

## **Poročilo o anketi o potrebah raziskovalne skupnosti po podatkih o aktualni jezikovni rabi**

*Projekt SLED - Spremljevalni korpus in spremljajoči jezikovni viri*

<http://sled.ijs.si/>

Delovni sklop 1, aktivnost 1.2: Priprava postopkov za izdelavo različnih podatkovnih zbirk z informacijami o besedah in besednih zvezah v sodobni slovenščini

Ljubljana, 24. januar 2022

### **1 Uvod**

V okviru projekta *Spremljevalni korpus in spremljajoči jezikovni viri* (SLED), ki ga od 1. 9. 2021 od 30. 10. 2022 financira Ministrstvo za kulturo Republike Slovenije, smo v decembru 2021 izvedli uporabniško anketo, s katero smo v skladu s cilji projektne aktivnosti 1.2 (Priprava postopkov za izdelavo različnih podatkovnih zbirk z informacijami o besedah in besednih zvezah v sodobni slovenščini) želeli preveriti, katerih podatkov o aktualni rabi jezika si raziskovalna skupnost želi in v kakšni obliki (različni sezname, kot so npr. kandidati za neologizme, besede in besedne zveze z najbolj izstopajočo rabo v določenem obdobju (dnevu, tednu, mesecu), izstopajoče besede in besedne zveze glede na vir ipd.).

V poročilu predstavljamo anketo in povzemamo pogloblitve ugotovitve.

### **2 O anketi**

Anketa je bila izdelana na platformi 1KA ([www.1ka.si](http://www.1ka.si)), sestavljena pa je bila iz 9 vprašanj: med temi je bilo 5 vsebinskih, 4 pa so zbirala demografske podatke (spol, starost, področje delovanja).

Anketa je bila diseminirana po e-poštnih seznamih slovenskih jezikoslovnih raziskovalnih skupnosti (npr. SlovLit ter e-poštni seznam Slovenskega društva za jezikovne tehnologije) ter po družbenem omrežju Facebook (na uradni strani Centra za jezikovne vire in tehnologije Univerze v Ljubljani ter v neformalnih jezikoslovnih uporabniških skupinah, kot je *Prevajalci, na pomoč!*).

Anketa je bila zaključena 24. januarja 2022, zbrala pa je 100 odgovorov (tj. 100 v celoti izpolnjenih vprašalnikov). Vsebinska vprašanja so bila naslednja:

1. Kateri podatki bi vas pri orodju, ki spremlja jezikovno rabo, najbolj zanimali? Spodaj so navedena vprašanja (skupaj s primeri v oglatih oklepajih za lažjo predstavbo) - označite, koliko bi vas po vaši oceni odgovori na vprašanja zanimali (oz. bi vam koristili pri delu).
2. Kateri način prikaza podatkov se vam zdi pomemben?
3. Ali ste mnenja, da bi vam orodje s podatki o aktualni jezikovni rabi koristilo pri delu?

4. Če imate dodatne predloge oz. scenarije, kaj bi vas zanimalo o aktualni jezikovni rabi, jih zapišite v spodnji okvir.
5. Ali ste pred to anketo že slišali za CJVT-jev Jezikovni sledilnik? (<https://viri.cjvt.si/sledilnik>)

### 3 Zajeti vzorec populacije

Vzorec, ki ga je zajela anketa, zajema predvsem osebe ženskega spola (82 %, Tabela 1), manjši delež pa je moških (18 %).

Spol	Frekvenca	Delež
Ženski	82	82 %
Moški	18	18 %
Drugo/nebinarno	0	0 %
Ne želim odgovoriti	0	0 %

**Tabela 1:** Udeleženci\_ke ankete po spolu.

Po starosti (Tabela 2) vzorec zajema predvsem generacije med 26. in 55. letom starosti (80 % vseh udeleženk\_cev).

Starostna skupina	Frekvenca	Delež
Do 15 let	0	0 %
Od 16 do 25 let	3	3 %
Od 26 do 35 let	33	33 %
Od 36 do 45 let	15	15 %
Od 46 do 55 let	32	32 %
Od 56 do 65 let	12	12 %
66 let ali več	5	5 %

**Tabela 2:** Udeleženci\_ke ankete po starostnih skupinah.

Večina udeleženk\_cev je zaposlenih bodisi v javnem sektorju (61 %, Tabela 3) bodisi je samozaposlena (20 %). Le manjši delež ima še študentski status (3 %), nekaj pa jih je zaposlenih v podjetjih (6 %), upokojenih (4 %) ali v iskanju zaposlitve (5 %).

