

RABA KRATKEGA NEDOLOČNIKA: KORPUSNI PRISTOP

Sara MOŽE

Amebis, d.o.o., Kamnik

Može, S. (2013): Raba kratkega nedoločnika: korpusni pristop. Slovenščina 2.0, 1 (1): 155-175.

URL: http://www.trojina.org/slovenscina2.0/arhiv/2013/1/Slo2.0_2013_1_08.pdf.

V prispevku so predstavljeni izsledki podrobne korpusne študije o rabi kratkega in dolgega nedoločnika. Uvodnemu delu sledita najprej krajši pregled literature na temo kratkega nedoločnika, nato pa še predstavitev problematičnih struktur z nedoločniki, ki so bile določene pri analizi korpusa pisnih izdelkov *Šolar*. V tretjem razdelku je opisana metoda izločanja najpogostejših polnopomenskih glagolov, ob katerih se pojavlja nedoločnik, iz polmilijonskega, oblikoskladenjsko in skladenjsko označenega učnega korpusa, ki je nastal znotraj projekta *Sporazumevanje v slovenskem jeziku*. Zadnji del prispevka je posvečen vplivu žanra in prenosnika na izbiro kratkega nedoločnika; za vsak izločen glagol je bila opravljena kvantitativna analiza deležev kratkih in dolgih nedoločnikov v korpusih *FidaPLUS* in *GOS*, ki prinaša nove izsledke o dejanski rabi kratkega nedoločnika v pisnih in govorjenih besedilih.

Ključne besede: kratki nedoločnik, korpus, slovenska skladnja, FidaPLUS, korpus GOS

1 UVOD

Govorci slovenščine se vsak dan tako pri pisni kot pri govorni komunikaciji pogosto srečujejo z dilemo, ali je v izbrani strukturi pravilno uporabiti kratki ali dolgi nedoločnik. Čeprav se z razlago, katera oblika je slovnično ustrezna, srečujejo že v šolskih klopeh, postane po kratki analizi vprašanj, ki jih odrasli govorci slovenščine postavljajo na spletnih forumih, kot je denimo *Medovernet*, kmalu jasno, da jim raba nedoločnika še zmerom povzroča večje preglavice. Cilj pričujoče raziskave je na podlagi korpusnih podatkov, ki odražajo realno jezikovno rabo, ugotoviti, kdaj prihaja do odklonov od

ustaljene norme in kateri dejavniki vplivajo nanje. Z analizo korpusnih podatkov se bodo preverjale naslednje hipoteze:

1. Kratki nedoločnik se najpogosteje rabi v govorjenem jeziku.
2. Najredkeje se kratki nedoločnik pojavlja v časopisih, revijah in knjigah, saj so ti viri navadno lektorirani, poleg tega pa gre za formalna besedila.
3. Dolgi nedoločnik prevladuje na Gorenjskem, Koroškem in južnem Štajerskem, kratki nedoločnik pa v ostalih slovenskih regijah.¹
4. Kratki nedoločnik se redkeje pojavlja v javnih, pogosteje pa v nejavnih govornih situacijah.

Po predstavitvi korpusnih podatkov v jedrnem delu prispevka sledi še sklep, v katerem bodo povzete glavne ugotovitve.

2 KRATEK PREGLED LITERATURE

Slovenski jezik pozna dve obliki nedoločnika – dolgega in kratkega. Kratki nedoločnik, ki ga Toporišič (1966) imenuje tudi *pogovorni nedoločnik*, je v literaturi opredeljen kot pogovorna oblika zbornega (dolgega) nedoločnika. Oblikovno je enak namenilniku, vendar pa se od njega včasih razlikuje v izgovoru, npr. *hočem spáti* (dolgi nedoločnik) / *spàt* (kratki nedoločnik) – *grem spát* (namenilnik) (Žagar 1993: 289).² Norma določa, da kratkega nedoločnika ne smemo rabiti v zbornem pisnem jeziku, in ga obravnava kot prvino neformalnega, govorjenega jezika. Z drugimi besedami: »pogovornega nedoločnika ne smemo zapisovati, govorimo ga pa zelo pogosto« (Žagar 1993: 289). Kratki nedoločnik je v normativnem priročniku prvič omenjen leta 1935, in sicer v Breznik-Ramovševem pravopisu, kjer je opredeljen kot pogovorna oblika; v prejšnjih pravopisih, npr. Breznikovem iz leta 1920, o tem še ni sledu. Vodušek (1959) Ramovševo odločitev, da se v normativnem priročniku

¹ Povzeto po Tominec (1955).

² Za dodatne informacije o glasoslovnih značilnostih namenilnika in kratkega nedoločnika prim. Toporišič (1958, 1966).

omeni tudi kratki nedoločnik, razlaga takole:

Da se je Ramovš zavzel za vpeljavo teh oblik – četudi pogojno – v uradni pravopis, pa je bila posledica njegovega novega, samostojnega študija slovenskega jezika, pri katerem je odkril takoimenovani kratki nedoločnik kot prastaro, še predtrubarsko varianto dolgega nedoločnika in ugotovil, da je razmeroma redka raba tega nedoločnika pri Trubarju znamenje, da je že Trubar privzdignil naravni jezik in mu s svojo načelno rabo dolgih nedoločnikov oblekel nekakšen prisilni jopič. Sicer se tudi ta Ramovšev reformatorični poskus naše pravopisne in pravorečne prakse v poznejših, skrčenih izdajah njegovega in Breznikovega pravopisa ni nadaljeval, je pa pustil prav tako vidne sledove kakor Župančičev, in Ramovševi avtoriteti je treba pripisati, da se vsaj načelna omemba pojava najde tudi v našem zadnjem, še danes veljajočem pravopisu iz leta 1950. (Vodušek 1959: 198.)

