

# STANDARDI TRANSKRIBIRANJA IN OZNAČEVANJA NAREČNEGA KORPUSA GOKO

KLARA ŠUMENJAK

Univerza na Primorskem, Fakulteta za humanistične študije, Koper, Slovenija  
klara.sumenjakk@fhs.upr.si

V članku so predstavljena nekatera načela gradnje GOKO (Govorni korpus Koprive na Krasu), prvega slovenskega narečnega korpusa, ki je dostopen na <http://jt.upr.si/GOKO/>. Na kratko so opisani obseg korpusa, demografsko vzorčenje, posnetki in njihova enota ter označevanje korpusa. Osrednji del je posvečen standardom transkribiranja korpusa in izzivom, s katerimi se je bilo treba soočiti. Opisani so vsi trije nivoji, v katerih je bil korpus zapisan: a) v fonetičnem zapisu, ki upošteva vse glasoslovne značilnosti govora Koprive na Krasu, b) v poenostavljenem narečnem zapisu, kjer so prikazane samo temeljne glasoslovne značilnosti krajevnega govora in c) v poknjiženi različici, kjer je vsaka posamezna beseda zamenjana s svojo knjižno ustreznico, na ravni besedne zveze in stavka pa se ohranjajo posebnosti govorjenega jezika. V zaključku so predstavljeni predlogi za izboljšave korpusa in možnosti za njegovo uporabo. Njegova vrednost ni zgolj v dialektoloških raziskavah, temveč tudi v izobraževanju in ohranjanju kulturne in jezikovne dediščine.

DOI  
[https://doi.org/  
10.18690/um.ff.4.2024.6](https://doi.org/10.18690/um.ff.4.2024.6)

ISBN  
978-961-286-882-6

**Ključne besede:**  
narečni/dialektološki  
korpus,  
GOKO,  
transkripcija,  
narečje Koprive na Krasu



Univerzitetna založba  
Univerze v Mariboru

DOI  
[https://doi.org/  
10.18690/um.ff.4.2024.6](https://doi.org/10.18690/um.ff.4.2024.6)

ISBN  
978-961-286-882-6

**Keywords:**  
dialectal/dialectological  
corpus,  
GOKO,  
transcription,  
Kopriva na Krasu dialect

# GOKO DIALECT CORPUS TRANSCRIPTION AND ANNOTATION STANDARDS

KLARA ŠUMENJAK

University of Primorska Faculty of Humanities, Koper, Slovenia  
[klara.sumenjak@fhs.upr.si](mailto:klara.sumenjak@fhs.upr.si)

The article presents some principles of the construction of GOKO (Govorni korpus Koprive na Krasu), the first Slovene dialect corpus accessible at <http://jt.upr.si/GOKO/>. It briefly describes the corpus scope, the demographic sampling, the recordings and their units as well as the corpus annotation. The central part is dedicated to the standards of the transcription of the corpus and the challenges that had to be overcome. All three levels in which the corpus was recorded are described: a) in phonetic transcription, which captures all the phonological features of the Kopriva na Krasu dialect, b) in simplified dialect transcription, which represents only the basic phonological features of the regional speech, and c) in the literary version, in which each individual word is replaced by its literary counterpart, while retaining the features of the spoken language at the phrase and sentence level.



## 1 Uvod<sup>1</sup>

Korpus GOKO (Govorni korpus Koprive na Krasu), ki je dostopen na spletni strani <http://jt.upr.si/GOKO/>, je leta 2013 nastal kot eden izmed ciljev doktorske disertacije *Opis govora Koprive na Krasu na osnovi dialektološkega korpusa*.<sup>2</sup> Gre za prvi slovenski narečni korpus, ki odpira številna vprašanja in možnosti za nadgradnjo.<sup>3</sup>

Načela gradnje korpusa GOKO lahko v grobem razdelimo na tri glavne teme: 1) izhodišča za terensko delo, 2) pravna podlaga za gradnjo korpusa in 3) izhodišča za gradnjo korpusa.

V prispevku je predstavljena zadnja tema, ki obsega velikost korpusa, demografsko vzorčenje, posnetke in kriterije za njihov izbor ter označevanje korpusa in transkripcijo, ki sta osrednji temi prispevka. V prispevku so predstavljeni tudi nekateri pomisleki in predlogi za izboljšavo korpusa, ki so se porodili v desetih letih od njegovega nastanka.

## 2 Izhodišča za gradnjo korpusa

### 2.1 Obseg korpusa

Obseg korpusa je odvisen od njegovega namena (več o tem npr. Vintar 2008; Zemljarič Miklavčič 2008, Verdonik in Zwitter 2020). Ker je korpus GOKO nastal kot poskus gradnje narečnega korpusa zgolj ene vasi, je njegov obseg v primerjavi z ostalimi govornimi korpusi (prim. Artur, Gos 2.0, IRISS in SPOG) izjemno majhen, saj obsega le približno eno uro posnetkov (na terenu je bilo zbranih približno 42 ur posnetkov), kar ustreza približno 10.000 besedam. Kljub majhnosti pa je za poskus gradnje takega korpusa število besed dokaj reprezentativno, saj poleg predstavitve posebnosti na leksikalni ravnini omogoča tudi analizo na oblikoslovni<sup>4</sup> in (delno) na skladijski ravnini. Tak korpus je dobro izhodišče za izdelavo ustrezne metodologije za nadaljnjo gradnjo drugih narečnih korpusov, predvsem pa prikaže izzive, s katerimi smo se soočali pri njegovem nastajanju (Šumenjak 2013a).

---

<sup>1</sup> Prispevek je nastal v okviru raziskovalnega projekta ARIS Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642).

<sup>2</sup> Korpus je realiziral dr. Jernej Vičič.

<sup>3</sup> Več o korpusu gl. Šumenjak (2012, 2013a in 2013b).

<sup>4</sup> V člankih Šumenjak (2016 in 2019) je predstavljena oblikoslovna analiza koprivskih samostalnikov, ki je nastala s pomočjo korpusne metode.

## 2.2 Demografsko vzorčenje

Demografski podatki v korpusu so: starost, spol informanta, kraj snemanja, zaposlitev in informacija o tem, ali je informantov mož/žena iz druge vasi, državljanstvo in socialni status.<sup>5</sup>

Zagotovo pa lahko vključimo tudi druge demografske kriterije, ki so pomembni pri gradnji specifičnih korpusov – npr. kraj bivanja, kraj rojstva, verska pripadnost idr. (Zemljarič Miklavčič 2008: 59), zagotovo pa bi lahko vključili dodatne metapodatke, v skladu s tistimi, ki se pripravljajo za projekt Mezzanine (prim. tudi Verdonik in Zwitter Vitez 2020).

