

## PRIPRAVA GRADIVA IN STANDARDIZACIJA NIVOJEV ZAPISA ZA POTREBE DIALEKTOLOŠKEGA KORPUSA GOKO

Klara Šumenjak

Fakulteta za humanistične študije, Koper

UDK 811.163.6'282:003.035

V prispevku je predstavljenih nekaj smernic za standardizacijo nivojev zapisa in pripravo gradiva, ki jih je avtorica upoštevala pri gradnji dialektološkega korpusa GOKO (*Govorni korpus Koprive*). Prispevek lahko predstavlja izhodišče za razpravo o poenotenju zapisovanja narečnih besedil za potrebe (referenčnega) dialektološkega korpusa, ki bi lahko nastal ob sodelovanju več strokovnjakov in z ustrezno dopolnitvijo in nadgraditvijo korpusa GOKO.

standardizacija transkripcije, narečni korpus, dialektološki korpus, GOKO

The paper presents some guidelines for the standardisation of transcription and the preparation of dialect material which the author followed while constructing the dialectological corpus GOKO (*Kopriva Speech corpus*). The contribution can serve as a starting-point for further discussion about the standardisation of transcription of dialect discourse. Cooperation with other experts could lead to the construction of a dialectological reference corpus based on an upgraded and expanded version of the GOKO corpus.

standardisation of transcription, dialectological corpus, dialect corpus, GOKO

### Uvod

Korpus GOKO (*Govorni korpus Koprive na Krasu*)<sup>1</sup> je nastal kot eden izmed ciljev doktorske disertacije *Opis govora Koprive na Krasu na osnovi dialektološkega korpusa*. Gre za realizacijo modela dialektološkega korpusa, ki bi z ustrezno nadgraditvijo lahko služil kot izhodišče za gradnjo referenčnega dialektološkega korpusa. Zaradi želje po čim širši dostopnosti gradiva je besedilo v korpusu zapisano v treh različicah, tj. v fonetični, poenostavljeni in poknjženi varianti, ter opremljeno z zvočnimi posnetki. Pri zapisovanju besedila je bilo vodilno vprašanje, do kolikšne mere poenostaviti zapis. Veliko avtorjev, predvsem slovstvenih folkloristov/

etnologov in dialektologov (prim. Ivančič Kutin 2011; Karničar 2008; Kenda Jež 2011; Klinar idr. 2012; Smole 1994; Stanonik 2001; Škofic 2006; Zemljak Jontes idr. 2002), se ukvarja z vprašanjem zapisa narečnih besedil in njegove standardizacije,<sup>2</sup> žal pa še niso našli povsem enotne rešitve. V prispevku so predstavljeni predlogi za pripravo gradiva (izbiro zvočnih posnetkov in določitev njihove osnovne enote analize) in nekateri predlogi za standardizacijo različnih nivojev zapisa za potrebe dialektološkega korpusa, ki sem jih uporabila pri gradnji korpusa GOKO. Korpus je dostopen na spletnem naslovu <http://jt.upr.si/GOKO>.<sup>3</sup>

<sup>1</sup> O korpusu GOKO gl. tudi Šumenjak 2012; Šumenjak, Vičič 2012.

<sup>2</sup> Že Štrekelj se je leta 1906 ukvarjal z enakim vprašanjem, ko je sestavil *Navodila in vprašanja za zbiranje in zapisovanje narodnih pesmi, narodne godbe, narodnih plesov in šeg, ki se nanašajo na to*. Njegovo delo pa »ni samo priročnik za poenostavljen zapis besedil ljudskih pesmi, ampak prava študija fonetike vseslovenskih narečnih govorov« (Klobčar, Smole 2011).

<sup>3</sup> Za realizacijo modela korpusa je poskrbel strokovnjak za informacijske tehnologije dr. Jernej Vičič.

### Posnetki in enota analize korpusa GOKO

Na terenu je bilo v letih 2009–2012 zbranih približno 42 ur posnetkov.<sup>4</sup> Ker je šlo za pilotsko raziskavo, je bilo v korpus GOKO zajetih okrog 60 minut s korpusnega vidika čim bolj primernih posnetkov, zato ti niso bili zajeti linearno.

