



Univerzitetna založba
Univerze v Mariboru



Stanje in perspektive uporabe govornih virov v raziskavah govora

Mira Krajnc Ivič
urednica



Univerza v Mariboru

Filozofska fakulteta

Stanje in perspektive uporabe govornih virov v raziskavah govora

Urednica

Mira Krajnc Ivič

Julij 2024

Naslov <i>Title</i>	Stanje in perspektive uporabe govornih virov v raziskavah govora <i>Status and Perspectives of the Use of Speech Resources in Speech Research</i>
Urednica <i>Editor</i>	Mira Krajnc Ivič (Univerza v Mariboru, Filozofska fakulteta)
Recenzija <i>Review</i>	Andreja Žele (Univerza v Ljubljani, Filozofska fakulteta) Tamara Gazdić-Alerić (Univerza v Zagrebu, Pedagoška fakulteta)
Jezikovni pregled <i>Language editing</i>	Mira Krajnc Ivič (slovenščina), Melita Zemljak Jontes (slovenščina), Katarzyna Konczewska (ruščina), Miloslav Vondráček (češčina), Gjoko Nikolovski (hrvaščina), Tadeja Tement (angleščina)
Tehnični uredniki <i>Technical editors</i>	Jan Perša (Univerza v Mariboru, Univerzitetna založba) Mira Krajnc Ivič (Univerza v Mariboru, Filozofska fakulteta) Ina Vivian Likar (Univerza v Mariboru, Filozofska fakulteta, študentka) Marina Bajić (Univerza v Mariboru, Univerzitetna založba)
Oblikovanje ovitka <i>Cover designer</i>	Samo Kramberger
Grafika na ovitku <i>Cover graphic</i>	Samo Kramberger
Grafične priloge <i>Graphic material</i>	Viri so lastni, razen če ni navedeno drugače. Krajnc Ivič (urednica), 2024
Založnik <i>Published by</i>	Univerza v Mariboru Univerzitetna založba Slomškov trg 15, 2000 Maribor, Slovenija https://press.um.si , zalozba@um.si
Izdajatelj <i>Issued by</i>	Univerza v Mariboru Filozofska fakulteta Koroška cesta 160, 2000 Maribor https://www.ff.um.si , ff@um.si
Izdaja <i>Edition</i>	Prva izdaja
Vrsta publikacije <i>Publication type</i>	E-knjiga

Dostopno na <http://press.um.si/index.php/ump/catalog/book/898>
Available at

Izdano Maribor, junij 2024
Published at



© Univerza v Mariboru, Univerzitetna založba
/ University of Maribor, University Press

Besedilo / *Text* © Krajnc Ivič (urednica), 2024

To delo je objavljeno pod licenco Creative Commons Priznanje avtorstva 4.0 Mednarodna. / *This work is licensed under the Creative Commons Attribution 4.0 International License.*

Uporabnikom je dovoljeno tako nekomercialno kot tudi komercialno reproduciranje, distribuiranje, dajanje v najem, javna priobčitev in predelava avtorskega dela, pod pogojem, da navedejo avtorja izvirnega dela.

Vsa gradiva tretjih oseb v tej knjigi so objavljena pod licenco Creative Commons, razen če to ni navedeno drugače. Če želite ponovno uporabiti gradivo tretjih oseb, ki ni zajeto v licenci Creative Commons, boste morali pridobiti dovoljenje neposredno od imetnika avtorskih pravic.

<https://creativecommons.org/licenses/by/4.0/>

CIP - Kataložni zapis o publikaciji
Univerzitetna knjižnica Maribor

808(0.034.2)

KRAJNC Ivič, Mira

Stanje in perspektive uporabe govornih virov v raziskavah govora
[Elektronski vir] / urednica Mira Krajnc Ivič. - 1. izd. - E-knjiga. -
Maribor : Univerza v Mariboru, Univerzitetna založba, 2024

Način dostopa (URL): <https://press.um.si/index.php/ump/catalog/book/898>

ISBN 978-961-286-882-6 (Pdf)

doi: 10.18690/um.ff.4.2024

COBISS.SI-ID 201876739



Javna agencija za znanstvenoraziskovalno
in inovacijsko dejavnost Republike Slovenije



SLAVISTIČNO DRUŠTVO
MARIBOR

**Knjigo je sofinancirala Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost
Republike Slovenije.**

Naslov projekta: Temeljne raziskave za razvoj govornih virov in tehnologij za slovenščino –
Mezzanine

Šifra projekta: J7-4642 Temeljne raziskave za razvoj govornih virov in tehnologij za slovenščino

ISBN 978-961-286-882-6 (pdf)

DOI <https://doi.org/10.18690/um.ff.4.2024>

Cena Brezplačni izvod
Price

Odgovorna oseba založnika prof. dr. Zdravko Kačič,
For publisher rektor Univerze v Mariboru

Citiranje Krajnc Ivič, M. (ur.) (2024). *Stanje in perspektive uporabe govornih*
Attribution virov v raziskavah govora. Univerza v Mariboru, Univerzitetna založba. Doi 10.18690/um.ff.4.2024

Kazalo

	Uvod Mira Krajnc Ivič	1
I Vloga in pomen raziskovanja govornega diskurza		7
1	60 let pozneje – pomen analize govornega diskurza <i>60 Years Later – the Significance of Spoken Discourse Analysis</i> Mojca Smolej	9
II Ob gradnji govornih virov		25
2	Nadgradnja digitalne slovarske baze za slovenščino in slovenskega oblikoslovnega leksikona Sloleks s podatki o govornjeni slovenščini: načrti in cilji <i>Extending the Digital Dictionary Database of Slovene and the Sloleks Morphological Lexicon of Slovene With Spoken Slovene Data: Plans and Goals</i> Jaka Čibej, Nejc Robida, Simon Krek	27
3	Skladenjska drevesnica govornjene slovenščine: stanje in perspektive <i>Spoken Slovenian Treebank: Current Situation and Perspectives</i> Kaja Dobrovoljc	41
4	Prednosti in slabosti dvotirnega zapisovanja govora v slovenskih govornih virih <i>Advantages and Disadvantages of Two-Tier Speech Transcription in Slovene Speech Resources</i> Darinka Verdonik, Mitja Trojar, Andreja Bizjak	63
5	Poslušati med vrsticami: parlamentarni govor in njegovi zapisi <i>Listening Between the Lines: Parliamentary Speech and its Transcripts</i> Ina Poteko, Marko Stabej, Kaja Jošt	81
6	Standardi transkribiranja narečnega korpusa GOKO <i>GOKO Dialect Corpus Transcription and Annotation Standards</i> Klara Šumenjak	103

7	Standardizacija prekmurske transkripcije samoglasnikov: študija primera <i>Standardisation of the Prekmurje Dialectal Transcription of Vowels: a Case Study</i> Melita Zemljak Jontes, Mihaela Koletnik	121
8	Transkribiranje v sociolingvističnih raziskavah in korpusih govornjene slovenščine <i>Transcription in Sociolinguistic Research and Corpora of Spoken Slovene</i> Maja Bitenc	151
9	Jezikovni modeli v jezikoslovni analizi: Programi za prepoznavanje govora <i>Language Models for Spoken Corpus Preparation: Speech Recognition Software</i> Teodor Petrič	169
10	Predlog izdelave korpusa humorja v govoru za slovenščino <i>Spoken Slovene Corpus of Humor: Draft Proposal</i> Mira Krajnc Ivič, Špela Antloga	195
11	Tvorba korpusů mluveného jazyka <i>Creation of Spoken Language Corpora</i> Miloslav Vondráček	221
III Vidiki raziskovanja govora in govornjenega diskurza		239
12	Mi i naši, oni i njihovi u politici: Osobne deikse u govorima hrvatskih saborskih zastupnika <i>We and Our, They and Their in Politics: Person Deixes in the Speeches of Croatian Parliamentarians</i> Goranka Blagus Bartolec	241
13	Fonološka zmožnost bosansko govorečih priseljenk in priseljencev <i>Phonological Competence of Bosnian-Speaking Immigrants</i> Jana Lovrec Srša, Gjoko Nikolovski	259
14	Večkodnost v žanrski analizi literarnega branja <i>Multimodality in the Genre Analysis of Literary Reading</i> Branislava Vičar, Katja Plemenitaš	277
15	Govor in govorna komunikacija v učnih načrtih za osnovno šolo in gimnazijo ter v katalogih znanj <i>Speech and Speech Communication in Curricula for Elementary Schools, Grammar Schools and in Catalogues of Knowledge</i> Simona Pulko, Melita Zemljak Jontes	299

16	<p>Spletna aplikacija Svojljni pridevniki iz prevzetih priimkov (SPiPP) kot vir za raziskave govora <i>Web Application Possessive Adjectives from Adopted Personal Proper Names (SPiPP) as a Resource for Speech Research</i> Irena Stramljič Breznik</p>	317
17	<p>Анализ спонтанной устной речи как способ исследования стратификационной вариативности языковых кодов на польско-белорусском пограничье <i>Analysis of Spontaneous Spoken Language as a Method for Investigating the Stratification Variability of Language Codes in the Polish-Belarusian Borderland</i> Katarzyna Konczewska</p>	333
18	<p>Konteksti snemanja govornenega diskurza v sociolingvistiki <i>Contexts of Recording Speech in Sociolinguistics</i> Maja Bitenc</p>	351
IV Med govornim in pisnim diskurzom		369
19	<p>Jezik influencera u kontekstu novih, novih medija <i>The Language of Influencers in the Context of New New Media</i> Borko Baraban, Snježana Barić-Šelmić</p>	371
20	<p>Komentarji novic Regionalobala.si med govornim in pisnim diskurzom <i>Regionalobala.si News Comments Between Spoken and Written Discourse</i> Maša Rolih</p>	385
V Govor v gledališki umetnosti		405
21	<p>Uporaba mikrofenomenološkega intervjuja pri raziskovanju igralčevega govora <i>The Use of Micro-Phenomenological Interview in Researching Actors' Speech</i> Martin Vrtačnik</p>	407
22	<p>Raziskovanje govornenega umetniškega jezika <i>Researching Artistic Speech</i> Nina Žavbi</p>	423

Uvod

MIRA KRAJNC IVIČ

Znanstvena monografija *Stanje in perspektive uporabe govornih virov v raziskavah govora* predstavlja rezultate prvega leta dela v raziskovalnem projektu *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik* (J7-4642) kot tudi rezultate raziskovalcev, ki se ukvarjajo z govorom v drugih raziskovalnih projektih, s skupnim glavnim ciljem strateškega in učinkovitega razvoja prostodostopnih govornih virov. Ti viri so nujno potrebni za poglobljene in širše veljavne raziskave govora in govorjenega diskurza v številnih disciplinah: fonetiki, fonologiji, dialektologiji, oblikoslovju, skladnji, leksikologiji in leksikografiji, sociolingvistiki, pragmatiki, hkrati pa tudi za področje govornih tehnologij.

Spontani govor in nasploh govorjeni diskurz predstavlja primarno človekovo jezikovno manifestacijo, še dodatno izpostavlja raznolikost jezika, prinaša nove podatke in izkazuje prvine, ki v pisni rabi niso prisotne. Raziskovanje govora in govorne komunikacije je nujno za celovito poznavanje in razumevanje jezika. Pogosto predstavlja ključni most za interdisciplinarno povezovanje jezikoslovja z drugimi vedami. Pomanjkanje govornih virov in pripadajoče raziskovalne infrastrukture sta danes osrednji oviri pri raziskavah govora in govorne komunikacije. Govorni viri so v primerjavi s pisnimi redki in slabo razviti. To je

predvsem posledica dejstva, da ustvarjanje govornih virov zahteva več truda, sistematičnega dogovarjanja o standardizaciji npr. zapisovanja govora, kot ustvarjanje pisnih, deloma pa tudi dejstva, da je pisno besedilo bilo in je še vedno v večji meri v središču zanimanja jezikoslovne vede.

Cilj znanstvene monografije je pregled stanja in opredelitev potreb po govornih podatkih in pripadajoči raziskovalni infrastrukturi v različnih disciplinah humanistike in družboslovja. Monografija prinaša 22 prispevkov, ločenih v pet poglavij:

- 1) Vloga in pomen raziskovanja govornega diskurza,
- 2) Ob gradnji govornih virov,
- 3) Vidiki raziskovanja govora in govornega diskurza,
- 4) Med govornim in pisnim diskurzom ter
- 5) Govor v gledališki umetnosti.

Posamezno poglavje vsebuje različno veliko prispevkov 32 raziskovalcev in raziskovalk govora, govornega diskurza in govornih virov doma ter v tujini.

Prvo poglavje *Vloga in pomen raziskovanja govornega diskurza* prinaša ugotovitve Mojce Smolej, o začetkih analize govorne slovenščine in pomenu tovrstnih analiz tudi s stališča kodificiranja norme pisnega oz. knjižnega standardnega jezika. Avtorica ugotavlja, da šele poznavanje zakonitosti govornega jezika lahko olajša ali pojasni marsikatero težavo, nejasnost ali »nenaravnost,« ki se pojavlja v pisnem jeziku.