## 2.3 Posnetki

Med letoma 2009 in 2012 je bilo v Koprivi na Krasu s šestnajstimi informanti, starimi od 10 do 93 let, posnetih približno 42 ur pogovorov. V korpus je zajetih okrog 60 minut posnetkov šestih informantov, starih od 71 do 93 let. V korpus so bili vključeni posnetki, ki so bili tehnično in vsebinsko ustrezni in pretežno monološki – s čim manj vključevanja izpraševalca.<sup>6</sup>

Zaradi enostavnejšega upravljanja s posnetki so ti razdeljeni na manjše enote in shranjeni v formatu MP3 (Šumenjak 2013a: 53).

## 2.4 Enota

Vsako prekodiranje govorne besede v pisano je hkrati interpretacija, odvisna tudi od zapisovalca. Zato se pri prekodiranju (narečnega) govora moramo odločiti, kaj je osnovna enota analize, saj se ta bistveno razlikuje od enote v pisnem jeziku, kjer je osnovna enota poved, zamejena z veliko začetnico in s končnim ločilom (Šumenjak 2013a, Šumenjak 2013b). Če je v pisnem jeziku osnovna enota jasna, je v govornem jeziku ta težje določljiva, ker »izrek 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

---

<sup>5</sup> Podatka o državljanstvu in socialnem statusu govorca sta nerelevantna za naš korpus, saj je prvi predpogoj za izbiro informanta ta, da je informant aktivni govorec izbranega narečja, tudi status vseh informantov, vključenih v korpus, je bil enak, njegova določitev pa je lahko netočna in subjektivna, a smo jih vseeno vključili, če bi se morda kdaj korpus nadgrajevalo in združevalo z drugimi korpusi.

<sup>6</sup> Natančnejše o kriteriju zajemanja posnetkov v korpus gl. Šumenjak 2013b.

pogoste rabe prirednih veznikov, nedokončanih stavčnih vzorcev ipd.« (Zuljan Kumar 2007: 32) Čeprav osnovno enoto lahko imenujemo *izrek*, *govorjeni odstavek*, *izjava* ali kako drugače, je vsem skupno to, da se osnovne enote spontanega govora ločujejo na podlagi prozodičnih lastnosti in premorov v govoru. V GOKO je bila v daljših izrekih, kjer govorec ni delal premorov, postavljena meja med posameznimi izjavami ob govorčevem vdihu (prim. tudi Zemljarič Miklavčič 2008: 132 in Verdonik 2006: 50).

## 2.5 Raba ločil ter velikih in malih začetnic

Da bi zapisanemu tekstu lažje sledili, je prekodirano besedilo opremljeno z ločili in velikimi začetnicami. Ločila (predvsem vejice in pike) so zapisana bolj arbitrarno, kot je to značilno za zapisan jezik, saj so postavljena tudi ob daljšem premoru oz. ob popravljenem zareku. Končna ločila pa so postavljena tudi takrat, ko naj bi izrek predstavljal neko zaključeno misel (Šumenjak 2013a 36–38), saj vemo, da ločila pri zapisovanju govora »samo približno prenašajo fonetične in prozodične informacije, ki jih v govorjenem jeziku opravlja intonacija.« (Zuljan Kumar 2007: 32).

## 2.6 Diskurzne in druge oznake

Nekateri akustični ali drugi dogodki so lahko ključni za razumevanje in interpretacijo govorenega diskurza, zato bi bilo smiselno razpolagati tudi z video posnetki, a je zbiranje in analiza takšnega gradiva veliko bolj zahtevno in zamudno (prim. Maurer-Lausegger 2010).

Čeprav priporočila za TEI predvidevajo tudi označevanje prozodičnih in neverbalnih dogodkov v govornih besedilih (prim. Zemljarič Miklavčič 2008: 97–98), je bilo tako označevanje za GOKO nerelevantno in precej zamudno – na voljo pa je tudi posnetek, kjer so akustični dogodki slišni (Šumenjak 2013a: 40).

Kljub zapisanemu pa je bilo besedilo vseeno opremljeno z nekaterimi oznakami, ki so relevantne za razumevanje vsebine.

To je (po Šumenjak 2013a: 56–59):

<neraz/> nerazumljivo. Pri transkribiranju govora, ne moremo vedno natančno ugotoviti, kaj je govorec povedal, saj lahko prihaja do različnih interferenc. Take dele bi lahko izpustili, a

zaradi težnje po čim bolj natančnem označevanju korpusa je na mestih, kjer kot zapisovalka nisem razumela, kaj je informant/informantka povedal/povedala zapisan <?>. Ta znak lahko označuje besedo, besedno zvezo ali stavek. Uporabnik ima možnost izbrati, ali je pri prikazu zadetkov iskanja ta oznaka vidna ali ne.

Primer zapisovanja s posebnimi oznakami: *ra'zuməš, in <neraz/> 'tir jə bu 'nutər*

Zapis, viden uporabniku korpusa: *ra'zuməš, in <?> 'tir jə bu 'nutər*

<nedokon value='się...?'/> začetek nedokončane besede. Zapisovalci se ponavadi odločijo, da ne zapisujejo mest zarekov (prim. Ivančič Kutin, 83). Sama teh podatkov nisem želela ne izpustiti in ne enačiti s pojavom, ko je celotna beseda ali besedna zveza nerazumljena. V večini primerov gre za samopopravljanje govorca, ko ta želi najprej povedati določeno stvar, nato pa si premisli in pove drugo, včasih pa informant želi nekemu podatku dodati še eno informacijo, na primer: Ker se mi je zdelo, da bi tudi ti podatki lahko pomembni, sem jih vseeno vključila v označevanje. Največkrat je šlo za samopopravljanje pripovedovalca, ki je najprej želel povedati eno stvar, nato pa si premisli in je povedal nekaj drugega. V nekaterih primerih pa je informant želel povedano informacijo dopolniti, na primer:

Primer: *ne 'vięm, <nedokon value='się...?'/> 'anix 'siędəm 'metrou*

Isti primer, kot je viden uporabniku: *ne 'vięm, <się...> 'anix 'siędəm 'metrou*

Pojasnilo k primeru: Informant je najprej želel povedati, da je spomenik visok sedem metrov, a ker ni bil prepričan v točnost podatka, je pred števniki dodal nedoločni člen *enib*, ki izraža približnost povedanega. V korpusu je to označeno tako, da so zabeleženi izrečeni glasovi, čeprav niso tvorili cele besede (npr. <się...>, ko je informant želel najprej izreči besedo *sedem*). Uporabnik ima možnost izbrati, ali je pri prikazu zadetkov iskanja ta oznaka vidna ali ne.

