

NAVODILA ZA SNEMANJE ZA PORTAL GOVORJENA SLOVENŠČINA

Darinka Verdonik, Andreja Bizjak, Gregor Donaj, Boštjan Makarovič, Cristina Contero Almagro

Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru

Januar 2025

Vsebina

1	Uvod	2
2	Uporaba posnetkov	3
2.1	Uporabniki	3
2.2	Anonimnost govorcev	3
2.3	Obdelava posnetkov	3
2.4	Dostopnost posnetkov	3
2.5	Izjava o varstvu osebnih podatkov za portal Govorjena slovenščina (Člen 13 in 14 GDPR) ..	4
3	Kako snemati	6
3.1	KOGA – kogarkoli, ki govori slovensko	6
3.2	KJE – v tistem prostoru (doma)	6
3.3	KAJ - vsakdanji govor	6
3.3.1	Družabni pogovor	6
3.3.2	Intervju	7
3.3.3	Diskusija	7
3.3.4	Navodila	7
3.3.5	Pripoved	7
3.3.6	Razlaga	7
3.3.7	Namigi za teme govora	8
3.4	S ČIM – avdio ali video, visoka kakovost posnetka	8
3.4.1	Dodatna navodila za snemanje	8
3.5	KOLIKO – med 5 in 90 minut	10
4	O nas	11
4.1	Povezani sedanji in pretekli projekti	11
5	Pogoji sodelovanja	12
6	Literatura	14
	Priloga 1: Privolitev za objavo posnetka mojega govora in z njim povezanih osebnih podatkov na portalu Govorjena slovenščina	15
	Priloga 2: Okoliščine snemanja	16

1 Uvod

Želiš ohraniti posnetke svojega narečja kot del kulturne dediščine za prihodnje generacije?

Želiš prispevati k razvoju orodij, kot sta razpoznavanje govora ali strojno prevajanje, tudi za slovenščino in njena narečja?

Snemaš govor za lastno raziskavo ali projekt in potrebuješ pravno ali tehnično podporo?

Te zanimajo razlike v govoru ljudi iz različnih krajev, opazuješ posebnosti in jih želiš bolje razumeti?

Pridruži se nam in postani del naših prizadevanj kot občanski znanstvenik, znanstvenica!

Slovenščino govorimo na različne načine – z različnimi naglasi, besedami in izrazi. Čeprav o tem pogosto razpravljamo, v resnici vemo zelo malo, saj nimamo dovolj zanesljivih podatkov. Za pisno slovenščino obstajajo obsežne zbirke besedil s spleta, medijev in drugih virov, medtem ko je posnetkov vsakdanjega govora premalo, da bi zajeli vso raznolikost govorne slovenščine in ohranili podatke o njej tudi za prihodnje rodove. Ta portal je namenjen zbiranju posnetkov vsakdanjega govora, da bi bolje spoznali pestrost svojega jezika in podprli razvoj digitalnih orodij zanj.

Posnetki bodo uporabljeni za raziskave in razvoj. Glavni uporabniki bodo raziskovalci, ki preučujejo slovenski govorni jezik, in razvijalci jezikovnih tehnologij za slovenščino.

Občanski znanstveniki, znanstvenice, ki oddajo skupaj 150 minut posnetkov ustrezne kakovosti, prejmejo praktično nagrado kot zahvalo za sodelovanje.

2 Uporaba posnetkov

2.1 Uporabniki

Uporabniki posnetkov bodo večinoma raziskovalci, ki preučujejo slovenski govorni jezik v jezikoslovju ali sorodnih družboslovnih in tehničnih vedah.

Uporabniki bodo tudi razvijalci jezikovnih in govornih tehnologij, kot so razpoznavanje govora, strojno razumevanje pomena govora in strojno prevajanje govora.

Uporabniki bodo lahko tudi vsi ostali, ki bi želeli posnetke uporabiti za svoje raziskave ali razvoj tehnologij za slovenski jezik.

2.2 Anonimnost govorcev

Govorci bodo označeni z anonimno kodo, ki ne bo razkrivala njihove identitete. Podatek o imenu in priimku govorca in kontaktni podatki ne bodo javno objavljeni skupaj s posnetkom, ampak bodo dostopni samo urednikom zbirke za potrebe upravljanja baze oziroma pravic posameznikov in nagradnega tekmovanja.