Druga oseba, ki jo v povezavi s kratkim nedoločnikom omenja Vodušek, je Župančič, ki je to obliko po koncu 1. svetovne vojne vpeljal v gledališče, ker je želel igralce naučiti pogovornega, naravnega jezika:³

Takoj v prvih letih po koncu prve svetovne vojske in po ustanovitvi Jugoslavije se je Župančič kot tedanji vodja Slovenskega narodnega gledališča zavzel v nasprotju s predpisano historično izreko za modernejšo in bolj resnično slovensko odrsko govorico. Učil je igralce na primer izgovarjati glagolski nedoločnik brez končnega i. (Vodušek 1959: 198.)

Pozneje je sicer to tendenco opustil in se vrnil k dolgemu nedoločniku, je pa s svojimi dejanji pustil globok vtis na gledališko sceno; kratki nedoločnik so kot prvo pogovornega jezika vpeljali tudi v mariborskem gledališču, kjer se je obdržal dalj časa kot v ljubljanskem. »Tako smo prišli do smešnega položaja, da je Ljubljana, ki je na območju kratkega nedoločnika, govorila dolgega, Maribor pa narobe« (Bajec 1955: 162). Župančičevo odločitev Bajec razlaga

³ Potreba po pogovornem jeziku se je pokazala po odpravi elkanja. Do takrat je za zborna slovenščina veljalo pravilo *Govori, kakor pišeš*, kar je za javni govorni položaj pomenilo obvezno rabo dolgega nedoločnika (Bajec 1955: 161–162).

takole:

Sam je navedel za vzrok, da ima slovenščina premalo samoglasnikov in da je zato vsakega škoda. Najbrž pa ga je k temu nagnilo spoznanje, da se je samo z nedoločnikom bore malo približal vsakdanjemu govoru in da govoriti ni nič manj papirno kakor govoriti, pogovorna oblika je namreč reducirani govorit. [...] Da sam kratki nedoločnik še ne daje govoru naravnega značaja, je gotovo (Bajec 1955: 162–163).

Zanimiva je tudi dialektološka opredelitev kratkega nedoločnika. Po Tomincu (1955) namreč kratki nedoločnik poznajo le nekatera slovenska narečja:⁴

Na Krasu, Notranjskem, Dolenjskem, na južnem Štajerskem poznajo v glavnem kratki nedoločnik brez i, na Gorenjskem, Koroškem, na vzhodnem Štajerskem pa dolgi nedoločnik z i. V knjižnem jeziku je danes v rabi dolgi nedoločnik. (Tominec 1955: 182.)

Če na kratko povzamemo: dilema, ali se sme kratki nedoločnik uporabljati v zbornem knjižnem jeziku, v normativnih priročnikih ne obstaja – uporabljati je dopuščeno le dolgi nedoločnik. Hkrati je jasno, da priročniki ne uravnavajo rabe kratkega nedoločnika v zasebnem diskurzu oziroma v pokrajinskem in narečnem jeziku. Edina sporna točka je v tem primeru raba oblike v javnem diskurzu oziroma predvsem v medijih. Jezikoslovci, ki pojav omenjajo v svojih prispevkih, nastopajo izrazito proti tovrstni rabi (prim. Bajec 1955). Vidali, ki je zagovornik rabe zborne izreke ne le na javnih prireditvah in v šolah, temveč tudi na radiu, o kratkem nedoločniku pravi naslednje: »Res je, da živi poleg dolgega tudi kratki nedoločnik v slovenskih narečjih, toda če smo se v pisavi dosledno odločili za dolgi nedoločnik, čemu bi ga potem ne izgovarjali?« (Vidali 1947: 732).

⁴ Tominčeva trditev je v skladu z Bajčevo razlago zgodovinskega razvoja dolgega in kratkega nedoločnika: »Praslovanščina je imela dva nedoločnika, daljšega na -ti, krajšega na -t in mehki polglasnik. Ko je le-ta odpal, sta se glasila nositi in nosit. Slovenska narečja so posplošila tega ali onega, gorenjščina n. pr. dolgega, dolenjščina kratkega« (Bajec 1955: 163).

Jezikoslovci so tako ostro nastopali proti rabi pogovornih prvin v skoraj vseh komunikacijskih situacijah (izjema je le zasebni diskurz), nekateri so v prispevkih zagovarjali svoj prav z ognjevitostjo, ki se današnjemu bralcu zdi nenavadna.⁵ Pri branju strokovne literature iz prve polovice 20. stoletja postane kmalu jasno, da je bil kratki nedoločnik v slovenskem prostoru občutljiva tema.

3 ANALIZA KORPUSNIH PODATKOV

3.1 Identifikacija problema: analiza korpusa Šolar

Korpus *Šolar* (Rozman in dr. 2010) je skoraj milijonski korpus šolskih pisnih izdelkov, ki je nastal v okviru projekta *Sporazumevanje v slovenskem jeziku*.⁶ Korpus vsebuje besedila, ki so jih slovenski osnovnošolci in srednješolci samostojno tvorili pri pouku, in učiteljske popravke, kar omogoča hitro in učinkovito identifikacijo in analizo jezikovnih napak, ki jih pri tvorjenju besedil delajo rojeni govorci slovenščine. Vsi popravki so bili označeni glede na tip učiteljevega komentarja (besedilni ali likovni komentar) in napake (napaka zapisa, besedišča, oblike ali skladnje) (prim. Rozman in dr. 2010). V okviru projekta je bil na podlagi analize posameznih kategorij učiteljskih popravkov nato oblikovan nabor specifičnih slovničnih problemov, s katerimi se soočajo učenci in dijaki, in ta je služil kot osnova za pisanje nove korpusne pedagoške slovnice (prim. Kosem in dr. v tisku).