Ob gradnji govornih virov je drugo in najboljše poglavje namenjeno, raziskavam ob gradnji govornih virov, že obstoječih, načrtovanih ali takih, ki jih je treba nadgraditi. V tem segmentu Jaka Čibej, Nejc Robida in Simon Kreka v *Nadgradnja digitalne slovarske baze za slovenščino in slovenskega oblikoslovnega leksikona Sloleks s podatki o govornih slovensčini: načrti in cilji* predstavljajo načrte in cilje za dopolnjevanje jezikovnih virov s podatki o tipično govornem besedišču. Predstavljeni so tudi poglavitni izzivi, kot so problematika kanoničnih oblik, tematika nestandardnih fonemov, nestandardnih izgovorjav standardnih besednih oblik ter tematika nestandardne morfologije. Rešitve teh izzivov bodo omogočile sistematično polnjenje obstoječih jezikovnih virov s tipično govorno leksiko. Kaja Dobrovoljc v *Skladenjska drevesnica govorne slovenščine: stanje in perspektive z vidika zasnove, vsebine in dostopnosti* predstavlja slovensko skladiščno drevesnico SST kot prvi skladiščno razčlenjeni korpus govorne slovenščine z uravnoteženim in

reprezentativnim naborom besedil referenčnega korpusa govorne slovenščine Gos. Darinka Verdonik, Mitja Trojar, Andreja Bizjak v *Prednosti in slabosti dvotirnega zapisovanja govora v slovenskih govornih virih* kritično prevprašujejo utemeljenost dvotirnega zapisovanja govora glede na uveljavljene prakse, zahtevan dodaten trud in glede na njegove prednosti. Zapis govora opredelijo kot prevod iz izvorno multimodalnega kanala komunikacije, v katerem verbalno izraženi pomen sooblikujejo glas in način govora, govorica telesa in situacija, v kateri poteka komunikacija, v eno, pisno modalnost. S prakso zapisovanja govora se ukvarjajo tudi Ina Poteko, Marko Stabej in Kaja Jošt v *Poslušati med vrsticami: Parlamentarni govor in njegovi zapisi*, ki predstavljajo prakso zapisovanja plenarnih sej v Državnem zboru Republike Slovenije med letoma 2010 in 2022 z namenom dokumentiranja poteka teh sej. V ospredju prispevka Klare Šumenjak *Standardi transkribiranja narečnega korpusa GOKO* je gradnja prvega slovenskega narečnega korpusa. Opisani so obseg korpusa, demografsko vzorčenje, posnetki in njihova enota ter označevanje korpusa. Vrednost korpusa GOKO je tudi v izobraževanju in ohranjanju kulturne in jezikovne dediščine. Mihaela Koletnik in Melita Zemljak Jontes v *Standardizaciji prekmurske transkripcije samoglasnikov: študija primera* pa preverjata primernost načrtovanega postopka standardizacije slovenske narečne transkripcije, in sicer (1) preverita fonetično transkripcijo prekmurskega gradiva za SLA za issln. *ā, issln. *ǎ in issln. *-ǎ; (2) opravita primerjavo s fonetičnim zapisom v dialektološki literaturi in dostopnem gradivu; (3) izbrane primere še eksperimentalnofonetično analizirata. O praksah in standardih pri zapisovanju govora v sociolingvističnih raziskavah razpravlja Maja Bitenc v *Transkribiranje v sociolingvističnih raziskavah in korpusih govorne slovenščine*. V teh raziskavah namreč ni standardnih transkripcijskih načel, temveč so ta načela vedno odvisna od teoretske usmeritve raziskovalca, predmeta in namena raziskave. Posebno pozornost namenja problematiki zapisovanja variant fonema /v/, polglasnika, specifičnih narečnih glasov in premen po zvonečnosti. Vidik laičnega uporabnika pri pretvorbi govora v pisno obliko preizkuša Teodor Petrič v *Jezirovni modeli za pripravo govornega korpusa: programi za prepoznavanje govora*. Avtor preizkuša programska orodja Razpoznavalnik, Microsoft Word Prepiši, Vosk/Kaldi in OpenAI Whisper in sistematičnost svoji ugotovitev zagotavlja z upoštevanjem meril: preprostost uporabe, časovni prihranek, morebitni stroški, zagotavljanje anonimnosti govorcev in kakovosti pretvorbe glede na deleže besednih napak, število zamenjav, vstavitev in izpustov. Ker je s številnih segmentov pomemben del govornega diskurza humor – čeprav po teorijah superiornosti omejen z nacionalnimi mejami in s časom v slovenistiki še vedno malo in v okrnjeni pojavnosti raziskovan – Mira Krajnc Ivič, Špela Antloga v *Predlog izdelave korpusa humorja v govoru*

za slovenščino predstavita teorije humorja in analizo obstoječih tujejezičnih korpusov humorja in kot prvi korak k izdelavi tovrstnega korpusa izvedeta anketo o dojemanju izbranih pasaj kot humornih/smešnih. Poglavje zaključuje prispevek Miloslav Vondraček *Tvorba korpusu mluveného jazyka*, v katerem avtor poroča o sodelovanju študentov pri gradnji češkega korpusa govornega jezika in o vprašanjih, s katerimi so se ukvarjali pri pridobivanju zvočnih posnetkov in njihovem zapisovanju.

Tretje poglavje *Vidiki raziskovanja govora in govornega diskurza* prinaša delne korpusne analize obstoječih, prostodostopnih ali priložnostno tvorjenih korpusov govornih besedil. Poglavje začena Goranka Blagus Bartolec s prispevkom *Mi i naši, oni i njihovi u politici: Osobne deikse u govorima hrvatskih saborskih zastupnika*, v katerem avtorica opisuje skladenjske in pomenske značilnosti osebnih deiktikov glede na reference, na katere se nanašajo, in ugotavlja, koliko raba teh deiktikov temelji na značilni podobi o nas in o njih kot polarizirani strani.

Prispevke v nizu priložnostno tvorjenih in ne prosto dostopnih govornih virov začena prispevek o poučevanju jezika kot drugega ali tujega. Gjoko Nikolovski, Jana Lovrec Srša v *Fonološka zmožnost bosansko govorečih priseljenc in priseljencev* obravnava fonološko zmožnost neslovensko govorečih priseljenc in priseljencev iz Bosne in Hercegovine, ki živijo v Mariboru. Opravljena analiza temelji na posnetkih govornih besedil in registrira vsaj osem najpogostejših težav, npr. izgovor polglasnika, izgovor vzglasnega v- pred (ne)zvnečim soglasnikom, težave z naglasnim mestom. Branislava Vičar in Katja Plemenitaš v *Večkodnost v žanrski analizi literarnega branja* ob izseku iz literarnega branja lezbične literature preučujeta součinkovanje in interakcijo različnih semiotskih kodov. Njuna študija med drugim razkrije, kako se določena dejanja uresničujejo v vzajemni kontekstualizaciji glasnega branja z nejezikovnimi kodi, kot so glasba, rokovanje s predmetom, drža telesa, pogled, obrazni izrazi, kretnje idr. Večkodne ali multimodalne analize tako govornega kot pisnega diskurza pa bi morale najti prostor tudi v izobraževalnem sistemu, saj je eden temeljnih namenov jezikovnega pouka razvijanje sporazumevalne zmožnosti, tj. praktično in ustvarjalno obvladovanje vseh sporazumevalnih dejavnosti in jezikovnosistemskih osnov. Tako je za celotno podobo in razumevanje vloge ter pomena govora bistvena tudi kritična presoja učnih načrtov za osnovno šolo in gimnazijo ter katalogov znanj za slovenščino. To sta naredili Simona Pulko, Melita Zemljak Jontes v *Govor in govorna komunikacija v učnih načrtih za osnovno šolo in gimnazijo ter v katalogih znanj za slovenščino*. Avtorici je sicer predvsem zanimalo, kolikšen je poudarek na razvijanju pravorečne zmožnosti in na

razvijanju zapisovalne zmožnosti govora, tudi v različnih načinih transkribiranja glede na stopnjo izobraževanja. Z enim od težjih vprašanj v slovenščini, tj. s tvorbo svojilnih pridevnikov iz prevzetih osebnih lastnih imen, se ukvarja Irena Stramljič Breznik v *Spletna aplikacija svojilni pridevniki iz prevzetih priimkov (SPIPP)* kot vir za raziskave govora. Ta prispevek spletni pripomoček in aplikacijo SPIPP predstavlja kot jezikovni vir za nadaljnje razvijanje govornih pripomočkov, konkretno kot postopno nadgradnjo eBralca tako, da bi bil specializiran za izgovarjavo tudi tujejezičnih lastnoimenskih besed in njihovih oblik, prilagojenih slovenščini.

Sledita prispevka, ki govorne fenomene raziskujeta z vidika sociolingvistike. Prvi je prispevek Katarzynie Konczewske *Анализ спонтанной устной речи как способ исследования стратификационной вариативности языковых кодов на польско-белорусском пограничье*. Avtorica v njem analizira spontani govor kot način preučevanja stratifikacijske variabilnosti jezikovnih kodov na večetničnem, večkulturnem in večjezičnem mikrookolju na obeh straneh poljsko-beloruske meje, tj. na območju s še danes ohranjeno razredno strukturo, nastalo v 16. stoletju. Pri analizi je bila pozornost med drugim usmerjena na jezikovne manifestacije socialne slojevitosti in značilnosti dojemanja svojega jezika s strani govorcev posameznih kodov. Drugi prispevek v tem nizu je prispevek Maje Bitenc *Konteksti snemanja govorjenega diskurza v sociolingvistiki*. Raziskava predstavlja sociolingvistični intervjuji kot metodološki pristop k preučevanju govorne variantnosti pri posameznem govorniku in v govorni skupnosti. Pri tem je poudarek na kontinuumu od narečne do standardne varietete, vključujoč jezikovno zmožnost tako v najbolj akrolektalnem kot bazilektalnem govornem stilu informantov. Ta kontinuum se opazuje z ustvarjanjem različnih kontekstov snemanja govorjenega diskurza.

Kot že sam naslov poglavja pove, četrto poglavje *Med govorjenim in pisnim diskurzom*, prinaša prispevka, ki analizirata spletna besedila. Tehnološki napredek digitalne in omrežene družbe je namreč omogočil razvoj novih novih medijev. Ti niso botrovali le nastanku novega tipa občinstva, temveč tudi novemu tipu vplivnežev, ki za razliko od vplivnežev nekoč svojo priljubljenost in vpliv gradijo zaradi razvoja tehnologije (nove besedilne skupine, npr. vblogi, tviti). Borko Baraban in Snježana Barić-Šelmić v *Jezik influenceru u kontekstu novih novih medija* opisujeta in navajata vzorec jezikovnih odstopanj izbranih vplivnežev v pisnih spletnih objavah s ciljem jasne določitve razmerja med novimi novimi mediji glede rabe hrvaškega knjižnega jezika. Maša Rolih v *Komentarji novic Regionalobala.si med govorjenim in pisnim diskurzom* pa analizira komentarje ob novicah na izbranem portalu. Avtorica ugotavlja, da so zbrani

komentarji večinoma pisni, a njihov jezik odstopa od jezikovne norme, zato poleg zapisa po izgovoru, leksike različnih socialnih zvrsti, rabe emotikonov in gif-ov ti komentarji vsebujejo tudi prvine besedilne semantike in skladenjske, značilnejše za govornjeni diskurz.

Posebno poglavje je namenjeno *Govoru v gledališki umetnosti*. Tu Martin Vrtačnik v *Uporaba mikrofenomenološkega intervjuja pri raziskovanju igralčevega govora* prikazuje vlogo mikrofenomenološkega intervjuja pri ustvarjanju igralčevega glasovnega sloja. Ta intervju namreč omogoča pridobivanje podatkov o nezavednih procesih pri ustvarjanju odrskega govora, tj. omogoča, da se igralec zave svoje subjektivne izkušnje in jo opiše. Drugi prispevek v tem sklopu je prispevek *Raziskovanje govornega umetniškega jezika* avtorice Nine Žavbi. Avtorica združuje slušnozaznavno in akustično analizo odrskega govora nasploh in konkretno pri uprizoritvi Cankarjevih *Hlapcev* leta 2015 v Slovenskem stalnem gledališču Trst. Analiza govora potega glede na razmerje med dramskim in uprizoritvenim besedilom ter kot govorna izvedba na odru. S tem pristopom povezuje znanost in umetnost.

Znanstvena monografija *Stanje in perspektive uporabe govornih virov v raziskavah govora* z raznovrstnimi in različnimi prispevki ne le dokumentira trenutno stanje znanja, temveč tudi postavlja temelje za prihodnje raziskave in izdelave govornih virov.



I

Vloga in pomen
raziskovanja
govorjenega diskurza

60 LET POZNEJE – POMEN ANALIZE GOVORJENEGA DISKURZA

MOJCA SMOLEJ

Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
mojca.smolej@ff.uni-lj.si

V prispevku je podan kratek pregled začetkov raziskav spontano govornega jezika na Slovenskem. Pred slabimi 60 leti je enega prvih prispevkov objavila Breda Pogorelec. Izhajajoč iz začetkov preučevanja govornega diskurza, so v prispevku podani tudi nekateri premisleki, zakaj je nujno, da se s preučevanjem nadaljuje in nadgrajuje predhodne raziskave. Podani so s stališča opisne slovnice tako govornega kot pisnega (knjižnega) jezika. Razumevanje zakonitosti govornega jezika lahko namreč pomaga pri razumevanju in reševanju nekaterih nejasnih ali pomanjkljivo opredeljenih slovničnih opisov (predpisov) knjižnega jezika. Neizodostno opredeljene ali opisane slovnične zakonitosti knjižnega jezika lahko posledično povzročajo tako stisko na ravni poučevanja slovenščine kot prvega in drugega/tujega jezika, prav tako pa tudi na ravni rabe jezika pri splošnih govornih (torej nejezikoslovcih).

DOI
[https://doi.org/
10.18690/um.ff.4.2024.1](https://doi.org/10.18690/um.ff.4.2024.1)

ISBN
978-961-286-882-6

Ključne besede:
govornji diskurz,
Breda Pogorelec,
skladnja,
diskurzni označevalci,
knjižni (standardni) jezik



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.1](https://doi.org/10.18690/um.ff.4.2024.1)

ISBN
978-961-286-882-6

Keywords:
spoken discourse,
Breda Pogorelec,
syntax,
discourse markers,
the standard language

60 YEARS LATER – THE SIGNIFICANCE OF SPOKEN DISCOURSE ANALYSIS

MOJCA SMOLEJ

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
mojca.smolej@ff.uni-lj.si

This paper gives a brief overview of the beginnings of research on spontaneous spoken language in Slovenia. One of the first papers was published almost 60 years ago by Breda Pogorelec. The first section of the paper delves into the beginnings of spoken language research, while in the second section some considerations are made as to why it is necessary to continue and build on previous research. They are presented from the point of view of the descriptive grammar of both spoken and written (literary) language. Understanding the regularities of spoken language can help to understand and resolve some unclear grammatical descriptions (rules) of literary language. Insufficiently defined or described grammatical regularities of the literary language can therefore cause problems both at the level of teaching Slovene as a first and second/foreign language and at the level of language use by general speakers (i.e. non-linguists).