Posnetki in zapisi bodo povezani s podatki o okoliščinah snemanja in anonimnimi kodami govorcev.

2.3 Obdelava posnetkov

Posnetki bodo del govorne baze slovenščine. K posnetkom bodo lahko dodani zapisi govora, ki ga slišimo na posnetkih.

Zapisi govora bodo lahko dodatno označeni s podatki o izgovoru, oblikah besed, stavčnih strukturah, pomenu, čustvenem tonu in podobno.

2.4 Dostopnost posnetkov

Posnetki iz govorne baze slovenščine bodo dostopni prek nacionalnega repozitorija za jezikovne vire in tehnologije Clarin.si. Bazo bo mogoče prenesti pod eno od odprtih licenc, kot je Creative Commons.

Dostop do posameznih izjav v posnetkih bo mogoč tudi prek specializiranih iskalnikov za jezikovne vire, imenovanih konkordančniki. Primer takšnega dostopa je govorni korpus Gos, ki je dostopen prek konkordančnikov CJVT, Clarin noSketch Engine ali Clarin Kontext.

2.5 Izjava o varstvu osebnih podatkov za portal Govorjena slovenščina (Člen 13 in 14 GDPR)

Upravljaivec osebnih podatkov na portalu Govorjena slovenščina je Inštitut za elektroniko in telekomunikacije, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru (UM FERl), Koroška cesta 46, 2000 Maribor.

Kontakt pooblaščenice osebe za varstvo podatkov je na voljo tukaj: <https://www.um.si/o-univerzi/dokumentno-sredisce/varstvo-osebni-podatkov/>

Portal Govorjena slovenščina je namenjen zbiranju posnetkov vsakdanje slovenščine in njihovi javni objavi z namenom uporabe za raziskave govora in razvoj digitalnih orodij za slovenski jezik ter kot arhiv jezikovne kulturne dediščine.

UM FERl obdeluje naslednje osebne podatke uporabnikov portala:

- ime, priimek in email registriranega uporabnika, ki naloži posnetek na portal
- kontaktni e-poštni naslov in druge kontaktne podatke
- avtomatsko določen ID vsakega oddanega posnetka
- število govorcev na posnetku
- kraj in regija snemanja
- prostor, kjer so bili govorci (npr. doma v dnevni sobi...)
- žanr in tip govora (npr. družabni pogovor)
- datum snemanja
- kanal govora
- podatki o napravi, s katero je bilo izvedeno snemanje
- opis situacije, v kateri je nastal posnetek
- ključne besede, ki so lahko v pomoč pri iskanju, opisu ali kategorizaciji
- zapis govora.

Od navedenih podatkov podatek o imenu in priimku in kontaktni podatki ne bodo javno objavljeni skupaj s posnetkom, ampak bodo dostopni samo urednikom zbirke za potrebe upravljanja baze oziroma pravic posameznikov ter analitike in nagradnega tekmovanja.

UM FERl dodatno obdeluje naslednje podatke govorcev na posnetkih:

- ime in priimek
- spol
- starost
- izobrazba v 4 stopnjah (OŠ ali manj, SŠ, višja ali visoka šola, fakulteta ali več)
- prvi jezik

- dvojezičnost od rojstva
- dodaten prvi jezik v primeru dvojezičnosti
- sedanji kraj in občina bivanja
- kraj in občina bivanja v otroštvu
- dodaten kraj in občina daljšega bivanja
- druge geografske in jezikovne podatke, ki izhajajo iz zgornjih podatkov.

Od navedenih podatkov podatek o imenu in priimku ne bo javno objavljen skupaj s posnetkom, ampak bo dostopen samo urednikom zbirke za namene upravljanja pravic govorcev.

Posnetki se na portalu zbirajo na podlagi privolitve govorcev. Privolitev kadar koli prekličete, ne da bi to vplivalo na zakonitost obdelave podatkov, ki se je na podlagi privolitve izvajala do njenega preklica.

Posnetki se nadalje lahko dodajo na portal na podlagi zakonitih interesov raziskovanja govora, razvoja digitalnih orodij za slovenski jezik in potrebe po arhiviranju jezikovne kulturne dediščine.

