

TEST POZNAVANJA SPLOŠNIH BESED V SLOVENŠČINI MED UDELEŽENCI MLADINSKE POLETNE ŠOLE SLOVENŠČINE

Matej KLEMEN,¹

¹Center za slovenščino kot drugi in tuji jezik, Filozofska fakulteta, Univerza v Ljubljani

V prispevku predstavljamo pilotno izvedbo testa poznavanja splošnih besed v slovenščini. S testom smo želeli ugotoviti, v kolikšni meri govorci slovenščine kot drugega in tujega jezika poznajo splošne besede v slovenščini in kakšen je pri testiranih napredek v poznavanju splošnega besedišča po udeležbi na poletni šoli slovenščine. Zanimalo nas je, ali bi bilo tak test mogoče uporabiti kot alternativo časovno potratnejšim uvrstitvenim testom, da bi z njim udeležence tečajev slovenščine kot drugega in tujega jezika razvrstili v skupine, ki so homogene po jezikovnem znanju. Po vzoru testov za angleščino in grščino kot neprva jezika smo pripravili t. i. da/ne test poznavanja najfrekventnejših 4000 splošnih besed v slovenščini, ki smo ga v dveh ponovitvah izvedli med udeleženci 17. Mladinske poletne šole slovenščine.

Ključne besede: besedišče, test, slovenščina kot drugi in tuji jezik, študentski prispevek

1 UVOD¹

Pri organizaciji tečajev slovenščine se srečujemo z izzivom, kako udeležence hitro in učinkovito razporediti v skupine, ki so po jezikovnem znanju čim bolj homogene. Tradicionalno jih na Centru za slovenščino kot drugi in tuji jezik Filozofske fakultete Univerze v Ljubljani razvrščamo na podlagi uvrstitvenega testa in uvrstitvenega intervjuja, s katerima ocenimo sporazumevalno jezikovno zmožnost udeležencev tečajev. Ob velikem številu udeležencev, zlasti ob začetku večjih prireditev (npr. poletnih šol, ki se jih udeležuje po več kot 100 tečajnikov), tak način uvrščanja v skupine pomeni veliko dela in terja veliko časa.

Kot odgovor na podobne težave je v osemdesetih letih 20. stoletja Meara s sodelavci za angleščino razvil Eurocentres Vocabulary Size Test, ki ga je kasneje še dopolnjeval (Meara in Jones, 1988; Meara in Miralpeix, 2016). Test

¹ Prispevek je nastal v sklopu doktorskega študija digitalnega jezikoslovja na Univerzi v Ljubljani.

temelji na predpostavki, da je poznavanje besedišča povezano s splošno jezikovno zmožnostjo. Gre za enostaven in hiter test, pri katerem se testirani za posamezne besede v tujem jeziku odločajo, ali jih poznajo ali ne.² Kot eno od prednosti testa Meara in Miralpeix (2016, str. 115) omenjata tudi možnost prilagoditve za različne jezike.

V prispevku predstavljamo pilotno izvedbo testa poznavanja splošnega besedišča v slovenščini in njegovo uporabnost pri uvrščanju udeležencev tečajev slovenščine kot drugega in tujega jezika (SDTJ) v skupine, ki so homogene po jezikovnem znanju.

2 TEST POZNAVANJA SPLOŠNIH BESED V SLOVENŠČINI

2.1 Zgledi za pripravo testa besedišča

Test za poznavanje besed v slovenščini je bil pripravljen po vzoru testa V_YesNo (Meara in Miralpeix, 2016), ki izhaja iz omenjenega Eurocentres Vocabulary Size Test, in testa obsega slovarja, kot sta ga pripravila Milton in Alexiou (2010).

Test V_YesNo preverja poznavanje 10.000 najpogostejših besed v angleščini. Gre za test v digitalni obliki, v katerem so testiranemu predstavljene posamezne besede brez konteksta. Testirani za besedo označi, ali njen pomen pozna ali ne.³ Ob pravih besedah so v test vključene tudi v angleščini neobstoječe besede (nebesede), kar naj bi preprečevalo goljufanje oz. ugibanje. Test obsega 100 pravih besed (po 10 besed za vsakih 1000) in 100 nebesed. Ob koncu testa se izpiše rezultat, ki je ocena, koliko besed testirani pozna. Doseči je mogoče največ 10.000 točk. Če testirani označi, da pozna nebesedo, se njegov končni rezultat zniža, odbitek točk za posamezno napačno prepoznavo nebesede pa je odvisen od tega, kako testirani odgovarja na vprašanja s pravimi besedami. Testirani, ki pravilno prepoznajo večino pravih besed, so za posamezno napačno prepoznavo nebesede kaznovani blažje, medtem ko so tisti, ki pravilno prepoznajo le nekaj pravih besed, za napačno prepoznavo nebesed kaznovani strožje (Meara in Miralpeix 2016).

² V angleščini se je takih testov prijelo poimenovanje *yes/no test*.

³ Test je dostopen na: https://www.lognostics.co.uk/tools/V_YesNo/V_YesNo.htm.