1 Uvod¹

Leta 1965 je M. Rupel v Jezikovnih pogovorih, v katerem je objavljeno eno prvih del o spontano govornem jeziku oz. govornem diskurzu, zapisal:

»V govoru je najhujša napaka zanikernost. Če govorimo na zborovanju, na sestanku, v šoli itd., bi se morali potruditi, da bi izrekli jasno in razločno brez narečnih primesi, brez neknižnih besed, v gladkih stavkih. A le poslušajte naše ljudi, tudi visoko izobražene, kako zanikerno se izražajo, kakor bi sedeli v gostilni / .../.« (JP 1965: 55)

V istem delu lahko prav tako preberemo:

»Zato ne verjemimo tistim, ki se jim zdi, da bi bilo treba v dobi tehnike in vesoljskih poletov tudi jezik nekako mehanizirati. Jezik ni samo sredstvo za sporazumevanje in nima dosti opraviti z znanstvenimi obratci. Enako pomembna je njegova druga vloga, da izraža človekovo doživljanje, in pri tem gre zmerom tudi za lepoto, eno človekovih največjih vrednot.« (JP 1965: 67)

Dvom v smiselnost in zmožnost »mehanizacije«, torej računalniške obdelave jezika je izrazil J. Gradišnik. Oba citata povzemata in povezujeta tematiki, na kateri se bo navezoval prispevek: a) delni pregled začetkov analize govornega slovenščine in b) pomen analize govornega diskurza danes. Predvsem slednje bo predstavljeno z zornega kota raziskovanja jezika in njegovih slovničnih zakonitosti, torej opisne slovnice. V zadnjih dveh desetletjih sicer gramatikografija sama po sebi ni več v ospredju raziskovanja, temeljni cilj slovničnega opisa se je namreč premaknil v smer, za katero je J. Gradišnik menil, da ni mogoča oz. da je nesmiselna. Različna programska orodja za analizo jezika, zbirke jezikovnih korpusov, raziskave govornega jezika so spremenili koncept in namen kateregakoli slovničnega opisa. Izčrpen in ustrezen slovnični opis lahko danes temelji na poznavanju tako pisnega kot govornega jezika vseh ravni. Če se hitro razvijajo orodja za kvalitetno računalniško obdelavo slovenščine in s tem zamejena množica (iz)rabe programske infrastrukture za digitalno rabo slovenskega jezika, je delno ob strani pristalo raziskovanje temeljnega slovničnega ustroja slovenskega jezika. Dandanes si po vsej verjetnosti enega brez drugega ne moremo in ne smemo zamišljati. Kvalitetna računalniška obdelava jezika temelji na kvalitetnem poznavanju slovničnega ustroja jezika. V sodobnem času pa velja tudi obratno, zato je nujno, da se oba vidika raziskovanja slovenskega jezika razvijata hkrati, v sodelovanju. Poleg tega pa je, kot

¹ Prispevek je nastal v okviru raziskovalnega projekta ARIS Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

je bilo že napisano, nujno hkrati preučevati tako pisni kot govorjeni jezik. Šele oboje skupaj lahko poda celovit vpogled v ustroj jezika.

2 B. Pogorelec – pionirka obravnave govorenega diskurza

B. Pogorelec je leta 1965 v že omenjenem delu *Jezikovni pogovori* objavila prvo razpravo, ki v polnosti obravnava spontano govorjeni jezik. To pomeni, da se dotakne tako glasoslovnih kot oblikoslovnih, skladenjskih in drugih značilnosti govorenega jezika. Razpravo *Vprašanja govorenega jezika* lahko povsem upravičeno uvrstimo med pionirske. Resda je pred tem izšla kopica jezikoslovnih razprav, ki se nanašajo na govorjeni jezik, vendar predvsem s stališča zborne izreke. Razprave A. Breznika, F. Ramovša, B. Voduška, A. Bajca, M. Rupla, J. Toporišiča in F. Bezlaja obravnavajo največkrat »le« pravopisna oz. pravorečna vprašanja, delno pa v središče preučevanja postavljajo odrski jezik, tudi tu skoraj izključno v povezavi z izreko.² Prevladovalo je mnenje, da so med pisnim in govorenim jezikom razlike predvsem na ravni izreke in besedišča.

*»Njegove (pogovorni jezik) značilnosti so manj v gramatičnih oblikah kot v besednem zakladu in izgovorjavi.«
(Vodušek 1953: 4)*

*»Slovinci imamo poleg najvišje, t. i. zborne zvrsti knjižnega jezika, še pogovorno. Razlika med obema zvrstema je v izreki, oblikoslovlju, skladnji in besedju s frazeologijo vred. Najbolj očitna je razlika prav v izreki.«
(Toporišič 1970: 55)*

Kot je bilo že omenjeno, je prispevek B. Pogorelec (1965) prvi, ki celostno zaobjame tako glasoslovne, oblikoslovne kot skladenjske in druge značilnosti (npr. okoliščine) spontano govorenega jezika.

»Govor /pa/ vedno spremlja okolje. Tudi motivika tistega, kar govorimo, se navadno močno razlikuje od motivike, obravnavane v pisnem sporočilu. Osnovna razlika med govorenim in pisanim jezikom, razlika v zvezab in besedju, je največkrat pogojena prav z razliko v namenu in z razliko v okoliščinah sporazumevanja v govoru ali v pisavi.« (1965: 150).

² Npr. M. Rupel: Slovensko pravorečje, 1946; A. Bajec: Kako je rasel naš knjižni jezik, 1951; B. Vodušek: O Slovenskem pravopisu in o jezikovnih načelih, 1951; B. Vodušek: Kakšen naj bo naš odrski jezik? Ali imamo pogovorni jezik na odru in ali ga imamo v življenju?, 1953; A. Bajec: O pogovornem jeziku, 1955; F. Bezlaj: O slovenskem jeziku, 1961.

Na skladijske posebnosti govornega jezika pa je B. Pogorelec v resnici opozorila že leta 1964 v svoji doktorski disertaciji *Veznik v slovenščini*, ki je prav tako pionirsko delo, saj gre za prvo disertacijo s področja sodobnega knjižnega jezika. Bibliografija doktorskih disertacij univerze in drugih visokošolskih in znanstvenih ustanov v Ljubljani med leti 1920 in 1968 (Kokole 1969: 135–136)³ namreč kaže, da je bilo uspešno obranjenih le devet nalog s področja slovenskega jezikoslovja, med vsemi jezikoslovnimi doktorskimi deli pa sta le dve, ki obravnavata sodobni knjižni jezik: *Oris slovenskega knjižnega izgovora* Franceta Bezlaja in *Veznik v slovenščini* Brede Pogorelec. V doktorskem delu B. Pogorelec npr. vseskozi opozarja na nekonsistentno, nezadostno in nepopolno obravnavo skladijskih razmerij v /S/lovenski slovnici, ki sloni izključno na analizi literarnih besedil. Pravkar napisano je prisotno še sedaj (gl. npr. *Slovenska slovnica* 2000: 636–652) in povzroča nemalo težav predvsem pri učenju slovenščine kot prvega jezika. Zaradi neupoštevanja skladijskih zakonitosti govornega jezika in prevelike naslonitve izključno na pisni oz. literarni jezik, je v /S/lovnici opisan jezik, ki je delno umetelen in katerega jezikovne zakonitosti se učenci največkrat učijo na pamet, ne da bi jih razumeli. Raven izražanja podrednih medstavčnih skladijskih razmerij v govornem jeziku je slabo raziskana, kar je velika pomanjkljivost, saj prav poznavanje zakonitosti govornega jezika lahko olajša ali pojasni marsikatero težavo, ki se pojavlja v pisnem jeziku. Primer pravkar napisanega so časovna, vzročna in načinovna podredna skladijska razmerja, ki so tipična predvsem za knjižni (literarni) jezik, v govornem jeziku so ta razmerja izražena največkrat s prirednimi skladijskimi sredstvi.

Poleg opozarjanja na skladijske in druge posebnosti spontano govornega jezika je B. Pogorelec vseskozi opozarjala tudi na nenehno spreminjanje govornega jezika, ki lahko prav tako povzroča težave pri oblikovanju splošnih slovnicih zakonitosti govornega jezika.

»Pogovorni jezik sproti nastaja in je pod stalnim vplivom dveh jezikovnih plasti [...]. Zanimivo je, da povzroča to prepletanje vplivov, da tudi pri enem samem govorniku pogovorni jezik ni vedno dosledna oblika s trdno slovnico in drugimi značilnostmi novitega jezikovnega sistema. Te nedoslednosti in nenehno individualno spreminjanje povzročajo, da je za slovenski pogovorni jezik na tej stopnji razvoja zelo težko določiti splošne

³ Anton Bajec: Nominalni akcent v rezijanskem narečju (1921), Rudolf Kolarič: Vokalizem in akcent vzhodnoštajerskega (prekmurskega) narečja s posebnim ozirom na jezik dajnkovskih spisov (1923), Mirko Rupel: Literarni in jezikoslovni odnosi med slovenskimi Artikuli, Tbinga 1562, in havaškimi Artikuli s cirilskimi in glagolskimi črkami, Tubinga 1562 (1923), Franc Tomšič: Jezik v Janeza Svetokriškega Sacrum promptuarium (1930), Fran Bezljaj: Oris Slovenskega knjižnega izgovora (1939), Tine Logar: Horjulski govor (1941), Jakob Rigler: Akcentska in glasoslovna analiza južnonotranjskih govorov med Snežnikom in Slavnikom (1960), Breda Pogorelec: Veznik v slovenščini, (1963), Martina Orožen: Zgodovinski razvoj futuralnih in modalnih gramatičnih oblik v knjižni slovenščini od 16. do 19. stoletja (1965).

zakonitosti in jih ločiti od trenutnih in enkratnih posebnosti jezikovnega izraza posameznikov.» (1965: 134-135)

Zanimivo je, da podobne ugotovitve nekaj desetletij kasneje poda tudi ena prvih slovníc govornega in pisnega jezika hkrati *Grammar of Spoken and Written English*.⁴ Ker je govorni jezik vsaj skladenjsko težje opisljiv, je toliko težje poiskati skupne definicije govornega in pisnega jezika. Na pravkar napisano slovnica večkrat opozori.

»Ali lahko določimo neko večjo skladenjsko neodvisno enoto slovnice govornega jezika, primerljivo s stavkom v pisni angleščini? /.../ Bistvena razlika je /.../ v tem, da govorni jezik poteka v realnem času in je podvržen omejitvam delovnega spomina /.../. Slovnico govornega jezika lahko imenujemo »dinamična« /.../. V nasprotju s tem je slovnica pisnega jezika bolj arhitekturna v smislu, da ima napisani stavek statični obstoj.« (Longman 1999: 1066)

»Neodvisne povedi ustrezajo temu, kar je v drugih slovnica h splošno opredeljeno kot stavki. Ta izraz je ustrežnejši, saj je težko podati dobro jezikoslovno opredeljeno stavka, ki se enako dobro nanaša na pisni jezik in spontani govor.« (Longman 1999: 202)⁵

Iskanje skupnega opisa obeh temeljnih plasti jezika, ki bo zadosten in dovolj vključujoč, je po vsej verjetnosti v današnjem času preseženo oz. upoštevano le na ravni kontrastive, ki pa ne pomeni le spoznavanja zakonitosti in značilnosti enega ali drugega, pač pa predvsem prepoznavanje tistih jezikovnih značilnosti, ki se kažejo pretežno v eni plasti, vendar pa imajo nezanemarljiv vpliv tudi na drugo plast. Opis npr. skladenjskih značilnosti govornega jezika je tako nujen za vsaj približno popoln opis pisnega, knjižnega jezika.

Raziskovalna pot, ki jo je pred slabimi 60 leti začrtala B. Pogorelec, tako še vedno ostaja le delno prehojena. Delno zaradi – kot je poudarila že B. Pogorelec – težko ulovljivih slovnicih zakonitosti spontano govornega jezika, delno pa tudi zato, ker, preprosto rečeno, gramatikografija nekoliko drsi v ozadje jezikoslovnih raziskav in jezikoslovja nasploh.

⁴ D. Bieber in drugi, 1999: *Grammar of Spoken and Written English*. Longman.

⁵ »Can we define some major syntactically independent unit of spoken grammar comparable to the sentence in written English? /.../ The crucial difference here is /.../ that spoken language takes place in real time, and is subject to the limitations of working memory /.../. We may refer to the grammar of speech as »dynamic« /.../. In contrast, the grammar of writing is more architectural, in the sense that a written sentence has a static existence.« (Longman 1999: 1066).

»Independent clauses correspond to what is generally defined as sentences in other grammars. This term is preferred, as it is difficult to give a good linguistic definition of a sentence which applies equally well to writing and spontaneous speech.« (Longman 1999: 202)

2.1 Živa Gruden in diskurzni označevalci

Če na hitro pregledamo slovnične raziskave kateregakoli spontano govornega jezika, lahko ugotovimo, da so v ospredju predvsem raziskave diskurznih označevalcev in drugih funkcijsko podobnih jezikovnih sredstev, konverzacijske analize, raziskave, povezane s prozodijo, ter raziskave, ki obravnavajo »kršitev« ali neupoštevanje kodificirane norme knjižnega jezika.

Na Slovenskem je pionirka⁶ na področju diskurznih označevalcev Živa Gruden, ki je leta 1975 na pobudo B. Pogorelec napravila obsežno diplomsko raziskavo govornih signalov v govornem jeziku tržaških Slovencev. Pri analizi je izhajala iz raziskave Krystyne Pisarkowe *Składnia rozmowy telefonicznej* (1975). Avtorica v razpravi o skladnji telefonskih pogovorov uporabi termin govorni signal,⁷ s katerim označi vse morfološko in funkcijsko različne besede, ki v telefonskih pogovorih, kjer imata govorca na razpolago zgolj glasovna sredstva, nadomeščajo kretnje in mimiko neposrednega pogovora ter grafična sredstva pisanega sporočila. Kot ugotavlja Ž. Gruden, se večina teh sredstev uporablja tudi v neposrednem pogovoru.

Delo Ž. Gruden je kljub skoraj 50-letnemu časovnemu preteku še vedno lahko za zgled ustreznega morfološkoskladenjskega pristopa k analizi najrazličnejših diskurznih označevalcev, govornih ali členitvenih signalov.

2.2 V. Gjurin in interesne govornice

Kot temeljni prispevek k raziskovanju govornega slovenščine je nujno omeniti še *Interesne govornice sleng, žargon, argo* avtorja Velemirja Gjurina (1974), ki ga je pripravil pod mentorstvom J. Toporišiča. Gjurin razmeji pomenski obseg v naslovu omenjenih terminov kot tri različne kategorije nekonvencionalnega jezika. Vse tri opiše in jih med seboj primerja oz. izpostavi njihove razlike. Razprava je na svojem področju vsekakor pionirska, saj je preusmerjala zorni kot jezikoslovnega raziskovanja z diahronega, narečnega in zbornega oz. knjižnega pogleda. Pojmovanje

⁶ Pred Ž. Gruden sta se diskurzni označevalci dotaknila že B. Pogorelec (1965) in J. Toporišič (1970). Prva (1965: 141-143) analizira rabo govornih sredstev v spominih Nade Kraigherjeve, torej v pisanem besedilu, in sicer v monologu, drugi (1970: 67) pa na kratko omeni zveze z ogovorilnimi glagolskimi izrazi in posebne zveze s členki, ki bi jih danes uvrstili med diskurzne označevalce.

⁷ Pri telefonskih pogovorih K. Pisarkowa (in po njej v svojem delu Ž. Gruden) loči: konativne, fatične, psevdokonativne, psevdofatične signale in interpunkciji ekvivalentne signale.

interesnih govoric⁸ je bilo pred navedeno razpravo razumljeno zelo ohlapno, saj so bili v slovenskem jezikoslovju termini žargon, sleng in argo uporabljeni sinonimno. Danes so v učbenikih za slovenščino tako v osnovni kot srednji šoli vsi trije termini stalnica, so pa razumljeni nekoliko omejeno oz. samosvoje, mestoma celo zavajajoče in napačno.