Skladno z zgoraj opredeljenim namenom se posnetki ter z njimi povezani podatki hranijo za nedoločen čas, do umika privolitve s strani govorcev na posnetku oziroma do prejema prepovedi osebe, ki je po zakonu, ki ureja dedovanje, zakoniti dedič prvega ali drugega dednega reda. V primeru umika privolitve oziroma izvrševanja pravice do pozabe bo UM FERl upravljavce, ki so s portala prenesli zbirko posnetkov, z razumnimi tehničnimi sredstvi obvestil, da posameznik, na katerega se nanašajo osebni podatki, od njih zahteva, naj izbrišejo morebitne povezave do teh osebnih podatkov ali njihove kopije.

Uporabniki posnetkov bodo predvsem raziskovalci slovenskega govornega jezika v jezikoslovju in sorodnih vedah ter razvijalci jezikovnih in govornih tehnologij, ki želijo ponuditi podporo za slovenski jezik.

Upravljavec ne namerava prenesti posnetkov in z njimi povezanih metapodatkov v tretjo državo zunaj EU, bodo pa posnetki na voljo za prenos prek spletnega portala od koderkoli na svetu.

Od upravljavca lahko zahtevate dostop do osebnih podatkov in drugih informacij o obdelavi, popravek ali izbris osebnih podatkov, v določenih okoliščinah pa tudi omejitev obdelave in prenosljivost podatkov ter ugovarjate nadaljnji obdelavi. V primeru, da niste zadovoljni z obdelavo osebnih podatkov, se lahko pritožite Informacijskemu pooblaščenca RS, <https://www.ip-rs.si>

3 Kako snemati

3.1 KOGA – kogarkoli, ki govori slovensko

Na enem posnetku lahko govori ena oseba ali poteka pogovor med dvema ali tremi osebami.

Govorci so lahko ljudje vseh starosti, vključno z otroki in starejšimi. Lahko opravljajo različne poklice in prihajajo iz različnih okolij, živijo v katerikoli regiji Slovenije, mestu ali vasi, zamejstvu, so Slovenci v tujini ali tujci, ki govorijo slovenščino. Če gre za naročeno snemanje, lahko naročnik določi bolj natančne zahteve glede govorcev.

Vse osebe, ki sodelujejo v posnetku, morajo vedeti za snemanje in se strinjati z uporabo posnetkov. Pred snemanjem morajo izpolniti in podpisati privolitev za objavo posnetka. Za otroke in mladoletne mora soglasje podpisati starš ali skrbnik.

Govorci naj govorijo sproščeno, tako kot v vsakdanjem življenju. Ni pomembno, ali govorijo narečno, pogovorno, zborna, slengovsko, žargonsko ali slovenščino kot tuji jezik.

Vsak posameznik lahko sodeluje samo v enem posnetku. Če sodeluješ tudi ti kot občanski znanstvenik, znanstvenica, naj tvoj govor na vseh oddanih posnetkih skupaj ne presega dveh ur.

3.2 KJE – v tistem prostoru (doma)

Prostor snemanja naj bo čim bolj tih, najbolje doma. V ozadju naj ne bo glasbe, govorjenja drugih ljudi, hrupa prometa ali podobnih motenj. Za sproščeno vzdušje je lahko med snemanjem na voljo kava, čaj ali kakšen drug napitek.

Snemanje lahko poteka tako, da so vse osebe v istem prostoru, kjer se pogovarjajo v živo (osebni stik). Lahko pa poteka pogovor na daljavo prek telefona, tablice ali računalnika (avdio ali video na daljavo). Prav tako lahko ena oseba snema svoj govor sama, brez prisotnosti drugih ljudi.

Podatke o prostoru, času, kraju, načinu snemanja in drugih okoliščinah vpiši v obrazec o okoliščinah snemanja.

3.3 KAJ - vsakdanji govor

Vsebina govora naj bo anonimna. Če so omenjene druge osebe, naj bodo samo z imenom, brez priimka, razen če gre za javne ali slavne osebe.

3.3.1 Družabni pogovor

Sproščen pogovor med dvema ali tremi osebami, ki se družijo. Pogovor je lahko resen ali zabaven, vendar naj bo spontan in naraven, brez vnaprejšnje priprave.

Pomembno je, da se snemane osebe vnaprej dogovorijo, o katerih temah bodo govorile, in izberejo samo tiste, ki jih želijo deliti z javnostjo.