<polglas/> ponavadi označuje informantovo obotavljanje. Uporabnik ima možnost izbrati, ali je pri prikazu zadetkov iskanja ta oznaka vidna ali ne, prikazana pa je kot polglasnik z dolžino (ə:).

Primer: *in <polglas/> 'tu jə <polglas/> 'čudno*

Isti primer, kot je viden uporabniku: *in ə: 'tu jə ə: 'čudno*

<nav> začetek navedka, </nav> konec navedka. Uporabnik korpusa to vidi zapisano z znakoma » in «. Čeprav sprva ni bilo predvideno ločevanje premege govora znotraj pripovedovanja, je bilo kasneje ugotovljeno, da se lahko večkrat zgodi, da informant pri navajanju besed druge osebe spremeni register in barvo glasu ter uporablja besede, ki niso značilne za koprivski govor, in zato taka besedila niso bila vključena v nadaljnjo analizo govora, ker niso bila več reprezentativna za obravnavani govor.

Primer: *zay'liədəm v'se t'ri 'təm pɾɛt, pɾɛd u'rati pɾɛd ɣos'tilno, <nav> P'ridi, p'ridi! </nav> jəs 'tečɛm čɛ 'ɣor. <nav> 'O, 'rauno p'raɥ, da si p'rišla, se ɣ'remo iɣ'rat ɥ sa'lɔn, </nav>*

Isti primer, kot je viden uporabniku: *in 'pol zay'liədəm v'se t'ri 'təm pɾɛt, pɾɛd u'rati pɾɛd ɣos'tilno, <»> P'ridi, p'ridi! <«> 'jəs 'tečɛm čɛ 'ɣor. <»> 'O, 'rauno p'raɥ, da si p'rišla, se ɣ'remo iɣ'rat ɥ sa'lɔn, <«>*

[ ] – med oglatima oklepajema so zabeležene izjave in neverbalni glasovi izpraševalke, saj so lahko vplivale na potek pripovedi informanta, niso pa vključene v nadaljnjo analizo govora, saj prihaja izpraševalka iz drugačnega krajevno govornega okolja.

Uporabnik ima možnost izbrati, ali je pri prikazu zadetkov iskanja ta oznaka vidna ali ne.

Primer: *'misləm, də 'vięš, kə'ku so 'təkrət [Aha] 'pisəli, 'ne, [Aha] šte'vilkə in <ə:> i'mię [Aha].*

Isti primer, kot je viden uporabniku: *'misləm, də 'vięš, kə'ku so 'təkrət [Aha] 'pisəli, 'ne, [Aha] šte'vilkə in <ə:> i'mię [Aha].*

<smeh/> in podobni neverbalni glasovi so lahko označeni opisno <smeh> ali fonetično <ha ha>. Ker je težko natančno določiti število zlogov medmetov, ki se lahko razlikujejo pri vsakem izreku, in ker je težko zapisati njihovo zvočno podobo, so v korpusu taki neverbalni glasovi označeni z oznako <smeh/>.

Uporabnik ima možnost izbrati, ali je pri prikazu zadetkov iskanja ta oznaka vidna ali ne.

Primer: *mət t'rięmi 'lipəmi in t'rięmi do'linəmi, də jə zakə'pana <smeh/> <nar> š'tięrna </nar> ce'kinou*

Isti primer, kot je viden uporabniku: *mət t'riɛmi 'lipəmi in t'riɛmi do'linəmi, də jə zakə'pana <smeh> š'tiɛrna ce'kinou*

<premor/> znak označuje vsak tihi premor, ki traja več kot 2 sekundi.

Primer: *in je <premor/> blo nalo'ženo, 'piše 'nutər*

Isti primer, kot je viden uporabniku: *in je <premor> blo nalo'ženo, 'piše 'nutər*

Primer besedila z vsemi dodatnimi oznakami:

*'Aaa, 'uasu, 'tu je pa <ə:> d'ryɣa z'ɣodba, 'tu so <ə:> nər'dili 'Aɥbərɕi, 'Aɥbərɕi so 'tu nər'dili, so pərpe'lali 'uəsla in pər'vɛzəli zə ta 'mali z'ɣun jin mi 'dali an 'kɥəs <?> se'na, 'ne, in kə je zɣ'rabu, jə <ə:> zəjɣɣu, 'ne, pe je zaz'ɣənu, 'polej so 'Kup'riɥci š'li pəɣ'lɛdət, kə'du z'ɣəne, 'kaj z'ɣəne, 'ne, 'pol so 'vidli, da jə 'uasu, in 'pol so z'vɛdli, ɣ'du je 'uəsla pər'pelaɥ, 'ne, ɣa so pe'lali ɥ 'Aɥber jən 'təm ɥ 'Aɥberi zə'bili c'vek mi ɥ ɣ'lavu, 'ne, [U!] 'ja, in 'polej <ə:> 'Aɥbəlci so c'vekərji.*

Primer besedila brez dodatnih oznak:

*'Aaa, 'uasu, 'tu je pa d'ryɣa z'ɣodba, 'tu so nər'dili 'Aɥbərɕi, 'Aɥbərɕi so 'tu nər'dili, so pərpe'lali 'uəsla in pər'vɛzəli zə ta 'mali z'ɣun jin mi 'dali an 'kɥəs se'na, 'ne, in kə je zɣ'rabu, jə zəjɣɣu, 'ne, pe je zaz'ɣənu, 'polej so 'Kup'riɥci š'li pəɣ'lɛdət, kə'du z'ɣəne, 'kaj z'ɣəne, 'ne, 'pol so 'vidli, da jə 'uasu, in 'pol so z'vɛdli, ɣ'du je 'uəsla pər'pelaɥ, 'ne, ɣa so pe'lali ɥ 'Aɥber jən 'təm ɥ 'Aɥberi zə'bili c'vek mi ɥ ɣ'lavu, 'ne, 'ja, in 'polej 'Aɥbəlci so c'vekərji.*

### 3 Transkripcija

Transkripcija govornih besedil je bila najzahtevnejše in kompleksnejše vprašanje, ki smo se mu posvetili pri gradnji korpusa GOKO, niti danes pa ni mogoče določiti smernic, ki bi bile splošno sprejete.