Test, s katerim sta Milton in Alexiou (2010) želela oceniti obseg besedišča pri učencih grščine, je podoben testu V_YesNo. Prav tako gre za *da/ne test*, besede pa zajema izmed prvih 5000 najpogostejših lem v grškem nacionalnem korpusu (Hellenic National Corpus). Za vsakih 1000 besed je izbranih po 20 besed, tem pa je dodanih še 20 v grščini neobstoječih besed. Besede so testiranemu na papirju predstavljene v seznamu, brez besedilnega konteksta, odločiti pa se mora, ali jo »pozna ali zna uporabiti« (Milton in Alexiou, 2010, str. 318). Če testirani označi, da obstoječo besedo pozna, se njegov odgovor točkuje s 50 točkami, najvišje število točk je 5000. Če pa testirani označi, da pozna nebesedo, se od končnega rezultata za vsak tak odgovor odšteje 250 točk, izgubi lahko torej največ 5000 točk. Končno število točk naj bi predstavljalo oceno, kolikšen je obseg besedišča pri testiranem (Milton in Alexiou, 2010).

Predstavljena testa sta bila kot model za pripravo testa poznavanja splošnih besed v slovenščini izbrana predvsem, ker omogočata enostavno distribucijo in hitro odgovarjanje. Tako smo lahko novo pripravljeni test izvedli med udeleženci Mladinske poletne šole slovenščine (MPŠ), ne da bi z raziskavo pretirano posegli v pouk.

2.2 Priprava testa poznavanja splošnih besed v slovenščini

Pri pripravi testa poznavanja besed v slovenščini smo sledili principom omenjenih testov. Pripravili smo *da/ne test*, ki izhaja iz Referenčnega seznama pogostih splošnih besed za slovenščino (Pollak in sod., 2020). V tem seznamu, ki je nastal s prekrivanjem najpogostejših lem iz štirih slovenskih besedilnih korpusov (Kres, GOS, Janes, Šolar 2.0), je zbranih 4768 pogostih splošnih lem (Arhar Holdt in sod., 2020). Za pripravo testa je bilo upoštevanih prvih 4000 lem, saj zadnja tisočica lem v seznamu ni popolna. Za vsakih zaporednih 1000 lem je bilo naključno izbranih 20 lem, skupaj torej 80. Ker smo želeli, da je izbor čim bolj enakomeren, je bila izmed vsakih zaporednih 50 lem naključno izbrana ena. Pri preverjanju izbora so bile leme, ki lahko pripadajo različnim besednim vrstam (npr. *raven*, ki je lahko samostalnik ali pridevnik), nadomeščene z drugimi iz istega frekvenčnega ranga. V nasprotju z nekaterimi testi besedišča, ki vključujejo samo samostalnike, glagole in pridevnike (gl. Kremmel in Schmitt, 2017, str. 1–2), so bile kot pri Meari (1992) vključene tudi

druge besedne vrste (npr. veznik *kajti*, prislovi *tako*, *večinoma* in *podobno*, ki so v izboru za prvih 1000 najpogostejših splošnih besed). Poleg izbranih besed smo v test vključili tudi v slovenščini neobstoječe besede, ki smo si jih izmislili in na videz delujejo kot slovenske (npr. *posminati*, *čembrita*, *deptanjski*). Razmerje med besedami in nebesedami je bilo kot v testu Milтона in Alexiou (2010) 5 : 1. Z izbranimi 80 obstoječimi besedami in 16 nebesedami test tako vsebuje 96 vprašanj.

Zaradi enostavnejše izvedbe testiranja, zbiranja in analize odgovorov smo se odločili za digitalno obliko testa. Test poznavanja besed v slovenščini smo naložili v didaktično orodje za izvedbo različnih kvizov Socrative.⁴ Pri vsakem vprašanju se je testiranemu prikazalo enako navodilo (*Če veste, kaj beseda pomeni, izberite TRUE. Če ne veste, kaj beseda pomeni, izberite FALSE. Če niste prepričani, izberite FALSE.*), ob njem pa beseda in gumbi *true*, *false* in *submit answer* (Slika 1).⁵ Besede so se vsakemu testiranemu prikazovale v naključnem vrstnem redu, kar je pomenilo, da so bile bolj in manj frekventne besede med seboj pomešane, kot za enega od testov besedišča priporoča Nation (2012).

Pred izvedbo je bil test preizkušen na 5 govorcih slovenščine kot prvega jezika, ki so vseh 16 besed ustrezno prepoznali kot nebesede. Dve verziji testa z različnimi 80 besedami, na enak način izbranimi iz Referenčnega seznam pogostih splošnih besed, in 16 nebesedami sta bili preizkušeni na skupini 5 govorcev SDTJ, ki so na obeh testiranjih dosegli podoben rezultat.

⁴ <https://www.socrative.com>

⁵ Ker testa nismo izvajali pri popolnih začetnikih v slovenščini (gl. razdelek 4), je bilo navodilo samo v slovenščini, saj smo predvidevali, da ga bodo testirani razumeli. Zaradi prednastavljenega formata in jezika kviza gumbov v angleščini nismo mogli preimenovati. Prav tako ni bilo mogoče dodati dodatnega gumba *Ne vem*, ali *Nisem prepričan/a*, s katerim bi bilo mogoče znižati stopnjo ugibanja pri odgovarjanju (Zhang, 2013).

Slika 1: Prikaz vprašanja v orodju Socrative.

4 of 96

Če veste, kaj beseda pomeni, izberite **TRUE**.
Če ne veste, kaj beseda pomeni, izberite **FALSE**.
Če niste prepričani, izberite **FALSE**.

hvala

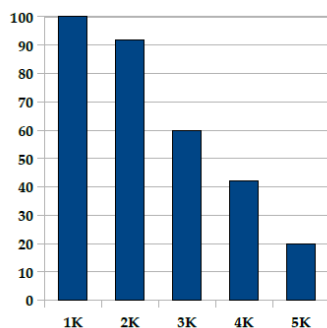
☐ T True ☐ F False

SUBMIT ANSWER

3 HIPOTEZE

S testom poznavanja najpogostejših 4000 splošnih besed v slovenščini smo želeli preveriti, ali tako kot za govorce drugih tujih jezikov tudi za govorce SDTJ velja, da v procesu učenja prej in v večji meri poznajo bolj frekventne besede kot manj frekventne. Predvidevamo, da bomo za udeležence MPŠ, med katerimi smo test izvedli, prišli do padajočega profila poznavanja besedišča, kot ga predstavlja Meara (1992; Slika 2) (hipoteza 1).