Korpus *Šolar* ima pomembno vlogo tudi v pričujoči raziskavi, saj je analiza širšega nabora problemov, ki so vezani na rabo nedoločnika, omogočila ne le celostni vpogled v problematiko rabe obeh oblik, temveč tudi natančnejšo določitev predmeta proučevanja. V povezavi z rabo nedoločnika je bila namreč identificirana široka paleta jezikovnih problemov, ki jih je mogoče kategorizirati v naslednje skupine:

⁵ Prim. Urbančičev komentar na rabo besede *moči* (1962: 203–204).

⁶ <http://www.korpus-solar.net>

- I. **Oblikoslovne napake:** učenec uporabi slovnično nepravilno obliko, npr. *dosežti* namesto *doseči*.
- II. **Napake besednega reda:** problematika je pogosto vezana na naslonski niz, npr. pri kombinaciji prostega morfema *se* in nedoločnika (*Zato je veliko bolje se prilagoditi in ljubiti, kar imamo* → *Zato se je veliko bolje prilagoditi in ljubiti, kar imamo*).
- III. **Slogovne napake:** v spodnjih zgledih ne gre za slovnične napake, temveč zgolj za slogovno manj ustrezne besede in strukture, ki jih učitelj pogosto popravlja glede na lastne preference. Znotraj kategorije je mogoče zaslediti zglede, v katerih učitelji različno popravljajo enake ali podobne strukture. Nekaj primerov slogovnih popravkov:
- a. **menjava samostalnika z nedoločnikom**, npr. Dekle si je želelo smrt → Dekle si je želelo umreti;
 - b. **menjava strukture z nedoločnikom z osebno glagolsko obliko**, npr. Povedali so ji da z njim naj neha se dopisovat → Povedali so ji da naj se ne dopisuje;
 - c. **menjava osebne glagolske oblike s strukturo z nedoločnikom**, npr. Hodil je v vrtec, ko je bil star štiri leta → Hoditi je začel v vrtec, ko je bil star štiri leta;
 - d. **odvečen nedoločnik**, npr. Sin in Nevesta sta hotla it v tujino → Sin in nevesta sta hotla v tujino.
- IV. **Skladenjske napake:** učenec uporabi strukturo, ki krši normo slovenskega knjižnega jezika. Sem lahko uvrstimo:
- a. **izpuste nedoločnika**,⁷ npr. Tako kot je v orkestru potrebno prilagoditi številu violin število trobent, moramo ljudje živeti v sožitju in slogi ter pripravljeni pomagati drugim → Tako kot je v orkestru potrebno prilagoditi številu violin število trobent, moramo ljudje živeti v sožitju in slogi ter biti pripravljeni pomagati drugim;
 - b. **rabo strukture tipa za jesti**, npr. Luciji so se začela odpirati vrata za

⁷ Z gotovostjo ne moremo trditi, da je učenec dejansko napravil napako; nenavadno se namreč zdi, da bi rojeni govorec slovenščine zavestno uporabil tovrstno strukturo. Učenec je najverjetneje naredil lapsus.

- začeti na novo → Luciji so se začela odpirati vrata za nov začetek;
- c. **hiperkorekcije**, npr. Otroci so počasi začeli odhajati od doma in šli delati, iskati službo → Otroci so počasi začeli odhajati od doma in šli delat, iskat službo;
- d. **raba kratkega nedoločnika namesto dolgega**, npr. Jekyll se s tem ni hotel sprijaznit → Jekyll se s tem ni hotel sprijazniti.

V Šolarju so kratki nedoločniki označeni kot napake oblike. Pri analizi te kategorije je bilo vsaki pojavitvi napake pripisana še oznaka, ki se navezuje na specifičen jezikovni problem, kar omogoča učinkovito izločanje podkorpusev z relevantnimi zgledi. Na podlagi specifičnih oznak je bilo tako mogoče izločiti kar 149 zgledov rabe kratkega nedoločnika namesto dolgega, kar v primerjavi s pogostnostjo pojavljanja ostalih slovničnih napak v korpusu kaže na to, da je pisna raba nedoločnika pereč problem za šolajočo se mladino.

3.2 Izločanje polnopomenskih glagolov ob nedoločniku: analiza učnega korpusa

Analiza podatkov iz korpusa Šolar je pokazala, da imajo rojeni govorniki slovenščine težave pri rabi nedoločnika na skoraj vseh jezikovnih ravneh, tj. na oblikoslovni, skladenjski in besedilni, in da je izbira napačne oblike nedoločnika relativno pogosta napaka, zato jo je vredno podrobneje proučiti.

Preden lahko začnemo preverjati hipoteze, ki smo jih postavili na začetku, moramo določiti, ob katerih polnopomenskih glagolih se nedoločnik pojavlja. Toporišič (1984: 336–358) strukture z nedoločniki deli v šest skupin:

- naklonski izrazi, ki izražajo namero (*nameravati*, *imeti namen*, *biti namenjen*, *misлити*, *kaniti*, *skušati*, *poskušati* ipd.), hotenje (*hoteti*, *želeти*, *morati*, *marati*, *mikati*, *biti voljan* ipd.), nujnost oziroma potrebo (*morati*, *smeti*, *biti sila*, *biti prisiljen*, *biti dolžan*, *biti nujno*, *imeti dolžnost* ipd.), ukazovanje (*zapovedati*, *prepovedati*, *ukazati*, *veleti*, *naročiti*, *biti primoran*, *biti zapovedan* ipd.) in dopustnost (*imeti pravico*, *biti upravičen*, *imeti oblast*, *smeti* ipd.);
- izrazi duševnega stanja (*bati se*, *sramovati se*, *biti pripravljen* ipd.);

