerimi bi se učenci pri svoji lastni jezikovni produkciji lahko srečali. Korpusni podatki pa kažejo objektivno, statistično sliko, s katerimi težavami so se učenci *dejansko* srečali, kar med drugim omogoča:

1. prioritizacijo učnih vsebin glede na dejansko pogostost in razširjenost posameznih jezikovnih napak,
2. izhajanje iz konkretne jezikovne težave (reševanje jezikovnih problemov od spodaj navzgor namesto od zgoraj navzdol),
3. izhajanje iz avtentičnega jezika (namesto uporabe izmišljenih primerov za zglede in vaje).

Vprašanju, kako korpusno analizo jezikovnih napak čim kvalitetneje prenesti v poučevalno prakso, se med leti 2010 in 2013 intenzivno posvečamo v sklopu projekta *Sporazumevanje v slovenskem jeziku* pri pripravi pedagoške korpusne slovnice. Ker je (kot pri številnih drugih raziskavah) prehod od teoretičnih ugotovitev k praktični uporabi zahteven, dolgotrajen, pa tudi kreativen in zato za nadaljnji razvoj zanimiv proces, pričujoče delo zaključujemo s to tematiko.

8.2 Pedagoška korpusna slovnica

Kot pove že samo ime, je pedagoška korpusna slovnica jezikovni vir, namenjen za uporabo pri poučevanju oz. učenju slovnice, pripravljen pa je na osnovi podatkov iz besedilnih korpusov. Poleg korpusa Šolar, na katerega se osredotoča ta knjiga, smo pri pripravi pedagoške

slovnice uporabljali še korpus pisnih besedil Gigagida⁴¹ in korpus govorjene slovenščine GOS.⁴²

Ker so ravno korpusni podatki tisto, kar dela novi jezikovni vir drugačen, je za začetek smiselno poudariti, da sta korpusni in pedagoški pogled na jezik v resnici precej različna, v določeni meri se celo izključujeta. Če korpusna slika na eni strani kaže jezikovno rabo v vsej njeni heterogenosti, večplastnosti in kompleksnosti, pa pedagoška situacija na drugi strani zahteva preglednost, predvidljivost in pa vsaj v izhodišču tudi jasne smernice glede tega, kaj je v jeziku prav in kaj narobe, kaj se sme in kaj ne.

Pedagoška korpusna slovnica skuša oba konca povezati, kolikor je to mogoče. V idealnem primeru bi morala biti pri učenju jezikovnih pravil vzpostavljena jasna povezava med teorijo in prakso, jezikovno normo in jezikovno rabo, abstraktnim in konkretnim, na osnovi česar bi učenci lahko pravila standardne slovenščine najprej spoznali in utrdili, nato pa tudi ustrezno relativizirali. Zadnji navedeni korak je pomemben zlasti zato, da učenci uzavestijo, da jezik ni črno-bela šahovnica absolutnega *prav* in absolutnega *narobe*, ampak da prinaša za različne situacije in kontekste različne izbire. Premišljena uporaba korpusnih podatkov pri jezikovnem pouku lahko pomaga narediti korak v to smer.⁴³

8.2.1 Nekaj izbranih problemov poučevanja slovenščine

V tem poglavju na kratko predstavljamo štiri probleme poučevanja slovenščine, na katere smo bili pri snovanju pedagoške slovnice posebej pozorni oz. smo se jim pri zasnovi novega vira poskušali izogniti.⁴⁴

1. Pri izdelavi večine gradiv niso upoštevani podatki o tem, kako pogosto do določenih jezikovnih napak pri pisni produkciji učencev v resnici prihaja. Avtorji gradiv se zanašajo na lastne izkušnje z jezikovno rabo oz. jezikovnimi napakami in predvidevanja, s čim bi učenci lahko imeli težave, oz. slednje povzemajo po drugih virih. Posledica tega dejstva je, da pri pouku vprašanja, na katera bodo učenci naleteli v vsakem spisu, enakovredno tekmujejo za čas za vprašanja, s katerimi se bodo učenci srečali zelo redko.⁴⁵
2. Mnoga gradiva v jezikovnih zgledih in slovnicih nalogah prinašajo izmišljene primere jezika, s pomočjo katerih naj bi učenci utrjevali posamezno jezikovno pravilo. Takšne vaje ne omogočajo dobrega prenosa teoretičnega znanja v prakso, saj se avtentična jezikovna raba od konstruiranih primerov običajno precej razlikuje.
3. Ko vsi učenci pri pouku ali doma rešujejo iste slovnicihne naloge, se individualnih jezikovnih težav, ki jih imajo učenci, pogosto ne do-
taknemo. Možnosti za individualizacijo v tem smislu so pri pouku

41 Korpus Gigafida vsebuje slovenska pisna besedila, nastala med letoma 1990 in 2011; obsega več kot milijardo besed. Korpus je dostopen na strani <http://www.gigafida.net>, izčrpn pa je opisan v monografiji Logar Berginc idr. (2012).

42 Korpus GOS je prvi slovenski korpus govorjenega jezika, obsega okrog 120 ur posnetkov govora v različnih življenjskih situacijah. Korpus opisuje monografija Verdonik in Zwitter Vitez (2011), dostopen je na strani <http://www.korpus-gos.net>.

43 O možnostih in omejitvah uporabe korpusov v razredu gl. npr. Krek in Arhar Holdt (2010).

44 Problemi so bili identificirani v sklopu priprav na snovanje pedagoške korpusne slovnice. Del rezultatov predstavlja projektno poročilo Arhar Holdt idr. (2011). S šibkimi in močnimi mesti poučevanja slovenščine se bolj sistematično ukvarja Rozman idr. (2012).

45 Takšna situacija ni značilna le za poučevanje slovenščine, ampak za učenje jezikov nasploh. Za angleški prostor o tem piše denimo A. Tsui (2004).

slovenščine precej omejene: učitelji vlagajo čas v popravljanje pisnih izdelkov učencev, učenci pišejo popravo, ni pa seveda realno pričakovati, da bodo učitelji za vsakega učenca posebej pripravljali specializirane jezikovne vaje oz. mu znova in znova individualno razlagali rešitve za vsak specifičen jezikovni problem.

4. Kljub temu da učni načrti predvidevajo usposabljanje učencev za samostojno reševanje jezikovnih problemov, je v resnici le od časa, razgledanosti in osebnih preferenc učitelja odvisno, na kakšen način bo v pouk vključil spoznavanje posameznih jezikovnih virov. Veliko obstoječih referenčnih priročnikov je za učence prezapletenih za intuitivno razumevanje.⁴⁶ Za ustrezno interpretacijo predstavljenih jezikovnih podatkov bi bilo torej učence treba posebej usposobiti, kar je seveda časovno precej potratna naloga.

Potrebno je poudariti, da naš namen nikakor ni kritiziranje obstoječih gradiv ali celo učiteljev, saj pred nastopom korpusnega jezikoslovja podatki, ki bi situacijo lahko izboljšali, niso bili na voljo. Razen tega je potrebno vzeti v ozir, da so trenutno tako učitelji slovenščine kot avtorji gradiv postavljeni v zahtevno situacijo, saj morajo sami premoščati vrzel, ki se pojavlja med slovenščino, opisano v obstoječih jezikovnih priročnikih, in sodobnim jezikom, ki ga poznajo in uporabljajo učenci. Pedagoška korpusna slovnica naj bi to povezavo vsaj deloma olajšala.

8.2.2 Cilj pedagoške korpusne slovnice

Z upoštevanjem do sedaj povedanega je bila pedagoška korpusna slovnica zasnovana kot brezplačno dostopen spletni vir, ki:

- se osredotoča na realne jezikovne probleme šolajoče se populacije,
- sloni na metodološko sodobnih analizah slovenščine,
- temelji na avtentičnem jeziku, prinaša izključno avtentične zglede in vaje,
- omogoča vsakemu učencu, da se osredotoči na točno tista in le tista poglavja, s katerimi ima sam težave,
- učenca neposredno in posredno uči uporabe spletnih jezikovnih virov in jezikovnih tehnologij za slovenščino,
- na jezikovna pravila gleda s praktičnoupornega zornega kota in odgovarja na vprašanje „kdaj mi bo učenje slovnice v življenju sploh prišlo prav“.

Kako smo te smernice prenesli v prakso, razlagajo poglavja v nadaljevanju.

8.2.3 Struktura slovnice poglavja

V skladu s splošnim tehnološkim razvojem družbe je pedagoška korpusna slovnica zasnovana kot spletni vir – Pedagoški slovnice portal.⁴⁷ Na portalu so zbrana posamezna slovnice poglavja, od katerih je vsako posvečeno enemu samemu konkretnemu jezikovnemu problemu, ki je bil identificiran pri analizi korpusa Šolar. Teme so npr. Zapis zanikanih glagolov (gl. razdelek 3.4 v 3. poglavju), Uporaba besed *nobeden* in *noben* (gl. razdelek 4.2.3 v 4. poglavju), Sklanjanje samostalnika *otrok* (gl. razdelek 5.2.11 v 5. poglavju) in podobno.