Kriteriji za zajemanje posnetkov v korpus so bili:

- 1) tehnična ustreznost posnetkov: prvi pogoj, da sem vključila posnetek v korpus, je bil ta, da se je informanta razločno slišalo, da ni bilo šumenja in drugih zunanjih interferenc;
- 2) vsebinska ustreznost posnetkov: čeprav Zakon o varstvu osebnih podatkov zvočnega snemanja izrecno ne ureja, sem se v izogib morebitnim pravnim ali etičnim kršitvam odločila, da v korpus ne vključim obrekovanja ali drugih spornih delov besedil;
- 3) faza snemanja: v korpus niso bili vključeni posnetki iz začetnih minut snemanja, saj zaradi prisotnosti snemalne naprave informant na začetku snemanja še ni popolnoma sproščen, zato so lahko taki posnetki nereprezentativni;
- 4) vključevanje izpraševalca: v korpus so zajeti le posnetki, ki vsebujejo malo izpraševalčevih posegov v pripoved, kajti ti lahko bistveno spremenijo njen potek.

Ko je bilo gradivo zbrano in posneto, ga je bilo treba ustrezno izbrati, transkribirati in označiti. Vsako prekodiranje govornih besede v pisano pa predstavlja hkrati tudi interpretacijo, ki je odvisna od zapisovalca. Pri prepisovanju (narečnega) govora se moramo odločiti, kakšna bo osnovna enota za analizo spontanega govora oz. narečnega diskurza.

V pisnem jeziku je osnovna strukturna enota poved oz. del besedila, ki se začne z veliko začetnico in konča s končnim ločilom. V govornem jeziku pa je poved nekoliko

težje določljiva. Različni avtorji enoto govornega besedila poimenujejo kot izrek (Zuljan Kumar 2007: 32), izjavo (Verdonik 2006: 51) ali kot govornjeni odstavek (Vitez, Zwitter Vitez 2004: 10). Sama sem se zaradi bolj ustaljene rabe odločila za termin izrek, ki se »v govornem jeziku (kot uresničitev povedi v kontekstu) navadno sestoji iz niza enot, ki bi jih z vidika tradicionalne slovnične delitve na povedi in znotraj njih na posamezne stavčne enote težko klasificirali zaradi pogoste rabe prirednih veznikov, nedokončanih stavčnih vzorcev ipd.« (Zuljan Kumar 2007: 32). Raziskovalci (spontanega) govora pa se strinjajo, da se osnovne enote spontanega govora ločujejo na podlagi prozodičnih lastnosti in premorov v govoru. V daljših izrekih, kjer govorec ni delal premorov, sem poskusila postaviti mejo med posameznimi izreki ob govornem vdihu (prim. Zemljarič Miklavčič 2008: 132; Verdonik 2006: 50). Uporabnik lahko v korpusu dostopa zgolj do posameznih izrekov, ne pa tudi do njihovega besedilnega okolja.

### Prekodiranje in smernice za transkripcijo GOKO

Prekodiranje je »prenos žive govornice v pisani medij« (Smole 1994: 143), pri tem pa lahko zaradi interpretacije ali lažje berljivosti in lažjega razumevanja povedanega pride do opuščanja narečnega in približevanja knjižnemu. Obstaja veliko vrst prekodiranja besedila: od najbolj natančnega, ki upošteva vse glasoslovne značilnosti določenega krajevnega govora, tj. fonetični zapis, do bolj enostavnega, ki se približuje knjižni različici. Slovenski jezikoslovci kljub številnim prizadevanjem še niso povsem uskladili načina zapisovanja govornega besedila oziroma njegovega standardiziranja, ko gre za natančnost fonetičnega zapisa. Povsem nedvoumna ni niti terminologija, ki označuje