Poleg omenjenih treh jezikoslovcev (Pogorelec, Gruden, Gjurin) so se analize govornega jezika postopoma lotevali tudi drugi, med drugimi tudi J. Toporišič, in sicer v že omenjeni razpravi *Slovenski pogovorni jezik* (1970).

Toporišič loči med a) pogovornim jezikom, ki se »goji predvsem v pogovoru«, in b) navadno govorico,⁹ tudi občevalni jezik, ki naj bi bil/a v precejšnji meri narečno obarvan/a, razen v srednje izobrazbenih krogih.

V razpravi je lahko tudi že na osnovi tvorjenja terminov, s katerimi so poimenovane posamezne plasti slovenščine, izluščiti idejo »elitizma«. Prav tako to izpričuje tudi sama navedba primerov, okoliščin, v katerih naj bi rabili pogovorni jezik.

»V pogovornem jeziku se sporazumevata strokovnjaka o katerikoli problematiki; tako govorita potnika v vlaku, avtobusu, letalu ali na ladji; tako se potniki sporazumevajo s sprevodnikom, prodajalcem vozovnic ali vstopnic, kupec s prodajalcem, bolnik z zdravnikom, nadrejeni s podrejenim na delovnem mestu (v tovarni, pisarni); tako se kramlja v kavarni ali privatnem stanovanju o zadevah kulture, civilizacije, politike in tehnike /.../.«¹⁰ (1970: 56–57)

Pogovorni jezik, predvsem splošno pogovorni jezik naj bi se opiral na govor Ljubljane in tako upošteval »višje« glasoslovne, oblikoslovne, naglasne in skladenjske standarde.

»Zagotovo so precej slabi tudi strokovnjaki, ki se v TV oglašajo v stalnih rubrikah, konkretno mislim na zdravnike, higienike (ali kaj so) ali pa na vremenoslovce, ne nazadnje na miličnike. Vsi ti ljudje obupno vekajo, zekajo, sekajo, kekajo (in morda tudi bekajo). [...] [I]z novejšega časa pa bodi omenjen predstavnik s telesnokulturne fakultete na Kodeljevem, ki je svojo dolgočasno slovenščino prav tako pestril s ke/ve/zekanjem. (Ubogi normalni študentje, ki ga morajo poslušati leta in leta!) [...] Prava žalost res, kaj vse delamo s svojim lastnim jezikom iz same ljube nebržižnosti!« (Toporišič 2007 (1989): 152)

⁸ Termin interesne govorice je predlagal J. Toporišič.

⁹ J. Toporišič je v svoji razpravi glede razumevanja/opredelitve govornega jezika pri B. Pogorelec napisal:

»B. Pogorelec ne loči dosledno govornega jezika od pogovornega, zato ni mogoče natančno povedati, katere značilnosti je treba pripisati pogovornemu jeziku.« (1970: 58).

¹⁰ Iz napisanega je tudi razvidno, da trditev pričakovano izhaja iz strukturalistične opredelitve različnih socialnozvrstnih oblik jezika, in ne funkcijske (npr. FDG).

Ob koncu kratkega pregleda začetkov obravnave govornega jezika lahko zaključim, da se je sodobna jezikoslovna stroka seveda od opisanih predpostavk povsem odmaknila (ostaja pa po vsej verjetnosti še na ulici, v šolah, medijih itn.). V ospredju je poudarjanje enakopravnosti različnih variant in nujnost opisa vsake izmed njih. Dihotomije med knjižnim in govornim ni več zaznavati. Zaznavati pa je drugačno vrsto dihotomije: digitalno jezikoslovje oz. digitalna raba jezika vs. »nedigitalna« gramatikografija, ki ima za cilj predvsem opis slovničnega ustroja jezika. Kot je bilo izpostavljeno že na začetku, je upati, da je omenjena dihotomija le navidezna. Obe področji morata sodelovati in se medsebojno podpirati. Eno brez drugega sicer lahko deluje, vendar okrnjeno.

3 Pomen analize govornega jezika

V opis slovničnega ustroja slovenskega jezika morajo biti vključene vse plasti jezika, tako pisna kot govorna. Zaradi ne vključevanja slovničnih značilnosti govornega jezika je /S/lovnica čedalje bolj oddaljena od govorca slovenščine. Analiza govornega jezika je ključnega pomena tudi takrat, kadar je nek določen slovnični pojav pogostejši v govornem jeziku, manj prisoten pa je v pisnem. Če želimo celostno zajeti slovnični ustroj jezika, moramo v analizo, v opis vključiti tudi obrobne ali manj pogoste slovnične prvine.

Pravkar napisano bo ponazorjeno z izbranimi skladenjskimi primeri. Pred samo navedbo se bomo na kratko ustavili še pri treh citatih, ki nakazujejo, da je povezovanje med različnimi vejami raziskovanja jezika nujno. Kvalitetna računalniška obdelava jezika mora po vsej verjetnosti temeljiti na kvalitetnem poznavanju slovničnega ustroja jezika, slednje pa je danes mogoče predvsem z izčrpnim korpusnim pristopom in dobro računalniško obdelavo jezika. Prav tako pa je v analizo, obdelavo nujno vključevati vse spektre jezika.

Leta 2006 sta T. Erjavec in N. Ledinek zapisala:

»Najbolj celovita je Slovenska slovnica (Toporišič, 1984), najizčrpnjša opisa različnih vidikov slovenske skladnje pa sta Nova slovenska skladnja (Toporišič, 1982) ter Vezljivost v slovenskem jeziku (s poudarkom na glagolu (Žele, 2001) /.../. Vendar pa za slovenščino še vedno ne obstaja nobena strogo formalna, računalniška (tj. primerna za računalniško obravnavo jezika) in obenem izčrpana slovnica.« (2006: 2).

Po vsej verjetnosti lahko tudi 18 let kasneje zapišemo, da kvalitetne, izčrpne slovnice, ki bi jo lahko izrabili za različne nadaljnje namene, še vedno nimamo. S tem sta najbrž delno povezani tudi naslednji trditvi.

»/S/ istemi korpusnega označevanja /so/ oblikovani za ciljno rabo in vzpostavljajo uporabnostne rešitve, zato jih je smiselno dojemati zgolj kot pripomoček, ki olajšuje jezikoslovno analizo podatkov, in ne kot končni rezultat jezikoslovne analize. Zaradi zahtev avtomatske analize jezika pri vzpostavljanju sistema označevanja namreč (lahko) približa do jezikovnoteoretičnih poenostavitev.« (Ledinek 2018: 105)

»Za slovenščino sicer obstaja namensko razvit skladenjski razčlenjevalnik (Dobrovoljc et al. 2012), s pomočjo katerega je mogoče graditi skladenjsko označene korpuse slovenščine, vendar pa večina jezikoslovcov, razumljivo, nima potrebnega specializiranega znanja za tovrstne postopke. Predstavljene možnosti zato v največji meri ostajajo neizkoriščene.« (Ledinek 2018: 107)

Napisano nakazuje neke vrste protislovje, tudi že omenjeno dihotomijo, kar pa ni nujno slabo. Bolj kot iskanje enotnega, enega jezika, je nujno dopolnjevanje in nadgrajevanje.

V nadaljevanju bo navedenih nekaj skladenjskih primerov, ki bi jih bilo treba na novo opisati, in sicer s hkratno obdelavo tako govornega kot pisnega (ne le knjižnega) jezika. Na vse skladenjske kategorije je bilo sicer že nekajkrat opozorjeno v različnih razpravah, vendar je očitno za splošno uveljavljanje in razširjanje na novo opredeljene jezikoslovne pojme, kategorije nujno oblikovati novo /S/lovnico.

a) Skladenjska razmerja in skladenjski pomen

Kot je bilo že omenjeno, so časovna, vzročna in načinovna podredna skladenjska razmerja, ki so tipična predvsem za knjižni (literarni) jezik, v govornem jeziku izražena največkrat s prirednimi skladenjskimi sredstvi. Tudi SS (Toporišič 2000: 433) navedene skladenjske pomene uvršča le pod kategorijo podredja. Nasprotno pa velja za protivna pomenska razmerja. SS (Toporišič 2000: 432) jih uvršča izključno med priredna skladenjska razmerja. Obdelava govornih, najrazličnejših publicističnih in literarnih besedil kaže, da je protivnost velikokrat izražena s podrednimi skladenjskimi sredstvi in da bi bilo nujno razumevanje skladenjskih in pomenskih razmerij dopolniti in nadgraditi.

Zaključimo a) da bi bilo nujno korpusno izluščiti prav vsa vezniška oz. konektorska sredstva, b) natančno analizirati vse njihove skladenjske in pomenske vloge, pri čemer bi v analizo morali vključiti konstrukcijski pristop, in c) pri razvrščanju

analiziranih konektorskih sredstev bi morali ločevati med skladenjskim in pomenskim razmerjem, kar nadalje pomeni, da bi morali razširiti oz. dopolniti/izpopolniti shemo pomenskih razmerij in posledično shemo strukturnih vlog obvezno stavčnoočlenskih odvisnikov, neobvezno stavčnoočlenskih odvisnikov in nestavčnoočlenskih odvisnikov. Ob vsem tem bi bilo nujno v analizo vključiti tudi jezikovnostilistični vidik, ki bi nakazoval dvofunkcijskost oz. neke vrste homonimijo (dvojno ali celo trojno pomensko in ilokucijsko vrednost) iste strukture.¹¹

Ločevanje skladenjskega in pomenskega razmerja pomeni, da bi se izognili tradicionalnemu ločevanju med prirednimi vezniškimi sredstvi, za katera so tipična točno določena pomenska razmerja (npr. vezalno, ločno, protivno itd.), in podrednimi vezniškimi sredstvi, za katera so značilna druga pomenska razmerja kot pri podrednih vezniških sredstvih (npr. načinovno, posledično, vzročno itd.). Ločevanje med skladenjskimi in pomenskimi razmerji je izhodiščno seveda nujno, vendar s temeljnima poudarkoma, da lahko isto konektorsko sredstvo izraža različna pomenska razmerja in da lahko isto pomensko razmerje izražajo različna konektorska sredstva (npr. tako podredna kot priredna). To nadalje pomeni, da bi bilo, izhajajoč iz korpusne analize, nujno razširiti oz. dopolniti obravnavo skladenjskih in pomenskih vlog (stavčnih in nestavčnih) obveznih in neobveznih členov in družljivih ter priredno povezanih stavčnih konstrukcij.

a1) Stavčne drevesnice

S predhodnim je povezano tudi razumevanje odvisnosti in povezanosti posameznih stavčnih ali besednozveznih enot znotraj povedi ali, širše gledano, besedila.

Analiza preprostega primera, ki se velikokrat pojavi v osnovnošolskih in srednješolskih učbenikih slovenščine ter maturitetnih polah za slovenščino pokaže, da razumevanje skladenjskih odnosov ni vedno preprosto oz. da je zdajšnja obravnava tovrstnih struktur zelo pomanjkljiva.

Primer:

- Povej mi, kaj te teži, ker bi ti rada pomagala.

¹¹ Zveza stavkov *Medtem ko Janež spi, France dela* lahko izraža zgolj časovno razmerje (istočasnost dveh dejanj), lahko pa zaradi drugačnih kontekstualnih danosti poudarja protivnost, v katero je sicer vključena tudi časovnost, le da je ta delno ali popolno zakrita (Ata dela, mama pa spi.).

Z razumevanjem skladijskih razmerij in skladijskih struktur, kakršno je npr. v SS (2000: 652) in Novi slovenski skladij, je navedeni primer skoraj nemogoče skladijsko razčleniti, čeprav je preprost, kar zopet potrjuje, da bi bila nadgradnja oz. povsem nova analiza obravnave skladijskih razmerij in skladijskih pomenov nujna. V navedenem primeru sta prisotni dve vrsti odvisnikov. Predmetni odvisnik je na ravni vezljivosti določilo, kar pomeni, da je propozicija odvisnika del propozicije matičnega stavka. Vzročni odvisnik je na ravni vezljivosti oz. družljivosti dopolnilo, saj njegova propozicija ni del matičnega stavka. Vzročni odvisnik je tako pravzaprav dopolnilo sklopa matičnega stavka in predmetnega odvisnika.¹²

b) Besedni red

Pred dobrimi 40 leti je J. Toporišič napisal:

»Eno zelo zanimivih poglavič skladij so gotovo vprašanja o besednem redu. Ker je tudi eno najtežjih, se ne čudimo, da v slovenščini, ki je skladijsko manj raziskana, še ni dognano.« (Toporišič 1982: 161).

Žal je Toporišičevo mnenje glede obravnave in poznavanja besednega reda (predvsem stalnega) še vedno aktualno. Še vedno je poznavanje zakonitosti besednega reda slovenskega jezika, ki bi zajemalo tako pisni kot govorni diskurz, le delno. Toporišičeva slovnica (2000: 667) in posamezne razprave (npr. Žele 2018, Zuljan Kumar 2003 in 2019) se dotikajo (le) nekaterih osnovnih vprašanj, temeljitih raziskav, ki bi se poglobile v prosto in stalno stavo, še ni.

¹²

S shemo odvisnosti bi navedeni primer lahko ponazorili kot:



oz. po teoriji S-struktur:

S/s1/s2

ki pa skladijskih in pomenskih razmerij ne prikazuje ustrezno. Primerjava obeh načinov prikaza podanega primera (shema odvisnosti vs. S-strukture) potrjuje, da bi morali teorijo S-struktur zelo dopolniti oz. jo iz šolskih načrtov celo odstraniti in nadomestiti s shemo odvisnosti (stavčne drevesnice).

Besedni red je zagotovo eno izmed tistih skladenjskih poglavij, ki zahtevajo upoštevanje zakonitosti vseh ravni jezika. Čeprav se zdi, da je stalna stava dobro poznana, izčrpno obdelana, je v resnici daleč od tega.