Slovnice poglavja so napisane v preprostem, plastičnem in konkretiziranem jeziku, povedi so kratke in enostavno strukturirane, jezikovni termini so uporabljani le, če je to neobhodno potrebno. Tisti težji pojmi, ki se v posameznem poglavju pojavijo (običajno gre za terminologijo, ki poimenuje besedne vrste, se pravi *samostalnik*, *glagol*, *predlog* itd.), so znotraj poglavja tudi razloženi, kar pomeni, da ima učenec vse informacije, ki jih potrebuje za razumevanje vsebine, neposredno pri roki. Čeprav je rdeča nit vsakega poglavja pojasniti izbrani jezikovni problem, pa so učencu na voljo tudi abstraktnejše vsebine, ki obravnavani problem umeščajo v širše področje slovnice in na ta način povezujejo konkretno z abstraktnim. Ta del poglavja je na voljo v ločenem oknu, do katerega učenec dostopa s klikom na posebno povezavo. Zainteresiranim učencem so torej zahtevnejše vsebine enostavno dostopne, vendar pa z njimi ne prekinjamo primarne razlage jezikovnega problema.

Notranja struktura vsakega poglavja je sestavljena iz več razdelkov. Glavnino slovnice gesla predstavlja **razlaga**, v kateri je jezikovni problem predstavljen. Razlaga problema je zgrajena po principu od bolj znanega k manj znanemu, nove vsebine se na spletni strani odpirajo postopoma, da učenci niso preobremenjeni s količino novih informacij. Vsi zgledi rabe v razlagi so realni in izvirajo iz enega od treh uporabljenih korpusov. Zgled iz korpusa Šolar uporabljamo takrat, ko ponazarjamo specifične obravnavane jezikovne probleme (npr. do kakšnih napak tipično prihaja pri rabi namenilnika, gl. razdelek 5.2.1 v 5. poglavju), zgled iz korpusa Gigafida in Gos pa takrat, ko opisujemo splošne značilnosti pisanja ali govorjenja v slovenščini (npr. s katerimi glagoli se namenilnik tipično pojavlja). Ker korpus Gos prinaša posnetke govorne slovenščine, lahko te zgled učencem tudi poslušajo.

Razlago dopolnjujejo krajše slovnice naloge, s katerimi lahko učenec sproti preverja, ali je snov razumel. Reševanje nalog učenca obenem aktivira in motivira, saj zgolj branje vsebin na spletu hitro postane monotono in nezanimivo. Zadnja točka vsake razlage pa je napotilo, kako si učenec v zvezi z obravnavanim problemom lahko pomaga. Nekatera poglavja prinašajo specializirane mnemotehniko, ki pomagajo učencu, da si zapomni določeno jezikovno pravilo, druga napotilo na jezikovni vir, v katerem lahko učenec najde odgovor na svoje vprašanje.

⁴⁷ Portal bo na voljo konec leta 2013, napredek pa lahko spremljate na <http://www.slovenscina.eu>. Razlog za spletno obliko je tudi želja po brezplačni dostopnosti. Zavedamo se dejstva, da nimajo vsi učenci in dijaki na voljo dostopa do interneta, vendar smo presodili, da je lažje in hitreje pripraviti kakovostna tiskana gradiva na osnovi spletnega vira kot pa obratno.

48 Primer napake *Nihče nesme povzročati bolečin drugim* je bil denimo uvrščen med napake zapisa, natančneje zapisa skupaj oz. narazen, znotraj tega pa opredeljen kot jezikovni problem pisanja nikalnice skupaj z zanikanim glagolom (gl. tudi razdelek 4.4 v 4. poglavju).

Vsako slovnično poglavje prinaša še nekatere druge, nekoliko krajše razdelke. **Povzetek** na kratko združuje ključne točke razlage in tako omogoči učencu, da še enkrat strnjeno vidi glavne, najpomembnejše točke v zvezi z obravnavanim jezikovnim problemom. Rubrika **Pozor** prinaša posebna opozorila, tipično povezana z regionalno vezanimi vidiki obravnavanega problema, izjemami k predstavljenim pravilom ali posebnostmi na ravni odnosa med govorjenim in pisnim jezikom. Učencem, ki bi radi o obravnavani temi izvedeli kaj več, je namenjena rubrika **Za navdušene**, v kateri je predstavljena izbrana zanimivost, povezana z jezikovnim problemom, tipično v obliki statističnih podatkov o slovenščini, pogosto pa tudi v povezavi z uporabo spletnih jezikovnih virov in jezikovnih tehnologij. Nazadnje so v razdelku **Vaje** zbrane vse slovnične naloge, ki so se pojavile med razlago, s tem da lahko učenec na tem mestu rešuje veliko število primerov (npr. 200 primerov za posamezno nalogo), program pa sproti preverja njegovo uspešnost in spremlja njegov napredek.

8.3 Pilotna slovnična poglavja

Upoštevajoč časovno-finančne okvire projekta *Sporazumevanje v slovenskem jeziku* je bilo določeno, da se v prvem koraku pripravi trideset slovničnih poglavij, poleg tega pa še do deset dodatnih poglavij, ki učijo uporabe spletnih jezikovnih virov. Ta poglavja imamo lahko za pilotna poglavja, pri izdelavi katerih smo razvili postopke kvantitativne in kvalitativne obdelave jezikovnih podatkov, na njihovi osnovi razvili oblikovno podobo Pedagoškega slovničnega portala (skupaj s sistemom za upravljanje spletnih vsebin), v zaključni fazi projekta pa bodo uporabljena tudi za evalvacijo portala in pripravo smernic za njegov nadaljnji razvoj.

8.3.1 Seznam slovničnih vsebin

Pri izbiri jezikovnih problemov za pilotna slovnična poglavja smo izhajali iz analize napak v korpusu Šolar. Kot kaže Tabela 12 v razdelku 2.3 v 2. poglavju, je bilo med analizo identificiranih več kot 35.000 napak, ki so bile nato ročno razvrščene v kategorije, podkategorije in na zadnji ravni v specifične jezikovne probleme.⁴⁸ Med analizo je bilo identificiranih 692 specifičnih jezikovnih problemov, kar je na prvi pogled visoka številka, vendar se je treba zavedati, da se številni od teh problemov v korpusu pojavijo zelo redko (medtem ko so drugi zastopani zelo pogosto).

Teme za pilotska slovnična poglavja smo izbrali na način, da bi skupaj kar čim bolj predstavljali heterogeno množico identificiranih jezikovnih težav. To pomeni, da se pri izbiri nismo opirali le na to, kako pogosta je določena težava med mladimi, ampak tudi na to,

za kakšne vrste težavo gre. Med problemi smo želeli imeti: (I) probleme z različnih jezikovnih ravnin, (II) različne vrste problemov s stališča razvoja postopkov za avtomatsko pridobivanje jezikovnih podatkov iz besedilnega korpusa, (III) tako splošne kot regionalno vezane probleme, (IV) tako probleme, pri katerih jezikovna raba sledi jezikovni normi, kot tiste, kjer ni tako, (V) preprostejše probleme, ki jih je mogoče razložiti v enem samem slovničnem poglavju, ter na drugi strani kompleksnejše probleme, ki zahtevajo delitev vsebine na več poglavij.

Tabela 33 prikazuje, kateri problemi so bili glede na gornje kriterije izbrani, skupaj s primerom za vsak problem in številom tovrstnih napak v korpusu Šolar. Tisti problemi, ki se eksplicitno omenjajo v predhodnih poglavjih te knjige, so opremljeni tudi z referenco na relevantno poglavje. Opozoriti pa je treba, da smo pri navedenih tridesetih problemih rezultate analize napak na ravni zapisa, besedišča in oblike obravnavali združeno, zato je mogoče, da je katera od frekvenc v tej tabeli drugačna od frekvenc, navedenih v prejšnjih poglavjih.

Tabela 33: Jezikovni problemi, izbrani za pilotska slovnična gesla

Št.	Jezikovni problem	Primer	Število napak	gl. še pogl.
1	Raba glagolov <i>morati</i> in <i>moči</i>	<i>Mislím, da bi v vsaki družini moglo vladati zaupanje, prijaznost, pomoč in pa tudi kdaj trda roka, če je to res potrebno.</i>	72	4.2.6
2	Preglas	<i>Laert Hamleta zadane z zastrupljenim mečom, nato si v ruvanju nože zamenjata in nadaljujeta.</i>	25	5.2
3	Raba variantnih oblik predlogov <i>z/s</i>	<i>Svoje telo seveda krepimo s veliko gibanja, predvsem na soncu.</i>	381	3.5.2
4	Raba variantnih oblik predlogov <i>k/h</i>	<i>Odnesel me je h socialni delavki, kamor me je položil na zločljivo posteljo, kjer sem se zbudil.</i>	22	3.5.2
5	Zapis predloga v s črko <i>u</i>	<i>Janez je u mestu zelo veliko razmišljal o njej kako je lepa.</i>	11	3.5.2
6	Raba odvečnega pomožnega glagola <i>biti</i> v prihodnjiku	<i>Vem, da se vedno lahko zanesem nate v dobrem in slabem, da boš bil ob meni.</i>	29	6.2.3.3
7	Sklanjanje samostalnikov moškega spola na <i>-r</i>	<i>V Mariborju živi še danes, kot svobodna umetnica.</i>	11	5.2
8	Sklanjanje samostalnikov moškega spola na <i>-o</i>	<i>Bronja se zaljubi v Džurota on je samo pomočnik v avtomehanični delavnici, je tudi precej mlajši od Bronje.</i>	13	5.2
9	Besede s sklopom <i>lj</i>	<i>Povezanost pomeni sožitje, medsebojno spoštovanje in prijateljstvo narodov.</i>	140	3.5
10	Besede s sklopom <i>nj</i>	<i>Romeo pride cel objokan do Julije in še jo zadnič poljubi in objame.</i>	127	3.5
11	Sklop <i>lj/nj</i> v pridevniki na <i>-ski</i>	<i>Njun življenski položaj je bil zelo težak.</i>	68	3.5
12	Sklop <i>nj</i> v pridevniki na <i>-nji</i>	<i>Tudi v današnjih časih se to pogosto dogaja.</i>	51	3.5