- izrazi ocene (*splačati se, biti vreden, biti koristen* ipd.);
- fazni glagoli (*začeti, nehati* ipd.);
- glagoli zaznavanja (*slišati, videti* ipd.);
- glagol *biti* ob vršilcu dejanja ali nosilcu stanja večinoma v dajalniku, pri čemer nedoločnik izraža moranje (npr. *iti mi je domov*), možnost (npr. *še zmerom je bilo slišati bobnenje*) ali negotovost (npr. *Kaj storiti?*).⁸

Pri določanju seznama polnopomenskih glagolov, ki jih bomo v raziskavi upoštevali, bi se lahko oprli na zgornji seznam, vendar se zdi zavoljo statistične obdelave podatkov, ki bo sledila v nadaljevanju, smiselno izluščiti tiste glagole, ki se najpogosteje pojavljajo v korpusih in tako v največji meri odražajo današnjo jezikovno realnost.⁹

Seznam najpogostejših polnopomenskih glagolov je bil izločen iz približno polmilijonskega učnega korpusa (Arhar 2009), ki ga sestavljajo vzorčeni odstavki iz korpusa FidaPLUS. Učni korpus je nastal v okviru projekta *Sporazumevanje v slovenskem jeziku* in je namenjen učenju statističnih modelov označevalnikov in razčlenjevalnikov (prim. Dobrovoljc in dr. 2012).¹⁰ Korpus sicer ni obsežen, vendar prinaša številne prednosti:

1. **zanesljivost:** vse oznake so bile v učnem korpusu večkrat ročno preverjene, zato so podatki bistveno bolj zanesljivi kot v korpusih večjega obsega, za katere se navadno uporablja orodja za samodejno jezikoslovno označevanje;
2. **raznovrstnost jezikoslovnih oznak:** učni korpus je v celoti označen z oblikoskladenjskimi oznakami, delno pa tudi skladenjsko razčlenjen (pribl. 200.000 pojavnic), kar pomeni, da lahko hkrati iščemo po

⁸ Čeprav bi bilo vredno proučiti tudi rabo oblik ob glagolu *biti* in povedkovem določilu oziroma vršilcu dejanja ali nosilcu stanja v dajalniku, je pričujočo raziskavo smiselno omejiti na polnopomenske glagole.

⁹ Pri korpusnem iskanju specifičnih besed in struktur, ki jih zasledimo v slovnica, npr. Toporišičevi (1984), namreč pogosto ni zadetkov.

¹⁰ Korpus in skladenjski razčlenjevalnik sta dostopna na povezavi http://razclenjevalnik.slovenscina.eu/Programska_oprema.aspx.

osnovni obliki ali vseh oblikah besede, oblikoskladenjskih oznakah in skladenjskih povezavah. Podatki, ki jih na tak način pridobimo, so veliko bolj zanesljivi, saj lahko hitro in zanesljivo poiščemo vse strukture, v katerih nastopa nedoločnik kot podredni del glagolske zveze, ne glede na razdaljo med elementoma ali tip vrinjenih elementov;¹¹

3. **orodje Označevalnik:**¹² v okviru projekta je nastalo orodje *Označevalnik*, ki omogoča hitro in preprosto pregledovanje in označevanje povedi, popravljanje napačnih oznak, napredno iskanje po korpusu in izvažanje podatkov, kar pomeni, da je z uporabo naprednih iskalnih pogojev mogoče učinkovito izločiti seznam relevantnih zgledov in jih izvoziti v zelenem formatu.

Iz korpusa je bilo izločenih 1334 relevantnih zgledov, od tega jih je bilo 10 (0,75 %) slovnično nepravilnih, kot prikazuje:

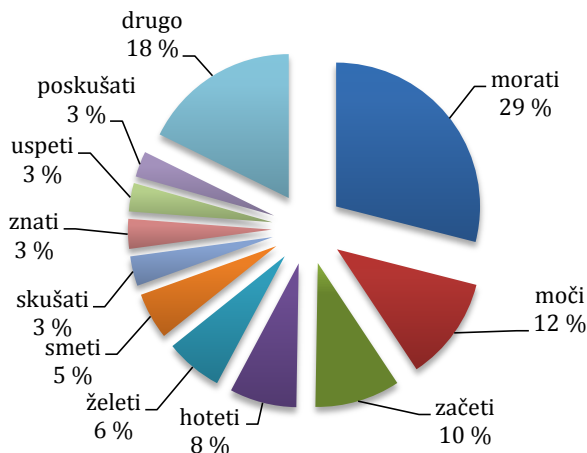
	OBLIKA	ZGLED	ŠT. POJAVITEV	ODSTOTEK VSEH ZGLEDOV
USTREZNO	dolgi nedoločnik	<i>moram reči</i>	1324	99,25
NAPAČNO	kratki nedoločnik	<i>probam prešaltat</i>	10	0,75
	SKUPAJ		1334	100,00

Tabela 1: Zveze polnopomenskega glagola in nedoločnika v učnem korpusu.

¹¹ Pri analizi oblikoskladenjsko označenih korpusov (npr. FidaPLUS) se je pokazalo, da je pri izločanju povedi, v katerih je med elementoma večja razdalja, delež nerelevantnih zgledov nenavadno visok; kljub avtomatičnemu izločanju ločil in besed, ki se med njima ne smeta pojavljati, je v vzorcu še zmeraj zaslediti zveze tipa *prišel je z namenom poslušati*, ki so sicer slovnično pravilne, ampak za pričujočo raziskavo nerelevantne. Tovrstnih in podobnih zvez ni mogoče avtomatično izločiti iz seznama zadetkov.

¹² Avtor programa je Janez Brank.

Pri izločanju polnopolomenskih glagolov so bili upoštevani le slovnično pravilni zgledi. Ob nedoločnikih je bilo zaslediti kar 64 različnih glagolov, od teh je bilo za nadaljnjo analizo izbranih deset najpogostejših, tj. *morati*, *moči*, *začeti*, *hoteti*, *želeli*, *smeti*, *skušati*, *znati*, *uspeti* in *poskušati*. Rezultate prikazuje Slika 1:



Slika 1: Polnopolomenski glagoli ob nedoločniku glede na pogostost pojavljanja v učnem korpusu.

