Metapodatki o posnetkih in govorcih v govornih virih: primer baze Artur

Darinka Verdonik,* Andreja Bizjak,* Andrej Žgank,* Simon Dobrišek†

* Fakulteta za elektrotehniko, računalništvo in informatiko, Univerza v Mariboru
Koroška 46, 2000 Maribor
darinka.verdonik@um.si, andreja.bizjak1@um.si, andrej.zgank@um.si
† Fakulteta za elektrotehniko, Univerza v Ljubljani
Tržaška 25, 1000 Ljubljana
simon.dobrisek@fe.uni-lj.si

Povzetek

Ob združevanju različnih govornih jezikovnih virov se pojavljajo težave, ki izhajajo iz vsebinske nezdružljivosti zabeleženih metapodatkov o posnetkih, govorcih oz. govoru nasploh (npr. tip govora, vrsta govornega dogodka, lokacija in čas snemanja, spol, izobrazba, regija govorca). Ti metapodatki se zajemajo po eni strani zato, da omogočajo preverjanje uravnoteženosti govornega vira glede na različne govorce in govorne situacije, po drugi strani pa zato, da omogočajo razvrščanje govornih podatkov v kategorije, potrebne bodisi za jezikoslovne analize bodisi za učenje algoritmov razpoznavanja govora ipd. Najpogostejše razlike med zabeleženimi metapodatki o posnetkih in govorcih v obstoječih prosto dostopnih govornih virih za slovenščino so v kategorizacijah vrste govora in lokacije snemanja oziroma v kategorizacijah in oznakah regije govorca. Različne kategorije se pojavljajo tudi v zvezi s starostnimi in izobrazbenimi skupinami govorcev. Veliko vrst metapodatkov se pojavlja samo v posameznih virih, v drugih pa ne. Prispevek poleg pregleda razlik podaja tudi predloge za njihovo premostitev.

Metadata on recordings and speakers in spoken language resources: The case of the Artur database

When merging data from different spoken language resources, problems arise due to incompatibility of metadata on recordings, on speakers or on speech in general (e.g., information about the speech type or speech event, time and place of the recording, the gender, education, region of speaker). These metadata are captured on the one hand to ensure the balance of speech samples according to different speakers and speech situations, and on the other hand to enable the classification of speech data into categories needed either for linguistic analysis or for learning speech recognition algorithms. The most common differences in metadata on recordings and on speakers in the existing freely available speech resources for Slovene relate to categorizations of the type of speech and the location of the recording as well as to categorizations and designations of the speaker's region. Different categories also emerge in relation to age and educational groups of speakers. Many types of metadata are recorded only in particular resources. In addition to reviewing the differences we also give some suggestions how to overcome them.

1. Uvod

Govorni jezikovni viri so pomembni tako za razvoj jezikoslovja in celostno poznavanje jezika kot tudi za razvoj govornih tehnologij, kot je razpoznavanje ali sinteza govora. Poleg posnetkov in zapisa govora vsebujejo običajno tudi manjše ali večje število podatkov o tem, kje, kdaj, kako so posnetki nastali in kakšne so lastnosti govorcev glede na spol, starost, izobrazbo ipd. Čeprav Text Encoding Iniciative – TEI vključuje tudi standardizacijska priporočila s področja govora, pa so vsebinske odločitve, katere kategorije tovrstnih podatkov zajeti in kako podrobno jih opisati, zelo odvisne od vrste gradiva in namena govornega vira. Tako se ob združevanju virov, nastalih v različnih časovnih obdobjih z delno različnimi cilji in vključujoč različne tipe govora, pojavljajo težave, ki izhajajo iz vsebinske nezdružljivosti popisanih podatkov o posnetkih, govorcih oz. govoru nasploh.

S ciljem, da se tovrstne težave v prihodnje zmanjšajo, bomo v tem prispevku pregledali, potrebe po katerih podatkih so se pojavljale v različnih vedah, s poudarkom na dosedanjih slovenskih govornih virih (poglavji 2 in 3), podrobneje predstavili strukturo teh podatkov na primeru govorne baze Artur, ki predstavlja najnovejši in hkrati najobsežnejši in najbolj heterogen govorni vir za slovenščino ta trenutek (poglavje 4), ter izpostavili tiste vrste podatkov, kjer so vsebinska razhajanja največja, in podali predlog za njihovo uskladitev (poglavje 5).

2. Metapodatki o posnetkih in govorcih v govornih korpusih

Korpus GOS je predstavljal enega prvih večjih projektov, namenjenih zagotovitvi obsežnejšega govornega vira za raziskave slovenskega jezika. Izdan je bil leta 2011 v obsegu ca. 112 ur posnetkov in je sledil za tisti čas aktualnim korpusnojezikoslovnim prizadevanjem po dopolnjevanju referenčnih pisnih korpusov z referenčnimi govornimi korpusi (npr. Burnard, 2007; Allwood et al., 2000; Oostdijk et al., 2002; Pořízka, 2009). Njegov namen je bil torej predvsem zagotoviti podatke o govorjeni slovenščini za leksikografske, slovnične in druge jezikoslovne raziskave, za poučevanje slovenščine, za poklicne govorce ali pisce oz. tudi za širšo zainteresirano javnost. Vseboval je kolikor mogoče reprezentativen nabor različnih govornih situacij, s ciljem, da bi zajeli vzorčne primere različnih govornih situacij in različnih govorjenih diskurzov, demografsko reprezentativen vzorec govorcev slovenskega jezika in tiste govorne situacije, v katerih so uporabnika jezika najbolj pogosto produktivno ali pa samo receptivno udeleženi (Verdonik in Zwitter Vitez, 2011: 17).

GOS je bil poleg transkripcij dopolnjen tudi s posnetki ter s številnimi podatki o posnetkih in govorcih (po katerih lahko uporabniki korpusa tudi filtrirajo zadetke). Podobna je praksa v drugih, tujih govornih korpusih. Običajni podatki o situaciji, ki je posneta, so datum, lokacija, vrsta interakcije, kontekst, tematika, udeleženci, trajanje, uporabljena oprema za snemanje, vir ipd. Podatki o

udeležencih so običajno identitifikacijska koda, starost, spol, narodnost oz. prvi jezik, regija oz. narečje, poklic, lahko pa tudi še mesto rojstva, trenutna lokacija, drugi jeziki ipd. (Zemljarič Miklavčič, 2008; Cresti in Moneglia, 2005; Ehmer in Martinez, 2014; Love et al., 2017).

V korpusu GOS so metapodatki o posnetkih vključevali (Verdonik in Zwitter Vitez, 2011):

- podatke o gradivodajalcu oz. viru posnetka,
- podatke o vrsti govora, institucionalnem okviru, govornem dogodku, prosti opis govorne situacije in število aktivnih udeležencev govornega dogodka,
- podatke o času in kraju snemanja, pri čemer je bil kraj snemanja opredeljen tako z imenom kraja kot umeščen v širše (registrsko) območje.