Primeri

- zasebni pisni komentar (pogovor), wikipedija (Janes): Kdaj je bila res ustanovljena Univerza v Mariboru, 2. julija ali 19. septembra 1975. Ugibam, da je bil 2. julija sprejet zakon o ustanovitvi, svečana razglasitev pa 19. septembra. Imam prav? **Sem** se pravkar vrnil iz ene sejne sobe na Rektoratu UM, kamor sem šel na hitro pogledat razstavo dokumentov o zgodovini UM.
- zasebni pisni komentar (pogovor), wikipedija (Janes): Umetniki pa s krogom (ali z drugim upodobitvenim objektom) pojasnijo ali hočejo pojasniti vse. **Me** zanima, kakšne ilustracije so v znamenitih Laundau-Lifšicovih učbenikih.
- intervju v lektorirani reviji (DiŽ), namenjeni širokemu krogu bralcev: Življenje skupaj pa se mi ne zdi v redu, izkušnja iz moje primarne družine ni tako dobra. **Bi** si želel manj vpliva none. **Sem** si moral izboriti vladavino v hiši, saj je bila nona šefica [...].

kažejo na neskladje med spontanim (govorjenim in spontano zapisanim) jezikom na eni strani in zbornim, knjižnim jezikom na drugi. Nujno bi bilo tako preučiti ne le stavo v govornih besedilih, pač pa tudi stavo v umetnostnih in nekaterih publicističnih besedilih, kjer bi tako stalna kot prosta stava zagotovo izkazovali posebnosti v primerjavi z besednim redom, kakršnega opisuje SS. Začetna stava naslonk bi lahko npr. dobila svoje mesto tudi v opisni slovnici slovenskega (knjižnega) jezika. Seveda pa bi morali za to predhodno poglobljeno raziskati najrazličnejše besedilne žanre.

c) Poročani govor

Obravnava poročanega govora je še eno izmed skladenjskih poglavij, ki bi nujno morale dobiti prenovljen opis, saj je obravnava v SS (2000: 653–655) daleč od realne rabe. Povezano je z obravnavo skladenjskih struktur oz. skladenjskimi razmerji.

Samosvoje, posebne (delno še neanalizirane) oblike poročanega govora so prisotne v spontanem govoru in drugih oblikah diskurza (publicističnem,¹³ literarnem itd.). V govorjenem diskurzu so se zaradi različnih okoliščin, ki pogojujejo tvorjenje besedila (npr. prozodija, sočasnost tvorjenja in prejetanja besedila, neverbalni znaki govornika, drugačne skladenjske in širše besedilne zakonitosti govorjenega besedila v primerjavi s pisnim, knjižnim jezikom itd.), razvile oblike, ki »vdirajo« tudi v druge zvrsti diskurzov (npr. pisni, knjižni jezik). Zaradi pravkar napisanega bi bilo nujno celostno in temeljito korpusno izluščiti vse pojavne oblike poročanega govora in tako prenoviti, dopolniti opis, kakršnega ponuja SS. Vse »nove« ali še »nezaznane« vrste poročanega govora pa poleg posebnosti na strukturni oz. oblikovni ravni kažejo posebne, njim lastne značilnosti tudi na metabesedilni ravni (npr. približevanje besedilnega sveta prejemniku, združevanje časovnih dimenzij prvotnega in drugotnega govornega dogodka, stopnjevanje konativne vloge, stopnjevanje dramatičnosti pripovedovanega dogodka itd.). Vse to pomeni, da si temeljite analize kateregakoli jezikovnega pojava, predvsem znotraj govorjenega diskurza, brez kvalitetne računalniške obdelave oz. kvalitetne programske infrastrukture, ki bi omogočala zajem vseh naštetih okoliščinskih danosti, ne moremo zamisliti.

Za konec obravnave izbranih skladenjskih struktur lahko ponovno izpostavimo, da opazovanje rab in različnih skladenjskih oblik ni zanimivo in nujno le s stališča primerjave med katerimkoli pisnim in (spontano) govorjenim jezikom, pač pa predvsem s stališča kodificiranja norme pisnega oz. knjižnega ali standardnega jezika. Poznavanje zakonitosti govorjenega jezika lahko olajša ali pojasni marsikatero težavo, nejasnost ali »nenaravnost,« ki se pojavlja v pisnem jeziku. Analiza govorjenega jezika je ključnega pomena tudi takrat, kadar je nek določen slovnični pojav pogostejši v govorjenem jeziku, manj prisoten pa je v pisnem. Če želimo celostno zaobjeti slovnični ustroj jezika, moramo v analizo, v opis vključiti tudi obrobne ali manj pogoste slovnične prvine. Temeljito poznavanje slovničnih zakonitosti govorjenega jezika je nujno tudi s stališča kontrastivne analize. Če je npr. opuščanje spremnega stavka ali pa zamenjava glagola rekanja z glagolom mišljenja ali pa celo drugo besedno vrsto v slovenskem jeziku pričakovana, je npr. v francoščini to zaznamovano oz. celo nesprejemljivo. Pravkar napisano seveda predstavlja nemalo prevodoslovnih zagat, ki pa jih je s spoznavanjem in uzaveščanjem razlik med spontano govorjenim in pisnim jezikom na eni strani ter med različnimi tujimi jeziki na drugi strani lahko uspešno premagovati.

¹³ V poročevalskih besedilih je poročani govor dobro raziskan. Gl. npr. Tomo Korošec (1998) in Klavdija Bec (2016).

4 Sklep

Danes si temeljitega opisa slovničnega ustroja jezika ne moremo in ne smemo predstavljati brez vključevanja vseh zvrsti jezika (npr. kateregakoli pisnega brez govornega), prav tako pa je nujno vključevanje znanja, doprinosa vseh vej jezikoslovja. Tudi tu enega brez drugega ne bi smelo biti, saj so v nasprotnem primeru dobljeni rezultati omejeni, poenostavljeni ali morda celo napačni. Kvalitetna računalniška obdelava jezika mora temeljiti na kvalitetnem poznavanju slovničnega ustroja jezika. Velja tudi obratno, zato je nujno, da se oba vidika raziskovanja slovenskega jezika razvijata hkrati in v sodelovanju.

Cilj – s stališča opisne gramatikografije – bi moral biti skupen, in sicer izčrpna slovnica, ki bi lahko odgovorila na vse zahteve računalniške obravnave jezika, obenem pa tudi zahteve splošnega uporabnika jezika, šolajočih se itd.

Literatura

- Douglas BIEBER in drugi, 1999: *Grammar of Spoken and Written English*. London: Longman.
- Claire BLANCHE-BENVENISTE, 2007: Corpus de langue parlée et description grammaticale de la langue. *Langue et société* 121/122. 129–141.
- Velemir GJURIN, 1974: Interesne govornice sleng, žargon, argo. *Slavistična revija* 22/1, 65–81.
- Janez GRADIŠNIK, 1965: Hvalnica lepi slovenščini. *Jezikovni pogovori*. Ur. France Vurnik. Ljubljana: Cankarjeva založba. 60–67.
- Živa GRUDEN, 1975: *Govorni signali. Poglavje iz govornega jezika tržaških Slovencev*. Diplomsko delo. Ljubljana: Univerza v Ljubljani.
- Jože KOKOLE, 1969: *Biografije in bibliografije univerzitetnih učiteljev in sodelavcev*. Ljubljana: Univerza v Ljubljani.
- Nina LEDINEK, Tomaž ERJAVEC, 2006: Slovenska odvisnostna drevesnica: prvi rezultati. http://nl.ijs.si/is-ltc06/proc/31_Erjavec_1of2.pdf (17. 5. 2023).
- Nina LEDINEK, 2018: Skladenjska analiza slovenščine in slovenski jezikoslovno označeni korpusi. *Jezik in slovstvo* 63/2/3, 103–116.
- Breda POGORELEC, 1964: *Veznik v slovenščini*. Doktorska disertacija. Ljubljana: Univerza v Ljubljani.
- Breda POGORELEC, 1965: Vprašanja govornega jezika. *Jezikovni pogovori*. Ur. France Vurnik. Ljubljana: Cankarjeva založba. 132–156.
- Mirko RUPEL, 1965: Zadnje leto. *Jezikovni pogovori*. Ur. France Vurnik. Ljubljana: Cankarjeva založba. 28–59.
- Jože TOPORIŠIČ, 1970: Slovenski pogovorni jezik. *Slavistična revija* 18/1/2, 55–70.
- Jože TOPORIŠIČ, 1982: *Nova slovenska skladnja*. Ljubljana: DZS.
- Jože TOPORIŠIČ, 2000: *Slovenska slovnica*. Maribor: Obzorja.
- Jože TOPORIŠIČ, 2007: *Jezikovni pogovori iz Sedem dni*. Maribor: Založba Pivec.
- Božo VODUŠEK, 1953: Kakšen naj bo naš odrski jezik? Ali imamo pogovorni jezik na odru in ali ga imamo v življenju? *Slovenski poročevalec* 14/296, 4.
- Danila ZULJAN KUMAR, 2003: Nekaj besednorednih posebnosti v nadiškem in briškem narečju. *Jezikoslovni zapiski* 9/2, 59–80.

- Danila ZULJAN KUMAR, 2019: Besedni red v slovenskem narečnem diskurzu. *Slovenski jezik* 12, 53–74.
- Andreja ŽELE, 2018: O aktualnostnočlenitveni stavi v slovenščini. *Jezik in slovstvo* 63/2/3, 59–73.



II

Ob gradnji
govornih virov

NADGRADNJA DIGITALNE SLOVARSKÉ BAZE ZA SLOVENŠČINO IN SLOVENSKEGA OBLIKOSLOVNEGA LEKSIKONA SLOLEKS S PODATKI O GOVORJENI SLOVENŠČINI: NAČRTI IN CILJI

JAKA ČIBEJ, NEJC ROBIDA, SIMON KREK

Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
jaka.cibej@ff.uni-lj.si, nejc.robida@ff.uni-lj.si, simon.krek@ff.uni-lj.si

V prispevku predstavljamo načrte in cilje za dopolnjevanje jezikovnih virov, kot sta Digitalna slovarska baza za slovenščino in Slovenski oblikoslovni leksikon Sloleks, s podatki o govornici slovenščini oz. natančneje o tipično govornem besedišču, in sicer predvsem za namene jezikovnotehnoloških potreb (npr. razpoznavalniki in sintetizatorji govora). Po kratkem pregledu sorodnih raziskav predstavimo gradivo, ki ga bomo uporabili za ta namen (korpora GOS in JANES), ter poglobljene izzive, na katere naletimo pri vključevanju nestandardnega besedišča v obstoječe vire, ki so bili do zdaj namenjeni predvsem pisni standardni slovenščini. Poleg problematike kanoničnih oblik (npr. lavfati/laufati) naslovimo npr. tudi tematiko nestandardnih fonemov ([ˈgrɔːza] vs. [ˈɦrɔːza]), nestandardnih izgovorjav standardnih besednih oblik (mislim [ˈmiːslim] → [ˈmiːsləm]) ter nestandardne morfologije (Mihatov, opravičavam). Opisane izzive bomo v okviru projekta MEZZANINE opisali, rešitve pa dokumentirali v smernicah, ki bodo omogočile sistematično polnjenje obstoječih jezikovnih virov s tipično govorno leksiko.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.2](https://doi.org/10.18690/um.ff.4.2024.2)

ISBN
978-961-286-882-6

Ključne besede:

Sloleks,
leksikon,
govornica slovenščina,
nestandardno besedišče,
korpora govornice
slovenščine



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.2](https://doi.org/10.18690/um.ff.4.2024.2)

ISBN
978-961-286-882-6

EXTENDING THE DIGITAL DICTIONARY DATABASE OF SLOVENE AND THE SLOLEKS MORPHOLOGICAL LEXICON OF SLOVENE WITH SPOKEN SLOVENE DATA: PLANS AND GOALS

JAKA ČIBEJ, NEJC ROBIDA, SIMON KREK

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
jaka.cibej@ff.uni-lj.si, nejc.robida@ff.uni-lj.si, simon.krek@ff.uni-lj.si

Keywords:
Sloleks,
lexicon,
spoken Slovene,
non-standard vocabulary,
corpora of spoken Slovene

This paper presents plans and goals for extending language resources such as the Digital Dictionary Database of Slovene and the Sloleks Morphological Lexicon of Slovene with data on spoken Slovene – particularly typically spoken vocabulary – for language technology purposes (e.g. speech recognition and synthesis). After a brief overview of related work, we present the material we will use for this purpose (the GOS and JANES corpora) as well as the main challenges we encounter when incorporating non-standard vocabulary into existing resources that have so far been mainly intended for written standard Slovene. In addition to the issue of canonical forms (e.g. *lavfati/laufati*), we also address the issues of non-standard phonemes ([^hgrɔ:za] vs. [^hirɔ:za]), non-standard pronunciations of standard word forms (*mislím* [^hmi:slím] → [^hmi:sləm]) and non-standard morphology (*Mihatov*, *opravičavam*). The challenges will be described in the framework of the MEZZANINE project, and the solutions will be documented in guidelines that will enable the systematic extension of existing language resources with typical spoken lexis.



1 Uvod¹

Slovensko jezikoslovje (tudi korpusno) se je do zdaj pri gradnji jezikovnih virov v veliki meri osredotočalo na pisni jezik, deloma tudi zato, ker je pridobivanje gradiva v pisni obliki enostavnejše in hitrejšo ter časovno in finančno manj zahtevno od pridobivanja zvočnih posnetkov, zamudnega transkribiranja in urejanja kompleksnih pravnih omejitev v zvezi z osebnimi podatki in avtorskimi pravicami. V zadnjem desetletju pa je konstanten napredek pri razvoju govornih tehnologij (npr. razpoznavalnikov in sintetizatorjev govora) vzrok za povečano potrebo po gradnji jezikovnih virov za govorno slovenščino. V slovenskem korpusnem jezikoslovju je bila ta v primerjavi s pisno (zlasti standardno) slovenščino deležna manj pozornosti: na voljo so že nekateri korpusi in podatkovne baze govorne slovenščine – npr. GOS v1.1 (Zwitter Vitez et al. 2015), skladiščno označeni Slovenian UD Treebank (Dobrovoljc & Nivre 2016), GOS-VL v4.2 (Verdonik et al. 2021) ter Artur v0.1 (Verdonik et al. 2022) in GOS v2.0 (Zwitter Vitez et al. 2023); zadnja dva sta bila izdelana v nedavno zaključenem projektu Razvoj slovenščine v digitalnem okolju (RSDO) –, precej manj pozornosti pa je bilo namenjene govornjeni slovenščini pri leksikonskih in leksikografskih virih, kot sta Slovenski oblikoslovni leksikon *Sloleks* (Čibej et al. 2022) in Digitalna slovarska baza slovenščine (Kosem et al. 2021)² *Sloleks*, ki v različici 3.0 vsebuje iztočnice, njihove pregibne oblike in podatke o njihovih izgovorjavah (v mednarodnih fonetičnih abecedah IPA in SAMPA), še ni bil razširjen s podatki, ki so tipični za govorno slovenščino. Pomanjkanje podatkov o npr. tipično govornem besedišču predstavlja oviro pri uspešni implementaciji jezikovnih tehnologij za slovenščino, kot so npr. razpoznavalniki govora; če določena beseda (npr. *zrihtati*, *poštimiti*) in podatki o njenem izgovoru niso vključeni v zaledne podatkovne baze, na katere se zanaša razpoznavanje govora, je razpoznavalnik bodisi ne razpozna ali pa jo razpozna napačno.³

¹ Prispevek je nastal v okviru raziskovalnega projekta *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik* (J7-4642) in raziskovalnega programa *Jezikovni viri in tehnologije za slovenski jezik* (P6-0411), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

² Tudi rastoči Slovar slovenskega knjižnega jezika (eSSKJ) se denimo opira na pisno jezikovno gradivo (npr. korpusi Gigafida 2.0, slWaC, KAS; Divjak Race in Gliha Komac 2022).