Status	Frekvenca	Delež
Dijak_inja	0	0 %
Študent_ka	3	3 %

Samozaposlen_a	20	20 %
Zaposlen_a v javnem sektorju (uprava, šolstvo, zdravstvo, sociala, kultura ...)	61	61 %
(Zaposlen_a v neprofitnem sektorju (društva, združenja ...))	0	0 %
Zaposlen_a v podjetju	6	6 %
Vzdrževanje gospodinjstva	0	0 %
Upokojen_a	4	4 %
V iskanju zaposlitve	5	5 %
Na porodniškem ali starševskem dopustu	0	0 %
Drugo [Nič od naštetega]	1	1 %

**Tabela 3:** Udeleženci\_ke ankete po statusu oziroma glavnem področju angažiranja.

Po področju delovanja (Tabela 4), pri katerem so udeleženci\_ke lahko izbrali\_e več možnosti, prednjačita lektoriranje (60 %) in prevajanje (46 %), visok delež pa imajo tudi ljubiteljsko raziskovanje jezika (38 %), strokovno in znanstveno pisanje (34 %), jezikoslovne raziskave (32 %) ter kreativno pisanje in blogerstvo (22 %). Skupno 40 % zajemajo tudi različne kategorije poučevanja jezika (slovenščina kot 1. jezik na osnovni ali srednji šoli, slovenščina kot 2. ali tuji jezik, jezikoslovni predmeti na višji/univerzitetni ravni). Vzorec nakazuje, da je anketa dobro zajela različna področja jezikoslovno-raziskovalnega udejstvovanja.

<b>Področje delovanja</b>	<b>Frekvenca</b>	<b>Delež</b>
Lektoriranje	60	60 %
Prevajanje	46	46 %
Poučevanje slovenščine kot 1. jezika - osnovna šola	8	8 %
Poučevanje slovenščine kot 1. jezika - srednja šola	7	7 %
Poučevanje slovenščine kot 2. ali tujega jezika	11	11 %
Predavanje jezikoslovnih predmetov na višji/univerzitetni ravni	14	14 %
Novinarstvo	8	8 %
Marketing	9	9 %
Pravo	7	7 %
Bibliotekarstvo	3	3 %

Javna uprava	14	14 %
Administracija	13	13 %
Menedžment	2	2 %
Beletristika	8	8 %
Kreativno pisanje, blogerstvo	22	22 %
Strokovno in znanstveno pisanje	34	34 %
Jezikoslovne raziskave	32	32 %
Leksikografija	16	16 %
Ljubiteljsko raziskovanje jezika	38	38 %
Drugo [Oblikovanje publikacij]	1	1 %
Drugo [Jezikovna politika]	1	1 %
Drugo [Poučevanje strokovnega predmeta risanje in slikanje]	1	1 %
Drugo [Poučevanje psihologije]	1	1 %
Drugo [Poučevanje matematike v SŠ]	1	1 %
Drugo [Folkloristika, etnolingvistika]	1	1 %

**Tabela 4:** Udeleženci\_ke ankete po področju delovanja (izbrali so lahko več možnosti).

#### 4 Vsebinska vprašanja

Iz podatkov v Tabeli 5 je razvidno, da se anketirancem\_kam vsi predlagani scenariji zdijo zanimivi: kategoriji "Zanima me" in "Zelo me zanima." namreč pri vsakem podvprašanju zajemata med 74 in 88 %. Po stopnji zanimanja najbolj izstopa zadnji scenarij, v katerem je mogoče primerjati trend rabe dveh ali več besed/besednih zvez (npr. *anticepilec*, *proticepilec*). Podobno jih zanima tudi, ali raba določene besede/besedne zveze v zadnjem obdobju glede na trende narašča ali pada.