Podatki o govorcih so zajemali:

- spol,
- starost, razdeljeno v 7 kategorij,
- izobrazbo, razdeljeno v 4 kategorije,
- regijo govorca, opredeljeno glede na registrsko območje, pri čemer je bila možnost opredelitve več regij v primeru, da je govorec več kot eno leto bival v različnih regijah (npr. zaradi študija, službe ipd.),
- prvi jezik govorca.

Korpusu GOS je v letih 2016–2019 v več izdajah sledila manjša govorna baza Gos Videolectures (Verdonik, 2018), ki je v nasprotju s korpusom GOS zajema področno omejeno gradivo javnih predavanj, dostopnih prek portala Videolectures.net. V svoji zadnji, četrti različici obsega skupno 22 ur posnetkov javnih predavanj, uravnoteženih glede na tematska področja družboslovja, humanistike, medicine, tehnike ter naravoslovja/matematike. Prav tako nastopajoči govorci enakomerno predstavljajo oba spola, starejše in mlajše govorce ter grobo opredeljene različne regije Slovenije.

Metapodatki o posnetkih in govorcih so sledili shemi, zastavljeni v korpusu GOS, vendar zaradi omejenega dostopa do informacij niso bili beleženi z isto natančnostjo. Če je bila starost govorcev v korpusu GOS deljena v 7 kategorij, je v Gos Videolectures samo v 2, pa še to predvsem na podlagi vizualnega vtisa, ne na podlagi neposredne, točne informacije. Prav tako ni bilo neposrednih podatkov o regiji govorca, ampak so bili pod to postavko zabeleženi slušni vtisi o značilnostih govora. Nekateri podatki pa niti niso bili opredeljeni, saj bodisi niso bili dostopni (izobrazba) bodisi niso bili relevantni (prvi jezik, ki je za vse govorce slovenščina). Ker je bila govorna baza Gos Videolectures namenjena tudi razvoju tehnologije razpoznavanja govora, so se pokazale potrebe še po beleženju kvalitete posnetka, ki je bila dodana zgolj kot subjektivna ocena transkriptorja na podlagi slušnega vtisa.

3. Metapodatki o posnetkih in govorcih v govornih bazah za razpoznavanje govora

Z vidika razvoja govornih tehnologij oziroma razpoznavalnikov govora je glavni razlog za zbiranje podatkov o govorcih in posnetkih predvsem ta, da se v govorni bazi zagotovi čim bolj ustrezna reprezentativna zastopanost vseh izrazitih govornih značilnosti, ki se spreminjajo med različnimi govorci in različnimi govornimi okoliščinami. Relevanten je torej katerikoli podatek o govorcu ali govornem posnetku, ki lahko nosi informacijo o govornih značilnostih samega govorca oziroma njegovih govornih okoliščinah, za katere se predpostavi, da imajo vpliv na akustične in jezikovne značilnosti posnetega govora. Z računskimi metodami obdelave signalov se namreč iz govornih signalov lahko izlušči različne govorne značilke, pri katerih se predpostavlja hierarhična razvrščenost pri njihovem odražanju tako nizkonivojskih anatomskih značilk človekovih govoril kot tudi višjenivojskih dialoških in semantičnih značilk.

Za razvoj samodejnih razpoznavalnikov govora je torej iz celotnega nabora metapodatkov smiselno ohraniti predvsem tiste, ki lahko prispevajo k boljšemu akustičnemu in jezikovnemu modeliranju govora. Pri razvoju govornih baz za razpoznavanje govora so bili tako v preteklosti metapodatki ključna informacija, na osnovi katere se je poskušala doseči ustrezna zastopanost vseh kategorij govorcev in govora, kot je bilo predvideno v specifikacijah. Glavni namen zbiranja teh metapodatkov je bil predvsem ta, da se v govorni bazi čim bolj realno odražajo okoliščine in scenariji možnih uporab samodejnih razpoznavalnikov govora (Kolář in Švec, 2008). Takšen pristop je zelo pomemben predvsem pri govornih bazah, ki obsegajo od vsaj nekaj 10 do več 100 ur govora oziroma govorcev.

Hiter tehnološki razvoj informacijsko-komunikacijskih sistemov je omogočil zbiranje in obdelavo vse večjih količin podatkov. Hkrati je prišlo tudi do izrazitega povečanja razpoložljivih računskih zmogljivosti sodobnih računalnikov, predvsem z razvojem zelo zmogljivih grafičnih procesnih enot (GPU), s katerimi se učinkovito izvajajo numerično zahtevni algoritmi t. i. globokega učenja (Gondi in Pratap, 2021). Posledica tega napredka je tudi ta, da so se za jezike z velikim številom govorcev začele pridobivati obsežne govorne baze, ki obsegajo tudi več kot 10.000 ur posnetkov govora. Tukaj gre praviloma za govorne baze, ki se pridobijo iz zelo različnih virov, kot so npr. razni mediji, spletne platforme, zvočne knjige idr. Zaradi velikega obsega takšnih baz se pridobljeni govorni posnetki pogosto ne označujejo in ne transkribirajo ročno. Za učenje razpoznavalnikov govora se potem uporabljajo nenadzorovani ali delno nadzorovani pristopi, ki ne zahtevajo ročno narejenih oznak in transkripcij govornih posnetkov (Hershey et al., 2017). Tako postane v večini primerov zelo obsežnih govornih baz dosledno uravnoteževanje govornih posnetkov na osnovi metapodatkov drugotnega pomena. Glede na zelo različne možne vire in načine zbiranja govornih posnetkov namreč pogosto tudi ni možno pridobivati relevantnih metapodatkov. V primerih, ko so metapodatki sicer na voljo, vendar jih je v govorni bazi težko uravnotežiti, pa pride v ospredje znamenit izrek Roberta Mercerja iz leta 1985, da ni boljših podatkov, kot je več podatkov.

Novi metapodatkovno neuravnoteženi pristopi k izdelavi govornih baz so dobili dodatno podporo pri postopkih globokega učenja, kjer se vse bolj pogosto uporabljajo metode samodejnega povečevanja obsega in plemenitenja učnih podatkov. Izvorni govorni posnetki se lahko tako s pomočjo sodobnih metod digitalne obdelave signalov modificirajo v različne simulirane oblike. Takšni osnovni pristopi so, denimo, pohitritve ali upočasnitve govora v izvornih govornih posnetkih. Z vidika metapodatkov, ki se navadno upoštevajo pri razvoju razpoznavalnikov govora, pa so se razvili tudi zahtevnejši pristopi, pri katerih se simulirajo različne snemalne okoliščine (npr. značilnosti kanala, nivo šuma, kodirniki, zvočna ozadja, prostor idr.). S takšnimi pristopi lahko učinkovito dopolnimo obseg izvornih govornih posnetkov

in uravnotežimo primanjkljaj določenih vrst govornih posnetkov (Karafiát et al., 2017).