³ Na težave pri razpoznavi tipično govornega besedišča smo npr. že naleteli pri projektu Online Notes, ki je namenjen razvoju sistema za samodejno prevajanje slovenskih predavanj v tuje jezike: <https://www.cjvt.si/online-notes/>

Če želimo tovrstne pojave vključiti v leksikon, moramo proučiti, kako sistematično navajati podatke o govorni slovenščini v pisni obliki tako, da so intuitivni uporabnikom in čim bolj odsevajo jezikovno rabo, hkrati pa so strojno berljivi ter neposredno uporabni za razvoj jezikovnih tehnologij. To praznino v jezikovnih podatkih nameravamo zapolniti v projektu MEZZANINE. Ena od nalog delovnega sklopa 4 je med drugim tudi leksikonska in leksikografska⁴ obravnava besedišča, ki se tipično pojavlja v govorni slovenščini, ne (oziroma zelo redko) pa v pisni.

2 Sorodne raziskave

Z govornim slovenskim jezikom so se do 60. let prejšnjega stoletja še največ ukvarjali dialektologi, kot raziskovalce govornega jezika pa Smolej (2012) pa v svoji raziskavi zvrstnosti besedil v spontanem govoru izpostavlja predvsem Bredo Pogorelec, Jožeta Toporišiča in Borisa Urbančiča. Od raziskav iz zadnjega desetletja, se jih večina ukvarja predvsem z žanrsko analizo (Lengar Verovnik 2010, Verdonik 2017), tudi kot s predpogojem pred primerjavo govornega in pisnega jezika (Zwitter Vitez 2016). Precej raziskav je tudi na temo socialnih zvrsti v govornem jeziku (Kalin Golob 2008, 2009; Poteko 2019, Rolih 2017, Race 2021). Omeniti je treba tudi analizo razlik med govorcami in govorkami glede na vrsto diskurza in druge metapodatke (Zwitter Vitez 2019, 2016). Tipično govorno besedišče, ki je osrednja tematika pričujočega članka, pa do zdaj še ni bilo sistematično raziskano z vidika kanoničnosti zapisa in vključevanja v jezikovne vire. Verdonik (2017) je v govoru denimo raziskovala diskurzne označevalce, kot so *ja, aba, aja, mhm, okej, no, eee, eem* itd., in izpostavila, da so „/o/vira /.../ tudi različni principi transkribiranja gradiva, zlasti pri segmentiranju na izjave in pravilih zapisa“ (Verdonik 2017: 95).

Obstaja že nekaj slovarskih virov, ki opisujejo tipično govorno besedišče (predvsem narečno): npr. slovar podjunske narečne leksike (Benko 2013), besednjak nadiškega narečja (Špehonja 2012), slovar oblačilnega izrazja v Kanalski dolini (Kenda Jež 2007),⁵ a gradivo ni bilo pridobljeno korpusno in ni strojno berljivo, kanonične oblike pa so bile pripisane od zgoraj navzdol, tj. brez vpogleda, kako govornici in govornice te besede dejansko zapisujejo. Ti vpogledi so postali mogoči šele s preselitvijo oz. razširitvijo dela nestandardne komunikacije na splet, ko je tipično

⁴ Še večji problem so višji nivoji obdelave s semantičnimi tehnologijami, saj še ni podatkov o pomenu tipično govornih besed. Tudi obogatitev DSB s pomenskimi podatki je v načrtu v projektu MEZZANINE, a se v tem prispevku omejujemo le na dopolnitev z oblikami in izgovorjavami.

⁵ Celoten nabor slovarskih virov z narečno leksiko je na voljo na povezavi: <http://bos.zrc-sazu.si/c/dial/>

govorjena leksika postala tudi zapisana in jo lahko najdemo npr. v tvitih, e-poštnih sporočilih in podobnih žanrih v spletni slovenščini (na to temo več v Čibej 2021, Michelizza 2015 ter Zwitter Vitez in Fišer 2018).

3 Korpusni podatki kot podlaga za analizo tipično govornjene leksike

Da bi proučili probleme, na katere naletimo pri vključevanju tipično govornjenega besedišča v digitalne jezikovne vire, smo se oprli na gradivo iz dveh korpusov. Kot vir za govornjeno slovenščino smo uporabili trenutno največji in najsodobnejši vir za govornjeno slovenščino GOS 2.0⁶ (Zwitter Vitez et al. 2023), ki obsega 320 ur transkribiranih posnetkov govora (uravnoveženega glede na različne tipe govornih dogodkov, npr. radijske in TV-oddaje, predavanja, zasebni pogovori, sestanki itd.; glej Robida et al. 2023: 35), zajetih med letoma 2007 in 2022, oz. približno 2,5 milijona pojavnic. Iz korpusa GOS 2.0 smo izvozili frekvenčni seznam lem, pri čemer smo izločili vse leme, ki so že vključene v Sloleks, in tiste, ki se v korpusu pojavijo le enkrat. Na ta način smo pridobili 3.879 lem, ki smo jih pregledali, označili pa smo tiste, ki so potencialno nestandardne oz. tipično govornjene (skupaj 503 leme). Na podlagi pregleda gradiva smo identificirali težave, ki jih je treba nasloviti v smernicah za vključevanje lem v jezikovne vire (več o tem v razdelku 4).

Ker iz korpusa GOS 2.0 pridobimo le transkribiran zapis govora, tj. zapis, kot so ga po določenih smernicah zapisali transkriptorji, ta ne odseva nujno zapisa, kakršnega bi uporabili govorci sami. Ker zlasti za problematiko kanoničnih oblik (glej razdelek 4.1) potrebujemo tudi podatke o dejanskem zapisu tovrstnega besedišča, smo na enak način kot pri korpusu GOS 2.0 izvozili seznam lem tudi iz korpusa spletne slovenščine JANES 1.0 (Erjavec et al. 2018); za namene preliminarne raziskave, ki jo opisujemo v tem članku, smo se omejili le na podkorpus tvitov, ki vsebuje 151 milijonov pojavnic oz. 10 milijon tvitov, ki so jih uporabniki omrežja Twitter (približno 9.000) napisali med letoma 2013 in 2017.

⁶ Gos 2.0 je najnovejša različica korpusa, ki je nastala v okviru projekta Razvoj slovenščine v digitalnem okolju z združitvijo korpusov Gos 1.1 (Zwitter Vitez et al. 2015), Gos VideoLectures 4.2 (Verdonik et al. 2021) in dela govorne baze Artur v0.1 (Verdonik et al. 2022).

4 Izzivi pri vključevanju tipično govorjene leksike v jezikovne vire

V tem razdelku na kratko predstavljamo glavne izzive, na katere smo naleteli ob pregledu frekvenčnih seznamov korpusov GOS 2.0 in JANES-Tviti 1.0. Rešitve za predstavljene dileme bodo opisane v smernicah, ki so načrtovane kot rezultat projekta MEZZANINE.

4.1 Kanonične oblike tipično govorjenega besedišča

Pri tipično govorjenem besedišču naletimo na težavo že na nivoju samega zapisa; ker se beseda pojavlja le v govoru in je v standardni slovenščini ni, načeloma tudi ne obstaja dogovorna standardna oblika (zlasti v primerih, ko beseda še ni izpričana v nobenem od obstoječih jezikovnih priročnikov). V takih primerih potrebujemo kanonične oblike, tj. oblike, ki dogovorno predstavljajo tipično govorjeno besedo v pisnem jeziku. Kanonične oblike so potrebne, da je v slovarske baze lahko enotno in sistematično vneseno besedišče govorjene slovenščine (in ne pride npr. do primerov, ko se *tavžent* in *tavžnt* obravnavata kot popolnoma ločeni iztočnici). Jezikovna raba v korpusu JANES-Tviti 1.0 kaže, da enoznačnega odgovora na vprašanje kanoničnih oblik ni: že če npr. opazujemo zapisovanje sklopa *au/av* in primerjamo besedi *lavfati/laufati* in *gravžati/graužati*, vidimo, da se uporabniki_ce odločajo za zapise, ki niso vedno medsebojno konsistentni: zapis *laufati* je neprimerljivo pogostejši kot *lavfati*, po drugi strani pa je zapis *graužati* mnogo redkejši od *gravžati* (Tabela 1).

Tabela 1: Absolutne frekvence lem *gravžati/graužati* in *lavfati/laufati* v korpusu JANES-Tviti 1.0

Lema (1)	Frekvenca (1)	Lema (2)	Frekvenca (2)
lavfati	182	laufati	4.330
gravžati	184	graužati	27

Vir: lasten

Nadaljnja sistematična analiza načinov zapisovanja v korpusih bo pokazala, kako lahko po eni strani zagotovimo čim bolj sistematično zapisovanje tipično govorjenega besedišča v strojno berljivih jezikovnih virih za slovenščino, po drugi strani pa poskrbimo, da je tudi iskanje po jezikovnih virih za uporabnike_ce intuitivno in omogočeno tudi s pomočjo nekanoničnih oblik – načelo, ki mu je sledil tudi korpus GOS z dvema nivojema zapisa – s pogovornim in standardiziranim (glej

Verdonik in Zwitter Vitez 2020: 57–58). To odpira tudi vprašanje splošne obravnave povezav med iztočnicami v Sloleksu, ki v trenutni različici (3.0) še ni zadovoljivo razrešeno: ni npr. še nobene povezave med iztočnicama *asortiment* in *asortima*; *zajedalec* in *zajedavec* (obstaja pa npr. povezava med *volilec* in *volivec*). Na enak način težavo predstavljajo večbesedne enote, ki v trenutno različico Sloleksa še niso vključene; dilema pisanja skupaj in narazen pa se pojavi tudi v tipično govornjem besedišču (npr. *kešpička* vs. *keš pička*; na obe različici naletimo v korpusu JANES-Tviti 1.0).

4.2 Sklopi

V govornjeni slovenščini pogosto naletimo na sklope, ki se v pisni standardni slovenščini zapisujejo narazen, precejšnja pa je tudi razlika med njihovo standardno in nestandardno izgovorjavo. Primeri, na katere naletimo v korpusu GOS 2.0, so npr. *nem* (ne bom), *toj* (to je), *vreji* (v redu), *nav* (ne bo), *daj* (da je), *nam* (ne bom), *kvauš* (kaj boš), *avte* (ali boste) in *navmo/naumo* (ne bomo). Upoštevati je treba, da se tovrstni primeri pojavljajo tudi v pisnem jeziku v nestandardni spletni slovenščini, kot jo izkazuje npr. korpus spletne slovenščine JANES 1.0.

ampak sej san še enkrat no morda da razčistmo pa da **navmo** predolgo tukaj ne (GOS 2.0)
Pa upam da se **navmo** spet zgrešil (JANES 1.0)

Sklopi so težavni za razpoznavalnike govora, saj so v zalednih podatkovnih bazah bodisi še povsem nepopisani (npr. *navmo*) ali pa so prekrivni z drugimi besednimi oblikami (*daj* kot glagol *dati*, *nem* kot pridevnik, *nam* kot zaimek). Problem predstavljajo tudi pri normalizaciji pisnih besedil v nestandardni spletni slovenščini (glej Čibej et al. 2016). Pri vključevanju tovrstnih elementov v Digitalno slovarsko bazo za slovenščino in Slovenski oblikoslovni leksikon Sloleks naletimo na več izzivov. Sloleks 3.0 še ne vsebuje podatkov o večbesednih enotah, zato dodajanje nestandardne izgovorjave [nem] za niz *ne bom* v trenutni strukturi še ni mogoče. V Sloleksu 3.0 so zgoraj omenjenim sklopom najbližje naslonske oblike zaimkov (npr. *zanj*, *nanjo*), ki so vključene kot ločene iztočnice (npr. iztočnica *name*, ki vsebuje pripadajoče oblike *name*, *nanj*, *nanjo*, *nanje* itd.). Pomembna razlika je ta, da so enobesedne različice v teh primerih prav tako standardne (*za njega – zanj*), v primeru *navmo* pa gre za nestandardno obliko. Digitalna slovarska baza je z vidika večbesednih enot trenutno omejena na kolokacije, sopomenske nize, frazeološke enote in podobno, ne pa na nepolnomenske nize tipa *ali boš*, *da je* itd. Določiti je

torej treba učinkovit način, na katerega lahko v bazo in leksikon dodajamo elemente, ki so v nestandardnem jeziku enobesedni, v standardnem večbesedni, izgovorjave enobesedne oblike pa ni mogoče sestaviti iz izgovorjav posameznih besednih delov. Možna rešitev so ločene iztočnice (ustrezno označene kot nestandardne) z dodano individualno izgovorjavo, a morajo biti na voljo povezave na ustrezne enobesedne komponente. Omeniti je treba, da so tudi pri sklopih problem kanonične oblike, kar je razvidno tudi iz nekonsistentnosti, ki jih najdemo pri standardiziranih zapisih iz korpusa GOS 2.0: npr. *pauš/nau/nou/auš* vs. *miuš/nevte/čevš/navmo/davš*, raznolike zapise pa najdemo tudi v korpusu JANES-Tviti 1.0.

4.3 Nestandardne izgovorjave standardnih oblik

Trenutna različica Sloleksa vsebuje le standardne izgovorjave besed, pogosto pa v govornjeni in pisni nestandardni slovenščini naletimo na nestandardne izgovorjave tudi pri besedah, ki imajo neposredne ustreznice v standardni slovenščini. Še zlasti je to očitno pri nekaterih pogostih nepolnopomenskih besedah, npr. *lahko* in *toliko*. V korpusu GOS 2.0 npr. ob pregledu različnih oblik z normalizirano obliko lahko najdemo 54 različnih nestandardnih oblik, med njimi *loh, lah, lahke, lohke, lehko, lahka, lahku, lohko, lohku, leko* in *lejko*. Nekatere od izgovorjav se precej razlikujejo od standardne in zato lahko povzročajo težave pri razpoznavi govora, zato je treba leksikon z njimi ustrezno dopolniti.