S transkripcijo govornih besedil so se ukvarjali tudi številni drugi slovenski avtorji, – tako slovstveni folkloristi/etnologi, dialektologi, fonetiki in korpusni jezikoslovci (prim. Grošelj idr. 2016, Ivančič Kutin 2011, Jurgec idr. 2005, Karničar 2008, Kenda Jež idr. 2007, Kenda Jež 2011, Klinar idr. 2012, Smole 1994, Stanonik 2001, Škofic 2006, Šumenjak 2013a, Šumenjak 2013b, Verdonik in Bizjak 2023, Verdonik idr. 2023, Verdonik in Zwitter Vitez 2020, Volk 2012, Zemljak Jontes idr. 2002, Zwitter



Vitez idr. 2009), odgovora, kako besedilo zapisovati, da bi zadovoljili potrebe vseh strok, pa (še) nimamo.

Vsak govorni dogodek je enkratno dejanje, interpretacij zapisa pa je lahko toliko, kolikor je zapisovalcev. Npr. dialektologe večinoma zanima zgolj fonetična transkripcija, kjer je natančno transkribirano besedilo bolj pomembno od količine gradiva, tako zapisano gradivo pa je navadno namenjeno predvsem dialektologom in drugim jezikoslovcem, medtem ko je gradivo s prilagojeno transkripcijo (prim. GOS, ŠOLAR ipd.) lahko namenjeno tudi oz. morda predvsem širši javnosti. Po drugi strani so za npr. sociolingvistično, folkloristično ali diskurzivno raziskovanje bolj pomembne informacije o kontekstu kot sama natančnost zapisa.

Posneto gradivo lahko prekodiramo v pisno obliko na več načinov: naprimer tako, da se ohrani natančne narečne glasoslovne značilnosti (fonetični zapis) ali pa se besedilo glasoslovno-pravopisno poknjži (ortografski zapis), a vseeno zapiše tudi nekatere izrazitejše narečne foneme (npr. polglasnik, zapisan kot *ə*, diftonge ipd.). Tako kljub poenostavljenemu fonetičnemu zapisovanju so še vedno vidne značilnosti na oblikoslovni, skladenjski in leksikalni ravni.

Tudi odločitev o stopnji prekodiranja oz. načinu transkripcije govornega narečnega besedila je odvisna predvsem od namena uporabe gradiva in samega naslovnika.

Za korpus GOKO je bilo gradivo prekodirano na tri načine, in sicer uporabljen je fonetični zapis, poenostavljeni narečni zapis in poknjžena različica (Šumenjak 2013a in Šumenjak 2013b). Ti trije načini (po Šumenjak 2013a in 2013b) so predstavljeni v nadaljevanju.

Pri zapisovanju govora se navadno izgubljajo pomembni akustični segmenti (npr. intonacija, poudarki, smeh, mimika, gestikulacija), ki jih pripovedovalec uporablja za ustvarjanje atmosfere, zato je pomembno, da je korpus opremljen s posnetki, ki delno nadomeščajo t. i. didaskalije ali opombe o redakciji, vsebini in sobesedilu<sup>7</sup>.

V slovenski dialektologiji se uporablja t. i. »nova nacionalna« fonetična transkripcija, ki je dogovorjen sistem znakov, ki poleg naglasa (jakostnega/tonemskega) zaznamuje tudi kakovost in kolikost glasov. Čeprav pri zapisovanju s fonetično

---

<sup>7</sup>Več o opombah glej Ivančič Kutin (2011: 83).

transkripcijo ohranimo največ značilnosti povedanega, je taka transkripcija izjemno zahtevna tudi za dialektologa, od njega zahteva večkratno natančno poslušanje gradiva in pri prekodiranju tudi ogromno znanja, zbranosti, potrpljenja in posluha.

Ker je korpus GOKO namenjen tako dialektologom kot tudi širši (nejezikoslovni) javnosti, je zbrano gradivo zapisano v treh različicah, opremljenih z zvočnimi posnetki, in sicer a) v fonetičnem zapisu, ki upošteva vse glasoslovne značilnosti govora Koprive na Krasu – ta oblika je namenjena jezikoslovni, zlasti dialektološki analizi zapisanega govora; b) v poenostavljenem narečnem zapisu, kjer so ohranjene temeljne glasoslovne značilnosti krajevnega govora (zapisani so npr. mesto naglasa, diftongi, polglasnik ...), in c) v poknjiženi različici, kjer je vsaka posamezna beseda zamenjana s svojo knjižno ustreznico, na ravni besedne zveze in stavka pa se ohranjajo posebnosti govorjenega jezika. Taka različica je nujna za iskanje po korpusu, saj bo uporabnik najverjetneje iskal po knjižni besedi, npr. *zgodba*, ne pa po ortografskem zapisu narečne besede, npr. *štorja*.

Glavno vprašanje pri poenostavljeni različici zapisa je (bilo), do kolikšne mere poenostaviti besedilo, da je različica še vedno reprezentativna za zapisani narečni govor, zato je pomembno pri zapisu ohraniti večino glasoslovnih posebnosti krajevnega govora, vendar pa jih zapisati z znaki, ki jih lahko berejo tudi laični uporabniki (torej s črkami knjižne abecede, ki jim lahko dodamo znak za polglasnik ipd.).

Komentirani primeri transkripcije besedila v korpusu (po Šumenjak 2013a in Šumenjak 2013b):

**Fonetični prepis** je nastal na podlagi analize fonološkega sistema govora Koprive (Šumenjak 2013c) in je pripravljen v skladu s pravili slovenske fonetične transkripcije. Tak zapis je kompleksen, zato je njegovo branje zahtevno in namenjeno predvsem naslovniku, ki tak zapis pozna, najverjetneje pa je neprimeren in prezahteven za širši krog bralcev.