<sup>4</sup> Že pred začetkom terenskega dela sem se odločila, da bom zbirala besedila, ki bodo pretežno monološka (informant in izpraševalec), saj je s korpusno analizo veliko lažje analizirati izrek ene osebe (monolog), kot dvo- ali večgovor, ki se pogosto razvije v klepet, posledično pa prihaja do številnih prekrivanj, hitrega menjavanja vlog in velikega števila nedokončanih izrekov (gl. tudi Rožac 2013).

način zapisovanja gradiva za potrebe govornega korpusa, saj se v literaturi za različne načine prenosa govornega besedila v zapisanega pojavljajo različni termini, npr. fonetična transkripcija, fonetični zapis, fonemska transkripcija, alofonska transkripcija, akustično-fonetična transkripcija, prozodična transkripcija, poenostavljena transkripcija, ortografska transkripcija, transliteracija, pogovorni zapis, poknjženi zapis, standardizirani zapis in knjižni zapis,<sup>5</sup> pomenske razlike med njimi pa niso vedno jasno vzpostavljene (npr. Zemljak Jontes idr. 2002; Zemljarič Miklavčič 2004, 2008; Zwitter Vitez idr. 2009).

Ker sem želela, da je korpus namenjen tudi širši javnosti, sem zbrano gradivo zapisala v treh različicah, ki so opremljene tudi z zvočnimi posnetki: a) **Fonetični zapis**, ki upošteva vse glasoslovne značilnosti govora Koprive na Krasu – ta oblika je namenjena jezikoslovni, zlasti dialektološki analizi zapisanega govora. b) **Poenostavljeni zapis**, kjer sem vendarle skušala obdržati vidne vsaj temeljne glasoslovne značilnosti krajevnega govora (zapisani so npr. mesto naglasa, diftongi, polglasnik).<sup>6</sup> Pomembno vprašanje, ki se pojavlja na tem mestu, je, do kolikšne mere poenostaviti besedilo; če bi popolnoma zanemarili zapis glasoslovnih značilnosti, namreč ne bi mogli več govoriti o koprivskem dialektološkem korpusu, ampak zgolj o korpusu, ki ima le nekaj zanimivih leksikalnih in skladenjskih narečnih potez. Zato sem se odločila, da pri transkripciji ohranim večino glasoslovnih posebnosti koprivskega govora, vendar jih zapišem z znaki, ki so bližji večini laičnih uporabnikov (torej knjižni abecedi). c) **Poknjženi zapis**, kjer je vsaka posamezna

beseda zamenjana s svojo knjižno ustreznico,<sup>7</sup> na ravni besedne zveze in stavka pa se ohranjajo posebnosti govornega jezika. Tak način zapisovanja je nujen za iskanje po korpusu, saj bo uporabnik najverjetneje iskal po knjižni besedi (npr. *vodnjak*), ne pa po ortografskem zapisu besede (npr. *šterna*). V vseh treh različicah se zapisujejo tudi posebne oznake besed (npr. li = lastno ime, nar = narečno).<sup>8</sup>

### Raba ločil ter velikih in malih začetnic

V pisnem jeziku velja normirana stava ločil, ki pa ni nujno enaka v zapisu govornega jezika. »Zapisi govornih besedil navadno vsebujejo ortografska znamenja (ločila), kot so pika, vejica, klicaj in vprašaj, vendar jih vanje postavi zapisovalec in samo približno prenašajo fonetične in prozodične informacije, ki jih v govornem jeziku opravlja intonacija.« (Zuljan Kumar 2007: 32) Napaka, ki jo pogosto počnemo zapisovalci, ko transkribiramo narečna (govorjena) besedila, je ta, da poskušamo tudi narečne pripovedi postaviti v strukture, značilne za pisni jezik – zaključene stavčne strukture (prav tam).