3.3.2 Intervju

Pogovor, kjer ena oseba postavlja vprašanja, druga (ali dve osebi) pa odgovarja. Vprašanja so lahko pripravljena vnaprej ali nastanejo med pogovorom. Odgovori naj bodo spontani, ne vnaprej pripravljene.

Intervju naj se izogiba temam, ki so za intervjuvane osebo občutljive ali neprimerne za javno objavo. Lahko je tudi raziskovalne narave.

3.3.3 Diskusija

Pogovor, kjer dve ali tri osebe delijo svoja mnenja, ideje in argumente o temah, ki jih izberejo vnaprej.

Pomembno je, da snemane osebe ne izberejo tem, o katerih ne želijo javno govoriti.

3.3.4 Navodila

Govor ene osebe ali pogovor med dvema ali tremi osebami, kjer prva oseba razlaga, kako nekaj narediti, izvesti ali doseči. V enem posnetku je lahko več različnih navodil.

Če gre za pogovor, naj druga (in tretja) oseba aktivno sodeluje, sprašuje in preverja, ali pravilno razume. Osebe si lahko izmenjujejo vlogo tistega, ki daje navodila.

Navodila naj obravnavajo teme, ki jih snemane osebe želijo deliti z javnostjo.

3.3.5 Pripoved

Prva oseba pripoveduje o svojih ali tujih dogodkih, doživetjih ali izkušnjah iz resničnega življenja, ne iz knjig ali filmov. V enem posnetku lahko pove več zgodb.

Če sta zraven še druga in tretja oseba, lahko tudi te izmenično pripovedujejo.

Pripovedi naj govorijo samo o dogodkih, doživetjih ali izkušnjah, ki jih snemane osebe želijo deliti z javnostjo.

3.3.6 Razlaga

Prva oseba opisuje ali razloži stvari, pojme, dogodke ali pojave, da bi jih drugi bolje razumeli. V enem posnetku lahko poda več različnih razlag.

Če gre za pogovor, naj druga (in tretja) oseba aktivno sodeluje, sprašuje in preverja, ali prav razume. Osebe si lahko izmenjujejo vlogo tistega, ki razlaga.

Razlage naj bodo o temah, ki jih snemane osebe želijo deliti z javnostjo.

3.3.7 Namigi za teme govora

hobiji	šport, rekreacija	glasba
potovanja	tehnologija	hrana
igre	vesolje	zdravje
serije, filmi	kultura	avtomobili, motorji
podnebje	slavni	okolje
mediji	šolstvo	zdravstvo
delo	znanost	otroci, otroštvo
prazniki, običaji	vsakdanji dogodki	promet, mobilnost
domače živali	starševstvo	aktualni družbeni dogodki

3.4 S ČIM – avdio ali video, visoka kakovost posnetka

Za snemanje lahko uporabiš telefon, računalnik ali drugo snemalno napravo. Naprava naj bo stabilno postavljena, mikrofona pa čim bližje govorcem, da bo njihov govor dobro slišen.

Posnetek je lahko avdio ali video. Priporočljivo je nastaviti čim višjo kakovost posnetka in, če je mogoče, uporabiti nestisnjeni format, na primer wav (priporočeno 44,1 kHz, PCM, 16-bit, mono).

Zelo pomembna je pravilna nastavitvev glasnosti. Pred snemanjem naj osebe povedo nekaj besed za preizkus zvoka.

Preden oddaš posnetek, se prepričaj, da je govor jasno slišen in da ni motečih šumov v ozadju.

3.4.1 Dodatna navodila za snemanje

PROSTOR SNEMANJA

Prostor, v katerem so govorci med snemanjem, naj bo miren in brez motenj. Pomembno je, da:

- niso prisotni drugi ljudje,
- so okna zaprta, da se ne sliši hrup od zunaj,
- so izklopljene vse naprave, ki oddajajo zvoke, kot so radio, ventilacijske in klimatske naprave,
- v prostoru ni odmeva - ta se lahko pojavi v zelo majhnih ali praznih prostorih.

Če ne sodeluješ v pogovoru, se med snemanjem umakni iz prostora ali bodi tiho v ozadju.

POSTAVITEV SNEMALNE NAPRAVE

Pomembno je, da:

- stoji snemalna naprava stabilno,
- je mikrofonski čim bližje obrazu govorcev (lahko uporabiš improviziran podstavek),
- med govorom mikrofona nihče ne premika.