Primer fonetičnega prepisa: *Smo 'tərko š'kuəde nər'dili čə'bijələm, 'ne, <nar> 'ma </nar> mula'rija 'təkət, ki smo blę, 'vięš, kə'ku je, <ə:> in smo 'tisti 'mięt, 'vięš, 'ku sə jə u'lijęku 'tisti 'mięt, 'ne ku z'dej, 'tu si 'mięla u'se po 'rokəx <?> smo š'li se u'mit in 'tako, 'ne, o'joj, sə smo 'mięl 'dobro, 'ja.*

Najbolj problematičen pri zapisovanju gradiva je bil diftong *iɛ*, saj se je glas *e* v drugem delu diftonga pri različnih informantih drugače realiziral, enkrat bolj *e*-jevsko, drugič bolj polglasniško. Enako težavo pri zapisovanju je zaradi različnih realizacij povzročal oslabljeni monoftong *ɐ* (Šumenjak 2013c).

Težavo bi bilo mogoče rešiti z ustreznim naborom posnetkov enako in podobno zapisanih fonemov, njihovo strojno analizo in določitvijo mej med različnimi realizacijami samoglasnikov (prim. Jurgec 2019, Robida 2020, Volk 2018).

Pri **poenostavljenem prepisu** moramo biti pozorni na stopnjo poenostavljanja, saj želimo ohraniti tipične glasoslovne značilnosti (koprivskega) govora, hkrati pa moramo zagotoviti, da je tak zapis sprejemljiv tudi za širši krog naslovnikov. Namenjen je uporabnikom, ki jih ne zanima zgolj vsebina povedanega, ampak želijo izvedeti tudi o nekaj osnovnih narečnih posebnostih zbranega gradiva. Zaradi tega je v GOKO ohranjen zapis diftongov in polglasnikov, niso pa upoštevane npr. različne kvalitete samoglasnikov, prav tako sta nezveneči in zveneči velarni pripornik *x* in *ɣ*, ki sta značilna za kraško narečje, zaradi lažje berljivosti zapisana s črkama *b* in *g*. Zapisano pa je tudi naglasno mesto, ki je označeno nevtralnimi diakritičnim znamenjem ' nad naglašnim samoglasnikom.

Seznam poenostavljenih zapisov fonemov v korpusu GOKO:

u → w  
 i → j  
 y → ü  
 ɐ → e  
 ɣ → g  
 x → h

Primer poenostavljenega prepisa:

*Smo tårko škwàde nàrdli čebièləm, nè, <nar> mà </nar> mularija tåket, ki smo ble, vièš, kəkù je, <a> in smo tisti mièt, vièš, ku sà jə wlièku tisti mièt, nè ku zdej, tù si mièla wsè po ròkəh <?> smo šli se wmit in tåko, nè, ojòj, sà smo mièl dóbno, já.*

**Knjižna (tj. na glasoslovni, oblikoslovni in leksikalni ravni poknjžena) različica** je pomembna predvsem kot izhodišče za iskanje po korpusu in za izdelavo diferencialnega slovarja, uporabniki korpusa pa se sami odločijo, ali se jim ta različica prikaže ali ne.

Primer poknjžene različice:

*Smo toliko škode naredili čebelam, ne, <nar> ma saj </nar> mularija takrat, ker smo bili, veš, kako je, <ə:> in smo tisti med, veš, kako se je vlekel tisti med, ne kot zdaj, to si imela vse po rokah <?> smo šli se umit in tako, ne, ojoj, se smo imeli dobro, ja.*

### 3. govorec3

#### Fonetični zapis:

In 'pòle si je, <ə:> zɣ'ljèda, də so 'tydi ɥ 'Pòstòjní 'dost 'dʲɛlli, zə 'Tərst zə 'marsikéj, 'ne ve'likix 'dʲɛl, 'ámpək so 'dʲɛlli, zə'stópəš, [Aha] <ə:> in tək'ku zɣ'ljèda, da jə 'jəmu ɥ'sak 'suɣ kamno 'lómčək in 'pač, 'kúkər jə š'lo, jə š'lo, 'ne, 'sej 'vięš [Aha], 'tu jə 'taka š'torje, 'təm ɥət 'təwžənt ɥəsəms 'tu 'de'viède'sedya du 'pól, 'kònc 'vojne, 'vięš 'təm, kə jə bla 'wəjska kònčana 'pərva.

#### Poenostavljeni zapis:

In 'pòle si je, <ə:> zglièda, də so t Poknjženo: zgodba 'təst zə 'marsikéj, 'ne velikih dièl, ámpək so dièlli, zə'stópəš, [Aha] <ə:> in təkú zglièda, da jə 'jəmu vsək súj kamnolómčək in pač, kúkər jə š'lo, jə š'lo, nē, séj viéš [Aha], tū jə táká š'torje, tēm wət 'təwžənt wəsəms'tu 'deviédesédga du 'pól, 'kònc 'vojne, viéš tēm, kə jə bla 'wəjska kònčana pərva.

#### Poknjženi zapis:

In potem se je, <ə:> zgleda, da so tudi v Postojni dosti delali, za Postojno, za Trst za marsikaj, ne velikih del, ampak so delali, razumeš, [Aha] <ə:> in tako zgleda, da je imel vsak svoj kamnolomček in pač, kakor je šlo, je šlo, ne, saj veš [Aha], to je taka zgodba, tam od tisoč osemstodevetdesetega do potem, 'kònc vojne, veš tam, ko je bila vojna končana prva.

**Slika 1: Primer iskanja po besedi *kònc* in prikaza pomena narečne besedne zveze v oblaku**

Vir: GOKO

Posebne oznake besed, kot so npr. li = lastno ime, nar = narečno ipd. je treba zapisovati ročno v vseh treh različicah, saj jih računalnik ne prepozna in zato ne označuje avtomatizirano. Večinoma so to besede, ki imajo v knjižnem jeziku bodisi drugačen spol bodisi drugačen pomen kot v obravnavanem govoru. Taki leksemi (v večini primerov gre za izposojenke, ki jih uporabnik navadno ne razume) so v korpusu prikazani drugače od ostalih besed (zapisani so ležeče in z drugačno barvo). Uporabnik lahko s pomikom kurzorja na neznano besedo prebere njen pomen, ki se prikaže v oblaku (gl. Sliko 1).

Slika 1 prikazuje rezultate iskanja besede *konec*. Ker gre za dialektološki korpus, je zato na prvem mestu prikazan fonetični prepis, nato je izpisana poenostavljena različica, na zadnjem mestu pa vidimo poknjženo varianto. Uporabnik se lahko glede na svoje potrebe sam odloči, katera različica bo prikazana med zadetki (lahko samo eno, dve ali vse tri). Z drugačno barvo so označene tudi narečne besede, s pomikom kurzorja nanje se v oblaku izpiše njihov pomen, kot je na sliki razviden primer za narečno besedo *štorja*, knj. zgodba.