Slika 2: Tipični profil poznavanja besedišča (vir: Meara, 1992, str. 6).



Želeli smo preveriti, ali bi test poznavanja najpogostejših 4000 splošnih besed v slovenščini lahko uporabili kot alternativo uvrstitvenim testom in/ali uvrstitvenim intervjujem. Predvidevamo, da se bo pokazala razlika v skupnem številu točk med posameznimi skupinami udeležencev MPŠ glede na jezikovno znanje, in sicer da bodo začetniki dosegli nižje rezultate kot nadaljevalci in

izpopolnjevalci (hipoteza 2). Predvidevamo, da bi na podlagi rezultatov testa poznavanja besedišča lahko večji del učencev razvrstili v skupine, ki bi bile sorazmerno homogene po jezikovnem znanju, primerljivo ustrezno kot na podlagi rezultatov uvrstitvenega testa (hipoteza 2.1).

Z dvema ponovitvama testa smo želeli poleg tega ugotoviti, kakšen je napredek v poznavanju splošnega besedišča po 40 urah pouka in dveh tednih življenja v slovenskem okolju. Predvidevamo, da testirani pri drugem testiranju dosežejo pomembno višji rezultat (hipoteza 3) in da bodo spremembe večje predvsem pri testiranih, ki so pri prvem testiranju dosegli nižji rezultat (hipoteza 3.1).

4 TESTIRANJE IN PODATKI

Za sodelovanje smo prosili udeležence Mladinske poletne šole slovenščine, ki jo obiskujejo mladostniki, stari od 13 do 18 let, za katere je slovenščina najpogostejše eden od prvih jezikov (zamejci, izseljenci), dediščinski ali tuji jezik.

Na 17. MPŠ je leta 2022 pouk potekal v devetih skupinah (v nadaljevanju so označene s številkami, npr. 4, 7, 9). Udeleženci MPŠ so bili vanje razdeljeni na podlagi rezultatov uvrstitvenega testa, deloma pa tudi na podlagi drugih dejavnikov (npr. prvega jezika). Skupine so bile po stopnji jezikovnega znanja torej različne, sosednje skupine (npr. skupini 8 in 9) pa se v jezikovnem znanju med seboj niso vedno nujno močno razlikovale.

Testiranje je bilo izvedeno dvakrat: 5. julija 2022 in 15. julija 2022, prvi in zadnji dan pouka na MPŠ. Učitelji so udeležencem MPŠ dali navodila, udeleženci pa so test izpolnjevali na svojih pametnih telefonih. Podatkov o času reševanja program ni beležil, ocenjujemo pa, da so testirani test zaključili v približno 5 do 12 minutah. Ker je bilo za test V_YesNo ugotovljeno, da ni zanesljiv za začetnike (Meara in Miralpeix, 2016, str. 118), smo se odločili, da popolnih začetnikov v slovenščini v testiranje ne bomo vključili, in tako na MPŠ 2022 test izvedli le v skupinah od 3 do 9. Iz različnih razlogov udeleženci v nekaterih skupinah testa niso reševali v obeh testiranjih.

V prvem testiranju (T1) je sodelovalo 56 udeležencev MPŠ: 3 od teh testa niso dokončali, za 5 pa nismo dobili podatka, v katero skupino so bili razporejeni na

podlagi uvrstitvenih testov. V drugem testiranju (T2) je sodelovalo 54 udeležencev MPŠ: 2 testa nista izpolnila do konca, za 8 pa nismo dobili podatka o skupini. V analizo posameznega testiranja so bili vključeni samo odgovori oseb, ki so na test odgovorile v celoti in za katere smo imeli podatek o skupini (za T1 je bilo takih 48, za T2 pa 44), za primerjalno analizo pa so bili upoštevani samo odgovori oseb, ki so sodelovale v obeh testiranjih in bile v obeh v isti skupini (takih je bilo 25).⁶

Poleg rezultatov testiranja smo za testirane, ki so sodelovali v T1, pridobili število točk, ki so jih dosegli na uvrstitvenem testu pri nalogah za preverjanje bralnega razumevanja in oblikoskladenjske zmožnosti. Za večino smo dobili tudi podatek, kako je bilo njihovo jezikovno znanje ocenjeno na podlagi omenjenih nalog uvrstitvenega testa in dodatne naloge za pisanje z oznakami Z, Z+, Z/N, N–, N, N+, N+/I, I–, I, ki pomenijo (boljši ali slabši) začetnik, nadaljevalec in izpopolnjevalec.⁷ Ker je naš vzorec majhen in je posamezno oznako lahko dobil tudi le en testirani, smo oznake razdelili v tri glavne: Z (začetnik), N (nadaljevalec) in I (izpopolnjevalec).⁸

5 REZULTATI

V podatkih je bilo pri obstoječih besedah vsakemu odgovoru, pri katerem je testirani označil, da besedo pozna (*true*), pripisanih 50 točk, vsak napačen odgovor pa je dobil 0 točk. Posebej je bilo točkovanih tudi 16 nebesed. Vsaki, ki je bila prepoznana kot slovenska beseda (testirani je označil *true*), je bilo pripisanih –250 točk. Vsak »pravilen« odgovor pa je dobil 0 točk. Izračunano je bilo skupno število točk za vsakih tisoč besed, skupno število točk za prave besede, skupno število točk za neprave besede in skupno korigirano število točk (od točk za obstoječe besede so bile odštete točke za nebesede, ki so bile prepoznane kot obstoječe besede).⁹

⁶ Nekateri udeleženci MPŠ so bili lahko po prvih dneh glede na svoje želje in/ali jezikovne zmožnosti prerazporejeni v druge skupine.