Kljub temu da so se nekateri avtorji odločili, da v zapisu spontanega/narečnega govora ne pišejo velikih začetnic (npr. Rožac 2013; Weiss 1998) ali celo zanemarijo stavčna ločila (npr. Volk 2011), sem se sama zaradi boljše preglednosti in lažje berljivosti odločila v korpusu uporabiti tako ločila kot velike začetnice. Ločila – predvsem pike in vejice – so postavljena nekoliko arbitrarno: poleg tistih vejic, ki so v skladu s stavo ločil po pravopisu, so postavljene vejice tudi takrat,

<sup>5</sup> Zdi se, da sta izraza *zapis* in *transkripcija* premalo razlikovalna, saj se pojavljata tudi sinonimno.

<sup>6</sup> Več gl. Šumenjak 2012: 76–78; Šumenjak, Vičič 2012: 81.

<sup>7</sup> Opozoriti moram, da za tak prepis dialektologij uporabljajo tudi izraz *prevod* (Smole 1994: 152; Škofic 2001), vendar pri mojem zapisu ne gre za prevod na vseh jezikovnih ravneh, temveč samo na ravni posamezne besede, saj korpus šteje mesta besed in jih preverja glede na mesto v besedilu. To pomeni, da je npr. sedma beseda poknjžene različice enaka sedmi besedi v poenostavljeni in sedmi besedi v fonetični varianti. Zaradi tega ne moremo govoriti ne o prevodu ne o knjižni različici, saj besedilo ni prilagojeno knjižnemu jeziku na vseh jezikovnih ravneh, vsebuje pa tudi besede, ki niso knjižne (npr. nedoločne člene). Zaradi naštetega sem se odločila, da tretjo različico imenujem *poknjženi zapis*.

<sup>8</sup> Več gl. Šumenjak Vičič 2012: 81–83.

ko je npr. informant naredil krajši premor ali ko se je zmotil in nato popravil; pika pa je postavljena tudi takrat, ko se je zdelo, da bi lahko izrek predstavljal smiselno in pomen-sko zaključeno enoto, tj. ob govorčevem vdihu.

### Označevanje neverbalnih dogodkov

Priporočila za TEI (Text Encoding Initiative)<sup>9</sup> sicer predvidevajo označevanje nekaterih prozodičnih in neverbalnih oznak v govorjenih besedilih (prim. Zemljarič Miklavčič 2008: 97–98), vendar sem se sama odločila, da kinezičnih dogodkov (kimanje, skomiganje z rameni ipd.) ne označujem, saj bi to le otežilo gradnjo korpusa in ne bi bilo relevantno za opis koprivskega govora (razen pri skladenjski analizi). Teh podatkov tudi ni mogoče izluščiti z avdioposnetkov, kar pomeni, da bi moral korpus vsebovati neke vrste

didaskalij z opisi kinezičnih dogodkov (prim. Ivančič Kutin 2011: 82). Neverbalni glasovi, kot so kašljanje, znaki obotavljanja, premori, smeh, ki se jih sliši na posnetkih, so v korpusu označeni. Uporabnik korpusa lahko izbere, ali se mu te oznake prikažejo v korpusu ali ne (gl. sliko 1).

### Prilagajanje zapisa za potrebe korpusa

Pri oblikovanju načel transkribiranja besedil za korpus je treba upoštevati mednarodne standarde in že obstoječa priporočila ter jih prilagoditi lastnim potrebam oziroma specifikam zapisovanja glasov v izbranem slovenskem narečju.

V mednarodnem jezikoslovju sicer prevladuje tendenca, da se uporablja fonetična transkripcija s simboli IPA (International Phonetic Alphabet), »ki je standard, ki ga za zapis fonemskih in fonetičnih informacij

The screenshot shows the GOKO corpus search interface. At the top, there is a navigation bar with 'Domov - Poizvedba CQP - Pomoč' and 'Korpus: c'. Below this, the search parameters are set to 'Iskana beseda: mama', 'Način iskanja: vse oblike besed', and 'Besedna vrsta: vseeno'. The search results are displayed in a table with columns for 'Potrdi', 'Oblika besedila', and 'Statistika: vsi zapisi: 8, zapisi po filtriranju: 8'. The first result is for the lemma 'mama', with the text 'Jaz nisem mogla verjeti mami, ki je pravila, mama je pravila mi to in je <e>, potem je zaspal, potem je šla hitro tega ubogega otroka podojit, saj je bila tako nesrečna potem; jaz ne vem, od koga je tista Pavlina, saj potem je ostala šivilja <e> in strašno pridna šivilja je bila nesrečna, bolj slabo hodila in tako, taka, je bila prizadeta.' The word 'verjeti' is highlighted in a box.