13	Zapis dvoustničnega u v izglasju deležnikov na -l	To je delav zaradi spoštovanja Božje besede in zaradi posebne ljubezni do materinščine.	29	3.5.2 5.2
14	Nestandardne glagolske oblike na -ve	Po plesu svi sa z Rozamundo obljubili, da si bove dopisovale.	13	5.2.10
15	Nestandardne glagolske oblike na -ma	Zelo se zabavama , saj igrava igrice.	16	5.2
16	Raba besed nobeden in noben	Noben od vojakov ga ne zapusti, vsi vzamejo v roke orožje .	4	4.2.3
17	Raba zaimkov nobeden in nihče	Bog pa mu je dal znak, da ga noben ne bi ubil.	56	4.2.3
18	Zapis skupaj/narazen – zanikani glagoli	Konča se tako, da velikani oziroma mlini zmagajo , Don kihot pa ugotovi, da nemore rešiti sveta.	548	3.4
19	Zapis skupaj/narazen – prislovi na ne-	Fatima spozna nekega fanta z imenom Adam , vanj se ne normalno zaljubi, spoznala pa ga je na neki zabavi, bil je frajer in jo je samo izkoriščal za njegove potrebe.	27	3.4
20	Stopnjevanje pridevnikov - opisno ali morfemsko?	Ta je s svojo nerazumskostjo eden od najsmušnejših likov, saj ga ta pogosto pripelje v različne situacije, ki se komično razpletejo.	14	5.2
21	Podvojeno stopnjevanje pridevnikov in prislovov	To je tudi ena od idej razsvetljenstva; nižji sloj je pametnejši in bolj spretnjši od plemstva.	7	6.2.3.3
22	Raba namenilnika	Pobriše mizo in se odpravi pomivati posodo dogodek je totalno miren in iz slušalk prihaja glas RAIA in ona ga neprično gleda za njegovim hrbtom.	110	5.2.1
23	Zapis kratkega nedoločnika pod vplivom govora	Jerma pa kljub zmagi nasprotne strani ostane na strani poražene in jo še vedno nekako poskuša širit med ljudmi, čeprav mu to ne uspeva, saj so ljudje že trdno prepričani o prednostih klerikalne stranke, vendar Jerma se nikoli ne vda in zato ga izženejo v Goličavo.	150	5.2.1
24	Glagoli na u- in v-	Večina ljudi si želi na različne načine uplivati na druge.	100	3.5
25	Besede na izs-	Naša družba je prvič v večih letih dobila „vodjo“, nekoga, ki je iztopal , ki je bil nekaj več kot vsi ostali, nekoga, ki je imel glavno besedo.	10	3.5.1
26	Raba naveznih oblik zaimkov	Ob branju napisov na mizi je videla eno izmed vprašanj in na njega odgovorila.	41	3.4 4.2.3
27	Raba nastonskih in naglasnih oblik zaimkov	Najstarejši sin, ki je svojemu očetu ves čas pomagal in mu bil pokoren, pa se zabave ni pridružil, saj mu oče ni nikoli priredil zabave.	5	4.2.3 6.2.1.5
28	Spreganje glagolov po vzorcu atematskih glagolov	Če vzamete svobodo nekomu, ki je s svobodo živel že od rojstva, ga razpolovite in s tem tudi umoriste .	19	5.2
29	Črkovanje primerov s podvojenimi črkami	Vendar župnik na koncu Jermanu vseno prizna, da je ravnal prav.	40	3.5.1
30	Sklanjanje besede otrok	Francka ostane sama s tremi otroci brez denarja.	39	5.2.11

Kot je razvidno iz Tabele 33, problemi posegajo predvsem na področje zapisa, oblikoslovja in besedišča, saj daje za te jezikovne ravnine uporaba korpusnih podatkov, ki so trenutno na voljo, najboljše rezultate. Za korpusno analizo skladenjske ravni jezika so potrebni skladenjsko označeni korpusni podatki, s katerimi v trenutku priprave pilotskih slovnčnih poglavij še ne razpolagamo.⁴⁹

49 So pa v pripravi, gl. npr. Dobrovoljc, Krek in Rupnik (2012).
50 <http://www.gigafida.net>
51 <http://www.korpus-gos.net>
52 <http://bos.zrc-sazu.si/sskj.html>
53 <http://www.slovenscina.eu/spletni-slovar>
54 <http://www.slovenscina.eu/sloleks>
55 <http://bos.zrc-sazu.si/sp2001.html>
56 <http://slogovni.slovenscina.eu>

8.3.2 Seznam jezikovnih virov

Nekaj poglavij na Pedagoškem slovnčnem portalu bo namenjenih tudi predstavitvi spletnih jezikovnih virov, za katere smo presodili, da lahko uporabniku ponudijo relevantne informacije o določenem segmentu slovenskega jezika. Med viri so npr. pisni korpus Gigafida,⁵⁰ govorni korpus GOS,⁵¹ spletna različica slovarja SSKJ,⁵² Spletni slovar slovenskega jezika SSSJ,⁵³ oblikoslovni leksikon Sloleks,⁵⁴ spletna različica Slovenskega pravopisa⁵⁵ in portal Slogovni priročnik.⁵⁶

Jezikovni viri bodo predstavljeni na nekoliko drugačen način, kot to velja za jezikovne probleme. V ospredju poglavja bodo povezave na videoposnetke, v katerih bo z uporabo realnih problemov razloženo, kako se posamezen od virov uporablja in katere jezikovne informacije prinaša (oz. tudi, katerih ne prinaša). Poglavja bodo vsebovala tudi posebna opozorila na močne in šibke točke oz. na nevarnosti neustrežne uporabe vsakega obravnavanega vira.

8.4 Morati ali moči?

V tem razdelku predstavljamo, na kakšen način so bili novi jezikovni podatki uporabljeni pri pripravi enega izmed pilotnih slovnčnih poglavij, in sicer tistega, ki se posveča zamenjevanju glagolov *morati* in *moči*. Zamenjevanje teh dveh glagolov je bilo med označevanjem napak v korpusu Šolar identificirano med napakami besedišča (gl. razdelek 4.2.6 v 4. poglavju) in, v manjši meri, napakami oblike.

8.4.1 Analiza napak

Korpusni podatki pokažejo, da je pri tem jezikovnem problemu ključen vpliv govornega (nestandardnega) jezika. Meja med sedanjsko obliko obeh glagolov je fonetično težko določljiva – v nekaterih narečjih reducirane sedanjske oblike za oba glagola kar sovpadajo. V korpusu GOS npr. za obliko *morš* najdemo 419 primerov (večina jih je iz ljubljanske (129) in mariborske (59) regije), od tega gre v 306 primerih za glagol *morati* in v 113 primerih za glagol *moči*.

Sovpadanje oblik v govoru odpira vprašanje, ali so imeli avtorji napak (431, 432) pri pisanju sploh uzaveščeno, da obstajata v slovenščini dva oblikovno podobna glagola z različnimi pomeni (pa so po pomoti

izbrali napačno možnost), ali pa tega dejstva še niso imeli uzaveščenega in so pri pisanju le sledili občutku „kaj dobro zveni“.

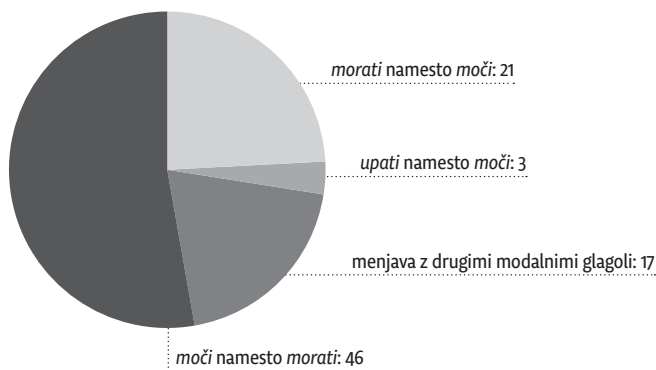
431 Imela sem občutek, da ti **morem** povedati kaj čutim do tebe oziroma, da se ti moram izpovedati.

432 Vendar ne moreš biti bogat če prvo nisi reven tudi veliko lešnikov **ne moraš** imeti, preden jih imaš malo.

Če želimo učencem omogočiti, da se tovrstnim napakam izognejo, jih moramo torej eksplicitno opozoriti, da gre v tovrstnih primerih za dva različna glagola in pa, da je zanašanje na občutek „kaj dobro zveni“ pri tem problemu lahko zavajajoče (da morajo ustrezno obliko izbrati zavestno in z upoštevanjem, kaj kateri od dveh glagolov pomeni).