⁷ Te oznake niso usklajene s stopnjami jezikovnega znanja po *Skupnem evropskem jezikovnem okviru* (SEJO).

⁸ Oznaka Z/N je bila prevedena v Z, N+/I pa v N.

⁹ Pri analizi rezultatov se ukvarjamo s točkami, ki so jih testirani dosegli pri odgovorih za besede, (opomba se nadaljuje na naslednji strani)

Pri T1 in T2 so testirani v povprečju v večji meri poznali frekventnejše besede in za vsakih naslednjih tisoč dosegli nekoliko nižji rezultat (Tabela 1). V obeh testiranjih se je pokazal padajoči profil poznavanja besedišča (Graf 1 in 2),¹⁰ ki se začne stopničasto spuščati šele pri poznavanju tretjih in četrtih tisoč besed. To nam potrdi tudi analiza variance (ANOVA) za odvisne vzorce, ki je pokazala, da se pri T1 in T2 rezultati za posameznih tisoč besed med seboj statistično pomembno razlikujejo (ob upoštevanju Greenhouse-Geissserjevega popravka T1: $F(2,48) = 29,9$, $p < 0,001$; T2: $F(2,37) = 31,9$, $p < 0,001$). Če s post hoc testom za T1 in T2 primerjamo povprečne dosežke za tisoč besed in upoštevamo Bonferronijev popravek, edino razlika med T1 1–1000 in T1 1001–2000 ($p_{\text{bonferroni}} = 1,0$) ter T2 1–1000 in T2 1001–2000 ($p_{\text{bonferroni}} = 0,8$) ni statistično značilna, vse druge razlike pa so statistično značilne (Tabela 2).

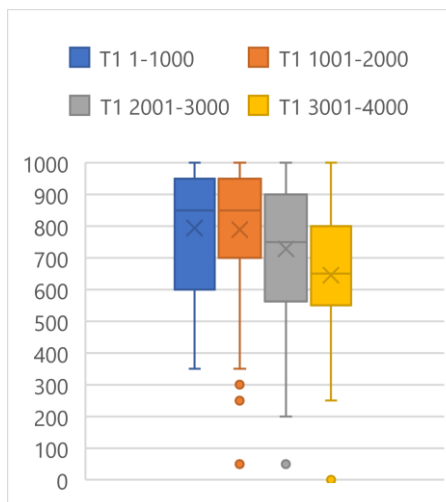
Tabela 1: Število točk, doseženih pri T1 in T2 za posameznih tisoč besed.

	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Minimum</i>	<i>Maksimum</i>
T1 1–1000	48	794	850	189	350	1000
T1 1001–2000	48	789	850	221	50	1000
T1 2001–3000	48	728	750	234	50	1000
T1 3001–4000	48	646	650	206	0	1000
T2 1–1000	44	842	875	166	400	1000
T2 1001–2000	44	816	900	225	250	1000
T2 2001–3000	44	738	800	248	100	1000
T2 3001–4000	44	684	700	252	50	1000

odgovore za nebesede večinoma puščamo ob strani. Kako pri rezultatu testa upoštevati nebesede, je med raziskovalci pogosto predmet razprave (Durrant in sod., 2022, str. 178). V predstavljeni test smo jih vključili, ker smo sledili formatu testa za angleščino oz. grščino in z namenom, da bi testirane odvrnile od ugibanja. Zdi se, da so jih testirani prepoznali kot take. Pri T1 in T2 bi odbitek točk za nebesede verjetno le malo vplival na povprečje (T1: $N = 48$, $M = -313$, $Mdn = -125$; T2: $N = 44$, $M = -455$, $Mdn = -125$).

¹⁰ Ker sta se skupini, ki sta sodelovali v T1 in T2, deloma razlikovali, rezultate prikazujemo v ločenih grafih. Rezultate za 25 testiranih, ki so sodelovali tako pri T1 kot pri T2, primerjalno prikazujemo kasneje.

Graf 1: Število doseženih točk za posameznih tisoč besed pri T1.



Graf 2: Število doseženih točk za posameznih tisoč besed pri T2.

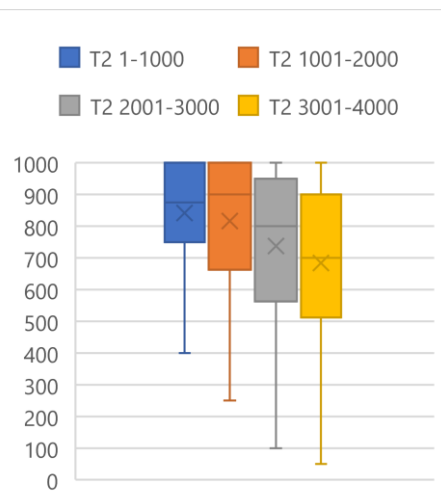


Tabela 2: Primerjava povprečnih dosežkov za posameznih tisoč besed.