Enako velja tudi za nekatere bolj sistematične in predvidljive nestandardne izgovorjave, npr. izpust končnega /i/ v deležniku na -l moškega spola množine – (*oni so*) *začeli* [za'ʧe:l]. Te bomo pridobili z analizo strojno izluščenih najpogostejših razlik med normaliziranimi in dejanskimi pojavnicami v govornjeni transkripciji v korpusu GOS 2.0 – ker je korpus majhen, ne bomo le neposredno dodajali izgovorjav, ki jih najdemo v njem, temveč bomo na podlagi njegovih podatkov poskusili določiti tipične vzorce, ki jih lahko nato v leksikonu apliciramo na izgovorjave različnih besed, ki spadajo v enako kategorijo. Za primer: med najpogostejšimi razlikami med normalizirano in dejansko pojavnico je denimo redukcija nenaglašene /i/ pri glagolih v prvi osebi ednine v sedanjiku, npr. *mislím* ['mi:slím] → ['mi:sləm], *sodím* ['so:dím] → ['so:dəm], *sovražím* [sɔv'ra:zím] → [sɔv'ra:zəm], *vožím* ['vo:zím] → ['vo:zəm]. Na enak način lahko nato tvorimo nestandardne izgovorjave tudi pri sorodnih oblikah, ki jih morda v korpusu GOS

2.0 ne najdemo, npr. *premislim* [prɛ'mi:sləm], *razsodim* [ra's:o:dəm], *zasovražim* [zasov'ra:zəm], *prevožim* [prɛ'vo:zəm].

4.4 Nestandardne izgovorjave standardnih oblik

Sloleks v različici 3.0 ne vsebuje oblik z nestandardnim oblikoslovjem. Izjema so redke nestandardne oblike, ki so bile sporadično dodane v leksikon v začetni različici kot primeri nestandardnega gradiva (za ponazoritev načina, na katerega format leksikona podpira nestandardne podatke) in za namene Slogovnega priročnika v okviru projekta Sporazumevanje v slovenskem jeziku. Primeri so npr. nestandardne oblike iztočnic "hči" in "mati" (npr. tožilnik "hčero", "hči"; tožilnik "mati"), nestandardne pregibne oblike lastnoimenskih samostalnikov (npr. Shakespeare, Shakespearea/Shakespeareja) in izlastnoimenskih pridevnikov ("Shakespearejev" vs. "Shakespeareov"). Tovrstne oblike so v leksikonu označene s kvalifikatorjem "nestandardno".

Ob dopolnjevanju leksikona s podatki o govornjeni slovenščini se odpira priložnost za nadaljnji razmislek o načinu opredeljevanja nestandardnosti v leksikonu, vzpostaviti pa je treba tudi enoten sistem, na katerega so nestandardne oblike povezane s standardnimi ustreznici. Za primer: iztočnici *Voltaireov* in *Voltaireov* sta v trenutni različici ločeni in povsem nepovezani, nestandardne in standardne oblike pa bo na enak način treba povezati tudi npr. v primerih z nestandardno podaljšavo osnove (npr. *Mibov* in *Mihatov*, *Mitjo*/*Mitja* in *Mitjata*).

Pregled frekvenčnih seznamov oblik in lem iz analiziranih korpusov razkrije več različnih pojavov nestandardnega oblikoslovja,⁷ a so mnogo pogostejši v korpusu JANES 1.0 kot v GOS 2.0 (kar je do neke mere pričakovano, saj je JANES 1.0 neprimerljivo večji). Poleg že omenjenih podaljšav osnove s *-t* pri samostalnikih (npr. *Mibata*, *Mihatov*) najdemo tudi npr. nestandardne podaljšave kratkih nedoločnikov in namenilnikov na *-č* s *-t* (*rečt*, *tečt*, *oblečt*) ter alternativne pregibne paradigme pri samostalnikih na *-elj* (*nudelj* - tožilnik množine *nudlje*/*nudeljne*) in pri glagolih na *-ovati*/*-avati* (npr. *obupujem*/*obupavam*):

⁷ Upoštevati je treba, da tudi potencialno standardne oblikoslovne variante niso vse vključene v Sloleks, npr. *stropi* vs. *stropovi*, ali pa so nestandardne oblike prekrivne s standardnimi, npr. *zrak* - *zraku* (rodilnik); dodajanje nestandardnih oblikoslovnih pojavov v leksikon bo torej potekalo vzporedno s standardnimi variantami, a se v prispevku z zgledi omejujemo le na nestandardne.

se **opravičavam** ker sn šele zdaj prišel domov..... (JANES 1.0, 2013, komentar na novice, rtvslo.si)

celo v kratkih rokavih se lahko sprehodiš po njej in **srečavamo** se na istih bregovih, ki pričujejo le vodo v bližini ... (JANES 1.0, 2014, komentar na novice, rtvslo.si)

Trenutno **dokončavam** piksno sonaxovega extreme 3 voska, nad katerim pa nisem pretirano navdušen. (JANES 1.0, 2009, forumsko sporočilo, avtomobilizem.com)

Na podoben način se pojavljajo tudi oblike na *-avlem*, *-avleš* itd.

ne ne **zajebavlem** ovi dela za štiristo evrof ti boš delo za tisoč štiristo evrof (GOS 2.0, 2009, klic prijatelju)

Se **opravičavlem**, če sem narobe dojela. (JANES 1.0, 2017, tvit, Twitter)

Pri teh se odpira tudi vprašanje, kako pripisati ustrezno (kanonično) normalizirano obliko in lemo. Da stvar še ni povsem razrešena, potrjujejo tudi nekonsistentnosti v transkripcijah govora za korpus JANES 1.0, kjer so normalizirane oblike za enake pojave pogosto različne od primera do primera, kot lahko vidimo v naštetih zgledih (normalizirana oblika je v oglatih oklepajih):

gda gledan letnice rojstev teh slavnih ljudi se vedno **spitavlen [spitaen]**, ka v pizdi delan s svojjin življenjon :/ (JANES 1.0, 2014, tvit, Twitter)

zaključil sem opolnoci na stacionu pri **Jakatu [Jakat]**;) (JANES 1.0, 2016, tvit, Twitter)

4.5 Nestandardni fonemi

Fonetični zapisi v mednarodni fonetični abecedi IPA ter njenem ekvivalentu SAMPA so v trenutni različici Sloleksa 3.0 prilagojeni standardni izgovorjavi – to velja tudi za nabor grafemov, ki jih uporabljamo za zapis izgovora. Glasov (fonemov in njihovih variant), ki jih v standardnem jeziku uporabljamo, lahko v Slovenski slovnici naštejemo 61, grafemov zanje pa 60. Vse te glasove (kot sta na primer favkalni in obstranski *d*) že imamo v naboru, z nekaterimi izjemami, npr. t. i. srednja samoglasnika *e* in *o*, ki jih navajata tudi Slovenski pravopis 2001 in 8.0, in višje polglasniško izgovorjeni kratki naglašeni *a* (SP 2001 in Slovenska slovnica), zapisan v paru s kratkim naglašnim *a* ([^la/^lΛ]). Pri vključevanju nestandardnega in tipično govornega besedišča pa bo s tega vidika potreben dodaten premislek o morebitnem

vključevanju nestandardnih glasov oziroma grafemov zanje v nabor vseh simbolov, ki jih uporabljamo za zapis izgovora. V posnetkih govornjene slovenščine v korpusu GOS 2.0 na primer lahko zasledimo tudi zveneči mehkonebni pripornik [ʏ], kadar govorec_ka besedo groza izreče kot ['ʎrɔ:za], ali pa izgovor z jezičkovim *r* (na primer ['ma:rka]). Govorci_ke pogosto tuje besede berejo s tujimi glasovi in izgovora ne poslovenijo, na primer izgovor mesta München, ki ga ne preberejo poslovenjeno kot ['mi:ŋhən], temveč z nemškim glasom *ü* (['my:ŋhən]), zato se odpira vprašanje, katere tuje glasove bi bilo smiselno dodati v nabor fonemov.

Nestandardno gradivo imamo v načrtu vključiti v jezikovne vire predvsem z vidika jezikovnotehnoloških potreb, zato je treba določiti, kateri fonemi sploh (zadostno) vplivajo npr. na uspešnost razpoznave govora; obenem pa tudi, kateri fonemi so dovolj pogosto uporabljeni in regionalno razpršeni, da jih velja vključiti. Za ta namen imamo v načrtu analizo kakovosti prepoznave govora v različnih narečjih, in sicer s pomočjo posnetkov in ročnih transkripcij spletne strani narecja.si; posnetki in transkripcije vsebujejo tudi metapodatke o regijah oz. narečnih skupinah, na podlagi statistične analize pa bomo poskušali ugotoviti, katere nestandardne foneme je smiselno vključevati v nabor in kako natančno mora biti opisana glasovna raven, da dobimo najoptimalnejšo razpoznavo govora in s tem najboljše mogoče samodejno transkribiranje.

5 Zaključek

V prispevku smo na kratko predstavili preliminarno analizo izzivov vključevanja tipično govornjenega besedišča v digitalne jezikovne vire na podlagi pregleda frekvenčnih seznamov lem iz korpusov GOS 2.0 in JANES-Tviti 1.0. V okviru projekta MEZZANINE bomo sistematično in na podlagi realne jezikovne rabe v smernicah opisali probleme in rešitve, ki smo jih implementirali, zato da lahko navajamo podatke o govornjeni slovenščini v pisni obliki tako, da so intuitivni uporabnikom_cam in čim boljše odsevajo jezikovno rabo; so strojno berljivi ter neposredno uporabni za razvoj jezikovnih tehnologij (kot so npr. razpoznavalniki in sintetizatorji govora); hkrati pa ne privedejo do neobvladljivosti digitalnih jezikovnih virov.

Rezultati raziskave bodo poleg že omenjenih smernic tudi novi različici Sloleksa oz. Digitalne slovarske baze, ki bosta obogatena s podatki o govornjeni slovenščini. Na tej točki velja ponovno omeniti, da je v okviru projekta MEZZANINE načrtovana tudi leksikografska obdelava tipično govornjenega besedišča z vidika pomenskih podatkov, kar bo še dodaten doprinos k temu v slovenskem prostoru trenutno še slabo raziskanemu področju.

Literatura

- Pravopis 8.0: Pravila novega slovenskega pravopisa za javno razpravo. Dostop 16. 6. 2022. na www.fran.si/pravopis8.
- Anja BENKO, 2013: *Strokovni narečni slikovni slovar – podjunska narečna leksika*. Dostop 31. 8. 2023 na <https://www.narecna-bera.si/>
- Jaka ČIBEJ et al., 2022: *Morphological lexicon Sloleks 3.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1745>.
- Jaka ČIBEJ, 2021: *Korpusna analiza in prepoznavanje regionalnih jezikovnih različic v spletni slovenščini (doktorska disertacija)*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Jaka ČIBEJ, Darja FIŠER, Tomaž ERJAVEC, 2016: Normalisation, Tokenisation and Sentence Segmentation of Slovene Tweets. *Proceedings of Normalisation and Analysis of Social Media Texts (NormSoMe) 2016, Language Resources and Evaluation Conference (LREC 2016)*. Portorož, Slovenia: 5–10.
- Duša DIVJAK RACE, Nataša GLIHA KOMAC, 2022: Rastoči slovar slovenskega knjižnega jezika (eSSKJ): organizacija in prikaz jezikovnih podatkov. *Leksikologija i leksikografija I : zbornik radova sa medunarodnog naučnog skupa "Leksikografski postupak u različitim tipovima referentnih djela": Sarajevo, 27.–28. maja 2022. godine*. Ur. Senahid Halilović. Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine. Dostop 18. 6. 2023 na <https://publications.anubih.ba/bitstream/handle/123456789/747/3.%20Divjak%20Race%20c%20D.%3b%20Gliha%20Komac%2c%20N..pdf?sequence=19&isAllowed=y>. 45–59.
- Kaja DOBROVOLJC, Joakim NIVRE, 2016: The Universal Dependencies Treebank of Spoken Slovenian. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia.
- Tomaž ERJAVEC, Nikola LJUBEŠIĆ, Darja FIŠER, 2018: Korpus slovenskih spletnih uporabniških vsebin Janes. Ur. Darja Fišer. *Viri, orodja in metode za analizo spletne slovenščine*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Monika KALIN GOLOB, 2008: SMS-sporočila treh generacij. *Slovenščina med kulturami*. Ur. Miran Košuta. 283–294.
- Monika KALIN GOLOB, 2009: Razpadajoči modeli: pogovorne zvrsti na javni prireditvi. *Slovenska narečja med sistemom in rabo*. *Obdobja* 26. Ur. Vera Smole. Ljubljana: Znanstvena založba Filozofske fakultete. 519–525.
- Karmen KENDA-JEŽ, 2007: *Shranli smo jih v bančah: slovarski prispevek k poznavanju oblačilne kulture v Kanalski dolini*. Dostop 31. 8. 2023 na <http://bos.zrc-sazu.si/c/dial/>.
- Iztok KOSEM, Simon KREK, Polona GANTAR, 2022: Semantic data should no longer exist in isolation: the Digital Dictionary Database of Slovenian. *19th EURALEX International Congress "Lexicography for Inclusion"*. Alexandroupolis, 2021.
- Tina LENGAR VEROVNIK, 2010: *Radijski novinarski dvogovorni žanri kot okvir jezikovnih izbir novinarjev (doktorska disertacija)*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani. Dostop 18. 6. 2023 na <https://repositorij.uni-lj.si/Dokument.php?id=114245&lang=slv>.