3.3 Vpliv žanra in prenosnika na izbiro nedoločnika

Kratki nedoločnik je v literaturi opredeljen kot pojav, ki je značilen samo za pogovorni jezik, zato se zdi smiselno proučiti, v kolikšni meri prenosnik in žanr dejansko vplivata na izbiro nedoločnika pri govorjenju in pisanju. Podatki, ki so predstavljeni v nadaljevanju, so bili izločeni iz korpusa FidaPLUS,¹³ ki med drugim prinaša tudi informacije o viru in vrsti besedila, v katerem se pojavlja iskani niz. Pri izločanju in obdelavi podatkov so bile glede na korpusne oznake, ki se nanašajo na vrsto prenosnika in žanr, oblikovane štiri kategorije: govorjena besedila, spletna besedila, revije in časopisi ter

¹³ <http://www.fidaplus.net>

leposlovje. Za vsakega od glagolov s seznama¹⁴ je bil izračunan delež odklona od norme, pri čemer so bili podatki obdelani za vsako kategorijo posebej.

Tabela 2 in Tabela 3 vsebujeta podatke za glagole, ob katerih je treba uporabljati dolgi nedoločnik.¹⁵ Po prvem pregledu podatkov se proti vsem pričakovanjem pokaže, da je kratki nedoločnik najpogostejše rabljen v leposlovju (1,23 %) in spletnih besedilih (0,98 %), manj pa v govorjenih besedilih (0,29 %), časopisju in revijah (0,12 %). Visok odstotek pojavljanja v leposlovju je mogoče razložiti, če si podrobneje pogledamo seznam zadetkov; najpogostejše se kratki nedoločniki pojavljajo v premem govoru, kar pomeni, da je avtor v pogovorih med pripovednimi osebami najverjetneje želel ohraniti realnost komunikacije in se je zato odločil za pogovorno različico nedoločnika. V tem primeru torej ne gre za napako, temveč za avtorjevo zavestno in utemeljeno odločitev. Prav tako je prek analize seznama zadetkov mogoče razložiti nepričakovano nizek odstotek pojavljanja kratkega nedoločnika v govorjenih besedilih; zajeten delež govorjenih besedil v korpusu namreč predstavljajo transkripcije sej Državnega zbora. Pričakovati gre, da se bodo govorci v tako formalni govorni situaciji poskušali izražati jezikovno korektno, pri čemer bodo zavestno skušali omejiti vdor pogovornih oblik. V manj formalnih situacijah, npr. v debatah na spletnih forumih, piscu oziroma govorcu ni treba strogo upoštevati norme, kar je skladno z nekoliko višjim odstotkom pojavljanja kratkega nedoločnika v spletnih besedilih – na seznamu zadetkov namreč prevladujejo ravno prispevki, ki so bili objavljeni na spletnih forumih. Hipotezo, ki se dotika vprašanja rabe kratkega nedoločnika v govorjenih besedilih, v tej točki ni mogoče ne potrditi ne ovreči,

¹⁴ Kot je razvidno iz spodnjih tabel, se je pri statistični obdelavi podatkov iz korpusa FidaPLUS potrdil prvotni vrstni red glagolov, ki je bil določen na podlagi pogostosti pojavljanja v učnem korpusu, kar kaže na zanesljivost metodologije, ki smo jo ubrali v prejšnjem razdelku.

¹⁵ Pri tem je treba opozoriti, da se je v korpusni poizvedbi iskalo le sosledje osebne glagolske oblike in nedoločnika, pri čemer je bil pogoj, da se vmes ne smejo vrvati drugi elementi, saj bi bila zanesljivost podatkov zaradi visokega deleža nerelevantnih zgledov postavljena pod vprašaj (prim. razdelek 3.2).

saj je za popolno sliko treba dodatno analizirati podatke iz govornega korpusa (prim. razdelek 3.4). Hkrati pa se je po pričakovanjih izkazalo, da se kratki nedoločnik najredkeje pojavlja v časopisih in revijah, kar potrjuje drugo hipotezo.

	GOVORJENA BESEDILA			SPLETNA BESEDILA		
	D	K	%	D	K	%
<i>morati</i>	3783	8	0,21	1541	15	0,96
<i>moči</i>	2064	7	0,34	5133	28	0,54
<i>začeti</i>	390	2	0,51	326	16	4,68
<i>hoteti</i>	900	4	0,44	92	11	10,68
<i>želei</i>	942	2	0,21	1545	11	0,71
<i>smeti</i>	605	2	0,33	352	3	0,85
<i>skušati</i>	158	0	0,00	57	0	0,00
<i>znati</i>	223	3	1,33	188	9	4,57
<i>uspeti</i>	144	0	0,00	72	0	0,00
<i>poskušati</i>	291	0	0,00	52	0	0,00
SKUPAJ	9500	28	0,29	9358	93	0,98

Tabela 2: Kratki in dolgi nedoločnik v govorjenih in spletnih besedilih (D = dolgi nedoločnik; K = kratki nedoločnik; % = odstotek kratkih nedoločnikov v vzorcu).

	REVIJE IN ČASOPISI			LEPOSLOVJE		
	D	K	%	D	K	%
<i>morati</i>	457480	610	0,13	25782	343	1,31
<i>moči</i>	273095	322	0,12	22974	270	1,16
<i>začeti</i>	208018	285	0,14	12639	224	1,74
<i>hoteti</i>	120650	216	0,18	18087	198	1,08
<i>želei</i>	139438	88	0,06	3871	5	0,13

<i>smeti</i>	97633	93	0,10	5269	77	1,44
<i>skušati</i>	56115	29	0,05	4240	9	0,21
<i>znati</i>	63822	199	0,31	4400	135	2,98
<i>uspeti</i>	82643	51	0,06	2483	5	0,20
<i>poskušati</i>	55105	38	0,07	2910	12	0,41
SKUPAJ	1553999	1931	0,12	102655	1278	1,23

Tabela 3: Kratki in dolgi nedoločnik v revijah in časopisih in v leposlovju (D = dolgi nedoločnik; K = kratki nedoločnik; % = odstotek kratkih nedoločnikov v vzorcu).