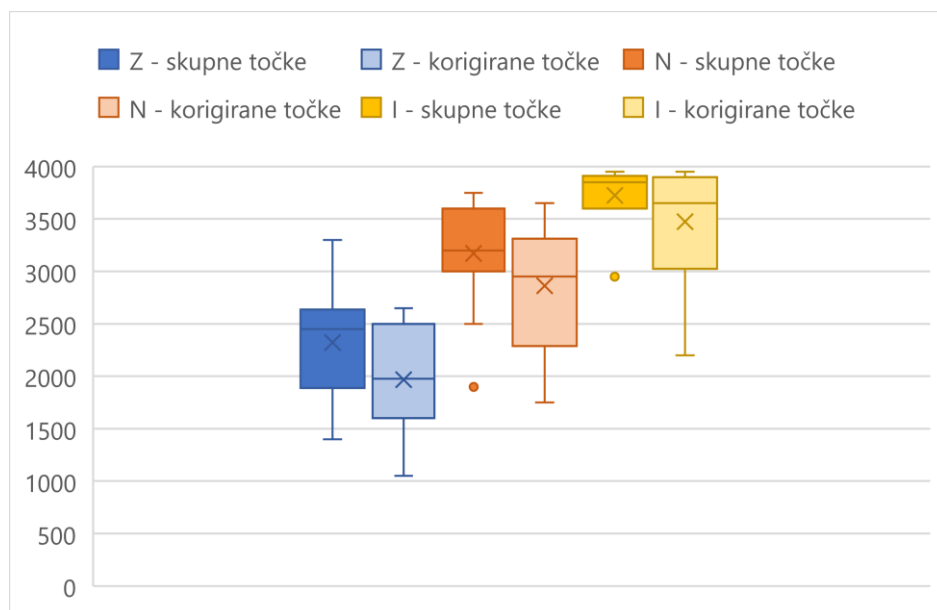
primerjava		razlika v povprečju	SE	df	t	p _{bonferroni}
T1 1–1000	- T1 1001–2000	5,21	19,9	47,0	0,261	1,000
	- T1 2001–3000	65,63	20,8	47,0	3,155	0,017
	- T1 3001–4000	147,92	20,6	47,0	7,190	< 0,001
T1 1001–2000	- T1 2001–3000	60,42	13,6	47,0	4,456	< 0,001
	- T1 3001–4000	142,71	15,2	47,0	9,400	< 0,001
T1 2001–3000	- T1 3001–4000	82,29	15,4	47,0	5,357	< 0,001
T2 1–1000	- T2 1001–2000	26,1	17,1	43,0	1,53	0,801
	- T2 2001–3000	104,5	21,6	43,0	4,84	< 0,001
	- T2 3001–4000	158,0	21,7	43,0	7,26	< 0,001
T2 1001–2000	- T2 2001–3000	78,4	15,0	43,0	5,22	< 0,001
	- T2 3001–4000	131,8	17,4	43,0	7,56	< 0,001
T2 2001–3000	- T2 3001–4000	53,4	14,5	43,0	3,69	0,004

Kot kažeta Tabela 3 in Graf 3 se glede na stopnjo jezikovnega znanja pri doseženem številu točk na T1 kaže razlika med testiranimi, ki so na različnih stopnjah jezikovnega znanja. Analiza variance (ANOVA) za neodvisne vzorce je pokazala, da se rezultati testa poznavanja splošnih besed v slovenščini, ki so jih testirani na različnih stopnjah jezikovnega znanja (Z, N, I) dosegli, med seboj statistično pomembno razlikujejo (T1 skupne točke: $F(2, 41) = 28,8$, $p < 0,001$; T1 korigirane točke: $F(2, 41) = 20,5$, $p < 0,001$).

Tabela 3: Število doseženih točk pri T1 za testirane glede na stopnjo jezikovnega znanja.

	poenostavljena oznaka stopnje	<i>N</i>	<i>M</i>	<i>Mdn</i>	<i>SD</i>	<i>Minimum</i>	<i>Maksimum</i>
T1	Z	12	2321	2450	549	1400	3300
skupne	N	22	3170	3200	449	1900	3750
točke	I	10	3725	3850	304	2950	3950
T1	Z	12	1967	1975	539	1050	2650
korigirane	N	22	2864	2950	566	1750	3650
točke	I	10	3475	3650	571	2200	3950

Graf 3: Skupno število točk in korigirane točke pri T1 glede na stopnjo jezikovnega znanja.



Izračunali smo povezanost med tem, v katero od devetih skupin na MPŠ so bili testirani uvrščeni, in rezultatom, ki so ga dosegli pri T1. Ugotovili smo, da je povezanost visoka in statistično značilna (T1 skupne točke: $r = 0,737$, $df = 46$, $p = < 0,001$; T1 korigirane točke: $r = 0,732$, $df = 46$, $p = < 0,001$). Ta povezanost je le nekoliko manjša, kot je povezanost med številom točk na uvrstitvenem testu in razvrstitvijo v skupino ($r = 0,842$, $df = 46$, $p = < 0,001$).