Podvprašanje	Sploh me ne zanima.	Me ne zanima.	Niti - niti.	Zanima me.	Zelo me zanima.	Ne vem.
Katere besede/besedne zveze so najznačilnejše za določeno obdobje v primerjavi z drugim obdobjem? [Npr. katere besede so se mnogo pogosteje uporabljale v februarju 2020 kot pa v februarju 2021?]	2 %	8 %	12 %	<b>48 %</b>	30 %	0 %

V katerem obdobju je določena beseda/besedna zveza najpogostejša? [Npr. ali je bila beseda "tajkun" res najpogostejša v obdobju 2008-2009?]	5 %	6 %	14 %	<b>44 %</b>	30 %	1 %
Ali raba besede/besedne zveze v zadnjem obdobju glede na trende narašča ali pada? [Npr. ali se "epidemija" uporablja vse pogosteje ali vse redkeje?]	2 %	5 %	5 %	42 %	<b>46 %</b>	0 %
V katerih besedilih (po tematiki) je beseda/besedna zveza pogostejša? [Npr. ali je "napadalec" res najpogostejši v besedilih o športu?]	1 %	5 %	9 %	37 %	<b>46 %</b>	2 %
Kakšna je razporeditev besede/besedne zveze po tematikah v določenem obdobju? [Npr. ali se besedna zveza "kolektivna imunost" pojavlja le v besedilih o medicini ali pa jo najdemo tudi drugod?]	2 %	4 %	12 %	<b>51 %</b>	30 %	1 %
Katera od dveh (ali več) besed/besednih zvez se v zadnjem obdobju pojavlja pogosteje? [Npr. ali v zadnjem obdobju v besedilih pogosteje najdemo "proticepilec" ali "anticepilec"?]	1 %	5 %	6 %	34 %	<b>53 %</b>	1 %

**Tabela 5:** Kateri podatki bi vas pri orodju, ki spremlja jezikovno rabo, najbolj zanimali? Spodaj so navedena vprašanja (skupaj s primeri v oglatih oklepajih za lažjo predstavo) - označite, koliko bi vas po vaši oceni odgovori na vprašanja zanimali (oz. bi vam koristili pri delu).

Dobre tri četrtine vprašanih je odgovorilo tudi, da bi jim podatki o aktualni jezikovni rabi koristili pri delu (kot prikazuje Tabela 6). Rezultati ankete torej potrjujejo, da jezikoslovno skupnost podatki o trendih jezikovne rabe zanimajo in da obstaja realna potreba po jezikovnem viru, ki tovrstne podatke prinaša sprotno in ažurno.

Odgovor	Frekvenca	Delež
Da	78	76 %
Ne	9	9 %

Ne vem	15	15 %
--------	----	------

**Tabela 6:** Ali ste mnenja, da bi vam orodje s podatki o aktualni jezikovni rabi koristilo pri delu?

V Tabeli 7 je prikazano, v kolikšni meri se anketirancem\_kam predlagani prikazi jezikovnih podatkov zdijo pomembni. Če združimo deleže kategorij "Pomembno" in "Zelo pomembno", dobimo deleže 79 % za grafe, 64 % za tabele s številskimi podatki in 87 % za sezname besed s padajočo/naraščajočo rabo. Anketiranke\_ce torej najbolj zanimajo preprosti sezname, najmanj pa napredne tabele s številskimi podatki.

Prikaz	Sploh ni pomembno.	Ni pomembno.	Niti - niti.	Pomembno.	Zelo pomembno.	Ne vem.
Grafi, ki prikazujejo trende jezikovne rabe.	1 %	4 %	16 %	<b>47 %</b>	32 %	0 %
Tabele s številskimi podatki.	0 %	6 %	29 %	<b>39 %</b>	25 %	1 %
Preprost seznam besed oz. besednih zvez z naraščajočo /padajočo rabo.	0 %	2 %	10 %	<b>50 %</b>	37 %	1 %
Drugo [npr. Opis, morda kombiniran s številskimi, Besedni grafi, ki prikazujejo kolokacijsko odvisnost med besedami]	6 %	0 %	6 %	12 %	6 %	<b>71 %</b>

**Tabela 7:** Kateri način prikaza podatkov se vam zdi pomemben?

Na voljo je bilo tudi vprašanje odprtega tipa: "Če imate dodatne predloge oz. scenarije, kaj bi vas zanimalo o aktualni jezikovni rabi, jih zapišite v spodnji okvir." V nadaljevanju navajamo predloge, razdeljene v kategorije.

### **Povezljivost z drugimi jezikovnimi viri in dostop do podatkov**

Za jezikovnotehnološko in računalniško skupnost bi bilo smiselno urediti dostop do podatkov o aktualni jezikovni rabi s pomočjo javnega API-ja, preko katerega lahko uporabnice\_ki v realnem času dostopajo do ažurnih podatkov.

"javni api za dostop do podatkov."