Pri zasledovanju osnovnega cilja, da govorna baza čim bolje odraža možne okoliščine in scenarije uporabe razpoznavalnikov govora, je smiselno postaviti določene prioritete pri upoštevanju metapodatkov in njihovi uravnoteženosti. Za razvoj splošnega samodejnega razpoznavalnika govora je tako priporočljivo upoštevati predvsem naslednje metapodatke:

- Oznaka govorca: enoznačno identificira vse posnetke istega govorca v bazi. To omogoča učinkovito izvajanje metod prilagajanja modela razpoznavalnika govora na posamezne govorce (npr. metode MLLR, SAT, iVector idr.) (Povey et al., 2008; Cardinal et al., 2015), kar lahko prispeva k znatnemu izboljšanju pravilnosti samodejnega razpoznavanja govora.
- Prvi jezik: samodejno razpoznavanje govora za določen jezik je navadno bistveno manj uspešno pri govorcih, ki jim ta jezik ni prvi. Zato se pri razvoju splošnega razpoznavalnika govora njihov govor navadno izloči iz učnega postopka in se potem izvajajo posebne prilagoditve splošnega razpoznavalnika takšnim govorcem.
- Narečna skupina (Draxler in Kleiner, 2017): metapodatek je še posebej pomemben v primerih spontanega nejavnega govora. V primeru izrazitega narečnega govora je namreč možno uporabiti različne pristope adaptacije razpoznavalnika govora na narečja govorcev, s čimer se lahko do neke mere odpravi poslabšanje rezultatov.
- Snemalne zvočne okoliščine (Zhang et al., 2018): imajo lahko bistven vpliv na zanesljivost samodejnega razpoznavanja govora. Njihov vpliv je delno možno tudi simulirati ali ga odstranjevati s postopki robustne obdelave in izboljševanja kakovosti govornih signalov.
- Spol in starost govorca: v primeru splošnega razpoznavalnika govora je pri tvorjenju akustičnega modela govora pomembna uravnoteženost govorcev po teh dveh kategorijah. Adaptacija razpoznavalnika govora na spol in starost govorca se sicer redko izvaja, saj se uporablja predvsem sprotno prilagajanje modela razpoznavalnika govora na posameznega govorca. Se pa ta informacija lahko uporabi pri razvoju in preizkušanju tovrstnih metod za ugotavljanje njihove odvisnosti od teh dveh metapodatkov.

Če predstavljeni metapodatki v neki govorni bazi niso na voljo, jih je z določeno zanesljivostjo možno tudi naknadno samodejno določiti z različnimi postopki samodejnega razpoznavanja govornih vzorcev, kot so postopki biometričnega razpoznavanja in grozdenje govorcev ali razpoznavanje prvega jezika govorca. Takšni naknadno samodejno določeni metapodatki seveda lahko vsebujejo tudi napake, kar je potrebno upoštevati pri njihovi uporabi.

4. Metapodatki o posnetkih in govorcih v govorni bazi Artur

Leta 2020 se je začel nacionalni projekt Razvoj slovenščine v digitalnem okolju,[1] ki sta ga sofinancirala Republika Slovenija in Evropska unija iz Evropskega sklada za regionalni razvoj. Operacija se je izvajala v okviru Operativnega programa za izvajanje evropske kohezijske politike v obdobju 2014–2020. Projekt je izvajal konzorcij 12 partnerjev, od tega 6 javnih raziskovalnih zavodov in 6 podjetij. Naslavljal je več sklopov jezikovnih tehnologij, med njimi tudi govorne tehnologije, kjer je bilo veliko pozornosti namenjene izdelavi govorne baze za razvoj razpoznavanja govora v obsegu 1000 ur. Pomanjkanje ustrezno velike, zahtevam razpoznavanja govora prilagojene in prosto dostopne govorne baze se je namreč pokazalo kot osrednja ovira pri razvoju razpoznavanja govora za slovenski jezik. Pri izdelavi govorne baze so sodelovali Univerza V Mariboru (FERI), Univerza v Ljubljani (FE in FRI), ZRC SAZU, Alpineon in STA. Vključuje 4 večje sklope različnih vrst govora: brani govor po pisnih predlogah (500 ur), javni govor (javni dogodki, mediji ipd. – 200 ur), parlamentarni govor (Državni zbor RS – 200 ur) in nejavni govor (terenski posnetki prosto govorjenih monologov in dialogov).

Podatki o posnetkih in govorcih so v bazi Artur organizirani kot tsv-datoteka in v obliki xml-zapisa po standardu TEI. V primerjavi s predhodnimi govornimi viri za slovenščino vključujejo predvsem zelo podroben popis tehničnih lastnosti posnetkov (npr. podatke o lastnostih izvornih posnetkov in tehnični opremi, uporabljeni za snemanje) ter vseh okoliščin, ki bi lahko na te lastnosti vplivale (od velikosti prostora snemanja, prisotnosti hkratnega govora vse do uporabe maske pri govorcih, ki je bila pogosta v času epidemije COVIDA-19).

Končni seznam metapodatkov o posnetkih v govornih bazi Artur je naslednji:

I. Identifikacijski podatki in kategorizacija posnetkov:

- ID-posnetka: je sestavljen iz imena baze (Artur), podatka o tipu govora (brani – B, javni – J, nejavni – N in parlamentarni govor – P), štirimestne identifikacijske številke govorca (Gxxxx), šestmestne identifikacijske številke posnetka (Pxxxxxx) ter podatka o vrsti datoteke (-avd). Pri posnetkih javnega govora, na katerih se običajno pojavlja večje število govorcev, je namesto štirimestne identifikacijske številke govorca navedba Gvecg (s pomenom *več govorcev*). Primer ID-posnetka: *Artur-N-G5134-P600134-avd*.
- Vrsta govornega dogodka: označuje, ali gre za javni, nejavni, parlamentarni ali brani govor (Žganec Gros in Vesnicer, 2020).
- Opisi govornih dogodkov oz. topiki: Pri parlamentarnem govoru je govorni dogodek vedno označen kot seja državnega zbora. Pri javnem govoru so govorni dogodki opredeljeni kot okrogle mize, intervjuji, nagovori na dogodkih, novinarske konference ipd. oziroma kot spletni dogodek, kadar gre za posnetke, posnete na daljavo. Pri branem govoru so govorni dogodki opisani kot branje vnaprej pripravljenih pisnih predlog ali kot dva različna tipa črkovanja. Izbrani nabor kratic so govorci črkovali z dodajanjem samoglasnikov (npr. *ef a ku*), vnaprej določene pare imen in priimkov pa z dodajanjem polglasnikov (npr. *jə o nə a sə*). Če je govorec črkoval na nepredviden način, je topik poimenovan kot črkovanje s samoglasniki (oz. soglasniki) z odstopanjem *(*npr. *ef fa ku*), če je med branjem tudi kaj

[1] https://www.slovenscina.eu/

dodal ali komentiral, pa kot črkovanje s samoglasniki (oz. polglasniki) s komentarjem. Pri nejavnem govoru sta za govorne dogodke uporabljeni oznaki prosti dialog med dvema sogovornikoma in prosti monološki govor – pri slednjem govorec prosto opisuje različne stvari, recimo svoj najljubši film. Za potrebe razvoja specializiranih razpoznavalnikov v projektu Razvoj slovenščine v digitalnem okolju so v bazi Artur opredeljeni še govorni dogodki, kjer je snemanje potekalo po vnaprej pripravljenih scenarijih z dveh področij: opisovanje obrazov in upravljanje pametnega doma.