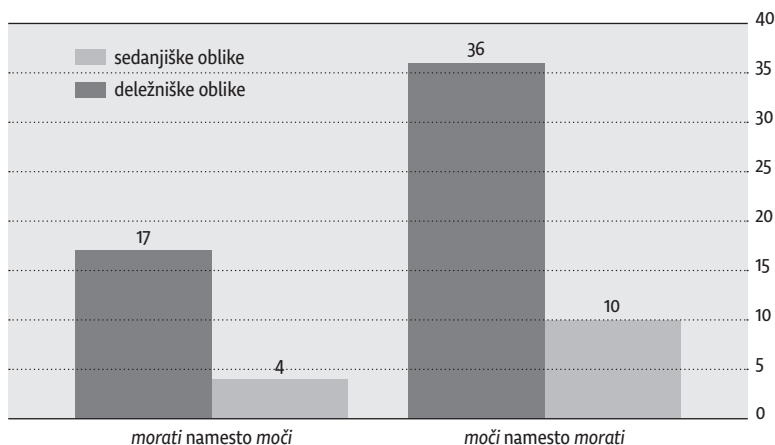
Z glagoloma *moči* in *morati* je v korpusu Šolar povezanih 87 primerov napak. Od tega gre pri večini za rabo glagola *moči* namesto *morati* (46 primerov), več kot pol manj je menjave *morati* namesto *moči* (21 primerov). V nadaljnjih 17 primerih gre za menjavo *moči* ali *morati* s katerim od drugih modalnih glagolov; ti primeri so za obravnavani jezikovni problem nerelevantni. Najdemo pa tudi 3 primere rabe besede *upati* namesto *moči*. Takšna raba se pojavlja le v goriški regiji.

Slika 6: Število različnih vrst jezikovnih napak, povezanih z glagoloma *moči* in *morati*



Kot kaže Slika 7, se največ napak pojavlja pri rabi deležnika, manj pa pri sedanjiških oblikah. Ta podatek lahko po vsej verjetnosti pripišemo dejstvu, da se sedanjik v šolskih spisih in podobnih nalogah v splošnem pojavlja nekoliko redkeje.

Slika 7: Primerjava števila napak pri sedanjiških in deležniških glagolskih oblikah glagolov *moči* in *morati*



Identificirana razmerja so vodilo za strukturiranje razlage ter izbiro zglede in vaj za utrjevanje znanja. V primeru, da zberemo dvesto avtentičnih povedi, v katerih bodo morali učenci popraviti napačno uporabljeni glagol, mora biti približno dve tretjini primerov takih, kjer je napačno uporabljen glagol *moči*, znotraj te množice pa približno petina s sedanjiško in preostale štiri petine z deležniško obliko. Na tak način posvetimo sorazmerno več časa tistim oblikam, ki so statistično najbolj problematične, s čimer povečamo možnost, da bo imelo utrjevanje znanja pozitiven prenos v prakso pisanja.

8.4.2 Razlaga problema

Na osnovi korpusnih ugotovitev je bila razlaga jezikovnega problema zasnovana v treh korakih.

8.4.2.1 Prvi korak

V prvem koraku razlage želimo pri učencih uzavestiti, da sta glagola *moči* in *morati* dva različna glagola, ki sta na pogled sicer podobna, vendar imata različen pomen. Da gre za dva glagola, je zapisano eksplicitno, nakazano pa je tudi s postavitvijo grafičnih elementov in izbiro barv (glagol *moči* je v celotnem poglavju povezan z modro barvo, glagol *morati* pa z oranžno).

Slika 8: Razlaga problema *moči* – *morati*, prvi korak

V tem poglavju bomo govorili o dveh **glagolih**. Prvi je glagol **morati** in drugi glagol **moči**.

Ta dva glagola sta si na pogled **podobna**. Vendar imata zelo **različen pomen**.

MORAM

pomeni **treba je**

- Za komentiranje **morate** biti prijavljeni.
- Solato in pijačo pa **sem moral** plačati sam.

NE MOREM

pomeni **ne gre**

- To je fenomenalno. **Ne morem** verjeti.
- **Ne bom mogla** prav dolgo blefirati.



Ker je nedoločnik *moči* po obliki precej oddaljen od osebnih glagolskih oblik, se v razlagi namesto nedoločnikov uporablja prva oseba ednine (*morati* → *moram*). Obenem se glagol *moči* večinoma uporablja v zanikani obliki, saj je takšna raba v slovenščini zanj izrazito tipična (*moči* → *ne morem*). Ta konkretizacija naj bi učencem pomagala lažje razumeti razliko v pomenu obeh glagolov. Temu so namenjeni tudi zgledi realne jezikovne rabe, izbrani iz korpusa Gigafida. Zaporedje izbranih zgledov poteka od bolj konkretnega k manj konkretnemu: prvi je v sedanjiku, drugi pa prinaša deležniško obliko.

Prvi korak razlage se zaključi z nalogo, pri kateri morajo učenci z miško potegniti nekaj povedi na ustrezno mesto – bodisi h glagolu *moči* ali glagolu *morati*. Ta naloga je na prvi pogled zelo preprosta, skoraj banalna, vendar je v tem koraku pomembna zato, da učenci avtomatizirajo povezavo med posameznimi glagolskimi oblikami: v tem oziru je problematičen predvsem glagol *moči*, pri katerem se nedoločniška, sedanjiška in deležniška oblika v korenu razlikujejo (*moči* – *morem* – *sem mogel*).

8.4.2.2 Drugi korak





V drugem koraku razlage želimo učence opozoriti, da se glagola *moči* in *morati* pogosto zamenjujeta, kar velja zlasti za govorjeni jezik. Predstavljeno je, do katerih napak v povezavi s tem problemom tipično prihaja, pri formulaciji česar so upoštevani rezultati analize, tj. na prvem mestu opozorimo na rabo glagola *moči* namesto *morati*.

Poleg opozorila o tipičnih napakah skušamo podati tudi razlog za zamenjevanje glagolov (*Zamenjave se dogajajo predvsem, ko govorimo*).

Verjetno tudi zato, ker *A* in *E* v govoru včasih podobno zvenita.). To pojasnilo je pomembno zato, da lahko učenci uzavestijo vpliv govornega jezika na pisanje. Slednjemu je posvečena tudi naloga (gl. Sliko 9), pri kateri učenci poslušajo posnetke govornje slovenščine in ugotavljajo, ali je v primerih glagol *moči* uporabljen ustrezno ali ne.

Slika 9: Razlaga problema *moči* – *morati*, govorni jezik

Zamenjave so zelo pogoste predvsem, ko **govorimo**. Nekaj takšnih zamenjav lahko slišiš v spodnjih štirih primerih.

in pol morem it v petek ob sedmih zjutraj na šihit še enkrat	
ja v bistvu pa otroka more spraviti domov veš to so pač obveznosti	
ker uradno bi jz zdele že trenutno moğu bit noter	
kam povej ti meni kam si moęla it da si ga našla	

Ko pri pisanju izbiramo med **MORAM** in **MOREM**, se ni dobro zanašati, kaj se "bolje sliši".
Treba je biti **pozoren in premisliti**, kaj obe besedi pomenita.

Drugi korak razlage zaključuje nasvet, da se pri pisanju v standardni slovenščini pri izbiri med glagoloma *moči* in *morati* ni dobro zanašati le na jezikovni občutek o tem, kaj bolje zveni, ampak se je treba odločiti zavestno, upoštevajoč pomen obeh glagolov. S to smernico želimo doseči, da se učenec zave, da sta glagola *moči* in *morati* problematična tudi (ali zlasti) v govornem jeziku in da s sledenjem jezikovni intuiciji ne bo nujno izbral ustrezne možnosti.

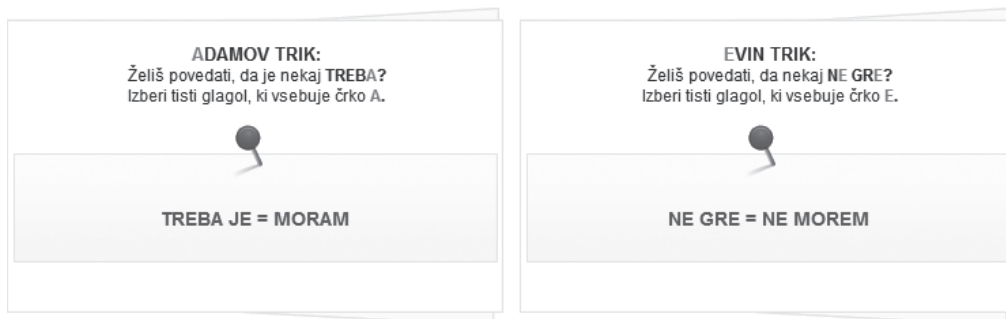
8.4.2.3 Tretji korak

V tretjem koraku je predstavljen „trik“ oz. mnemotehnika, s katero si lahko učenci pomagajo, da v dani situaciji uporabijo ustrezni glagol. Mnemotehnika se imenuje *Trik Adama in Eve*. Adamov trik temelji na črki A, ki jo najdemo tako v besedi *moram* kot tudi v pomensko primerljivi zvezi *treba je*. Evin trik temelji na črki E, ki se pojavlja tako v besedi *morem*, kot v *gre* oz. v nikalnici *ne*.

Adam in Eva pa se pojavljata tudi v naslovu slovničnega poglavja, ki se glasi: *Če Eva ne more po jabolka, mora pa Adam*. Tako deluje naslov poglavja kot „spominka“ – ob morebitnem ponovnem branju vsebine spomni učenca na mnemotehniko, ki jo poglavje prinaša, in na ta način dodatno izpostavlja oz. utrjuje nova znanja.

Slika 10: Razlaga problema moči – morati, mnemotehniki

Kako si lahko pomagamo, da ne zamenjamo **MORAM** in **MOREM**?
Uporabimo trik **Adama in Eve** iz naslova tega poglavja.



Pod vsakim od obeh trikov je na voljo tudi slovnična naloga, pri kateri mora učenec v stavek vpisati pravilno obliko glagola. Slovnični nalogi sta namenjeni avtomatizaciji povezovanja obeh glagolov z ustrežno „sopomenko“ (*treba je* vs. *ne gre*). Mnemotehnika se sicer navezuje le na prvo osebo ednine, vendar je predvideno, da bo s pomočjo zadostne količine premišljeno strukturiranih vaj učenec pravilo lahko splošil tudi na deležniške oblike.