NASTAVITVE SNEMANJA

Posnetek je lahko avdio ali video. Nastavi čim višjo kakovost posnetka in, če je mogoče, izberi nestisnjeni avdio format. Najpriporočljivejši je:

- WAV,
- mono,
- 44.1 kHz,
- pcm,
- glasnost snemanja med -6 dB in 0 dB.

Ustrezno glasnost najnatančneje preveriš s pomočjo grafičnega prikazovalnika na zaslonu snemalne naprave ali programa. Če je pretiho (-12 dB ali manj), povišaj občutljivost, če nivo glasnosti večkrat sega do 0 dB, občutljivost znižaj.

Za snemanje s pametnim telefonom si lahko namestiš katero od aplikacij za snemanje, npr. ASR Voice Recorder (<https://play.google.com/store/apps/details?id=com.nll.asr>).

UREJANJE POSNETKOV

Končna dolžina posnetka je lahko najmanj 5 in največ 90 minut. Nič pa ni narobe, če se sogovorniki pogovarjajo že pred začetkom in po koncu snemanja. Nasprotno, to lahko pripomore k sproščenemu vzdušju in je priporočljivo. V tem primeru odreži posnetek na začetku in na koncu, preden ga oddaš.

Če pa nastanejo med govorjenjem premori, dolgi več kot minuto, lahko izrežeš tudi te. Vendar naj bo takega rezanja posnetkov čim manj ali nič.

Ne izvajaj nobenih drugih načinov urejanja posnetka, ampak oddaj takšnega, kot je bil posnet.

Za urejanje posnetkov lahko uporabiš Audacity (<https://www.audacityteam.org/>) ali drugo primerno aplikacijo.

POŠLJI NAM TESTNI POSNETEK

V primeru dvomov nam predhodno pošlji kratek, ne več kot minuto dolg testni posnetek. Pošlji ga na govorjena-slo@um.si pod zadevo »Testni posnetek«.

3.5 KOLIKO – med 5 in 90 minut

En posnetek mora trajati najmanj 5 minut in največ 90 minut. Med govorom naj ne bo zelo dolgih premorov, ki bi trajali več minut.

Na začetku ali koncu lahko posnetek odrežeš in oddaš samo del posnetega govora. Rez naj bo narejen v premoru, pred začetkom nove teme ali izjave.

Naredi toliko posnetkov, kolikor najdeš različnih govorcev. Vsak govorec je lahko samo na enem posnetku. Če sodeluješ tudi ti, naj tvoj govor na vseh oddanih posnetkih skupaj ne presega dveh ur.

4 O nas

Portal Govorjena slovenščina je del projekta LLM4DH, ki se osredotoča na raziskave razvoja velikih jezikovnih modelov za slovenščino in njihovo uporabo v digitalni humanistiki. Prek portala je mogoče oddajati posnetke tudi v okviru sodelovanja z drugimi raziskovalnimi projekti.

4.1 Povezani sedanji in pretekli projekti

2024-2027 LLM4DH – Veliki jezikovni modeli za digitalno humanistiko

Vzpostavitev portala Govorjena slovenščina

2022-2025 Mezzanine – Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik
Učni govorni korpus ROG (Verdonik et al., 2024a)

2022-2024 SPOT – Na drevesnici temelječ pristop k raziskavam govorne slovenščine
Učni govorni korpus ROG (Verdonik et al., 2024a)

2020-2023 RSDO – Razvoj slovenščine v digitalnem okolju

Nastanek govorne baze za razpoznavanje slovenskega govora Artur (Verdonik et al., 2023b; Verdonik et al., 2023c; Verdonik et al., 2024c)

Nadgradnja referenčnega govornega korpusa Gos (Verdonik et al., 2023a; Verdonik et al., 2024b)

2008-2011 SSJ – Sporazumevanje v slovenskem jeziku

Nastanek referenčnega korpusa govorne slovenščine Gos (Zwitter Vitez et al., 2021; Verdonik et al. 2013)