3.4 Govor: posameznik in okoliščine

Korpus GOS (Miklavčič Zemljarič in dr. 2009; Verdonik in dr. 2010)¹⁶ je enomilijonski referenčni korpus govorne slovenščine, ki je nastal pri projektu *Sporazumevanje v slovenskem jeziku*. Kljub omejeni velikosti so se avtorji trudili »zajeti vzorčne primere različnih govornih diskurzov v različnih situacijah, zajeti govorni diskurz demografsko reprezentativnega vzorca govorcev, zajeti predvsem tiste govorne situacije, v katerih so uporabniki jezika najpogosteje produktivno-receptivno udeleženi« (Miklavčič Zemljarič in dr. 2009: 425), in tako zagotoviti uravnoteženost korpusa.¹⁷ Korpus prinaša tudi dve posebnosti: prva je povezava med zvokom in zapisom (pri pregledovanju konkordanc je mogoče zadetke tako prebrati kot poslušati) ter dvojni zapis govora (vsako besedilo je bilo transkribirano v pogovorni in standardni različici) (Verdonik in dr. 2010: 12).

Spletni konkordančni omogoča iskanje po izbrani obliki ali vseh oblikah besede, določanje besed v okolici in njihove oddaljenosti od iskane besede ter oblikoskladenjskih značilnosti iskane besede in besed v njeni okolici. Ko se prikažejo zadetki iskanja, jih je mogoče dodatno filtrirati glede na

¹⁶ www.korpus-gos.net

¹⁷ Za podrobnejše informacije o sestavi korpusa in načelih zajema gradiva prim. Zemljarič Miklavčič in dr. (2009).

metapodatke o diskurzih (regija snemanja, leto snemanja, tip diskurza, kanal in govorni dogodek) ter govorcih (spol, starost, izobrazba, regionalna pripadnost in prvi jezik) (Verdonik in dr. 2010: 13).

Na podlagi podatkov iz govornega korpusa smo preverjali dvoje, tj. vpliv govorčeve regionalne pripadnosti in okoliščin na izbiro vrste nedoločnika. Natančneje bomo preverili Tominčevo (1955: 182) trditev o rabi kratkega in dolgega nedoločnika v slovenskih regijah, na podlagi katere je bila oblikovana hipoteza, da bo v podatkih za govorce z Gorenjske, Koroške in vzhodne Štajerske prevladoval dolgi, povsod drugod pa kratki nedoločnik. V korpusu smo iskali sosledje osebne glagolske oblike in kratkega oziroma dolgega nedoločnika, pri čemer je bilo podatke mogoče filtrirati glede na regijo in registrsko območje govorca. Rezultati iskanja so prikazani v Tabeli 4 in Tabeli 5, v katerih so zbrani podatki za vseh deset glagolov, ki smo jih določili v razdelku 3.2.

REGIJA	Dolgi nedoločnik	Kratki nedoločnik	% kratkega nedoločnika v vzorcu
Notranjska	0	7	100,00
Posavska	1	66	98,51
Gorenjska	10	113	91,87
Dolenjska	16	83	83,84
Koroška	13	60	82,19
Primorska	15	65	81,25
Prekmurska	61	14	18,67
SKUPAJ	116	408	77,86

Tabela 4: Raba dolgega in kratkega nedoločnika v sedmih slovenskih regijah.

REGISTRSKO OBMOČJE	Dolgi nedoločnik	Kratki nedoločnik	% kratkega nedoločnika v vzorcu
ljubljsko	67	456	87,19
mariborsko	60	387	86,58
celjsko	25	89	78,07
goriško	13	45	77,59
SKUPAJ	165	977	85,55

Tabela 5: Raba dolgega in kratkega nedoločnika po Sloveniji glede na registrsko območje.

Iz podatkov je razvidno, da je v govorjenem jeziku kratki nedoločnik prevladujoča oblika povsod po Sloveniji razen v Prekmurju; proti pričakovanjem glede na Tominčevo ugotovitev je bilo zaslediti visok odstotek kratkih nedoločnikov na Gorenjskem (91,87 %), Koroškem (82,19 %) in v mariborskem registrskem območju (86,58 %), kjer naj bi po Tomincu (1955) prevladoval dolgi nedoločnik.¹⁸ Hipotezo, ki smo jo postavili pred zbiranjem in analizo podatkov, lahko torej ovrzemo.

Korpus GOS prinaša podrobno klasifikacijo tipov diskurza, na podlagi katere je mogoče pridobiti dragocene podatke o vplivu nekaterih okoliščin, tj. kanala in govornega dogodka, na pogostost rabe kratkega nedoločnika. Oblikovana je bila hipoteza, da se kratki nedoločnik redkeje pojavlja v javnih, pogostejše pa v nejavnih govornih situacijah. Tabela 6 prikazuje končni seštevek pojavitev za vseh deset glagolov, na podlagi katere lahko hipotezo potrdimo – kratki nedoločnik se v nejavnem diskurzu (93,71 %) pojavlja precej pogostejše kot v javnem (69,2 %).¹⁹ Najvišje odstotke je mogoče zaslediti v nejavnem zasebnem

¹⁸ Odstotki so ponekod celo višji kot v regijah, ki jih Tominec uvršča v območje kratkega nedoločnika.