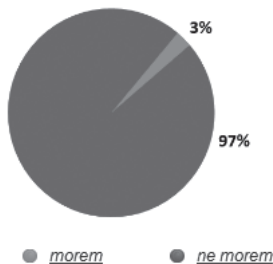
8.4.3 Za navdušene

Kot rečeno v razdelku 8.2.3, prinaša vsako slovnično poglavje tudi rubriko *Za navdušene*. Pri primeru, ki ga predstavljamo, je ta v celoti posvečena glagolu *moči*. Kot je že bilo omenjeno, se *moči* v razlagi uporablja primarno v zanikani obliki, kar sledi korpusnim podatkom o tipični rabi tega glagola. V rubriki *Za navdušene* pa imajo učenci možnost, da se seznanijo z rabo nezanikanega glagola *moči*. Ta oblika je sicer redkejša, se pa zelo pogosto pojavlja v stalnih besednih zvezah (npr. *če le morem*, *kaj pa morem*, *kadar morem* in podobno).

V drugem delu rubrike je razloženo, kako smo podatke o zanikanosti glagola *moči* pridobili. V tem razdelku so torej nakazane možnosti raziskovanja jezika, ki jih ponujajo besedilni korpusi.

Slika 11: Za navdušene, nezanimani glagol *moči*

Da je glagol *moči* običajno zanikan, smo izvedeli s pomočjo **korpusa Gigafida**. Spodnji graf prikazuje, kako pogosto se v korpusu beseda *morem* pojavi **skupaj z NE** in kako pogosto **brez**.



Oglej si graf.

Katera oblika je bolj pogosta, *morem* ali *ne morem*?

Če klikneš na povezavi pod grafom, si lahko ogledaš primere v **korpusu Gigafida**.

Med primeri z *morem* lahko najdeš tudi številne primere, kjer je glagol uporabljen napačno - pisec je uporabil glagol *moči*, kjer bi v resnici ustrezalo *morati*.

KAKO SO TI PODATKI LAHKO KORISTNI?

Če veš, da je glagol *moči* običajno zanikan, je ločevanje med *morati* in *moči* enostavnejše: na eni strani imamo **MORAM** in na drugi **NE MOREM**.

Učenci si lahko ogledajo grafično reprezentacijo jezikovnih podatkov, na osnovi katere jim je omogočeno razmišljanje o jeziku kot o sistemu, v katerem so določene pojavnosti bolj in druge manj pogoste oz. bolj ali manj tipične. Učenci imajo možnost izbrati tudi povezavo primerov realne jezikovne rabe v korpusu Gigafida, kar jih lahko motivira za nadaljnje raziskovanje obravnavanega problema. K raziskovanju jih poziva tudi besedilo ob grafu, ki opozarja, da se v korpusu pojavljajo tudi primeri rabe, ki odstopajo od jezikovne norme. Zainteresirani učenci bodo lahko pregledali poljubno število korpusnih primerov in s tem razvijali jezikovni občutek za rabo glagola *moči*.

8.4.4 Povzetek, Vaje in Goriška, pozor

Preostali razdelki obravnavanega slovnicega poglavja so Povzetek, Vaje in rubrika *Goriška, pozor*. Prva dva navedena razdelka prinašata že predstavljeno vsebino v strnjeni obliki, zato ju na tem mestu ne opisujemo natančneje. Rubrika *Goriška, pozor* pa prinaša ločeno opozorilo glede rabe glagola *upati* namesto *moči* v goriškem narečju. Tema je bila v poglavje vključena, čeprav tovrstne napake v korpusu Šolar niso posebej pogoste (pojavijo se le trije primeri tega tipa). Odločitev o vključitvi je bila sprejeta, ker smo želeli predstaviti tudi regionalno vezane jezikovne probleme: za učence, ki prihajajo s področja, za katerega je problem značilen, je opozorilo neposredno uporabno, za preostale pa prinaša dodatno zanimivost o rabi slovenščine.

8.5 Sklep

Predstavljene rešitve predstavljajo prvi poskus uporabe korpusnih podatkov pri pripravi učnih gradiv za poučevanje slovenščine. Korpusni podatki so bili uporabljeni za identifikacijo jezikovnih problemov, ki jih imajo pri pisanju učenci, vendar to še zdaleč ni edini doprinos. Ker celotna slovnična razlaga temelji na podatkih iz realne jezikovne rabe, je omogočeno zelo dobro povezovanje med teorijo in prakso. K temu denimo pomembno prispeva dejstvo, da so vsi uporabljeni zgledi avtentični, prav tako primeri, ki so uporabljeni za slovnične vaje. Obenem pa vključevanje posnetkov iz govornega korpusa omogoča primerjavo pisne in govorne slovenščine, kar je pri razlagi določenih jezikovnih problemov izredno dobrodošlo.

Izhajanje iz jezikovne rabe prinaša še eno pomembno prednost, namreč uvid v to, kako določeno jezikovno pravilo v sodobni slovenščini dejansko funkcionira. Obstoječi jezikovni priročniki rabo jezikovnih pravil pogosto prikazujejo kot premikanje po šahovnici, kjer so določene izbire bodisi absolutno pravilne ali absolutno nepravilne. Korpusni pristop pri pripravi gradiv pa omogoča, da učencem pokažemo nekoliko širšo sliko: da v različnih komunikacijskih situacijah uporabljamo različne jezikovne žanre in da ima neupoštevanje določenega pravila v posameznih kontekstih lahko različne komunikacijske posledice. Razlaga, ki se ne boji pokazati, kako se narečja razlikujejo od govorne standardne slovenščine ali kako se jezik SMS-ov in internetnih forumov razlikuje od pisne standardne slovenščine – obenem pa nestandardne rabe ne razglša *a priori* za neustrezno in nepravilno – lahko pri učencih veliko pridobi. S tem ko sprejmemo načine komuniciranja, ki so med mladimi prisotni, kot legitimne, novo snov povežemo z njihovo primarno komunikacijsko izkušnjo; obenem pa jim pokažemo, da je obvladanje standardne slovenščine (in torej pravil, ki v tem žanru veljajo) nujno za doseganje številnih pomembnih komunikacijskih ciljev, s čimer učenje pravil utemeljimo in osmislimo.

V zadnjem poglavju knjige smo predstavili, kako smo podatke iz korpusa Šolar uporabili v sklopu projekta *Sporazumevanje v slovenskem jeziku*. Seveda je pot, ki smo jo izbrali, le ena izmed možnosti, ki jo novi jezikovni viri (in nova jezikoslovna orodja za obdelavo velike količine jezikovni podatkov) omogočajo. Pričakovati je, da bodo avtorji učnih gradiv in učbenikov v prihodnosti poiskali nove možnosti uporabe korpusnih podatkov pri poučevanju slovenščine. Glede na to, da smo med analizo identificirali skoraj 700 različnih jezikovnih problemov, gradiva vsekakor ne bo prav kmalu zmanjkalo.

Povzetek

Monografija *Analiza jezikovnih težav učencev: korpusni pristop* opisuje analizo napak v korpusu Šolar (Rozman idr. 2012), skoraj milijon besed obsežni zbirki pisnih besedil osnovnošolcev in srednješolcev. Analiza napak je bila izvedena v okviru projekta *Sporazumevanje v slovenskem jeziku*, rezultati analize pa so bili uporabljeni pri pripravi gesel za pedagoško korpusno slovnico (Arhar Holdt idr. 2011).

Uvodno poglavje je namenjeno predstavitvi korpusov usvajanja maternega jezika, kamor spada tudi korpus Šolar, njihovih ključnih lastnosti in različnih analiz, opravljenih na tovrstnih korpusih. Kot ugotavljamo, so pisni korpusi usvajanja maternega jezika redki, podobno pa velja za analize napak, ki v korpusih usvajanja maternega jezika še niso bile izkoriščene v takšni meri kot v korpusih usvajanja tujega jezika, zlasti za namene izdelave jezikovnih učnih gradiv in virov.

V drugem poglavju podajamo opis korpusa Šolar, s poudarkom na lastnostih, pomembnih za analizo, npr. število in delež besedil po razredih/letnikih, regijah, šolah ipd., ločeno pa so podani tudi statistični podatki za tistih dobrih 56 % besedil v korpusu, ki vsebujejo označene napake učencev. Učiteljski popravki oz. oznake napake, ki so bile izhodišče za našo analizo, so klasificirani v štiri tipe, tj. napake zapisa, besedišča, oblike in skladnje, napake zapisa in napake skladnje pa so nadalje razdeljene še v pet oz. štiri podtipe. Vseh označenih napak v korpusu je 35.035, od tega je 61,1 % napak zapisa, 10,9 % napak besedišča, 10,3 % napak oblike in 17,7 % napak skladnje. Opisano je tudi orodje WordSmith Tools, ki smo ga uporabili pri analizi, ter nekateri postopki priprave in urejanja podatkov. V zaključku drugega poglavja sta razdelka, namenjena diskusiji o rabi terminov “napaka” in “jezikovna težava” v analizi ter predstavitvi različnih načinov podajanja zgledov za jezikovne težave, identificirane pri analizi.