II. Podatki o okoliščinah snemanja:

- Datum snemanja je zapisan v obliki »mesec leto« (npr. *april 2021*).
- Podatek o občini snemanja temelji na seznamu občin v Republiki Sloveniji v času snemanja (2020–2022).
- Prostor snemanja natančneje opredeljuje, kje je govorni dogodek posnet, na primer v stanovanju ali pisarni, studiu ali premičnem snemalnem studiu, v dvorani, parlamentu ali pa je snemanje potekalo v odprtem prostoru.
- Velikost prostora je razdeljena v tri kategorije: do 20 m^2, od 20 do 80 m^2 in nad 80 m^2.
- Prisotnost šuma označuje, ali se na posnetku občasno pojavlja šum v ozadju, kot je šelestenje, šumenje, prometni hrup, zvok ventilatorja ipd. Če se šum po osebni presoji validatorja posnetkov pojavlja v preveliki meri, je tak posnetek uvrščen v skupino izločenih posnetkov.
- Presluh se občasno pojavi pri 2-kanalnem snemanju nejavnega govora, ko je spontani pogovor dveh sogovornikov posnet z dvema ločenima mikrofonoma. Prisotnost presluha je označena, če se pri takem snemanju pogosto in jasno sliši govor govorca z drugega kanala.
- Pogost hkratni govor je zabeležen pri nejavnem govoru, ko je sneman zasebni pogovor med dvema sogovornikoma, ki pogosto hkrati govorita.
- Podatek o tem, ali govorec nosi masko, je bil aktualen v času epidemije COVIDA-19, ko je veliko javnih dogodkov potekalo ob uporabi obrazne maske. To pomembno vpliva na akustične značilnosti posnetka. Posamezni redkejši posnetki te vrste, ki so bili uvrščeni v bazo Artur, so zato ustrezno označeni.

III. Podatki o formatu izvornih posnetkov:

- Najpogostejši formati izvornih posnetkov so WAV, MP3 in M4A.
- Čeprav so vsi posnetki v bazi Artur pretvorjeni v enotni format WAV, 44,1 kHz, pcm, 16-bit, mono, so bili posamezni posnetki, pridobljeni iz nelastnih virov, posneti v drugačnih formatih. Kadar so bile informacije dostopne, je bil izvorni format posnetkov popisan glede na frekvenco vzorčenja, bitno hitrost in bitno ločljivost.

IV. Podatki o opremi, uporabljeni za snemanje:

- Najpogosteje uporabljene snemalne naprave za posnetke v bazi Artur so prenosni ali namizni računalnik, prenosni snemalnik, pametni telefon, kamera in diktafon.
- Podatki o tehničnih lastnostih snemalne opreme zajemajo: opis naprave (npr. *MacBook PRO, Asus Vivobook, Zoom H4n, Zoom H1n*), naziv operacijskega sistema (npr. *iOS 14.2.1, Windows 10*), podatek o morebitnem mešalniku zvoka (npr. *Focusrite Scarlett 2i2 3rd Gen*), adapterju in opisu njegovega modela (npr. *Yamaha Audiogram 6*), vrsti mikrofona (npr. *namizni, vgrajeni* ali *studijski mikrofon*), modelu mikrofona (npr. *Samson Q2U*) in snemalnem programu (npr. *Adobe Audition 12, Audacity 2.3.2, Premiere Pro 14.0, Zoom, MS Teams*).

V. Podatki o viru posnetkov:

- Vir posnetka je lahko lastni posnetek, ki ga je naredila ekipa govorne baze Artur namensko za to bazo – to so vsi posnetki branega in nejavnega govora. V primeru parlamentarnega in javnega govora pa gre za arhivsko ali drugo gradivo, pridobljeno od različnih gradivodajalcev: Državni zbor RS, STA, Arnes, ZRC SAZU, Univerza v Mariboru, SDJT, Radio Štajerski Val idr.
- Pri javnem govoru je za posnetke večkrat na voljo tudi spletna povezava do videa.

Mnogi metapodatki o posnetkih večkrat niso bili dostopni. To velja zlasti za posnetke, ki niso bili lastni, ampak pridobljeni iz drugih virov, torej pri javnem in parlamentarnem govoru. Posnetki so bili uvrščeni v bazo, tudi če so kakšni metapodatki o njih manjkali, saj zlasti za javni govor ne moremo pričakovati, da bodo že obstoječi posnetki dokumentirani z metapodatki tako podrobno, kot je to mogoče, kadar snemamo namenoma za uvrstitev posnetka v govorno bazo.

Končni seznam metapodatkov o govorcih v govorni bazi Artur je naslednji:

I. Identifikacijski in sociodemografski metapodatki:

- ID-govorca zajema ime baze (Artur), oznako vrste govornega dogodka (B, J, N in P) ter vnaprej določeno štirimestno identifikacijsko številko govorca (Gxxxx). Primer ID-govorca: *Artur-N-G5097*.
- Spol (moški, ženski, drugo) je minimalno določljiv metapodatek o govorcih, tudi ko govor ni bil posnet kot lastni vir in govorci svojih sociodemografskih podatkov niso sami posredovali.
- Izobrazba je ločena v 9 kategorij: osnovna šola – nedokončana; osnovna šola – dokončana; nižje poklicno izobraževanje; srednje poklicno izobraževanje; gimnazije, SSI in PTI; višješolski programi, VS in UNI programi (1. bolonjska stopnja); magisterij stroke (2. bolonjska stopnja); magisterij znanosti (pred bolonjsko reformo); doktorat znanosti.
- Metapodatek o starosti je razvrščen v skupine: 12–17 let, 18–29 let, 30–49 let, 50–59 let, 60+ let.

II. Metapodatki o regiji govorca:

- Občina stalnega bivališča vključuje tako občine v Republiki Sloveniji kot stalno bivališče v tujini.