Slika 1: Iskanje po lemi *mama* in na levi strani prikaz možnosti filtriranja korpusa GOKO  
Vir: GOKO

<sup>9</sup> Inicijativa za izdelavo standardnega zapisa besedil in priporočil o njegovem obsegu, strukturi, vsebini in načinu izdelave.

priporočata TEI in NERC<sup>10</sup>» (Zemljarič Miklavčič 2008: 106). Simboli IPA pa so zelo težko berljivi za računalnike, zato so v okviru evropskega projekta SAM (Speech Assessment Methodology) razvili transkripcijski sistem SAMPA (Speech Assessment Methods Phonetic Alphabet), ki je računalniku prijaznejši. Adaptacija je bila narejena tudi za slovenski jezik (prav tam), saj so se slovenski jezikoslovci že sredi 90. let prejšnjega stoletja želeli povezati in poenotiti računalniški zapis slovenskega govora (prim. Dobrišek idr. 1996), s čimer bi lahko standardizirali in poenotili tudi »raziskovalno delo na področju govornih jezikovnih tehnologij, ki se nanaša na fonetične označitve slovenskih govorjenih besedil« (prav tam). Slovenski korpusni jezikoslovci so se že zgodaj zavedali, da je pomembna »standardizacija računalniškega simbolnega fonetičnega zapisa slovenskega govora, ki bi tako ohranjal narodno in raziskovalno enotnost in nas s tem postavil ob bok svetovni eliti« (Zemljak Jontes idr. 2002: 159). M. Zemljak Jontes idr. so tako leta 2002 predlagali računalniški simbolni fonetični zapis slovenskega govora MRPA (Machine Readable Phonetic Alphabet), tj. računalniško pisavo, ki naj bi bila enostavna in prijazna za računalniško uporabo.

Sama sem se odločila, da za svoj narečni korpus uporabim slovensko fonetično transkripcijo,<sup>11</sup> ki jo je mogoče zapisati s pomočjo vnašalnega sistema ZRCola,<sup>12</sup> iz več razlogov: 1) glede na to, da bo korpus namenjen predvsem Slovincem, se zdi slovenska fonetična transkripcija najenostavnejša za branje; 2) če

bi želel kdo uporabiti gradivo za primerjalno analizo, je gradivo najbolj relevantno za slovanski prostor, ki pa to transkripcijo pozna in uporablja; 3) dialektološki korpus GOKO podpira pisavo ZRCola.<sup>13</sup>

V dokaz, da je slovenska fonetična transkripcija najustreznejša izbira, je v nadaljevanju eno od besedil iz korpusa GOKO zapisano v treh različicah: a) s slovensko fonetično transkripcijo; b) z mednarodno fonetično abecedo IPA; c) z abecedo MRPA.

a) Primer prepisa s slovensko fonetično transkripcijo

Smo 'tərko š'kʷəde nər'dili če'bʲeləm, 'ne, 'ma mula'rija 'təkət, ki smo blę, 'vʲeš, kə'ku je, in smo 'tisti 'mʲet, 'vʲeš, 'ku sə jə ʋ'lięku 'tisti 'mʲet, 'ne ku z'deʲ, 'tu si 'mʲela ʋ'se po 'rokəx smo š'li se ʋ'mit in 'tako, 'ne, o'joi, sə smo 'mʲel 'dobro, 'ja.