### 3.1 ZRCola<sup>8</sup>

Po desetih letih od začetka gradnje korpusa GOKO ostaja še precej odprtih vprašanj o načinu in vrsti transkripcije narečnih besedil, čeprav se trenutno zdi smiselno, da bi zaradi širše uporabnosti gradiva namesto slovenske fonetične pisave uporabili mednarodni zapis IPA. Vendar je besedilo, zapisano z IPA, zaradi zapisovanja nekaterih glasov (prim. č, ž, š ...), nekoliko težje berljivo. Ravno zaradi specifik slovenske fonetične transkripcije je Peter Weiss na Znanstvenoraziskovalnem centru SAZU v Ljubljani za jezikoslovne, predvsem dialektološke potrebe razvil vnašalni sistem ZRCola, ki deluje v Microsoftovem programu Word v operacijskem sistemu Windows. Gre za sistem za vnos znakov, ki ga poznajo in uporabljajo predvsem slovenski dialektologi, pisavo pa se lahko uporablja tudi v ostalih programih, ki uporabljajo oz. podpirajo Microsoftov format pisav (*TrueType*), zato se jo lahko uporablja tudi pri modernejših spletnih brskalnikih. Pisava je bila uporabljena tudi pri oblikovanju spletnih strani, ki prikazujejo vsebino korpusa. Na Sliki 2 lahko vidimo del kode CSS (*Cascading Style Sheets*), ki ima vključeno pisavo ZRCola, zraven kode pa moramo imeti datoteko s to pisavo, shranjeno na strežnik. Datoteka je dostopna na ZRC SAZU ([Http://zrcola.zrc-sazu.si/](http://zrcola.zrc-sazu.si/)) (Šumenjak 2013a: 48–50, Šumenjak in Vičič 2012).

---

<sup>8</sup> Besedilo je bilo pripravljeno z vnašalnim sistemom ZRCola (<http://zrcola.zrc-sazu.si/>), ki ga je na Znanstvenoraziskovalnem centru SAZU v Ljubljani (<http://www.zrc-sazu.si/>) razvil Peter Weiss.

```

@font-face {
  font-family:"00 ZRCola";
  src: url('zrcola.ttf');
      src: local('@'),
  url('zrcola.ttf') format('opentype');
}

```

Slika 2: Del kode CSS, ki vključuje pisavo ZRCola

Vir: Jernej Vičič, osebni arhiv

### 3.2 Možnosti zapisovalnih/prepisovalnih napak

Največ napak lahko nastane pri transkripciji in doslednosti zapisovanja. Ker je besedilo zapisano ročno, mora zapisovalec paziti, da ne izpusti besede ali ločila, saj ročni vnos ne omogoča enakega preverjanja kot avtomatski zapis. Doslednost je ključna tudi pri uporabi posebnih oznak, npr. `<nar>` in `</nar>`, saj izpust ene izmed oznak prepreči računalniku prepoznavanje označenih besed (Šumenjak 2013a: 53).

Primer napake v prepisovanju (slika 3) – v fonetičnem zapisu je vejica za besedo jaz *'mati*, ki je v ostalih zapisih ni.

Poizvedba:

**Fonetični zapis:**  
 'Joj, 'jas u'sa ve'sela, ka'ku bəm do'ma po'vjeđla, ja blo za p'lačet d'vej 'liri, u'si so š'li, 'samo 'jas 'ne, 'ni blo d'vej 'liri za me 'dat, 'ma so i'mjełi, ki u'sək d'ryji 'dan, 'mati, smo <rek...>, 'hona bi rekli [Aha] z 'dej, 'ne, <a:> 'ma še 'bəl se s'pounem, 'tərko m'lječka ja 'mjela in 'ma, 'ma, 'kaj d'vej 'liri bi in 'Lujze ja 'dieləw 'təm, pər *Ferari* u 'tisti <?>.

**Poenostavljeni zapis:**  
 Jój, jas vša vesela, kakú bəm domá povedla, ja blo za plačet dvěj liri. Vsi so šli, samo jas né, ni blo dvěj liri za me dát, má so jmieli, ki vsək drugi dan, mati smo <rek...>, nona bi rekli [Aha] zděj, né, <a:> má še bəl se spounem, tərko mlička ja miela **m** má, má, kaj dvěj liri bi in Lujze ja dieləw təm, pər *Ferari* w tisti <?>.

**Poknjženi zapis:**  
 Joj, jaz vsa vesela, kako bom doma povedala, je bilo za plačati dve liri. Vsi so šli, samo jaz ne, ni bilo dveh lir za mi dati, ampak so imeli, ker vsak drugi dan, mati smo <rek...>, babica bi rekli [Aha] sedaj, ne, <a:> ampak še bolj se spomnim, toliko mleka je imela **m** ma, ma, kaj dve liri bi in Lojze je delal tam, pri Ferariju v tisti <?>.

Slika 3: Primer napake v prepisovanju

Vir: GOKO

Ker računalniški algoritmi za iskanje in primerjanje temeljijo na natančnem ujemanju nizov znakov, je v zgornjem v primeru, ko je fonetičnem zapisu dodana vejica, ki je v ostalih prepisih ni, privedlo iskanje v tem izreku do napačnih rezultatov. Računalnik namreč temelji na predpostavki, da so mesta v besedilu enaka za vsak nivo zapisa, izhodišče iskanja pa je poknjžena različica. To pomeni da, če je beseda *in* v poknjženi različici na 67 mestu, bodo kot rezultat iskanja besede *in* prikazalo znake/besede, ki so na 67 mestu v ostalih dveh nivojih zapisa. Kot je razvidno iz slike 3, je zaradi dodatno zapisane vejice za besedo *in* v fonetičnem prepisu, napačno prikazan rezultat za iskanja besede *in*. Glede na to, da se nam prikaže tudi sobesedilo iskane besede, napako vidimo, vendar bi bilo bolje, da do takih napak ne bi prihajalo.

Takih napak je bilo veliko, predvsem zato, ker je bilo besedilo zapisovano v Wordu, kar se je takrat (tudi zaradi slovenske fonetične transkripcije) zdelo najenostavnejše. Sicer pa, če so imeli zapisovalci GOS-a težave pri dvotirnem sistemu zapisovanja (prim. Verdonik 2013 in Verdonik 2014), je nekako jasno, da bo zapisovanje besedila na treh nivojih še večji izziv, čeprav bi bilo to z današnjimi orodji nekoliko lažje.