Če opazujemo odgovore 25 udeležencev MPŠ, ki so sodelovali v obeh testiranjih, vidimo, da so pri drugem testiranju nekoliko bolje odgovarjali glede poznavanja v slovenščini obstoječih besed ($M = 3108 : 3296$, $Mdn = 3450 : 3750$), pri prepoznavanju nebesed pa so v povprečju dosegli malenkost slabši rezultat ($M = -320 : -440$, $Mdn = 0 : -250$) (Tabela 4). S t-testom za odvisne vzorce smo pokazali, da za te testirane obstaja statistično pomembna razlika med T1 in T2 pri skupnih točkah ($t(24) = 2,744$, $p = 0,011$) in rezultatih za besede 1–1000 ($t(24) = 2,113$, $p = 0,045$), 1001–2000 ($t(24) = 2,511$, $p = 0,019$), 3001–4000 ($t(24) = 2,343$, $p = 0,028$), za besede iz frekvenčnega ranga 2001–3000 pa razlika ni bila statistično značilna ($t(24) = 1,67$, $p = 0,108$). Prav tako razlika ni bila statistično značilna za korigirane točke ($t(24) = 0,583$, $p = 0,565$).

Tabela 4: Rezultati testiranih, ki so sodelovali pri T1 in T2.

	<i>T1</i>	<i>T1</i>	<i>T1</i>	<i>T2</i>	<i>T2</i>	<i>T2</i>
	<i>skupaj</i>	<i>nebesede</i>	<i>korigirane</i>	<i>skupaj</i>	<i>nebesede</i>	<i>korigirane</i>
	<i>točk</i>	<i>(odbitek)</i>	<i>točke</i>	<i>točk</i>	<i>(odbitek)</i>	<i>točke</i>
N	25	25	25	25	25	25
M	3108	–320	2788	3296	–440	2856
Mdn	3450	0	3050	3750	–250	3050
SD	883	547	998	828	634	917
Minimum	450	–2250	450	800	–2500	800
Maksimum	3950	0	3950	4000	0	3850

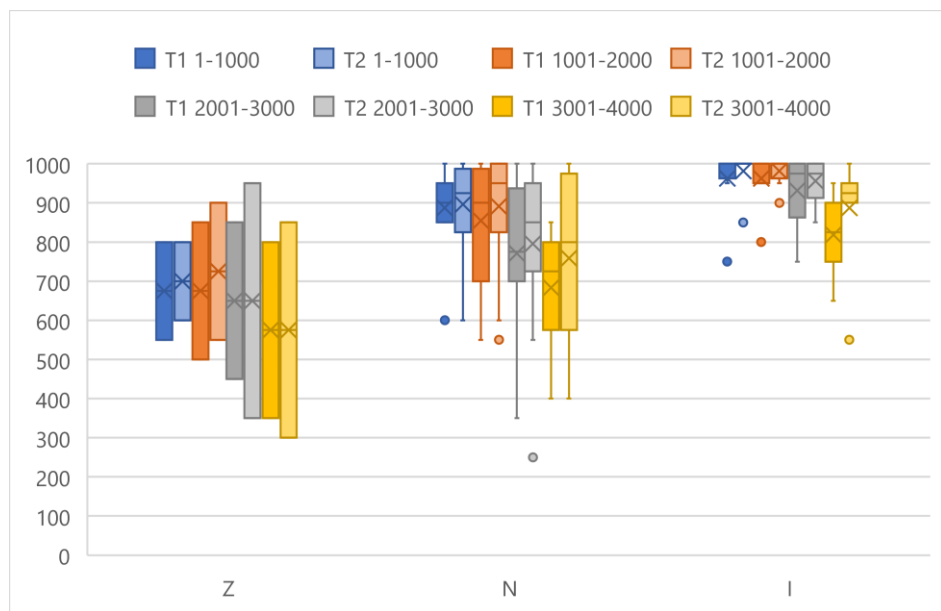
Ko smo pri testiranih, ki so sodelovali tako pri T1 kot pri T2, opazovali rezultate glede na njihovo splošno jezikovno znanje (Z, N, I), smo za posameznih tisoč besed opazili, da so rezultati višji ali podobni (Tabela 5, Graf 4). Za skupini N in I smo izvedli t-test za odvisne vzorce.¹¹ Niti za korigirane točke niti za posameznih tisoč besed nismo ugotovili, da bi bile vrednosti statistično značilno drugačne (p je povsod $> 0,05$). Izjema je bilo skupno število točk pri skupini I, ki se je statistično značilno spremenilo ($p = 0,013$).

¹¹ Za skupino Z t-testa nismo izvedli, saj sta test reševali le 2 osebi, katerih jezikovno znanje je bilo ob T1 označeno kot začetno.

Tabela 5: Rezultati za testirane, ki so sodelovali tako pri T1 kot pri T2, glede na splošno jezikovno znanje.

	<i>poenostavljena oznaka stopnje</i>	<i>T1 skupne točke</i>	<i>T2 skupne točke</i>	<i>T1 korigirane točke</i>	<i>T2 korigirane točke</i>
N	Z	2	2	2	2
	N	12	12	12	12
	I	8	8	8	8
M	Z	2575	2650	1450	1900
	N	3196	3342	2904	2904
	I	3675	3806	3581	3556
SD	Z	1025	1202	566	141
	N	539	672	621	741
	I	323	272	384	309

Graf 4: Število doseženih točk za posameznih tisoč besed za testirane, ki so sodelovali tako pri T1 kot pri T2, glede na splošno jezikovno znanje.



6 DISKUSIJA

V obeh testiranjih se je potrdilo, da v splošnem govorci SDTJ prej v procesu učenja in v večji meri poznajo frekventnejše besede. Pri udeležencih MPŠ, ki smo jih opazovali, z uporabljenim testom nismo ugotovili večje razlike v

poznavanju prvih in drugih tisoč besed, za nadaljnjih dvakrat tisoč besed pa se je pokazal padajoči profil. Tako se je hipoteza 1 le delno potrdila.