b) Primer prepisa z mednarodno fonetično abecedo IPA

fmo 'tərko f'kwəde nər'dili tʃe'bʲeləm, 'ne, 'ma mula'rija 'təkət, ki smo ble, 'vieS, kə'ku je, in smo 'tisti 'mjet, 'vieʃ, ku sə jə w'lieku 'tisti 'mjet, 'ne ku z'deɪ, 'tu si 'miela ʋ'se po 'rəkəx smo š'li se w 'mit in 'tako, 'ne, o'joi, sə smo 'miel 'dɔbrɔ, 'ja.

c) Primer prepisa z abecedo MRPA

SmO ''t@rkO Sk''w@dE n@r''dili tSE''biel@m, ''nE, ''ma mula''rija ''t@ket, ki smO ble, ''vieS, k@''ku je, in smO ''tisti ''mjet, ''vieS, ku s@ j@ w''lieku ''tisti ''mjet, ''nE ku z''dEI, ''tu si ''miela W''sE pO

<sup>10</sup> Projekt Network of European Reference Corpora za poskus standardizacije korpusnih zapisov v Evropi.

<sup>11</sup> V slovenski dialektologiji se uporablja t. i. nova nacionalna fonetična transkripcija, ki sledi osnovnim načelom transkripcije Slovanskega lingvističnega atlasa (OLA) (Kenda Jež 2011: 80–81), »tj. dogovorjen sistem znakov, ki zaznamuje kakovost in kolikost glasov, vključno z naglaševanjem (jakostnim/dinamičnim ali tonemskim)« (Smole 1994: 150). Taka transkripcija je tudi za dialektologa zelo zahtevna, saj od njega zahteva veliko znanja, zbranosti in potrpljenja. »Je pa edini način, ki ohranja največ prvin, s katerimi je pripovedovalec oblikoval pripoved« (Ivančič Kutin 2001: 61).

<sup>12</sup> Vnašalni sistem ZRCola je na Znanstvenoraziskovalnem centru SAZU v Ljubljani (www.zrc.sazu.si) razvil dr. Peter Weiss.

<sup>13</sup> Če imajo uporabniki že naloženo pisavo ZRCola, internetni brskalniki prikažejo strani, tudi če nimajo naloženih teh oznak. Sicer pa se ob prvem ogledu korpusa naloži samo pisava, ki omogoča pravičen prikaz posebnih znakov.



"rOk@x s'mO S"li sE w"mit in "takO, "nE, O"jOI, s@ smO "miel "dObrO, "ja.

### Zapisovanje z vnašalnim sistemom ZRCola in predlogi za hitrejšo delo

Čeprav vnašalni sistem ZRCola zadosti vsem potrebam zapisovanja narečnega gradiva,<sup>14</sup> je lahko njegova uporaba zamudna.

V koprivskem govoru ni veliko zapletenih dialektoloških znakov, ki bi se razlikovali od črk za zapis knjižnega jezika, vseeno pa jih je toliko, da bi nam delo, ki zahteva vsakokratno nizanje črk in zaslonskih diakritičnih znamenj za naglas in kvaliteto samoglasnikov in za soglasniške alofone, vzelo veliko časa. V nadaljevanju je prikazanih nekaj dialektoloških znakov, potrebnih za zapis koprivskega govora v fonetični in poenostavljeni transkripciji, ter opisan postopek (niz ukazov) v sistemu ZRCola, potreben za njihov zapis. Nekateri od znakov se izpišejo že s samim vtipkanjem (npr.  $\gamma$ ), drugi pa se prikažejo v obliki črk z ločenimi diakritičnimi znamenji, ki jih sestavimo šele z ukazom *Sestavi znake* v Wordovem dokumentu (npr.  $\check{z}$ ,  $\mu$ ).

Znaki in ukazi, potrebni za fonetično transkripcijo s sistemom ZRCola, so npr.: alt + shift + j =  $\gamma$ ; u in alt + shift + o =  $u_{\sim}$  =  $\underline{u}$ ; alt + z in alt + 2 =  $\check{z}$  =  $\check{z}$ .