5 Pogoji sodelovanja

- Sprejemam naslednje pogoje prispevanja posnetkov in nagradnega tekmovanja (v nadaljevanju: pogoji) na portalu Govorjena slovenščina (v nadaljevanju: portal), ki ga upravlja Inštitut za elektroniko in telekomunikacije, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru (v nadaljevanju: UM FERI):
 - Vsi posnetki, ki jih uporabnik naloži na portal (v nadaljevanju: posnetki), morajo izpolnjevati zahteve, opisane na naslovu <https://govorjena-slovenscina.um.si/kako-snemati/>.
 - Vsi posnetki morajo slediti tehničnim navodilom za snemanje, ki so na voljo na naslovu <https://govorjena-slovenscina.um.si/dodatna-navodila-za-snemanje/>.
 - Na posnetkih ne smejo obstajati avtorske ali sorodne pravice uporabnika, govorcev ali tretjih oseb. Če na posnetku uporabnik v nasprotju s to zahtevo ima avtorske ali sorodne pravice, kot so zlasti pravice izvajalcev, prenaša te pravice na UM FERI v celoti in trajno, vključno s pravico do reproduciranja, dajanja na voljo javnosti v digitalni obliki, distribuiranja v vsakršni obliki in predelave.
 - Uporabnik mora od vseh govorcev na posnetku pridobiti lastnoročno podpisane izjave v predpisanem formatu, ki so na voljo tukaj <https://govorjena-slovenscina.um.si/kako-snemati/> in katerih sken v formatu pdf mora uporabnik priložiti posnetku. Uporabnik jamči, da so bile izjave pridobljene prostovoljno, brez zavajanja, sile ali grožnje.
 - Posnetki ne smejo biti žaljivi, spodbujati rasnega, etničnega, verskega ali drugega sovraštva, ali hujskanja k nasilju in ne smejo vsebovati kaznivih dejanj.
 - Za eno od nagrad, kot jih UM FERI opredeli na naslovu <https://govorjena-slovenscina.um.si/nagrade/>, se lahko uporabnik poteguje, ko na portal naloži posnetke v skupnem obsegu najmanj 150 minut, če ti v celoti ustrezajo pogojem. Uporabniki, ki z UM FERI pogodbeno sodelujejo na drugi osnovi, za svoje prispevke niso upravičeni do nagrade.
 - V primeru hujših ali ponavljajočih se kršitev pogojev lahko UM FERI uporabniku onemogoči nadaljnji dostop do portala in kreiranje novih uporabniških računov ter odvzame morebitno pravico do nagrade.

- Izjava o varstvu osebnih podatkov za portal Govorjena slovenščina je na voljo na naslovu <https://govorjena-slovenscina.um.si/uporaba-posnetkov/>.

6 Literatura

Verdonik, D., Kosem, I., Zwitter Vitez, A., Krek, S., Stabej, M., 2013, Compilation, transcription and usage of a reference speech corpus: the case of the Slovene corpus GOS. *Language Resources and Evaluation*. 1031-1048. DOI: [10.1007/s10579-013-9216-5](https://doi.org/10.1007/s10579-013-9216-5).

Verdonik, D.; et al., 2023a, *Spoken corpus Gos 2.1 (transcriptions)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1863>.

Verdonik, D.; et al., 2023b, *ASR database ARTUR 1.0 (transcriptions)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1772>.

Verdonik, D.; et al., 2023c, *ASR database ARTUR 1.0 (audio)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1776>.

Verdonik, D.; et al., 2024a, *Training corpus of spoken Slovenian ROG 1.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1992>.

Verdonik, D., Dobrovoljc, K., Erjavec, T., & Ljubešić, N. (2024b). Gos 2: A New Reference Corpus of Spoken Slovenian. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 7825-7830). ISBN 978-2-493814-10-4. ISSN 2951-2093, ISSN 2522-2686. <https://aclanthology.org/2024.lrec-main.691/>.

Verdonik, D., Bizjak, A., Žgank, A., Maučec, M. S., Trojar, M., Gros, J. Ž., ... & Dobrišek, S. (2024c), Strategies for managing time and costs in speech corpus creation: insights from the Slovenian ARTUR corpus. *Language Resources and Evaluation*, 1-26. <https://link.springer.com/article/10.1007/s10579-024-09746-8#article-info>, DOI: [10.1007/s10579-024-09792-2](https://doi.org/10.1007/s10579-024-09792-2).

Zwitter Vitez, A.; Zemljarič Miklavčič, J.; Krek, S.; Stabej, M., Erjavec, T., 2021, *Spoken corpus Gos 1.1*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1438>.