Podobni profili za posameznih tisoč besed pri T1 in T2 so se pokazali za vse testirane skupaj (Grafa 1 in 2), prav tako pa tudi za skupine z različnim jezikovnim znanjem – Z, N in I (Graf 4), ki pa so se po pričakovanjih med seboj razlikovale v doseženem rezultatu: začetniki so dosegli nižji rezultat kot nadaljevalci in ti nižjega kot izpopolnjevalci (hipoteza 2). Ob primerjavi rezultatov testa poznavanja besedišča pri T1 in rezultatov, doseženih na uvrstitvenem testu, se je pokazalo, da je tudi v naši raziskavi korelacija med številom točk in rezultatom, doseženim pri T1, visoka ($r = 0,78$ za skupne točke; $r = 0,732$ za korigirane točke), kar je skladno z navedbami Meare in Miralpeix (2016, str. 129) za druge podobne raziskave. To pomeni, da bi lahko le na podlagi testa poznavanja najpogostejših 4000 splošnih besed v slovenščini podobno ustrezno razvrstili udeležence MPŠ v skupine, ki bi bile po jezikovnem znanju homogene (hipoteza 2.1).

Če bi za razvrščanje tečajnikov uporabili predstavljeni test poznavanja splošnih besed, bi ga bilo za natančnejše razporejanje v skupine verjetno smiselno dopolniti z nalogo za samostojno produkcijo, ki bi dala vpogled v posameznikovo jezikovno zmožnost bolj celostno (in s tem učitelju omogočila načrtovanje pouka, ki bo po meri učencev), tako kot so bile naloge za preverjanje bralnega razumevanja in oblikoskladenjske zmožnosti v uvrstitvenem testu za MPŠ 2022 dopolnjene z nalogo za pisanje daljšega besedila. Iz predstavljenega testa poznavanja besedišča namreč dobimo informacijo samo o omejenem delu jezikovne zmožnosti (prim. Read, 2000, str. 3–7).

Rezultati testiranj kažejo, da so testirani na različnih stopnjah jezikovnega znanja pri T2 poznali več besed, vendar razlike z izjemo skupnega števila točk pri skupini I niso bile statistično značilne. Tako hipotezo 3 lahko potrdimo le deloma. Zdi se, da je 40 ur pouka premalo, da bi lahko govorili o velikem napredku pri poznavanju besedišča. Če pogledamo razlike med povprečji za skupne točke za skupine Z, N in I pa se nam hipoteza, da bo do največjega napredka prišlo pri testiranih z nižjim rezultatom na T1 (hipoteza 3.1), ni potrdila: večji napredek se kaže pri testiranih, ki so v slovenščini nadaljevalci in izpopolnjevalci (razlike med povprečji skupnih točk za T1 in T2: Z – 75 točk,

N – 146 točk in I – 131 točk). Verjetno bi bila slika drugačna, če bi napredek v poznavanju besedišča merili pri popolnih začetnikih.

7 SKLEP

Predstavili smo test poznavanja splošnega besedišča, ki poznavanje besed preverja diskretno, izbirno in neodvisno od konteksta (gl. Read, 2000, str. 7–13). S testom smo med udeleženci 17. Mladinske poletne šole slovenščine preverjali, koliko besed od 4000 najpogostejših splošnih besed v slovenščini poznajo. Rezultati kažejo, da testirani v večji meri poznajo bolj frekventne besede kot manj frekventne: po drugih tisoč besedah se pokaže tipičen stopničasto spuščajoč se profil. Pilotna izvedba testa je pokazala, da bi ga lahko na MPŠ uporabili kot alternativo uvrstitvenemu testu za razvrščanje tečajnikov v skupine, ki bi bile homogene po jezikovnem znanju. Ali bi ga lahko uporabili tudi na drugih tečajih SDTJ, bi bilo treba še preveriti.

Pri interpretaciji rezultatov glede obsega in globine poznavanja besedišča moramo biti previdni. Test preverja le najosnovnejše poznavanje besede, in sicer ali določeno kombinacijo črk testirani prepozna kot besedo (prim. Meara, 1994, str. 6). O tem, v kolikšni meri besedo receptivno ali produktivno zna (na ravni oblike, pomena in rabe, prim. Nation, 2022, str. 54), tako ne moremo sklepati. Na podlagi rezultatov izvedenega testa tudi ne moremo z gotovostjo trditi, da testirani – tudi zaradi možnosti ugibanja in subjektivnosti pri odločanju, kaj testiranim pomeni, da *vejo*, kaj beseda pomeni – dejansko poznajo tolikšen del testiranih besed, kot jih pokažejo dosežene točke, t. i. *da/ne testi* namreč v večini primerov precenjujejo poznavanje besedišča pri testiranih (Durrant in sod., 2022, str. 178), na kar sta opozorila že Meara in Jones (1988). Opozoriti velja tudi, da so nekateri testirani kot govorci slovenščini sorodnih jezikov lahko zaradi podobnosti besed (napačno) sklepali, da poznajo pomen besed v slovenščini.¹² Uporabljeni test zajema v primerjavi z drugimi testi besedišča, ki segajo tudi do 10.000 najfrekventnejših besed ali dlje, iz sorazmerno majhnega seznama 4000 lem. Domnevamo, da so testirani poznali tudi manj frekventne besede, ki se niso uvrstile na

¹² Na podoben fenomen opozarja Meara (1992, str. 11) pri govorcih romanskih in germanskih jezikov, ki si z angleščino delijo del besedišča.