- Čim celovitejša demografska uravnoteženost govorcev branega in nejavnega govora je upoštevana tudi pri statistični regiji njihovega stalnega bivališča.
- Metapodatek o občini bivanja v otroštvu pokriva diahroni vidik morebitnih narečnih vplivov na govor govorca.
- Prvi jezik. Poleg govorcev, katerih prvi jezik je slovenščina, so v bazo Artur v manjši meri vključeni tudi govorci, katerih prvi jezik je hrvaščina, srbščina, makedonščina, bosanščina, ruščina, madžarščina idr. Podatek je izpolnjen samo pri govorcih, od katerih je pridobljen neposredno, pri javnih govorcih pa samo, če se lahko z veliko verjetnostjo sklepa, da je prvi jezik slovenski.
- Značilnosti govora se nanašajo na socialno zvrstnost jezika in so bile opredeljene s strani transkriptorja standardiziranega zapisa ali validatorja posnetkov. Namenjene so v pomoč pri morebitnem prilagajanju modelov razpoznavanja govora regionalnim značilnostim, prav tako so lahko v pomoč pri analizah zvrstnosti slovenskega govora. Niso pa mišljene kot točna strokovna opredelitev zvrsti govora govorca na posnetku. Ker je podrobna teorija socialne zvrstnosti za slovenščino (Toporišič, 2000) na empiričnem gradivu težko enoumno in robustno uporabljiva, je bila poenostavljena v tri osnovne kategorije: standardni jezik, pogovorni jezik in narečje. Glede na okoliščine govora je bilo predvideno, da se v javnem in parlamentarnem govoru pojavljata bodisi standardni jezik bodisi pogovorni jezik, pri čemer smo za pogovorni jezik šteli situacijo, ko so bili v govoru govorca pogosto prisotni sistematični glasoslovni pojavi, značilni za nestandardne zvrsti. Za standardni jezik pa je bil na primer označen tudi govor, ki je imel sicer prepoznavno regionalno obarvano melodiko, vendar je bil hkrati razviden zavesten večji odmik od vsakdanjega pogovornega jezika govorca proti standardnemu – to velja zlasti za govorce iz obrobja Slovenije ali drugih neosrednjih delov Slovenije. Razlike v izgovorjavi so bile zaznane tudi pri branem govoru, ki ga pa zaradi okoliščin (branje vnaprej napisanih povedi) težko ločimo na standardni in pogovorni jezik, zato sta bili pri branem govoru uporabljeni oznaki standardna izgovorjava in nestandardna izgovorjava. Predvsem v nejavnem govoru pa je lahko prisotna tudi oznaka narečje. V kolikor je bila izbrana, je dodana tudi oznaka o vrsti narečja, ki je določena na podlagi metapodatka o občini bivanja govorca v otroštvu.
- Zadnja oznaka se nanaša na opazne izgovorne težave. Pri posameznih govorcih se namreč pojavijo kakšne posebnosti, ki so povezane na primer z izgovorom glasov r, l ali podobno.

Navedeni metapodatki bodo v bazi Artur predstavljeni s slovenskimi poimenovanji kot tudi s prevodi v angleški jezik.

5. Razhajanja v metapodatkih o posnetkih in govorcih

Govorni korpusi, ki nastajajo za potrebe jezikoslovnih raziskav, in govorne baze, pripravljene za namene razpoznavanja govora, so praviloma zelo podobni govorni viri. Zato je smiselno, da se iščejo sinergijski učinki in se vsaj del gradiva uporabi v oba namena (Žgank et al., 2014). Tako se je že baza Gos Videolectures delala z mislijo na uporabo tudi za razpoznavanje govora (Verdonik, 2018), vendar je v metapodatkih še dokaj dosledno sledila zastavljeni shemi v korpusu GOS. Tudi v projektu Razvoj slovenščine v digitalnem okolju je bil iz velikega obsega posnetkov za govorno bazo Artur izbran primeren del za nadgradnjo govornega korpusa GOS. Ob tem pa se je v veliki meri ravno v zvezi z metapodatki o govorcih in posnetkih zgodilo precej razhajanj, ki so večinoma posledica bolj natančnega popisovanja podatkov, specifik ali pa namena baze, povzročajo pa težave ob združevanju gradiv. Katere vrste metapodatkov so take, pri katerih se najpogosteje pojavljajo različne odločitve?

5.1. Metapodatki o posnetkih

Obstajajo različne kategorizacije posnetega govora, saj se te praviloma izvedejo na podlagi tega, kaj vse neki govorni vir vsebuje. GOS je tako ločeval štiri tipe diskurza: javni informativno-izobraževalni, javni razvedrilni, nejavni nezasebni in nejavni zasebni. Če primerjamo to s kategorizacijo v bazi Artur, vidimo, da se tam pojavi še kategorija parlamentarni govor, manjka pa javni razvedrilni, ki se v Arturju tako rekoč ne pojavlja, pač pa se lahko celoten javni govor uvrsti kot javni informativno-izobraževalni. Prav tako ni nejavnega nezasebnega, ki se nanaša na različne uradovalne, storitvene, trgovalne in druge podobne nezasebne govorne situacije v vsakdanjem življenju. Je pa prisoten brani govor, ki se nanaša na zelo specifično, za namene snemanja posnetkov za bazo Artur ustvarjeno govorno situacijo, v kateri govorci berejo vnaprej pripravljene povedi eno po eno.

Poleg krovne kategorizacije posnetkov v manjše število krovnih kategorij se tako v korpusu GOS kot v bazi Artur uporabljajo še bolj podrobne opredelitve posnetega govora glede na govorni dogodek. V korpusu GOS je zabeleženih več kot 20 vrst govornih dogodkov, prav tako v bazi Artur, pri čemer pa jih je približno polovica namenjenih opredelitvi gradiva, ki je zelo specifično za potrebe razpoznavalnikov govora (črkovanje, področno specifični razpoznavalniki za pametni dom in opisovanje obrazov). Opredelitev vrste govornega dogodka je nadvse pomembna, saj omogoča po potrebi tudi naknadno prekategorizacijo zbranega gradiva ob združevanju različnih virov, zato je verjetno eden najbolj bistvenih metapodatkov o tipih posnetkov za vsak govorni vir, bolj pomemben kot širša, krovna kategorizacija, ki se lahko naknadno tudi spreminja na podlagi razvrščanja informacij o vrstah govornih dogodkov ali deloma tudi na podlagi informacij o viru.

Obvezna metapodatka o posnetkih v govornih virih sta čas in lokacija snemanja. Medtem ko so pri času lahko razhajanja samo v večji ali manjši natančnosti zabeleženega časa, pa se pri opredelitvi lokacije pojavljajo razlike, na katere enote se pri tem naslonimo. V korpusu GOS je bil ta metapodatek opredeljen dvojno: kot kraj, torej z imenom mesta ali vasi, ki pa skozi spletni konkordančnik ni dostopen zaradi varovanja identitete govorcev, in kot regija

snemanja, ki pa jo lahko opredelimo zelo različno. V korpusu GOS se je označila na podlagi registrskih območij. V bazi Artur je metapodatek o lokaciji zabeležen kot občina snemanja. V slovenskem kontekstu se zdi (glede na veliko število in razdrobljenost občin) informacija lokacije snemanja skozi občino ustrezen kompromis. V slovenskem podeželskem okolju lahko namreč navajanje točnega kraja z imenom vasi razkriva identiteto govorcev, enote, večje od občine (npr. upravna enota, registrsko območje ali statistična regija) pa niso več zadosti natančne in skladne z narečno razpršenostjo, ki je v Sloveniji pregovorno velika.