Znaki in ukazi, potrebni za poenostavljeno transkripcijo s sistemom ZRCola, so npr.: e in alt + shift + 6 =  $e^{\cdot}$  =  $\acute{e}$ ; u in alt + " in alt + shift + 6 =  $u^{\cdot}$  =  $\acute{u}$ .

Ko nanizamo vse ukaze, moramo v Wordovem dokumentu, ki ima naloženo ZRColo, izbrati funkcijo *Sestavi znake*, ki te znake sestavi v zeleni sestavljeni simbol (npr.  $\check{z}$ ,  $\mu$ ,  $\acute{e}$ ,  $\acute{u}$ ,  $\check{z}$ ).

Ugotovila sem, da si je bolje zamisliti posebne ukaze za prikaz znakov, ki so potrebni za zapis koprivskega govora, in jih zapisati s pomočjo *vstavljanja simbolov*; za vsak simbol je namreč mogoče ustvariti posebno bližnjico *ShortcutKey*, pod pogojem, da je na

računalniku nameščena ZRCola. Tak način omogoča neposredno vstavljanje izoblikovanih znakov, ki jih ni treba naknadno sestavljati, zato je dokument že med nastajanjem veliko bolj pregleden, kot če bi imeli na koncu razstavljene simbole, ki jih je treba še sestaviti, npr.  $\underline{u}$  = alt + u;  $\acute{u}$  = alt + n;  $\check{z}$  = alt + u;  $\acute{u}$  = ctrl + alt + u, kar zagotavlja tudi manjšo možnost napak. Za tak način zapisovanja je smiselno, da si vsak zapisovalec sam določi bližnjice, ki mu najbolj ustrezajo, saj se glasoslovne značilnosti različnih govorov lahko zelo razlikujejo, posledično se razlikujejo tudi znaki, s katerimi zapisujemo glasove.

### Kako naprej

V prispevku je bilo predstavljenih nekaj smernic, ki sem jih upoštevala pri zapisovanju besedil in pripravi gradiva za korpus GOKO. Ker gre za pionirski korpus, upam, da bo ta prispevek uporaben predvsem kot izhodišče in pobuda za skupno postavitve standardov zapisovanja narečnih besedil, ki bi jih lahko upoštevali tudi pri morebitni gradnji prvega referenčnega dialektološkega korpusa.

### Literatura

- DOBRIŠEK, Simon idr., 1996: Pobuda za standardizacijo računalniškega simbolnega fonetičnega zapisa slovenskega govora. Baldomir Zajc, Franc Solina (ur.): *Zbornik pete Elektrotehniške in računalniške konference ERK '96*. Ljubljana: Somaru. 247–250.
- IVANČIČ KUTIN, Barbara, 2011: *Živa pripoved v zapisu: kontekst tekstura in prekodiranje*. Ljubljana: Založba ZRC, ZRC SAZU.
- KARNIČAR, Ludvik, 2008: Fonetično zapisovanje narečnih etnoloških besedil. *Traditiones* 37/1. Ljubljana: Založba ZRC, ZRC SAZU. 155–167.
- KENDA JEŽ, Karmen, 2011: Fonetična transkripcija. Barbara Ivančič Kutin: *Živa pripoved v zapisu: kontekst tekstura in prekodiranje*. Ljubljana: Založba ZRC, ZRC SAZU. 79–83.
- KLINAR, Klemen idr., 2012: *Metode zbiranja hišnih in ledinskih imen: projekt FLU-LED v okviru Operativnega programa Slovenija–Av-*

<sup>14</sup> Za nadgrajevanje in sprotno posodabljanje nabora znakov (tudi na željo uporabnika) ažurno skrbi dr. Peter Weiss.