Glede na predlog iz ankete bi bilo smiselno razmisliti tudi o integraciji jezikovnega vira s podatki o aktualni jezikovni rabi v druge jezikovne vire, npr. v Slovenski oblikoslovni leksikon Sloleks in v korpus pisne standardne slovenščine Gigafida.

"vključitev trendov v besano (nimam pojma, ali je sploh izvedljivo)."

### **Primerjava sopomenskih različic besed oz. besednih zvez**

Anketirance\_ke zanima primerjava trendov aktualne jezikovne rabe zlasti za sopomenske različice: bodisi za pare tujka/domača beseda ali za sopomenke nasploh. Razmisliti je torej treba, na kakšen način lahko podatke o jezikovni rabi povežemo s podatki iz Slovarja sopomenk sodobne slovenščine.

"kolikšen delež tujk uspemo posloveniti in kako pogosto je rabljena slovenska različica v primerjavo z angleško, npr mail namesto e-pošta v pogovornem jeziku."

"primerjava določene besede/besedne zveze z njenimi sopomenkami. npr. uporaba oče vs. ata v danem okviru (obdobje, tematika...)"

### **Vključitev zgledov rabe**

Poskrbeti je treba, da imajo uporabniki\_ce poleg podatkov o trendih na voljo tudi zglede rabe, npr. s povezavo na korpus ali celo primer članka, v katerem je bila beseda zaznana.

"meni bi prišlo prav, če bi bilo možno na seznamu besed/besednih zvez videti tudi primer kakšne povedi, v kateri je bil ta izraz uporabljen (če to še ni predvideno). npr. s klikom ali pomikom miške čez besedo."

### **Primerjava po registru in pomenu**

Sledenje spremembam v registru ali pomenu sicer presega cilje projekta SLED, a odzivi anketirancev\_k kažejo, da velja razmisliti, kako implementirati spremljanje sprememb pomena skozi čas oz. stilne zaznamovanosti besed.

"kako se stroka izreka o modnih besedah in besednih zvezah, ki so napačne ali vsaj sporne (temu ni tako, naslavljeni probleme, uporabniki\_ce ...)?"

"sprememba stilne zaznamovanosti skozi čas"

"pomenski opis časovnega razvoja besed in besednih zvez"

### **Primerjava po slovničnih kategorijah**

Anketirance\_ke zanima tudi spremljanje trendov rabe ne le konkretnih besed, temveč tudi njihovih slovničnih struktur. Razmisliti je torej treba, ali orodje lahko omogoči tudi npr. spremljanje n-gramov oblikoskladenjskih oznak.

"dobrodošla bi bila možnost spremljanja skozi čas in žanre (vsako posebej in oboje skupaj) za celo kategorijo besed (npr. deležja, posamostaljenja ipd.). bistveno je, da je korpus žanrsko dovolj natančno opredeljen."

### **Spremljanje daljših enot**

Pri večbesednih enotah smo se pri sledenju jezikovni rabi do zdaj osredotočali predvsem na kolokacije in besedne n-grame, veljalo pa bi razmisliti, ali je mogoče spremljanje učinkovito razširiti tudi npr. na frazeme.

"če bi bilo možno, da je iskalni besednih zvez razširjen tudi na dolžino pregovorov, torej da je možno potem tudi iskati "daljše" besednozvezne tvorbe."

### **Spremljanje izgovora in oblike**

Pojavil se je tudi predlog po spremljanju izgovorjevalnih različic (npr. áneks, anéks), a to v pisnih korpusih ni izvedljivo. Pri pisnih različicah besed pa je mogoče spremljati npr. variantne oblike (npr. roditelj *hčer*, *hčera*).

"kakšne vse izgovorjevalne in pisne različice besed se pojavljajo in kako so pogoste."

## **5 Zaključek**

Rezultati ankete so pokazali, da se večina želja uporabnikov sklada z načrti projekta, dober odziv pa potrjuje potrebo po tovrstnih podatkih in interes javnosti. V prihodnje bi veljalo razmisliti predvsem o načinu vključitve spremljevalnega korpusa oz. podatkov o trendih jezikovne rabe v širšo sliko jezikovnih orodij.

*Projekt financira Ministrstvo za kulturo Republike Slovenije kot del Javnega razpisa za (so)financiranje projektov, namenjenih gradnji in posodabljanju infrastrukture za slovenski jezik v digitalnem okolju 2021–2022.*