Metapodatek o viru prinaša informacijo o izvornem nosilcu avtorskih pravic. Podobno kot za pisna besedila namreč tudi za govorjena besedila velja, da so njihovi tvorci hkrati tudi avtorji z avtorskimi pravicami[2] nad besedili in pogosto obstajajo pogodbene zaveze, da bo ta podatek v jezikovnem viru ustrezno naveden. Pri posnetkih govora se v zvezi z avtorskimi pravicami in navajanjem vira srečujemo s štirimi vrstami situacij: (1) Če gre za posnetek na terenu, ki je bil narejen za namene govornega vira in zajema avtentični govor v vsakdanjih situacijah, govorci prenesejo avtorske pravice praviloma na nosilca projekta, v katerem nastaja govorni vir. Praksa je, da je v takih primerih kot vir označeno *terenski/lastni posnetek*. (2) Če gre za posnetek, ki je bil predvajan prek radia ali televizije, so pogosto nosilci avtorskih pravic medijske hiše in so posledično te navedene kot vir. Tudi pri spletnih virih (npr. posnetki na Youtubu[3]) je treba pogosto urediti avtorske pravice z njihovim/-i nosilcem/-i in v metapodatkih ustrezno navesti vir. Če gre za spletne dogodke, ki jih sicer organizira in objavi neka institucija (npr. spletne konference, delavnice, seminarji), je pogosto treba urejati avtorske pravice z neposrednimi tvorci teh besedil. Pri tem se pojavi vprašanje, kako je najbolj smiselno definirati metapodatek o viru: kot posameznika/e, ki je/so pravice odstopil/-i in nastopa/-jo na posnetku, ali kot institucijo, ki je organizirala in objavila spletni dogodek. V bazi Artur je bila pri tovrstnih posnetkih izbrana druga možnost. (3) Določeni internetni viri že imajo urejene avtorske pravice na način, ki omogoča nadaljnjo uporabo, in sicer pod pogoji katere od licenc Creative Commons. Taka večja vira posnetkov v slovenščini sta portala Videolectures.net in Arnes Video. V takih primerih se v obstoječih bazah za slovenščino kot vir navaja kar ime portala. (4) Določena govorjena besedila niso avtorsko varovana. V skladu z 9. členom ZASP so taka »uradna besedila z zakonodajnega, upravnega in sodnega področja«. Čeprav še ni tovrstne sodne prakse ali doktrine, se lahko kot tovrstna med drugim štejejo govorjena besedila, ki nastajajo v Državnem zboru RS v okviru zakonodajnih postopkov. V tem primeru se kot vir v bazi Artur, kjer se pojavljajo tovrstni posnetki, navaja kar Državni zbor Republike Slovenije.

Druge vrste metapodatkov o posnetkih, kot smo jih predstavljali v poglavjih 2, 3 in 4, se v določenih govornih virih pojavljajo, v drugih ne, odvisno od specifičnega namena govornega vira. Pri združevanju govornih virov se lahko bodisi izpustijo bodisi ostanejo nedefinirani, če niso bili zabeleženi in niso na voljo.

5.2. Metapodatki o govorcih

Čeprav so metapodaki o govorcih manj raznovrstni kot metapodatki o posnetkih, pa se razlike, kako jih opredelimo, pojavljajo tako rekoč pri vseh kategorijah razen pri spolu.

Najzahtevnejše vprašanje je povezano s potrebo, da se zabeležijo različni regionalni vplivi na govor posameznika. V zvezi s tem sta problematični naslednji točki:

1. Opredelitev regionalnih vplivov na govor govorca ni nujno enoznačna. Tako so se na primer v dodatku h govornemu delu korpusa BNC (British National Corpus) iz leta 2014, v katerem so zajemali samo vsakdanje pogovore, prepustili govorcem, da so sami s svojimi besedami opisali svoj dialekt, in nato te opise preslikali v shemo statističnih teritorijskih enot Velike Britanije (Love et al., 2017). Tudi v slovenskih govornih virih se je uveljavila praksa, da se regija govorcev beleži skozi geopolitične, in ne geolingvistične kategorije. Razlog je bržkone ta, da lahko zanesljive geolingvistične kategorizacije naredi samo stroka, in to naknadno, na podlagi zbranih podatkov. V korpusu GOS so bile tako kategorije za regijo govorcev definirane na podlagi registrskih območij, ki jih je za Slovenijo skupno 11, k temu pa so bile dodane še kategorije za zamejske Slovence (Avstrija, Italija, Madžarska) in govorce, ki jim slovenščina ni prvi jezik (tujina). Taka razdelitev je izredno ohlapna in nenatančna v primerjavi s slovensko dialektalno razpršenostjo. Tudi sam koncept »regionalna pripadnost«, zveden na registrsko označbo na avtomobilu, se zdi neustrezen, čeprav ima za teren zelo koristno lastnost robustnosti. V bazi Artur se je zato iskala bolj natančna, enoznačna, enostavna in manj sporna opredelitev metapodatka, ki bi nosil informacije o regiji govorcev. Ker smo ime kraja, zlasti ko gre za podeželsko okolje, že izpostavili kot problematično zaradi potencialnega razkrivanja identitete govorca, je bila kot osnovna enota izbrana občina. Slovenija je v času zbiranja posnetkov za bazo Artur razdeljena na 212 občin. Prednost te kategorije je tudi ta, da je mogoče občine enostavno enoznačno preslikati na širše geopolitične enote – 12 statističnih regij Slovenije, kot jih v času nastajanja baze definira Statistični urad Republike Slovenije.
2. Marsikdo danes ne živi vse življenje v nekem omejenem geografskem prostoru, ki je govorno homogen, pač pa je veliko ljudi mobilnih, bodisi z dnevnimi/tedenskimi migracijami zaradi šolanja ali zaposlitve bodisi zaradi selitev. Slika regionalnih vplivov na govor govorca je zato lahko pri določenih posameznikih izredno kompleksna in hkrati včasih tudi zelo specifična. Korpus GOS je tako omogočal, da so govorci zase izbrali skupno tudi do pet »regionalnih pripadnosti«. Tako nastane precej kompleksna slika, saj

[2] Termin avtorske pravice tukaj uporabljamo za vse materialne avtorske pravice, druge pravice avtorja v skladu z ZASP in avtorski sorodne pravice, ki utegnejo nastati pri snemanju. O vprašanjih osebnostnih pravic in varstva osebnih podatkov, ki so prav tako pomembna za vsako uporabo posnetkov v govornih virih, tukaj ne razpravljamo, saj ni relevantno v kontekstu tega članka. Bralca samo opozarjamo, da uporabo posnetkov za govorne vire ovira tudi ta pravni vidik.