Vsa nadaljnja poglavja, z izjemo zadnjega, predstavljajo rezultate analize po tipih napak. Tretje poglavje se loteva napak zapisa, ki so v korpusu Šolar najpogostejše in med katerimi je daleč največ napak ločil (skoraj 72 %), pri čemer jih je večina vezana na rabo vejice. Še vedno pogoste so tudi napake črkovanja, napake v rabi male oz. velike začetnice in napake v pisanju skupaj oz. narazen. Med napakami črkovanja se jih največ nanaša na neustrezen zapis črke ali glasu v besedi, učencem pa težave povzroča tudi zapis sklopov *lj* in *nj*. Največ napak v rabi male oz. velike začetnice je vezanih na zapis lastnih osebnih, zemljepisnih in stvarnih imen ter na napačno rabo velike oz. male začetnice za ločili, precej problematičen pa je tudi zapis pridevnikov na -ski oz. -ški. Od napak v pisanju skupaj oz. narazen velja izpostaviti predvsem zapis zanikanih glagolov, pri čemer je najpogostejše zaslediti

napačno zapisane oblike glagola *biti*. Na drugi strani so napake v zapisu števil in napake v zapisu krajšav zelo redke.

Napak besedišča, ki so opisane v četrtem poglavju, je v korpusu Šolar slabih 11 % vseh napak. Precej napak besedišča je semantične narave (npr. napačna polnopomenska beseda, napačna pogovorna ali nezaznamovana beseda), pogoste pa so tudi napake v rabi zaimkov, predlogov, veznikov in glagolov. Pri rabi zaimkov imajo učenci največ težav s svojilnimi oz. povratnosvojilnim zaimkom, zaimkoma *ki* in *kateri* ter zaimkoma *nihče* in *noben*. Največ napak predlogov se nanaša na neustrezno rabo predlogov *v* oz. *na* in *iz* oz. *z/s*, pri veznikih pa je veliko napak v rabi *in* in *ter* ter protivnih veznikov (*vendar*, *ampak*, *toda* itd.). Med napakami glagolov je dobra tretjina takšnih, ki so povezane z zamenjavo predpinskih obrazil (npr. *iz-* in *-z*), skoraj tretjina napak pa je povezana z naklonskimi izrazi, večinoma z zamenjavami *morati* oz. *moči*. Napak v rabi pridevnikov, prislovov, členkov in samostalnikov je med napakami besedišča v korpusu Šolar precej manj.

Podobno pogoste kot napake besedišča so v korpusu napake oblike, ki jih predstavlja peto poglavje. Kot je pokazala analiza, imajo učenci največ težav z rabo nedoločnika oz. namenilnika, zelo pogoste pa so tudi napake v ujemanju, zlasti pri ujemanju povedka z osebkom (največkrat ujemanje v številu) ter pri rabi samostalnika. Veliko napak se nanaša tudi na rabo glagolov v dvojini, predvsem rabo preteklega deležnika, ter rabo samostalniških zaimkov, predvsem napačne oblike roditeljskega in menjavanje mestnika in dajalnika. Od bolj specifičnih težav učencev velja izpostaviti težave pri sklanjanju samostalnikov *mati* in *otrok*, dokaj problematična samostalnika sta tudi *hči* in *gospa*.

Med napakami skladnje, ki so predstavljene v šestem poglavju, je največ odvečnih in izpuščenih podatkov, sledijo jim napake besednega reda, najmanj pa je napak strukture. Analiza je pokazala, da gre v mnogih primerih za slogovne ali vsebinske popravke oz. spremembe poudarkov, medtem ko je dejanskih napak učencev tu malo. Najpogostejše, ki terjajo največ pedagoške in jezikoslovne pozornosti, so nekatere napake besednega reda, npr. v povezavi z naslonskim nizom ali za veznikoma *pa* in *vendar*. Težavna sta tudi položaj členkov *tudi*, *še*, *že* ter zaporedje pridevnikov in zaimkov v samostalniški besedni zvezi. V večstavčnih povedih učenci izpuščajo ali uporabljajo odvečne pomožne glagole, deležnike, nikalne členke, zaimenske predmete ali prosta morfema *se* oz. *si*. Enako velja za referenco pred odvisnikom – včasih učenci uporabijo odvečni zaimek ali prislov, včasih pa ga izpustijo. Med napakami strukture izstopa izražanje svojilnosti, pri kateri učenci namesto svojilnega pridevnika uporabijo zvezo s samostalnikom v drugem sklonu. Pri mnogih težavah, ki so v korpusu Šolar označene kot napake skladnje, pa gre zaradi narave označevalne sheme v resnici za probleme oblikoslovja, npr. pri dvojnem stopnjevanju pridevnikov ali prihodnjiku glagola *biti* z rabo deležnika (*bom bil*).

Sedmo poglavje predstavlja analizo povezanih napak, a se omeji na povezane napake oblike, ki so se izkazale za nadvse relevantne, medtem ko so bile ostale povezane napake analizirane v okviru analiz napak zapisa, besedišča in skladnje. Večina odkritih jezikovnih težav je bila odkritih že pri analizi napak oblike ali pa gre za slogovne popravke, novoodkritih jezikovnih problemov je malo. Med njimi najdemo sklanjanje samostalnika *starši*, ločevanje med prislovnimi določili, ki izražajo smer in statično lokacijo, rabo skupnih imen (npr. *družina* in *sloj*), ko imamo v mislih le posamezne člane skupine, ter zamenjevanje dajalniških in orodniških oblik samostalnikov ter pridevnikov.

Zadnje, osmo poglavje smiselno zaokroža monografijo, saj predstavlja pedagoški vir, pedagoško korpusno slovnico, ki bo rezultate analize prenesel v poučevalno prakso. Na začetku je opisan namen oz. cilj pedagoške korpusne slovnice, izpostavljenih je nekaj izbranih problemov poučevanja slovenščine, ki smo se jim pri snovanju vira skušali izogniti, opisana je tudi struktura slovničnega poglavja. Sledi predstavitev pilotnih slovničnih poglavij ter njihovih vsebin, ki bodo pripravljena v okviru projekta *Sporazumevanje v slovenskem jeziku*. Na koncu so na vzorčnem slovničnem poglavju za zamenjevanje glagolov *morati* in *moči* prikazani vsi koraki izdelave poglavja, od uporabe rezultatov analize do pisanja posameznih delov poglavja.

Summary

The monograph *The analysis of student language problems: a corpus-based approach* describes the analysis of errors in the Šolar corpus (Rozman et al. 2012), a nearly one-million-word collection of written texts, produced by elementary and secondary school students. The analysis was conducted within the *Communication in Slovene* project, and its results were used for the purposes of pedagogic corpus-based grammar (Arhar Holdt et al. 2011).

Chapter 1 makes an overview of first language learner corpora (i.e. corpora of texts produced by young native speakers), their key features and various analyses, conducted on them. The overall conclusion is that there are not many such corpora in existence. In addition, there are even fewer error analyses and therefore their potential is yet to be fully exploited, especially in terms of the development of language materials and resources. There are lessons to be learned from (foreign language) learner corpora where such analyses and applications are much more established.

Chapter 2 describes the Šolar corpus, focussing on the features, relevant for the analysis, e.g. the numbers and percentages of texts by year/grade, region, type of school etc. In addition, there is a separate section that presents the statistical information on the 56 % of texts, containing annotated student errors that were the point of departure for the analysis. There are 35,035 annotated errors in total in the corpus, made by the teachers of students: 61.1% of them are orthography errors, 10.9% vocabulary errors, 10.3% morphology errors, and 17.7% syntax errors. Orthography errors and syntax errors are formed of four and five sub-categories, respectively. The chapter also describes WordSmith Tools, software used in the analysis, and certain procedures of data preparation and manipulation. The chapter concludes with a discussion on the use of terms “error” and “language problem” in the monograph, and a description of different ways in which corpus examples of language problems, identified during the analysis, are presented.

The following chapters, with the exception of the last one, present the results of the analysis for each error category. Chapter 3 presents orthography errors, the most common type of error in the corpus. By far the commonest type of error in this category (nearly 72 %) are punctuation errors, with most of them associated with the use of a comma. Also frequent are spelling errors, errors in the use of lower or capital case, and errors in writing words together or apart. Most spelling errors are the result of incorrect use of a letter or sound in the word, while errors in the use of lower or capital case are mostly found in personal names and names of places or things. The errors in the use of numbers and abbreviations are rare.

Vocabulary errors, described in Chapter 4, represent nearly 11 % of all the errors in the corpus. A great deal of these errors are semantic (e.g. the use of wrong lexical word, the use of wrong informal or unmarked word), also frequent are errors in the use of pronouns, prepositions, conjunctions, and verbs. On the other hand, the errors in the use of adjectives, adverbs, particles, and nouns are much rarer.

Similarly common as vocabulary errors are morphology errors, described in Chapter 5. As the results show, the students have problems with the use of infinitive and supine, also common are errors in agreement, particularly subject-verb agreement (in most cases in number). In addition, many errors found are associated with the use of verbs in dual, especially part participle, and the use of nominal pronouns, with the incorrect use of genitive, and the mixing of instrumental and dative being particularly common.

Syntax errors are presented in Chapter 6 and most of them are associated with redundant or missing text, followed by word order errors, and erroneous structure errors. The analysis has revealed that in many cases, teacher corrections are stylistic or content-related, whereas there are not many real student errors. Real errors are associated with word order, e.g. when using clitics, certain conjunctions and particles, and several adjectives and pronouns in a noun phrase. In complex sentences students often omit or unnecessarily use auxiliary verbs, participles, negative particles etc. Similarly, students sometimes unnecessarily use, or forget to use, pronoun or adverb before a subordinate clause.