Priloga 1: Privolitev za objavo posnetka mojega govora in z njim povezanih osebnih podatkov na portalu Govorjena slovenščina

Spol (1) moški (2) ženski (3) drugo _____ (4) ne želim odgovoriti

Starost (1) _____ let (2) ne želim odgovoriti

Izobrazba (1) osnovnošolska ali manj (2) srednješolska (3) višješolska ali visokošolska strokovna (4) univerzitetna ali več (5) ne želim odgovoriti

Prvi jezik (1) _____ (2) ne želim odgovoriti

Dvojezičnost od rojstva (1) ne (2) da (3) ne želim odgovoriti **Če da, dodaten prvi jezik** _____

Sedanji kraj in občina bivanja _____ (v obliki: Kraj, Občina)

Kraj in občina bivanja v otroštvu _____ (v obliki: Kraj, Občina)

Dodaten kraj ali kraji bivanja Izpolnite, če ste leto ali več živeli, delali ali se šolali še v kakšnem tretjem kraju.

_____ (v obliki: Kraj, Občina; Kraj, Občina)

Dajem privolitev za objavo posnetka mojega govora in z njim povezanih osebnih podatkov na portalu Govorjena slovenščina, ki ga upravlja Inštitut za elektroniko in telekomunikacije, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, kot dela zbirke posnetkov, dostopnih javnosti za namen raziskav govora in razvoja digitalnih orodij za slovenski jezik ter kot arhiv jezikovne kulturne dediščine.

Nadaljnje informacije o obdelavi osebnih podatkov so na voljo v Izjavi o varstvu osebnih podatkov za portal Govorjena slovenščina, objavljeni na <https://govorjena-slovenscina.um.si/uporaba-posnetkov/>. Moje ime in priimek ne bosta objavljena na portalu skupaj s posnetkom in podatki o posnetku ter bosta dostopna le urednikom portala za namene upravljanja pravic govorcev. Privolitev lahko kadar koli prekličem, ne da bi to vplivalo na zakonitost obdelave podatkov, ki se je na podlagi privolitve izvajala do preklica.

Lastnoročni podpis: _____

Izjavljam, da na posnetku ne obstajajo moje avtorske ali sorodne pravice. Če v nasprotju s to izjavo na posnetku imam avtorske ali sorodne pravice, kot so pravice izvajalcev, prenašam te pravice na Inštitut za elektroniko in telekomunikacije, Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru, v celoti in trajno, vključno s pravico do reproduciranja, dajanja na voljo javnosti v digitalni obliki, distribuiranja v vsakršni obliki in predelave.

Ime in priimek: _____

Kraj in datum: _____

Lastnoročni podpis: _____

Priloga 2: Okoliščine snemanja

Vaša lastna oznaka posnetka _____

Vrsta snemanja (1) doniran posnetek (Posnetek oddajate kot občanski znanstvenik.)

(2) naročen posnetek (Posnetek oddajate v dogovoru z naročnikom.)

Kraj snemanja Pri snemanju na daljavo navedite vse kraje, kjer so govorcei.

Prostor/i, kjer so bili govorcei Pri snemanju na daljavo navedite vse prostore, kjer so govorcei.

_____ (npr. doma v
dnevni sobi, v sobi v študentskem domu, v pisarni)

Datum snemanja _____ (v obliki: 2025-01)

Kanal govora

(1) osebni stik (Osebe so med pogovorom v istem prostoru oziroma je snemalec v istem prostoru z osebo, ki jo snema.)

(2) avdio na daljavo (Osebe se pogovarjajo prek telefona ali spletne platforme brez videa oziroma se govorec snema sam v avdio načinu.)

(3) video na daljavo (Osebe se pogovarjajo na daljavo v video načinu oziroma se govorec snema sam v video načinu.)

Vrsta govora (1) družabni pogovor (2) intervju (3) diskusija (4) navodila

(5) pripoved (6) razlaga

Naprava, s katero ste snemali Opišite napravo čim podrobneje.

_____ (npr. prenosni
računalnik Asus ROG z zunanjim mikrofonom Samson Q2U)

Opis situacije V par stavkih opišite situacijo, v kateri je nastal posnetek, brez omenjanja

osebnih podatkov oseb na posnetku.

Ključne besede Navedite 3 do 6 pojmov, ki so lahko v pomoč pri iskanju, opisu ali kategorizaciji.