¹⁹ Nenavadno nizka pogostost pojavljanja kratkega nedoločnika v športnih prenosih (samo 4,48 %) občutno vpliva na končni delež v kategoriji, vendar bi bil odstotek kratkih nedoločnikov v javnem diskurzu nižji od odstotka v javnem diskurzu, tudi če rezultatov za (opomba se nadaljuje na naslednji strani)

diskurzu (94,02 %), kamor uvrščamo pogovore v družini in med prijatelji oziroma znanci, v nejavnem nezasebnem diskurzu (93,03 %), kamor spadajo formalni in neformalni delovni sestanki in razgovori, konzultacije na fakulteti, storitve, prodaja in svetovanje, ter v nekaterih zvrsteh razvedrilnega diskurza, npr. v moderiranem pogovoru na televiziji (98,48 %), moderirani televizijski oddaji (90,30 %), resničnostnem šovu (90,24 %) in moderiranem pogovoru na radiu (84,91 %).²⁰ Kratki nedoločnik se najredkeje uporablja v športnih prenosih (4,48 %); ker je odstotek bistveno nižji kot v ostalih podtipih in ker je število zadetkov razmeroma nizko, se je zdelo smiselno pregledati, ali se v zadetkih pojavlja samo en govorec, ki je pri rabi dolgega nedoločnika precej dosleden, kar bi postavilo pod vprašaj zanesljivost pridobljenih podatkov. Po dodatni analizi konkordanc je bilo mogoče določiti vsaj tri različne profile (moški, star od 25 do 34 let, iz okolice Ljubljane, srednješolska izobrazba; moški, star od 45 do 59 let, iz okolice Ljubljane, fakultetna izobrazba ali več; moški, star od 35 do 59 let, z Dolenjske, srednješolska izobrazba), kar dokazuje, da so v gradivu zbrane izjave vsaj treh različnih govorcev. Nekoliko nizek odstotek je zaslediti v novinarskih prispevkih (40 %), povsod drugod pa je odstotek kratkih nedoločnikov med 60 % in 80 %. Na podlagi podatkov lahko sklenemo, da je kratki nedoločnik prevladujoča oblika v govorjenem diskurzu (z izjemo športnih prenosov in novinarskih prispevkov) ter da tip diskurza vpliva na izbiro oblike.

Tip diskurza		Dolgi nedoločnik	Kratki nedoločnik	% kratkega nedoločnika v vzorcu
i. JAVNI		567	1274	69,20
	a. Informativno-izobraževalni:	392	833	68,00

športne prenose ne bi upoštevali v končnem seštevku (odstotek bi se minimalno zvišal na 71,16 %).

²⁰ Za dodatne informacije o jezikovnih značilnosti intervjujev v radijskih programih prim. Verovnik (2012).

televizija	<i>moderirani pogovor</i>	103	246	70,49
	<i>novinarski prispevek</i>	54	36	40,00
radio	<i>moderirani pogovor</i>	53	146	73,37
	<i>moderirani program</i>	46	91	66,42
osebni stik	<i>predavanja, tečaji ipd.</i>	136	314	69,78
b. Razvedrilni:		175	441	71,59
televizija	<i>moderirani pogovor</i>	1	65	98,48
	<i>moderirana oddaja</i>	13	121	90,30
	<i>resničnostni šov</i>	4	37	90,24
	<i>športni prenos</i>	64	3	4,48
radio	<i>moderirani pogovor</i>	8	45	84,91
	<i>moderirani program</i>	85	140	62,22
ii. NEJAVNI		82	1221	93,71
a. Nezasebni		29	387	93,03
b. Zasebni		53	834	94,02

Tabela 6: Raba kratkega in dolgega nedoločnika glede na tip diskurza, kanal in govorni dogodek.

4 SKLEP

Na podlagi analize podatkov iz korpusa FidaPLUS in korpusa govorjene slovenščine GOS je bilo mogoče ugotoviti, kateri dejavniki vplivajo na izbiro

napačne oblike nedoločnika. Potrjene so tri hipoteze: kratki nedoločnik se v govorjenem jeziku redkeje pojavlja v javnih, pogostejše pa v nejavnih govornih situacijah, v pisnem jeziku pa se najredkeje pojavlja v besedilih, ki so navadno lektorirana, tj. v časopisih in revijah. Edina izjema so literarna dela, v katerih se kratki nedoločnik pojavlja pogostejše, kot bi si pričakovali, vendar se je pri podrobnejši analizi pokazalo, da je odstotek odvisen predvsem od posameznih besedil, ki so zajeta v korpus (nekateri avtorji kratki nedoločnik namerno uporabljajo kot sredstvo za doseganje specifičnega slogovnega učinka). V nadaljevanju se je izkazalo, da korpus FidaPLUS ni primeren vir podatkov za proučevanje govorjenega jezika, saj vsebuje le manjši vzorec besedilnozvrstno enoličnih besedil, zato je bilo treba pogostost rabe kratkega nedoločnika preveriti še v govornem korpusu. Na podlagi podatkov iz korpusa GOS, ki vsebuje uravnotežen in reprezentativen vzorec govorne slovenščine, je bilo tako mogoče prvo hipotezo v celoti potrditi; kratki nedoločnik se namreč v vseh štirih kategorijah (javni informativno-izobraževalni, javni razvedrilni, nejavni zasebni in nejavni nezasebni diskurz) pojavlja v najmanj 68 % vseh zgledov. Pri proučevanju pisnih besedil iz korpusa FidaPLUS tako visokega odstotka ni bilo mogoče zaslediti. Edina hipoteza, ki jo lahko z gotovostjo ovrzemo, je povezana z vprašanjem regionalne variantnosti in temelji na Tominčevi trditvi, da »/n/a Krasu, Notranjskem, Dolenjskem, na južnem Štajerskem poznajo v glavnem kratki nedoločnik brez i, na Gorenjskem, Koroškem, na vzhodnem Štajerskem pa dolgi nedoločnik z i« (Tominec 1955: 182). Podatki iz korpusa GOS so namreč pokazali, da je kratki nedoločnik prevladujoča oblika povsod po Sloveniji razen v Prekmurju, kjer se povečini uporablja dolgi nedoločnik. Pri tem velja posebej opozoriti, da tega dela raziskave ne gre razumeti kot alternativo dialektološkim obravnavam, katerih metodologija in gradivo sta za pridobivanje podrobnih ter zanesljivih podatkov za to specifično problematiko nedvomno najustreznejša; naš namen je bil zgolj podati nekaj splošnih ugotovitev v zvezi z rabo dolgega in kratkega nedoločnika v slovenskih regijah, ki temeljijo na statistični analizi obstoječih korpusnih podatkov za sodobno govorjeno slovenščino.