#### 4 Sklep

Pred desetimi leti se je morda zdelo, da korpusno jezikoslovje in dialektologija nimata veliko skupnega, čeprav ni tako. Sicer korpusni pristop ne more nadomestiti tradicionalne dialektološke metode raziskovanja, a jo lahko smiselno dopolni in vsebinsko nadgradi, predvsem pa omogoča široko dostopnost gradiva. Korpus GOKO je prvi dialektološki korpus in kot tak ima številne možnosti za izboljšavo.

#### Iskanje po korpusu

Ker je bil korpus primarno namenjen zgolj avtoričini analizi in obdelavi podatkov ter ugotovitvi primernosti pristopa za analizo gradiva, je bila zanemarjena grafična podoba korpusa in s tem tudi iskalnika ter okrnjene njegove možnosti. Zagotovo bi bilo dobro, če bi se lahko v korpusu iskalo tudi po posameznih fonemih, tako bi se hitreje pokazala njihova distribucija in morda odstopanja od pričakovanega. Foneme bi se lahko tudi dodatno označilo, npr. njihov izvor, kar bi nam pomagalo tudi pri fonološkem opisu.

Ker po korpusu iščemo v poknjženi različici, ne moremo iskati po narečnih leksemih, zato bi bilo smiselno dodati dvojnico, npr. da bi po korpusu lahko iskali ne le po besedi *zgodba*, ampak tudi po njeni narečni ustreznici *štorja*.

Glede na to, da korpus ne vsebuje veliko besed in je bilo prvotno mišljeno, da se bo zaradi analize gradiva iskalo zgolj po slovničnih kategorijah, bi bilo za širšo javnost smiselno dodati seznam besed, ki so vključene v korpus, tako da uporabniku ne bi bilo treba ugibati, po kateri besedi je mogoče iskati.

### Transkripcija

Vprašanja transkripcije najverjetneje še nekaj časa ne bo mogoče rešiti, zato bi bilo morda bolje, da bi govorni korpusi za posamezne kraje čimprej začeli nastajati (najverjetneje bi se lahko oblikoval nek »osnovni korpus«, ki bi bil že vnaprej pripravljen in označen v skladu z ostalimi korpusi (GOKO, Gos, Šolar), kamor bi lahko uporabniki vnašali gradivo. Najbolje bi bilo, da bi bili taki korpusi zapisani zgolj v poenostavljeni obliki (če jih ne bi zapisovali dialektologi), kjer pa bi vseeno skušali ohraniti čim več narečnih potez. Posnetki pa bi bili na voljo dialektologom, ki bi lahko gradivo uporabljali tudi za dialektološke raziskave.

### Premisleki o diskurzni in drugih oznakah

Zagotovo pri korpusnem označevanju velja, da je bogato označen korpus boljši, saj zagotavlja bolj celovito in podrobno zbirko podatkov, posledično pa nudi več možnosti za nadaljnjo analizo. Tak način označevanja sem izbrala, ker sem želela postaviti osnovo za razvoj oz. gradnjo še drugih narečnih korpusov, do česar pa ni prišlo. Če bi se ponovno odločala (vedoč, da bo korpus GOKO edini dialektološki korpus), teh oznak ne bi vključila v zapisovanje, saj se v dosedanjih opravljenih raziskavah (Šumenjak 2016, Šumenjak 2019) niso izkazale za relevantne. Poleg tega so na voljo tudi zvočni zapisi, ki zajemajo diskurzne značilnosti in so lahko koristni za tiste, ki bi jih to lahko zanimalo.

Ker pa nam korpusi omogočajo široko dostopnost jezikovnega gradiva in tudi ohranjanje narečja ter s tem naše kulturne jezikovne dediščine, se velja vprašati, ali smo lahko do neke mere fleksibilni in zanemarimo določena pravila, ki jih je sicer treba upoštevati pri oblikovanju korpusa (npr. transkripcija, označevanje itd.) v korist oživljanja, ohranjanja in širjenja (slovenske) jezikovne raznolikosti. Taki



korpusi bi lahko postali most med znanstvenim raziskovanjem in praktičnimi prizadevanji za ohranjanje jezikovne različnosti tudi na vseh ravneh izobraževanja – na primer kot didaktični pripomočki v osnovnošolskem, srednješolskem in fakultetnem izobraževanju. Učiteljem in profesorjem bi lahko olajšali celostno obravnavo slovenskega jezika, kulture in zgodovine in jim tako pomagali približati tudi raziskovanje narečij.