Chapter 7 focusses on the analysis of related errors (errors that become an error only after a correction of another error is made), but remains limited to the related morphology errors, which proved to be extremely relevant. The other related errors were analysed together with orthography errors, vocabulary errors, and syntax errors, respectively. Most of the language problems, discovered in the analysis of related morphology errors, have already been detected during the analysis of morphology errors; there are in fact very few newly discovered language problems to report on (e.g. differentiating between adverbials of direction and adverbials of location, incorrect use of certain collective nouns, and differentiating between dative and instrumental forms of nouns and adjectives).

Chapter 8 concludes the monograph by presenting a pedagogic resource (pedagogic corpus-based grammar) that will transfer the results of the analysis to language pedagogy. First, the chapter describes the purpose and aim of the pedagogic grammar, addresses a few problems of teaching Slovene that the authors of the grammar tried to avoid, and presents the structure of a grammar section (part of the grammar dealing with a particular language problem). Next, sample sections and their contents that will be prepared as part of

the *Communication in Slovene project* are presented. The chapter concludes by using the section on modal verbs *morati* ('must') and *moči* ('can') to demonstrate all the steps in the preparation of a section, from utilizing the results of the analysis to writing individual parts of the section.

Literatura

- ANDERSEN, Gisle, 1997: *Pragmatic markers in teenage and adult conversation*. 18. konferenca ICAME. Neobjavljeni prispevek.
- ARHAR, Špela, GORJANC, Vojko, 2007: Korpus FidaPLUS: Nova generacija slovenskega referenčnega korpusa. *Jezik in slovnost* 52/2, str. 95–110.
- ARHAR HOLDT, Špela, KOSEM, Iztok, KRAPŠ VODOPIVEC, Irena, LEDINEK, Nina, MOŽE, Sara, ŠTRITAR, Mojca, SVENŠEK, Tanja, ZWITTER VITEZ, Ana, 2011: *Pedagoška slovnica pri projektu Sporazumevanje v slovenskem jeziku. Kazalnik 16 – Standard za korpusno analizo slovničnih pojavov*. Ministrstvo za šolstvo in šport Republike Slovenije.
- BALAŽIČ BULC, Tatjana, 2004: Jezikovni prenos pri učenju sorodnih jezikov (na primeru slovenščine in srbohrvaščine). *Jezik in slovnost* 49/3-4, str. 77–89.
- BOYD, Adriane, 2008: EAGLE: An Error-Annotated Corpus of Beginning Learner German. *LREC 2008 Proceedings*, str. 1897–1902.
- BRIŠKI, Anja, 2010: *Jezikovna zmožnost v slovenščini kot drugem jeziku v osnovni šoli*. Diplomsko delo. Ljubljana: Oddelek za slavistiko in Oddelek za slovenistiko Filozofske fakultete Univerze v Ljubljani.
- BUŠTA, Jana, HLAVÁČKOVÁ, Dana, JAKUBÍČEK, Miloš, PALA, Karel, 2009: Classification of Errors in Text. V Petr SOJKA, Aleš HORÁK (ur.). *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2009*. Brno: Masaryk University, str. 109–119. <http://nlp.fi.muni.cz/raslan/2009/papers/6.pdf> (december 2012).
- DOBROVOLJČ, Kaja, KREK, Simon, RUPNIK, Jan, 2012: Skladenjski razčlenjevalnik za slovenščino. V T. ERJAVEC, J. ŽGANEC GROS (ur.) *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Institut Jožef Stefan, str. 42–47.
- DULAY, Heidi, BURT, Marina, KRASHEN, Stephen, 1982: *Language Two*. New York, Oxford: Oxford University Press.
- EDGE, Julian, 1989: *Mistakes and Correction*. London, New York: Longman.
- ELLIS, Rod, 1994: *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- ERIKSSON, Anders, 1991: Aspects of Swedish speech rhythm. PhD thesis, University of Göteborg, Sweden.
- ERJAVEC, Tomaž, 2004: MULTEXT–East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. *Proceedings of the Fourth International Conference on Language Resources and Evaluation, LREC'04*. Lizbona: ELRA, str. 1535–1538.
- ERJAVEC, Tomaž, GORJANC, Vojko, HOLOZAN, Peter, STABEJ, Marko, 2001: Application to Slovene. V T. ERJAVEC (ur.) *Specifications and Notation for MULTEXT–East Lexicon Encoding*. MULTEXT–East / Concede.
- GRANGER, Sylviane, 2001: *International Corpus of Learner English: The ICLE Project*. <http://jupiter.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/download/icle.pdf>.
- GRANGER, Sylviane, DAGNEAUX, Estelle, MEUNIER, Fanny, PAQUOT, Magali, 2009: *International Corpus of Learner English, version 2*. Louvain-la-Neuve: Presses universitaires de Louvain.
- GRANGER, Sylviane, 2004: Computer Learner Corpus Research : Current Status and Future Prospects. V U. CONNOR in T. UPTON (ur.) *Applied Corpus Linguistics: A Multidimensional Perspective (Language and Computers)*. Amsterdam & Atlanta: Rodopi, str. 123–145.
- HANA, Jirka, ŠKODOVÁ, Svatava, ROSEN, Alexandr, ŠTINDLOVÁ, Barbora, 2010: Error-tagged Learner Corpus of Czech. *Proceedings of the Fourth Linguistic Annotation Workshop*. Uppsala: Association for Computational Linguistics, str. 11–19.
- HEWINGS, Martin, HEWINGS, Ann, 2002: "It is interesting to note that...": A comparative study of anticipatory "it" in student and published writing. *English for Specific Purposes* 21, str. 367–383.
- HOEY, Michael, 2005: *Lexical Priming: A new theory of words and language*. London: Routledge.
- HOLOZAN, Peter, 2012: Kako dobro programi popravljajo vejice v slovenščini. V T. ERJAVEC, J. ŽGANEC GROS (ur.) *Zbornik Osme konference Jezikovne tehnologije*. Ljubljana: Inštitut Jožef Stefan, str. 101–106.
- HORVATH, Jozsef, 1999: *Advanced Writing in English as a Foreign Language, A Corpus-based Study of Processes and Products*. Doktorska disertacija. http://www.geocities.com/writing_site/thesis/ (december 2012).
- HUNSTON, Susan, FRANCIS, Gill, 2000: *Pattern Grammar: A corpus driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.
- IZUMI, Emi, UCHIMOTO, Kiyotaka, ISAHARA, Hitoshi, 2004: SST speech corpus of Japanese Learners' English and automatic detection of learners' errors. *ICAME Journal* 28, str. 31–48.