Referenčni seznam pogostih splošnih besed za slovenščino, in tudi zato bi o obsegu slovarja testiranih lahko postavili le zelo približne ocene.¹³

Če bi želeli dobiti natančnejši vpogled v obseg besedišča za posamezne stopnje jezikovnega znanja po SEJO, bi morali pripraviti test drugačnega formata in vanj zajeti tudi manj frekventne besede. Za večjo zanesljivost testa pa bi morali vključiti tudi večje število besed.¹⁴

8 ZAHVALA

Zahvaljujem se vodjema Mladinske poletne šole slovenščine dr. Mihaeli Knez in dr. Damjani Kern Andoljšek ter vsem učiteljicam in učencem, ki so mi omogočili izvedbo testa. Za nasvete pri obdelavi rezultatov se zahvaljujem prof. dr. Anji Podlesek. Za pomoč se zahvaljujem mentorici znan. sod. dr. Špeli Arhar Holdt. Zahvaljujem se tudi recenzentom za koristne pripombe.

9 LITERATURA

Arhar Holdt, Š., Pollak, S., Robnik Šikonja, M. in Krek, S. (2020). Referenčni seznam pogostih splošnih besed za slovenščino. V D. Fišer in T. Erjavec (Ur.), *Zbornik konference Jezikovne tehnologije in digitalna humanistika* (str. 10–15). Inštitut za novejšo zgodovino. Pridobljeno 12. januarja 2021, https://nl.ijs.si/jtdh20/pdf/JT-DH_2020_Arhar-Holdt-et-al_Referencni-seznam-pogostih-splosnih-besed-za-slovenscino.pdf

Durrant, P., Siyanova-Chanturia, A., Kremmel, B. in Sonbul, S. (2022). *Research Methods in Vocabulary Studies*. John Benjamins Publishing Company.

Kremmel, B. in Schmitt, N. (2017). Vocabulary Levels Test. V *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons. <https://doi.org/10.1002/9781118784235.eelt0499>

Meara, P. (1992). *EFL vocabulary tests*. University College Swansea, Centre for Applied Language Studies. Pridobljeno 22. maja 2022, <https://www.lognostics.co.uk/vlibrary/meara1992z.pdf>

¹³ Na podlagi rezultatov tega testa lahko zelo grobo ocenimo, da začetniki (ki pa niso popolni začetniki v slovenščini), nadaljevalci in izpolnjevalci, ki smo jih opazovali, v slovenščini poznajo okoli 2000 najpogostejših splošnih besed, okoli 3000 oz. okoli 3500 besed od 4000 najpogostejših splošnih besed v slovenščini.

¹⁴ Po 30 besed oz. 3 % odstotke besed za posameznih tisoč (prim. Durrant in sod., 2022, str. 158).

- Meara, P. (1994). The complexities of simple vocabulary tests. V F. G. Brinkman, J. A. van der Schree in M. C. Schouten-van Parreren (Ur.), *Curriculum research: different disciplines and common goals* (str. 15–28). Vrije Universiteit. Pridobljeno 4. maja 2024, <https://www.lognostics.co.uk/vlibrary/meara1994a.pdf>
- Meara, P. in Jones, G. (1988). Vocabulary size as a placement indicator. V P. Grunwell (Ur.), *Applied Linguistics in Society* (str. 80–87). CILT.
- Meara, P. in Miralpeix, I. (2016). *Tools for Researching Vocabulary*. Multilingual Matters. <https://doi.org/10.21832/9781783096473>
- Milton, J. in Alexiou, T. (2010). Developing a vocabulary size test in Greek as a foreign language. V A. Psaltou – Joyce in M. Mattheoudakis (Ur.), *Advances in Research on Language Acquisition and Teaching: Selected Papers* (307–318). Greek Applied Linguistics Association. Pridobljeno 22. maja 2022, <https://www.enl.auth.gr/gala/14th/Papers/English%20papers/Milton&Alexiou.pdf>
- Nation, P. (2012). *The Vocabulary Size Test*. Pridobljeno 31. januarja 2024, <https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-tests/the-vocabulary-size-test/Vocabulary-Size-Test-information-and-specifications.pdf>
- Nation, I. S. P. (2022). *Learning Vocabulary in Another Language. Third edition*. Cambridge University Press.
- Pollak, S., Arhar Holdt, Š., Krek, S. in Robnik-Šikonja, M. (2020). *Reference List of Slovene Frequent Common Words*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1346>
- Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.
- Zhang, X. (2013). The I don't know option in the vocabulary size test. *TESOL Quarterly*, 47(4), 790–811. doi:10.1002/tesq.98

TEST OF KNOWLEDGE OF COMMON WORDS IN SLOVENE AMONG PARTICIPANTS OF THE YOUTH SUMMER SCHOOL OF SLOVENE

This paper presents a pilot test assessing the knowledge of the common general words in Slovene. The aim of the test was to find out to what extent speakers of Slovene as a second and foreign language know general words in Slovene and what progress the test participants made in their knowledge of general vocabulary after attending a summer school of Slovene. We were interested in whether such a test could be used as an alternative to the more time-consuming placement tests, in order to classify participants of Slovene as a second and foreign language courses into groups homogeneous in terms of language proficiency. Following the example of the tests for English and Greek as FL, we developed a yes/no test of the knowledge of the most frequent 4000 common words in Slovene, which was administered in two iterations to the participants of the 17th Youth Summer School of Slovene.

Keywords: vocabulary, test, Slovene as a second and foreign language, student paper

To delo je ponujeno pod licenco Creative Commons: Priznanje avtorstva-Deljenje pod enakimi pogoji 4.0 Mednarodna.

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International.

<https://creativecommons.org/licenses/by-sa/4.0/>