## Literatura

- Robert GROŠELJ idr., 2016: *Lipalja vas in njena slovenska govorica. 1. izd.* Ljubljana: Založba ZRC, ZRC SAZU.
- Barbara IVANČIČ KUTIN, 2011: *Živa pripoved v zapisu: kontekst tekstura in prekodiranje.* Ljubljana: Založba ZRC.
- Peter JURGEC idr., 2005: *Ovčja vas in njena slovenska govorica : raziskovalni tabor Kanalska dolina 2003 = Valbruna e la sua parlata slovena : stage di ricerca Val Canale 2003.* Ukve: Slovensko kulturno središče Planika, Kanalska dolina: = Centro culturale sloveno Stella alpina, Val Canale; Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.
- Peter JURGEC, 2019: Glasoslovne raziskave slovenskih narečij na Univerzi v Torontu. *Slovenski javni govor in jezikovno-kulturna (samo)zavest.* Ur. Hotimir Tivadar. (Obdobja 38.) Ljubljana: Znanstvena založba Filozofske fakultete. 59–69.
- Ludvik KARNIČAR, 2008: Fonetično zapisovanje narečnih etnoloških besedil. *Traditiones 37/1.* Ljubljana: Založba ZRC. 155–167.
- Karmen KENDA-JEŽ idr. 2007: *Sbranli smo jih v bančah : slovski prispevek k poznavanju oblačilne kulture v Kanalski dolini = contributo lessicale alla conoscenza dell'abbigliamento in Val Canale.* Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.
- Karmen KENDA JEŽ, 2011: Fonetična transkripcija. Ivančič Kutin, Barbara: *Živa pripoved v zapisu: kontekst tekstura in prekodiranje.* Ljubljana: Založba ZRC. 79–83.
- Klemen KLINAR idr. 2012: *Metode zbiranja hišnih in ledinskih imen: projekt FLU-LED v okviru Operativnega programa Slovenija - Avstrija 2007-2013.* Jesenice: Gornjesavski muzej.
- Herta MAURER-LAUSEGGER, 2010: Avdiovizualna dialektologija in multimedijska tehnika v dialektologiji. *Slovenska narečja med sistemom in rabo.* Znanstvena založba Filozofske fakultete. 319–24.
- Nejc ROBIDA, 2020: Slovenščina in sinteza govora Kaj je bilo, kaj bo in kaj je mogoče. *Varia XXVIII : zbornik prispevkov z XXVIII. kolokvija mladih jazykovedcov.* Ur. Jana Nemčeková, Patrik Petráš. Nitra: Univerzita Konštantína Filozofa. 229–235. Dostop 20. 8. 2023 na <https://www.juls.savba.sk/ediela/varia/28/Varia28.pdf>.
- Vera SMOLE, 1994: Folklorist med prevajanjem in zapisovanjem. *Traditiones 23.* Ljubljana: Založba ZRC. 143–154.
- Marija ŠTANONIK, 2001: *Teoretični oris slovstvene folklore.* Ljubljana: Založba ZRC.
- Jožica ŠKOFIC, 2006: Prevajanje govorenega narečnega besedila v pisani knjižni jezik. *Diabronija in sinbronija v dialektoloških raziskavah.* Ur. Marko Jesenšek. (Zora 41). Maribor: Slavistično društvo. 174–182.
- Klara ŠUMENJAK, 2012: Zasnova dialektološkega korpusa na primeru govora Koprive na Krasu. *Slavistika v regijah - Koper.* Ur. Boža Krakar Vogel. Ljubljana: Zveza društev Slavistično društvo Slovenije: Znanstvena založba Filozofske fakultete. 73–78.
- Klara ŠUMENJAK in Jernej VIČIČ, 2012: Izzivi pri izdelavi dialektološkega korpusa GOKO. *Slavistika v regijah - Koper.* Ur. Boža Krakar Vogel. Ljubljana: Zveza društev Slavistično društvo Slovenije: Znanstvena založba Filozofske fakultete. 79–87.

- Klara ŠUMENJAK, 2013a: *Opis govora Koprive na Krasu na osnovi dialektološkega korpusa*. Doktorska disertacija. Koper: Univerza na Primorskem, Fakulteta za humanistične študije Koper. Dostop tudi na <http://share.upr.si/fhs/PUBLIC/doktorske/Sumenjak-Klara.pdf>.
- Klara ŠUMENJAK, 2013b: Priprava gradiva in standardizacija nivojev zapisa za potrebe dialektološkega korpusa GOKO. *Družbena funkcija jezika (vidiki, merila, opredelitve)*. Ur. Andreja Žele. (Obdobja 32 – metode in zvrsti). Ljubljana: Znanstvena založba Filozofske fakultete. 443–449.
- Klara, ŠUMENJAK, 2013c: Glasoslovne značilnosti govora Koprive na Krasu (SLA T110). *Dialektološki razgledi*. Ur. Peter Weiss. Ljubljana: Založba ZRC. 149–162.
- Klara ŠUMENJAK, 2019: 1. in 2. sklanjatev samostalnikov ženskega spola v koprivskem govoru : uporabnost korpusne obdelave podatkov pri oblikoslovni analizi narečnega govora. *Annales: anali za istrske in mediteranske študije. Series historia et sociologia*. 29/2, 225–236.
- Klara ŠUMENJAK, 2016: Uporabnost korpusne obdelave podatkov pri oblikoslovni analizi narečnega govora : 1. sklanjatev samostalnikov moškega spola v koprivskem govoru. *Annales: anali za istrske in mediteranske študije. Series historia et sociologia*. 26/4, 741–750.
- Darinka VERDONIK, 2006: *Analiza diskurza kot podpora sistemom strojnega simultanege prevajanja govora*. Doktorska disertacija. Filozofska fakulteta Univerze v Ljubljani: Oddelek za slovenistiko.
- Darinka VERDONIK, 2014: Vprašanja zapisovanja govora v govornem korpusu Gos. *Jezikovne tehnologije : zbornik 17. mednarodne multikonference*. Ur. Erjavec, Tomaž, Žganec Gros, Jerneja. Ljubljana: Institut Jožef Stefan. 151–156.
- Darinka VERDONIK in Ana ZWITTER VITEZ, 2020: *Slovenski govorni korpus Gos*. 1. e-izd. Ljubljana: Znanstvena založba Filozofske fakultete.
- Darinka VERDONIK idr., 2013: Compilation, transcription and usage of a reference speech corpus: the case of the Slovene corpus GOS. *Language resources and evaluation*. 47/7, 1031–1048.
- Darinka VERDONIK in Andreja BIZJAK, 2023: *Pogovorni zapis in označevanje govora v govorni bazi Artur projekta RSDO*. Dostop 5. 4. 2024 na <https://dk.um.si/Dokument.php?lang=slv&id=170009&dn>.
- Darinka VERDONIK idr., 2023: *Standardizirani zapis v govorni bazi Artur projekta RSDO*. Dostop 5. 4. 2024 na <https://dk.um.si/Dokument.php?id=170007&lang=slv>.
- Špela VINTAR, 2008: *Terminologija: terminološka veda in računalniško podprta terminografija*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Jana VOLK, 2012: Using the ToBI transcription to record the intonation of Slovene. *Linguistica* 52. 169–186.
- Jana VOLK, 2018: Polglasnik v slovenskem spontanem govoru prebivalcev Obale. *Jezik in slovstvo*. 63/1, 19–37.
- Melita ZEMLJAK JONTES idr., 2002: Računalniški simbolni fonetični zapis slovenskega govora. *Slavistična revija* 50/2. 159–169.
- Jana ZEMLJARIČ MIKLAVČIČ, 2008: *Govorni korpusi*. Ljubljana: Oddelek za prevajalstvo.
- Danila ZULJAN KUMAR, 2007: *Narečni diskurz*. Ljubljana: Založba ZRC, ZRC SAZU.
- Ana ZWITTER VITEZ idr., 2009: Načela transkribiranja in označevanja posnetkov in referenčnem govornem korpusu slovenščine. *Infrastruktura slovenščine in slovenistike*. Ur. Marko Stabej. (Obdobja 28). Ljubljana: Znanstvena založba Filozofske fakultete. 437–442.