- strija 2007–2013*. Jesenice: Gornjesavski muzej.
- KLOBČAR, Teja, SMOLE, Vera, 2011: Ozadje sto let stare zbirateljske akcije Odbora za nabiranje slovenskih narodnih pesmi in Štrekljeva vloga v njej. Marko Jesenšek (ur.): *Globinska moč besede: red. prof. dr. Martini Orožen ob 80-letnici*. 208–218.
- ROŽAC, Tina, 2013: Besedilne vrste v vsakdanjih narečnih pogovorih. Andreja Žele (ur.): *Družbena funkcijskost jezika (vidiki, merila, opredelitve)*. *Obdobja* 32. Ljubljana: Znanstvena založba Filozofske fakultete. 351–357
- SMOLE, Vera, 1994: Folklorist med prevajanjem in zapisovanjem. *Traditiones* 23. Ljubljana: Založba ZRC, ZRC SAZU. 143–154.
- STANONIK, Marija, 2001: *Teoretični oris slovenske folklorne*. Ljubljana: Založba ZRC, ZRC SAZU.
- ŠKOFIC, Jožica, 2001: *Govorica jih izdaja: skica govora kovaške Krope*. Kropa: Tomco.
- ŠKOFIC, Jožica, 2006: Prevajanje govorenega narečnega besedila v pisani knjižni jezik. Mihaela Koletnik, Vera Smole (ur.): *Diahronija in sinhronija v dialektoloških raziskavah*. Maribor: Slavistično društvo. 174–182.
- ŠUMENJAK, Klara, 2012: Zasnova dialektološkega korpusa na primeru govora Koprive na Krasu. Boža Krakar Vogel (ur.): *Slavistika v regijah – Koper*. Ljubljana: Zveza društev Slavistično društvo Slovenije, Znanstvena založba Filozofske fakultete. 73–78.
- ŠUMENJAK, Klara, VIČIČ, Jernej, 2012: Izzivi pri izdelavi dialektološkega korpusa GOKO. Boža Krakar Vogel (ur.): *Slavistika v regijah – Koper*. Ljubljana: Zveza društev Slavistično društvo Slovenije, Znanstvena založba Filozofske fakultete. 79–87.
- ŠUMENJAK, Klara, 2013: *Opis govora Koprive na Krasu na osnovi dialektološkega korpusa*. *Doktorska disertacija*. Koper: Univerza na Primorskem, Fakulteta za humanistične študije Koper. <http://share.upr.si/fhs/PUBLIC/doktorske/Sumenjajak-Klara.pdf>
- VERDONIK, Darinka, 2006: *Analiza diskurza kot podpora sistemom strojnega simultanege prevajanja govora*. *Doktorska disertacija*. Filozofska fakulteta Univerze v Ljubljani: Oddelek za slovenistiko.
- VITEZ, Primož, ZWITTER VITEZ, Ana, 2004: Problem prozodične analize spontanega govora. *Jezik in slovstvo* 49/6. Ljubljana: Slavistično društvo Slovenije. 3–24.
- VOLK, Jana, 2011: *Intonacija v spontanem neformalnem govoru slovenskih govorcev na dvojezičnem področju Slovenske Istre*. *Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- WEISS, Peter, 1998: *Slovar govorov Zadrebke doline med Gornjim gradom in Nazarjami*. Ljubljana: Založba ZRC, ZRC SAZU.
- ZEMPLJAK JONTES, Melita idr., 2002: Računalniški simbolni fonetični zapis slovenskega govora. *Slavistična revija* 50/2. 159–169.
- ZEMPLJARIČ MIKLAVČIČ, Jana, 2004: Taksonomija besedilnih tipov za gradnjo govorenega korpusa. Erika Kržišnik (ur.): *Aktualizacija jezikovnozvrstne teorije na Slovenskem: členitev jezikovne resničnosti*. *Obdobja* 22. 503–522.
- ZEMPLJARIČ MIKLAVČIČ, Jana, 2008: *Govorni korpusi*. Ljubljana: Oddelek za prevajalstvo.
- ZULJAN KUMAR, Danila, 2007: *Narečni diskurz*. Ljubljana: Založba ZRC, ZRC SAZU.
- ZWITTER VITEZ, Ana idr., 2009: Načela transkribiranja in označevanja posnetkov v referenčnem govornem korpusu slovenščine. Marko Stabej (ur.): *Infrastruktura slovenščine in slovenistike*. 437–442.