[3] Sama licenca Youtube ne omogoča uporabe posnetkov za govorni vir.

dobimo poleg govorcev s samo eno regijo še precej govorcev z zelo različnimi kombinacijami regij, med katerimi pa posamezna kombinacija ne zajema veliko govorcev. Na koncu je za slednje najbrž smiselno zabeležiti samo eno skupno kategorijo »različni regionalni vplivi«, kot naredijo v korpusu C-ORAL-ROM (Cresti in Moneglia, 2005). V bazi Artur je bila opredelitev geografske mobilnosti skozi čas poenostavljena na dve vrsti metapodatkov, prva se nanaša na občino bivanja v otroštvu, druga na občino trenutnega stalnega bivališča. S tem se izgubi precej informacij o morebitni dodatni mobilnosti posamezne osebe, ki bi sicer bile pomembne za podrobno analizo govora posameznega govorca, vprašljivo pa je, koliko so relevantne za (kvantitativno) korpusno analizo ali za morebitno prilagajanje razpoznavalnika govorcem po regijah.

Določenemu delu govorcev slovenščina ni prvi jezik. Tudi to je podatek, ki je za govorni vir, če se v njem tovrstni govorci pojavljajo, zelo pomemben. Niti iz korpusa GOS niti iz baze Artur se nematerni govorci slovenščine niso izključevali, pač pa nasprotno – namenoma vključevali. S tem je v obeh virih bistven tudi metapodatek o prvem jeziku govorca.

Niti metapodatek o geografski pripadnosti govorca niti metapodatek o prvem jeziku pa še ne povesta, kakšen je dejansko govor nekega govorca v govornem viru z vidika socialnozvrstne delitve. Slednjo lahko ugotavljamo šele na podlagi (zlasti) slušne analize govora. Ne gre torej za metapodatek, ki ga zabeležimo na terenu, ampak za naknadno interpretacijo govornih podatkov. V korpusu GOS se ni delala, v bazi Artur pa je bila izražena tovrstna želja za potrebe razpoznavalnikov govora.

Izobrazba in starost govorcev sta metapodatka, preko katerih predvsem zagotavljamo ustrezno demografsko razpršenost govorcev, zajetih v govorni vir. Za posnetke javnega govora večinoma niti nista dostopna in posledično za velik del posnetkov v korpusu GOS in bazi Artur teh metapodatkov ni. Kjer pa sta na voljo, so skupine glede na starost in izobrazbo delno različno opredeljene in različno podrobne, kar otežuje združevanje virov. Minimalne kategorije starostnih skupin so po našem mnenju skupina najstnikov (okvirno do 19 let), skupina upokojencev (okvirno nad 60 let) in vse ostalo vmes. V kategoriji izobrazbe imamo 4-stopenjsko delitev v GOS-u in 9-stopenjsko delitev v Arturju. Po našem mnenju minimalna delitev je vsaj v dve skupini glede na to, ali je oseba zaključila izobraževanje po srednji šoli ali pa nadaljevala šolanje. Večja podrobnost metapodatkov o govorcih bi bila zanimiva verjetno predvsem za sociolingvistične raziskave, zato je potrebna ustrezna previdnost pred prehitrim posploševanjem v zelo grobe kategorije.

6. Zaključek

V prispevku smo obravnavali metapodatke o posnetkih in govorcih, ki se tipično uporabljajo v govornih jezikovnih virih. Osredotočili smo se na obstoječe prosto dostopne govorne vire korpusnega tipa za slovenski jezik, tj. referenčni govorni korpus GOS, bazo Gos Videolectures in govorno bazo v nastajanju znotraj projekta Razvoj slovenščine v digitalnem okolju, Artur. Govorni podatki iz teh treh baz namreč predstavljajo vir podatkov za razširitev referenčnega govornega korpusa GOS, ob tem pa se kažejo težave z združevanjem, ki med drugim[4] izhajajo tudi iz razlik v popisu in kategorizacijah metapodatkov o posnetkih in govorcih.

V prihodnje bi si želeli večjo homogenizacijo metapodatkov o posnetkih in govorcih zlasti tam, kjer gre za ključne metapodatke, ki so bistveni tako za spremljanje uravnoteženosti gradiv kot za kategoriziranje govornih podatkov. Pri posnetkih so taki ključni metapodatki: (1) opis govornega dogodka, ki mora biti zadosti podroben in se lahko razume v smislu govornih situacij, ki imajo večje število skupnih kontekstnih lastnosti, vključno z vrsto lokacije, vrsto razmerja med tvorci in naslovniki, namenom in kanalom komunikacije; (2) čas in lokacija snemanja, pri čemer je zlasti pri lokaciji pomembno, da je zadosti podrobna, npr. ime kraja ali občine, kjer poteka snemanje; (3) vir posnetka, pomemben zaradi korektne obravnave avtorskih pravic, v pomoč je lahko tudi pri sortiranju govornih podatkov po tipih, zaradi naknadnega dostopa do video vsebine pa je skoraj nujna tudi povezava do videoposnetka, če obstaja; (4) vedno koristni in zaželeni, a morda manj nujni pa so tudi vsi razpoložljivi podatki o snemalni opremi in tehničnih lastnostih posnetka. Pri govorcih so ključni metapodatki o: (1) identifikaciji, (2) spolu, (3) starosti, (4) prvem jeziku in (5) regiji/-ah, pri čemer mora biti slednja zadosti podrobno opredeljena (npr. na ravni kraja ali občine) in vsaj v grobem upoštevati tudi diahroni vidik. Pogosto prisoten je tudi metapodatek o (6) stopnji izobrazbe, medtem ko beleženje metapodatkov o poklicih, socialnem sloju ali pripadnostih različnim družbeno-kulturnim skupinam v slovenskih govornih virih do zdaj ni bilo prakticirano.