- JAMES, Carl, 1998: *Errors in Language Learning and Use: Exploring Error Analysis*. London, New York: Longman.
- KALIN GOLOB, Monika, 2005: *Jezikovna kultura in jezikovni kotic̃ki*. Ljubljana: Jutro.
- KILGARRIFF, Adam, KOSEM, Iztok, 2012: Corpus tools for lexicographers. V S. GRANGER, M. PAQUOT (ur.) *Electronic Lexicography*. Oxford: Oxford University Press, str. 31–56.
- KILGARRIFF, Adam, RYCHLY, Pavel, SMRZ, Pavel, TUGWELL, David, 2004: The Sketch Engine. V G. C. WILLIAMS, S. VESSIER (ur.) *Proceedings of the 11th Euralex International Congress*. Lorient, Francija: Universite de Bretagne–Sud, str. 105–116.
- KOTSINAS, Ulla–Britt, 1994: *Ungdomsprik*. Hallgren and Fallgren, Uppsala.
- KRANJC, Simona, 2004: Besedni red, usvajanje prvega in učenje drugega/tujega jezika. *Jezik in slovstvo* 49/3–4, str. 145–157.
- KREK, Simon, ARHAR HOLDT, Špela, 2010: Slovenski Besedilni Korpusi: Kako v Razred? *Sodobna Pedagogika* 61/1, str. 224–241.
- KURE, Ana, 2010: *Ocena napredka jezikovne zmožnosti učencev priseljencev po 30–urnem jezikovnem tečaju (sociološki in jezikoslovni vidiki)*. Diplomsko delo. Ljubljana: Oddelek za slovenistiko in Oddelek za sociologijo Filozofske fakultete Univerze v Ljubljani.
- LEECH, Geoffrey, 1992: Corpora and theories of linguistic performance. V J. SVARTVIK (ur.) *Directions in corpus linguistics*. Berlin: Mouton de Gruyter, str. 105–122.
- LOGAR BERGINČ, Nataša, GRČAR, Miha, BRAKUS, Marko, ERJAVEC, Tomaž, ARHAR HOLDT, Špela, KREK, Simon, 2012: *Korpusi slovenskega jezika Gigafida, KRES, ccGigafida in cckRES: gradnja, vsebina, uporaba*. Ljubljana: Trojina, zavod za uporabno slovenistiko, in Fakulteta za družbene vede.
- LÜDELING, Anke, WALTER, Maik, KROYMANN, Emil, ADOLPHS, Peter, 2005: Multi–level error annotation in learner corpora. *Proceedings of Corpus Linguistics 2005*. Birmingham.
- MACMILLAN ENGLISH DICTIONARY FOR ADVANCED LEARNERS, 2. izdaja. 2007.
- MARJANOVIČ UMEK, Ljubica, KRANJC, Simona, FEKONJA PEKLAJ, Urška, GRGIČ, Katja, 2007: Govor dekliv in dečkov. *Razprave – Razred za filološke in literarne vede* 20, str. 115–131.
- MARJANOVIČ UMEK, Ljubica, FEKONJA PEKLAJ, Urška, KRANJC, Simona, GRGIČ, Katja, 2008: The effect of children's gender and parental education on toddler language development. *European Early Childhood Education Research Journal* 16/3, str. 325–342.
- NICHOLLS, Diane, 2003: The Cambridge Learner Corpus – error coding and analysis for lexicography and ELT. V D. ARCHER, P. RAYSON, A. WILSON, T. MCENERY (ur.) *Proceedings of the Corpus Linguistics 2003 conference*. Lancaster: UCREL, str. 572–581.
- PALA, Karel, RYCHLY, Pavel, SMRZ, Pavel, 2003: Text Corpus with Errors. V V. MATOUŠEK, P. MAUTNER (ur.) *Text, Speech and Dialogue: 6th International Conference*. Berlin, Heidelberg: Springer, str. 90–97.
- POMIKÁLEK, Jan, RYCHLY, Pavel, KILGARRIFF, Adam, 2009: Scalling to billion–word plus corpora. *Advances in Computational Linguistics* 41, str. 3–13.
- PIRIH SVETINA, Nataša, 2005: *Slovenščina kot tuji jezik*. Ljubljana: Izolit.
- POŽGAJ HADŽI, Vesna, FERBEŽAR, Ina, 2001: Tudi to je slovenščina. *37. seminar slovenskega jezika, literature in kulture*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete, str. 57–68.
- RÖMER, Ute, 2009a: English in academia: Does nativeness matter? *Anglistik: International Journal of English Studies* 20/2, str. 89–100.
- RÖMER, Ute, 2009b: The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7/1, str. 141–163.
- ROWLAND, Caroline F., 2007: Explaining errors in English children's questions. *Cognition* 104/1, str. 106–134.
- ROWLAND, Caroline F., PINE, Julian M., LIEVEN Elena V., THEAKSTON, Anna L., 2005: The incidence of error in young children's Wh–questions. *Journal of speech, language and hearing research* 48/2, str. 384–404.
- ROZMAN, Katarina, 2006: *Napaka kot izhodišče in motivacija za širjenje jezikovne zmožnosti pri učenju slovenščine kot tujega jezika*. Diplomsko delo. Ljubljana: Oddelek za slovenistiko Filozofske fakultete Univerze v Ljubljani.
- ROZMAN, Tadeja, 2010: Vloga enojezičnega razlagalnega slovarja slovenščine pri razvoju jezikovne zmožnosti (doktorska disertacija). Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- ROZMAN, Tadeja, KRAPŠ VODOPIVEC, Irena, STRITAR, Mojca, KOSEM, Iztok, 2012: *Empirični pogled na pouk slovenskega jezika*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- ROZMAN, Tadeja, STRITAR, Mojca, 2010: Korpus usvajanja slovenščine. *Nova didaktika poučevanja slovenskega jezika: Sporazumevanje v slovenskem jeziku*. Ljubljana: str. 36–41. http://www.slovenscina.eu/Media/Kazalniki/Kazalniki5/Nova_didaktika_Sporazumevanje.pdf (december 2012).

- ROZMAN, Tadeja, STRITAR, Mojca, KRAPŠ VODOPIVEC, Irena, KOSEM, Iztok, KREK, Simon, 2010: *Nova didaktika poučevanja slovenskega jezika. Kazalnik 15 projekta Sporazumevanje v slovenskem jeziku*. Ministrstvo za šolstvo in šport Republike Slovenije. http://www.slovenscina.eu/Media/Kazalniki/Kazalnik15/Nova_didaktika_Sporazumevanje.pdf (december 2012).
- SCOTT, Mike, 2008: WordSmith Tools, version 5. Liverpool: Lexical Analysis Software.
- SINCLAIR, John McHardy, 1987: *Looking Up: An account of the COBUILD Project in lexical computing*. London and Glasgow: Collins ELT.
- SINCLAIR, John, 1991: *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- SINCLAIR, John, 2004: *Trust the Text: Language, Corpus and Discourse*. London: Routledge.
- SINCLAIR, John, 2005: Corpus and Text – Basic Principles. V M. WYNNE (ur.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, str. 1–16.
- STABEJ, Marko, 2011: Jezikovni potrošnik in potrošnica. *Sodobna pedagogika* 62/2, str. 102–113.
- STENSTRÖM, Anna–Brita, ANDERSEN, Gisle, HASUND, Ingrid Kristine, 2002: *Trends in Teenage Talk: Corpus compilation, analysis and findings*. Studies in corpus Linguistics 8. Amsterdam: John Benjamins.
- STRITAR, Mojca, 2012: *Korpusi usvajanja tujega jezika*. Ljubljana: Zveza društev Slavistično društvo Slovenije.
- TOPORIŠIČ, Jože, 1991: *Slovenska slovnica*. Maribor: Založba Obzorja.
- TOPORIŠIČ, Jože, 1992: *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.
- TOPORIŠIČ, Jože, idr., 2001: *Slovenski pravopis. Spletna izdaja*. Ljubljana: Inštitut za slovenski jezik Frana Ramoviša ZRC SAZU. Dostopno na naslovu <http://bos.zrc-sazu.si/sp2001.html>. Prevzeto 17. 12. 2011.
- TSUI, Amy Bik May, 2004: What teachers have always wanted to know – and how corpora can help. V J. SINCLAIR (ur.) *How to use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, str. 39–61.
- VERDONIK, Darinka, ZWITTER VITEZ, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Spletni viri

- baza CHILDES (Child Language Data Exchange System): <http://childes.psy.cmu.edu/>
- Cambridge Learner Corpus: <http://www.cambridge.org/gb/elt/catalogue/subject/custom/item3646603/Cambridge-International-Corpus-Cambridge-Learner-Corpus>
- korpus BASE (British Academic Spoken English): <http://www2.warwick.ac.uk/fac/soc/al/research/collect/base/>
- korpus BAWE (British Academic Written English): <http://www2.warwick.ac.uk/fac/soc/al/research/collect/bawe/>
- korpus COLT (The Bergen Corpus of London Teenage Language): <http://www.hd.uib.no/colt/>
- korpus Gigafida: <http://www.gigafida.net/>
- korpus Gos: <http://www.korpus-gos.net/>
- korpus LOCNESS: <http://www.uclouvain.be/en-cecl-locness.html>
- korpus LUCY: <http://www.grsampson.net/LucyDoc.html>
- korpus MICASE (Michigan Corpus of Academic Spoken English): <http://quod.lib.umich.edu/m/micase/>
- korpus MICUSP (Michigan Corpus of Upper-level Student Papers): <http://micusp.elicorpora.info/>
- korpus POW (Polytechnic of Wales): <http://khnt.hit.uib.no/icame/manuals/pow.htm>
- korpus Šolar: <http://www.korpus-solar.net/>
- korpus The Bank of English: <http://www.mycobuild.com/about-collins-corpus.aspx>
- Longman Learners' Corpus: <http://www.pearsonlongman.com/dictionaries/corpus/learners.html>
- oblikoslovni leksikon Sloleks: <http://www.slovenscina.eu/sloleks/>
- portal Slogovni priročnik: <http://slogovni.slovenscina.eu/>
- Slovar slovenskega knjižnega jezika (SSKJ): <http://bos.zrc-sazu.si/sskj.html>
- spletna stran projekta SACODEYL: <http://www.um.es/sacodeyl>
- spletna stran projekta Sporazumevanje v slovenskem jeziku: <http://www.slovenščina.eu/>
- Spletni slovar slovenskega jezika (SSS): <http://www.slovenscina.eu/spletni-slovar/>
- Slovenski pravopis 2001: <http://bos.zrc-sazu.si/sp2001.html>

Imensko kazalo

A

Andersen 18, 24
Arhar 23, 28
Arhar Holdt 13, 108, 109, 123, 126

B

Balažič Bulc 59, 78
Boyd 57, 78
Briški 20
Burt 78
Bušta 19, 25, 57, 78

D

Dobrovoljc 115
Dulay 78

E

Edge 59
Ellis 78
Eriksson 25
Erjavec 61

F

Ferbežar 57, 78
Francis 16

G

Gorjanc 23, 28
Granger 19, 20, 22, 24, 57, 78

H

Hana 57, 78
Hasund 18, 25
Hewings A. 25
Hewings M. 25
Hoey 16
Holozan 25, 26
Horvath 24
Hunston 16

I

Izumi 57, 78
Isahara 57, 78

J

James 38, 42, 59, 78

K

Kalin Golob 14
Kilgarriff 16, 23, 28, 38
Kosem 16, 28
Kotsinas 25

Kranjc 81

Krashen 78
Krek 109, 115
Kure 20

L

Leech 16
Logar Berginc 28, 109
Lüdeling 57, 78

M

Marjanovič Umek 26
Može 45, 78, 103

N

Nicholls 57, 78

P

Pala 57, 78
Pomikálek 16
Piriš Svetina 17, 42
Požgaj Hadži 57, 78

R

Römer 19, 25
Rowland 25
Rozman K. 20, 57, 58, 78
Rozman T. 13, 17, 19, 20, 26, 28, 45,
48, 49, 51, 57, 66, 78, 79, 98, 103,
109, 110, 123, 126
Rychlý 16, 57, 78

S

Scott 38
Sinclair 16, 21
Smrž 57, 78
Stabej 26
Stenström 18, 25
Stritar 20, 28, 57, 78

T

Toporišič 47, 58, 81
Tsui 109

U

Uchimoto 57, 78

V

Verdonik 28, 109

Z

Zwitter Vitez 28, 68, 109