VIRI

Korpus FidaPLUS. Dostopno prek: <http://www.fidaplus.net>.

Korpus GOS. Dostopno prek: <http://www.korpus-gos.net>.

Korpus Šolar. Dostopno prek: <http://www.slovenscina.eu/dat/korpusi/solar/SOLAR.zip>.

Paket korpusov in programske opreme za skladiščno razčlenjevanje.

Dostopno prek: http://razclenjevalnik.slovenscina.eu/Programska_oprema.aspx.

LITERATURA

Arhar, Š. (2009): Učni korpus SSJ in leksikon besednih oblik za slovenščino. *Jezik in slovstvo*, 53 (3–4): 43–56.

Bajec, A. (1955): O pogovornem jeziku. *Jezik in slovstvo*, 1 (6–7): 161–165.

Dobrovoljc, K., Krek, S., in Rupnik, J. (2012): Skladišni razčlenjevalnik za slovenščino. V T. Erjavec, in J. Žganec Gros (ur.): *Zbornik Osme konference Jezikovne tehnologije*: 42–47. Ljubljana: Institut Jožef Stefan.

Kosem, I., in dr. (v tisku): *Analiza jezikovnih težav učencev: korpusni pristop*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

Miklavčič Zemljarič, J., Stabej, M., Krek, S., in Zwitter Vitez, A. (2009): Kaj in zakaj v referenčni govorni korpus slovenščine. V M. Stabej (ur.): *Infrastruktura slovenščine in slovenistike*: 423–428. Ljubljana: Znanstvena založba Filozofske fakultete.

Rozman, T., Stritar, M., Krapš Vodopivec, I., Kosem, I., in Krek, S. (2010): *Nova didaktika poučevanja slovenskega jezika*. Dostopno prek: http://projekt.slovenscina.eu/Media/Kazalniki/Kazalnik15/Nova_didaktika_Sporazumevanje.pdf.

Sporazumevanje v slovenskem jeziku. Dostopno prek: <http://www.slovenscina.eu>.

Tominec, I. (1955): Dolgi in kratki nedoločnik. *Jezik in slovstvo*, 1 (6–7): 181–182.

Toporišič, J. (1966): Pogovorni nedoločnik. *Jezik in slovstvo*, 11 (8): 264–265.

- Toporišič, J. (1984): *Slovenska slovnica*. Maribor: Obzorja.
- Urbančič, B. (1962): O kriterijih pravilnosti v knjižni slovenščini. *Jezik in sloustvo*, 7 (7): 200–209.
- Verdonik, D., Zwitter Vitez, A., Romih, M., in Krek, S. (2010): Konkordančnik za govorni korpus GOS. V T. Erjavec in J. Žganec Gros (ur.): *Zbornik Sedme konference Jezikovne tehnologije*: 12–15. Ljubljana: Institut Jožef Stefan. Dostopno prek:
<http://nl.ijs.si/isjt10/JezikovneTehnologije2010.pdf>.
- Verovnik, T. (2012): *Radijska dvogovornost: jezikovne izbire novinarjev*. Ljubljana: Založba FDV. Dostopno prek: http://www.fdv.uni-lj.si/docs/zalozba/radijska-dvogovornost_jezikovne-izbire-novinarjev.pdf.
- Vidali, I. (1947): Več živega jezika. *Novi svet*, 2 (9–10): 729–733.
- Vodušek, B. (1959): Historična pisava in historična izreka. *Jezik in sloustvo*, 4 (7): 193–200.
- Žagar, F. (1993): *Slovenska slovnica in jezikovna vadnica*. Maribor: Obzorja.

THE SHORT INFINITIVE IN SLOVENE: A CORPUS-BASED APPROACH

The article presents the results of an in-depth, corpus-based study on the use of the short and long infinitive in Slovene. After a few introductory remarks followed by a brief theoretical overview of the topic, a categorised list of problematic infinitive structures based upon a previously performed analysis of *Šolar*, a corpus of student texts, is provided. *Šolar*'s data gave insight into the whole range of linguistic problems associated with the use of two forms and allowed for the subject of the study, i.e. 'full verb + infinitival complement' structures, to be precisely defined. The following section describes the method used to compile a shortlist of high-frequency full verbs co-occurring with infinitival complements from the half-million-word, morphosyntactically and syntactically annotated training corpus that was created within the *Communication in Slovene* project. Finally, the co-relation between genre, mode and the use of the two forms is examined in detail; based on a quantitative analysis of data extracted from corpora of written (*FidaPLUS*) and spoken language (*GOS*) for each full verb on the shortlist, new findings on the actual use of the short and long infinitive in both written and spoken texts are presented.

Keywords: short infinitive, corpus, Slovene syntax, FidaPLUS, GOS

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 2.5 Slovenija.

This work is licensed under the Creative Commons Attribution ShareAlike 2.5 License Slovenia.

<http://creativecommons.org/licenses/by-sa/2.5/si/>