V članku med drugim predstavljamo tudi podroben opis metapodatkov o posnetkih in govorcih v govorni bazi Artur. Pri določanju metaoznak se je pokazalo, da pri čisto vseh kategorijah vnosi niso enoznačni in enostavno določljivi. Pri metaoznakah, nanašajočih se na govorce, je bila največji izziv kategorija značilnosti govora, saj je bil odločevalec pogosto soočen z dilemo, ali je jezik še standardni ali pogovorni oz. ali je pogovorni ali narečje. Kot pišemo v poglavju 4, so bili sistematični glasoslovni pojavi, značilni za nestandardne zvrsti, odločilni kriterij, da gre za pogovorni jezik; in nasprotno, opazno prizadevanje govorca, da bi uporabljal standardni jezik, čeprav je v njegovem govoru še vedno mogoče zaznati regionalno obarvano melodiko, je bilo odločilno za oznako standardni jezik. Če je bil govor označen kot narečni, smo se za točno določitev vrste narečja oprli na podatek o občini bivanja v otroštvu. Preostali metapodatki o govorcih so bili bodisi pridobljeni neposredno od govorcev bodisi jih nismo določali. Izjema je spol govorca, ki smo ga določili na podlagi posnetka, tudi ko ni bilo neposredne informacije. Pri javnih govorcih, za katere nismo imeli neposrednih informacij, a smo lahko z veliko verjetnostjo sklepali, da je njihov prvi jezik slovenski, je lahko bil dodan tudi ta podatek. Veliko izzivov je bilo tudi pri pridobivanju metapodatkov o posnetkih, saj je v primeru, ko ni podatkov s terena, izjemno težko sklepati o vrsti in velikosti prostora snemanja ali identificirati podatke o datumu in občini

[4] Določene razlike so sicer tudi v pravilih zapisovanja govora. V tem članku se osredotočamo samo na metapodatke o posnetkih in govorcih.

snemanja dogodka ter nemogoče zagotoviti natančen tehnični popis snemalne opreme. V bazi Artur so bili ti metapodatki vpisani samo, ko so bili znani.

Baza Artur je prioritetno namenjena razvoju modelov razpoznavanja govora, vendar lahko s svojim izredno podrobnim popisom metapodatkov predstavlja izhodišče pri morebitni nadgradnji ali razvoju podobnih govornih virov v prihodnosti. Po zaključku, od novembra 2022 naprej, bo prosto dostopna prek repozitorija CLARIN.SI pod licenco Creative Commons.

7. Literatura

Jens Allwood, Maria Björnberg, Leif Grönqvist, Elisabeth Ahlsen in Cajsa Ottesjö. 2000. The spoken language corpus at the Linguistics Department, Göteborg University. *Forum Qualitative Social Research,* 1(3).

Lou Burnard, ur. 2007. *Reference guide for the British National Corpus (XML Edition)*. URL: http://www.natcorp.ox.ac.uk/XMLedition/URG/.

Patrick Cardinal, Najim Dehak, Yu Zhang in James Glass. 2015. Speaker adaptation using the i-vector technique for bottleneck features. V: *Proceedings of Interspeech 2015*, str. 2867–2871.

Emanuela Cresti in Massimo Moneglia, ur. 2005. *C-ORAL-ROM: Integrated reference corpora for spoken romance languages*. John Benjamins Publishing Company, Amsterdam, Philadelphia.

Christoph Draxler, Stefan Kleiner. 2017. A cross-database comparison of two large German speech databases. V: *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow, UK, 10.–15. avgust 2015. International Phonetic Association.

Oliver Ehmer in Camille Martinez. 2014. Creating a multimodal corpus of spoken world French. V: Sükriye Ruhi, Michael Haugh, Thomas Schmidt, Kai Wörner, ur., *Best Practices for Spoken Corpora in Linguistic Research*, str. 142–161. Cambridge Scholars Publishing, Newscastle upon Tyne.

Santosh Gondi in Vineel Pratap. 2021. Performance Evaluation of Offline Speech Recognition on Edge Devices. Electronics 2021, 10, 2697. MDPI, Basel, Switzerland.

John R. Hershey, Jonathan Le Roux, Shinji Watanabe, Scott Wisdom, Zhuo Chen in Yusuf Isik. 2017. Novel deep architectures in speech processing. V: *New Era for Robust Speech Recognition*, str. 135–164. Springer.

Martin Karafiát, Karel Veselý, Kateřina Žmolíková, Marc Delcroix, Shinji Watanabe, Lukáš Burget, Jan "Honza" Černocký in Igor Szőke. 2017. Training data augmentation and data selection. V: *New Era for Robust Speech Recognition*, str. 245–260. Springer.

Jáchym Kolář in Jan Švec. 2008. Structural Metadata Annotation of Speech Corpora: Comparing Broadcast News and Broadcast Conversations. V: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).

Robbie Love, Claire Dembry, Andrew Hardie, Vaclav Brezina in Tony McEnry. 2017. The spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics,* 22(3):319–344.

Nelleke Oostdijk, Wim Goedertier, Frank Van Eynde, Lou Boves, Jean-Pierre Martens, Michael Moortgat in Harald Baayen. 2002. Experiences from the Spoken Dutch corpus project. V: M. González Rodriguez, C. Paz Suárez Araujo, ur., *Proceedings of the third international conference on language resources and evaluation* (LREC'02), str. 340–347. Las Palmas, Kanarski otoki. ELRA.

Petr Pořízka. 2009. Olomouc corpus of Spoken Czech: Characterization and main features of the project. *Linguistik online*, 38(2). http://www.linguistik-online.de/38_09/porizka.html.

Daniel Povey, Hong-Kwang J. Kuo in Hagen Soltau. 2008. Fast speaker adaptive training for speech recognition. V: *Proceedings of Interspeech 2008*, str. 1245–1248.

Jože Toporišič. 2000. Slovenska slovnica. Založba Obzorja, Maribor.

Darinka Verdonik. 2018. Korpus in baza Gos Videolectures. V: Darja Fišer, Andrej Pančur, ur., *Zbornik konference Jezikovne tehnologije in digitalna humanistika,* str. 265–268. Znanstvena založba Filozofske fakultete, Ljubljana.

Darinka Verdonik in Ana Zwitter Vitez. 2011. *Slovenski govorni korpus Gos*. Trojina, zavod za uporabno slovenistiko, Ljubljana.

Jana Zemljarič Miklavčič. 2008. *Govorni korpusi*. Znanstvena založba Filozofske fakultete, Ljubljana.

Zixing Zhang, Jürgen Geiger, Jouni Pohjalainen, Amr El-Desoky Mousa in Wenyu Jin, Björn Schuller. 2018. Deep learning for environmentally robust speech recognition: An overview of recent developments. V: *ACM Transactions on Intelligent Systems and Technology (TIST) 9.5*, str. 1–28.

Jerneja Žganec Gros in Boštjan Vesnicer. 2020. Izbor fonetično uravnoteženih besedilnih predlog za bazo branega govora. V: Tanja Mirtič, Marko Snoj, ur., *Razprave II. razreda SAZU: 1. slovenski pravorečni posvet,* str. 111–119. Slovenska akademija znanosti in umetnosti, Ljubljana.

Andrej Žgank, Ana Zwitter Vitez in Darinka Verdonik. 2014. The Slovene BNSI broadcast news database and reference speech corpus GOS: Towards the uniform guidelines for future work. V: Nicoletta Calzolari, ur., *LREC 2014: proceedings of the Ninth International Conference on Language Resources and Evaluation*, str. 2644–2647, Reykjavik, Islandija. ELRA.