- Mija MICHELIZZA, 2015: *Spletna besedila in jezik na spletu: Primer blogov in Wikipedije v slovenščini*. Ljubljana: Založba ZRC.
- Ina POTEKO, 2019: Socialnozvrstna analiza govora slovenskih govorcev na YouTubu. *Slovenski javni govor in jezikovno-kulturna (samo)zvest. Obdobja 38*. Ur. Hotimir Tivadar. Ljubljana: Znanstvena založba Filozofske fakultete. 237–246.
- Duša RACE, 2021: *Pogovorni jezik: vrste in položaji (doktorska disertacija)*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani.
- Nejc ROBIDA, Kaja DOBROVOLJC, Luka TERČON, Darinka VERDONIK, 2023: *Gos 2.0 [Elektronski vir]: poročilo projekta Razvoj slovenščine v digitalnem okolju: aktivnost DS1.5*. Ljubljana: Univerza v Ljubljani, Center za jezikovne vire in tehnologije. Dostop 15. 6. 2023 na https://www.cjvt.si/rsdo/wp-content/uploads/sites/18/2023/06/RSDO_Kazalnik_Gos_v2.pdf.
- Maša ROLIH, 2017: *Slang in pogovorni jezik v spletni komunikaciji (doktorska disertacija)*. Koper: Fakulteta za humanistične študije.
- Mojca SMOLEJ, 2012: *Besedilne vrste v spontanem govoru*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Nino ŠPEHONJA, 2012: *Besednjak nediško-taljansko*. Dostop 31. 8. 2023 na <http://bos.zrc-sazu.si/c/Dial/Spehonja/index.html>
- Darinka VERDONIK et al., 2021: *Spoken corpus Gos VideoLectures 4.2 (transcription)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1444>.
- Darinka VERDONIK et al., 2023: *ASR database ARTUR 1.0 (transcriptions)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1772>.
- Darinka VERDONIK, 2017: Vpliv komunikacijskih žanrov na rabo diskurzivnih označevalcev. *Slovenske korpusne raziskave*. Ur. Špela Vintar. Ljubljana: Znanstvena založba Filozofske fakultete. Dostop 15. 6. 2023 na <https://ebooks.uni-lj.si/ZalozbaUL/catalog/download/30/81/852?inline=1>. 88–108.
- Darinka VERDONIK, Ana ZWITTER VITEZ, 2020: *Slovenski govorni korpus Gos*. Ljubljana: Znanstvena založba Filozofske fakultete. Dostop 12. 6. 2023 na <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/228/328/5306>.
- Ana ZWITTER VITEZ, 2019: Javni diskurz in profil govorcev: oblikoslovska in leksikalna analiza korpusa Gos. *Slovenski javni govor in jezikovno-kulturna (samo)zvest. Obdobja 38*. Ur. Hotimir Tivadar. Dostop 13. 6. 2023 na <https://centerslo.si/simpozij-obdobja/zborniki/obdobja-38/>. 255–267.
- Ana ZWITTER VITEZ, 2016: Specifike govorne slovenščine glede na formalnost sporazumevalnega položaja. *Toporiščeva obdobja*. Obdobja 35. Ur. Erika Križišnik in Miran Hladnik. Dostop 19. 6. 2023 na <https://centerslo.si/simpozij-obdobja/zborniki/obdobja-35/>. 351–359.
- Ana ZWITTER VITEZ, Darja FIŠER, 2018: Govorne prvine v nestandardni spletni slovenščini. *Viri, orodja in metode za analizo spletne slovenščine*. Ur. Darja Fišer. Ljubljana: Znanstvena založba Filozofske fakultete. Dostop 19. 6. 2023 na <https://ebooks.uni-lj.si/zalozbaul/catalog/download/111/203/2405-1?inline=1>. 254–272.
- Ana ZWITTER VITEZ et al., 2023: *Spoken corpus Gos 2.0 (transcriptions)*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1771>.
- Ana ZWITTER VITEZ, Jana ZEMLJARIČ MIKLAVČIČ, Simon KREK, Marko STABEJ, Tomaž ERJAVEC, 2021: *Spoken corpus Gos 1.1*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1438>.

SKLADENJSKA DREVESNICA GOVORJENE SLOVENŠČINE: STANJE IN PERSPEKTIVE

KAJA DOBROVOLJC

Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija

kaja.dobrovoljc@ff.uni-lj.si

Institut Jožef Stefan, Odsek za umetno inteligenco, Ljubljana, Slovenija

V prispevku predstavljamo drevesnico SST (angl. Spoken Slovenian Treebank), prvi skladijsko razčlenjeni korpus govorne slovenščine, ki vsebuje uravnotežen in reprezentativni nabor besedil referenčnega korpusa govorne slovenščine Gos z ročno pripisanimi podatki o lemah, besednih vrstah in oblikoslovnih lastnostih besed ter njihovimi odvisnostnimi skladijskimi razmerji. Konkretno drevesnica temelji na označevalni shemi Universal Dependencies (UD), ki si prizadeva za mednarodno poenoteno oblikoskladijsko označevanje besedil in se zaradi svoje interoperabilnosti, fleksibilnosti in naslavljanja širokega nabora slovničnih pojavov – tudi tipično govornih – vse pogosteje uporablja tudi za razčlenjevanje govornih besedil. Po predstavitvi zasnove, vsebine in dostopnosti obstoječe različice drevesnice SST v drugem delu prispevka predstavimo prve rezultate in načrte v povezavi z njenim nadaljnjim razvojem, kot sta razširitev z novimi besedili in nadgradnja smernic za označevanje tipično govornih pojavov.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.3](https://doi.org/10.18690/um.ff.4.2024.3)

ISBN
978-961-286-882-6

Ključne besede:
jezikoslovno označevanje,
skladijsko razčlenjeni
korpusi,
odvisnostna slovnica,
govorni jezik,
korpusno jezikoslovje



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.3](https://doi.org/10.18690/um.ff.4.2024.3)

ISBN
978-961-286-882-6

Keywords:
linguistic annotation,
syntactically parsed corpora,
dependency grammar,
spoken language,
corpus linguistics

SPOKEN SLOVENIAN TREEBANK: CURRENT SITUATION AND PERSPECTIVES

KAJA DOBROVOLJC

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
kaja.dobrovoljc@ff.uni-lj.si

Jožef Stefan Institute, Department for Artificial Intelligence, Ljubljana, Slovenia

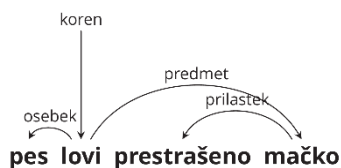
In this paper we present the Spoken Slovenian Treebank (SST), the first syntactically annotated corpus of spoken Slovene containing a balanced and representative set of transcriptions from the Gos reference corpus of spoken Slovene, with manually annotated lemmas, morphological features and syntactic dependencies. The treebank is based on the Universal Dependencies (UD) annotation scheme, which aims at harmonised corpus annotation across languages and is increasingly applied to spoken data due to its interoperability, flexibility and the coverage of a wide range of grammatical structures, including speech-specific phenomena. After summarising the design, content and accessibility of the existing version of the SST, the second part of this paper describes the first results of the ongoing development, which includes the extension of the corpus with new data and the improvement of speech-specific annotation guidelines.



1 Uvod¹

Jezikoslovno označeni korpusi, tj. besedilne zbirke, v katerih so besedam, stavkom, besedilom ali drugim oblikam jezikovnih podatkov pripisane opisne oznake različnih vrst, poleg svoje pomembne vloge na področju računalniške obdelave jezikov, kjer se tovrstni viri uporabljajo za učenje statističnih modelov in evalvacijo na njih temelječih orodij, predstavljajo enega temeljnih virov za razvoj slovničnih teorij in na podatkih temelječe jezikoslovne raziskave (Ide in Pustojevsky 2017).

Med najbolj splošno razširjenimi oblikami jezikoslovno označenih korpusov so skladijsko razčlenjeni korpusi, t. i. skladijske drevesnice, ki zgradbo stavkov opisujejo v obliki hierarhično urejenih grafov (skladijskih dreves), v katerih točke grafa označujejo posamezne besede, povezave med točkami pa skladijska razmerja med njimi. Med raznolikimi teoretskimi pristopi (gl. pregled v Kübler et al. 2009 in Ledinek 2014) danes prevladujejo predvsem skladijske drevesnice na temelju odvisnostne slovnice (Tesnière 1959, Mel'čuk 1988), ki skladijsko strukturo opisuje skozi binarne asimetrične relacije (odvisnosti) med posameznimi besedami. Preprost primer take razčlembे prikazuje slika 1, v katerem je povedek *lovi* analiziran kot jedro stavka, osebek in predmet pa kot njegova odvisna oz. podrejena elementa; prav tako je samostalniškemu jedru predmeta podrejeno njegovo določilo v obliki levega prilastka.



Slika 1: Primer odvisnostnega drevesa

Vir: lasten

Čprav imajo skladijsko razčlenjeni korpusi pisne slovenščine že precej dolgo zgodovino (Džeroski et al. 2006, Erjavec et al. 2010, Krek et al. 2020) in danes največja drevesnica slovenskega jezika obsega že preko 13.000 razčlenjenih povedi (Arhar Holdt et al. 2023), so se maloštevilni poskusi jezikoslovnega označevanja

¹ Prispevek je nastal v okviru raziskovalnega projekta Na drevesnici temelječ pristop k raziskavam govornene slovenščine (št. Z6-4617) in raziskovalnega programa Jezikovni viri in tehnologije za slovenski jezik (št. P6-0411), ki ju sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

govornega gradiva doslej osredotočali predvsem na organizacijo diskurza (npr. Verdonik 2020) in semantično analizo (npr. Antloga 2022), ne pa na slovnično analizo na nižjih ravneh, kot sta oblikoslovna in skladenjska analiza.

Da so tako označeni korpusi pomembni gradivni vir za raziskave govora, potrjujejo tudi številne govorne drevesnice za druge jezike, ki so nastajale od začetka 90. let prejšnjega stoletja, kot so korpusi Switchboard za angleščino (Godfrey et al. 1992), CGN za nizozemščino (van der Wouden et al. 2002), PDTSL za češčino (Hajič et al. 2008), NDC in LIA za norveščino (Øvrelid et al. 2018, Käsen et al. 2022), Rhapsodie za francoščino (Lacheret-Dujour et al. 2019) ter večjezični zbirki Verbmobil (Hinrichs et al. 2000) in CHILDES (MacWhinney 2014), če jih naštejemo le nekaj.

Kot odgovor na ta infrastrukturni manko v slovenskem prostoru je bila leta 2016 izdelana prva skladenjska drevesnica govorne slovenščine, drevesnica SST (angl. *Spoken Slovenian Treebank*), ki še danes predstavlja edini tovrstni jezikovni vir za govorno slovenščino, a njegova vsebina v slovenskem prostoru doslej še ni bila podrobneje predstavljena. Da bi zapolnili to vrzel in osvetlili doslej le delno izkoriščen metodološki potencial tega jezikovnega vira za bodoče jezikoslovne in jezikovnotehnološke raziskave govorne slovenščine, v nadaljevanju prispevka opišemo zasnovo, vsebino in dostopnost obstoječe različice drevesnice SST (Dobrovoljc in Nivre 2016), v drugem delu prispevka pa predstavimo tudi njeno aktualno nadgradnjo znotraj nacionalnega projekta, v okviru katerega bo drevesnica bistveno nadgrajena z vidika obsega in raznolikosti vsebovanih besedil.

2 Zasnova in vsebina drevesnice SST

Korpus, na katerem temelji drevesnica SST, je bil izdelan za potrebe označevanja in analize diskurznofunkcijskih stalnih besednih zvez v slovenskem govoru (Dobrovoljc 2018) in obsega nekaj manj kot 30.000 besed. Zasnovan je bil kot reprezentativni vzorec takratnega referenčnega korpusa govorne slovenščine, korpusa Gos 1.0 (Verdonik in Zwitter Vitez 2011; Zwitter et al. 2013), ki si je prizadeval za ohranjanje raznolikosti govornih dogodkov in govorcev referenčnega korpusa.

Natanćneje to pomeni, da je bil iz vsakega izmed 287 govornih dogodkov oz. besedil korpusa Gos s pomoćjo računalniškega programa vzorćen sorazmerni delež pojavnic glede na delež pojavnic tega besedila v referenćnem korpusu nasploh. Vzorec vsakega besedila vsebuje niz ene ali več zaporednih govornih vlog, torej neprekinjen in zakljućen govor enega ali več govorcev,² pri ćemer je bil zaćetek vzorćenja znotraj besedila, torej izbira prve vloge v vzorćenem nizu, doloćen naključno.

Kot prikazuje tabela 1, vzorćeni korpus vključuje enako število različnih govornih dogodkov in zelo podobno razmerje posameznih podzvrsti kot referenćni korpus Gos, tj. 33,5 % besedil javnega informativno-izobraževalnega diskurza, 23 % besedil javnega razvedrilnega diskurza, 28 % besedil nejavnega zasebnega diskurza in 15,5 % besedil nejavnega nezasebnega diskurza, zaradi omejenosti na kratke zaključene segmente posameznih govornih dogodkov pa vzorćeni korpus vsebuje nekoliko manjše število različnih govorcev (606) kot referenćni korpus (1.561). Posamezno besedilo vzorćenega korpusa tako v povprećju vsebuje 102 pojavnici, 11 izjav, 8 izmenjanih vlog in 2 različna govorca.

Tabela 1: Velikost in sestava obstojeće različice korpusa SST

Tip diskurza	Besedila	Govorci	Vloge	Izjave	Pojavnice
javni informativno-izobraževalni	129	263	703	959	9.899
javni razvedrilni	42	78	499	726	6.833
nejavni nezasebni	45	102	425	497	4.535
nejavni zasebni	71	163	833	1.006	8.221
SKUPAJ	287	606	2.460	3.188	29.488

Vir: Dobrovoljc 2018: 120

Pri izdelavi vzorca smo sledili zapisovalnim načelom izvirnega korpusa, kar pomeni, da so meje vlog, izjav in besed v vzorćenem korpusu enake tistim, ki so bile roćno doloćene ob nastanku referenćnega korpusa Gos, prav tako pa je bil podedovan tudi sam nabor (besednih in nebesednih) pojavnic.

² Edina izjema v algoritmu so bila besedila javnega informativno-izobraževalnega diskurza, pri katerih se je lahko vzorec po dosegu predvidenega števila pojavnic lahko konćal zgolj z zaključeno izjavo in ne nujno tudi zaključkom celotne vloge govorca. Zaradi monološke narave tovrstnega diskurza so namreć vloge posameznih govorcev pogosto zelo dolge ali obsega celotni govorni dogodek. Take nezaključene vloge sicer predstavljajo zgolj 5,5 % vseh vzorćenih vlog korpus.

3 Označevalna shema

Poleg podedovanih ročnih transkripcij in segmentacij izvornega korpusa so bili (standardiziranim) pojavnicam korpusa SST nato ročno pripisani še podatki o osnovni obliki (lemi), besedni vrsti in drugih oblikoslovnih lastnostih, ter skladenjski vlogi. Konkretno te oznake sledijo dvema označevalnima shemama: označevalni shemi MULTEXT-East za leme in oblikoskladenjske lastnosti ter shemi Universal Dependencies (UD) za oblikoslovnne lastnosti in odvisnostne skladenjske relacije, pri čemer slednja v slovenskih drevesnicah posredno vsebuje tudi prvo, kot podrobneje razložimo v nadaljevanju.

3.1 Shema Universal Dependencies

Universal Dependencies je mednarodno oz. medjezično usklajena shema za slovnično označevanje besedil na oblikoslovni in skladenjski ravni, ki je bila leta 2013 vzpostavljena z namenom, da z neposredno primerljivostjo označenih korpusov za čim več svetovnih jezikov omogoči napredek na področju razvoja jezikovnih tehnologij na eni strani ter kontrastivojezikoslovnih raziskav na drugi. Znotraj sheme UD je bil tako vzpostavljen univerzalni nabor jezikoslovnih kategorij (besednih vrst, oblikoslovnih lastnosti in odvisnostnih skladenjskih relacij) in smernic za njihovo pripisovanje, ki odslej omogoča enotno označevanje podobnih slovničnih pojavov v različnih svetovnih jezikih. Shema temelji na načelih odvisnostne slovnice, njena teoretska izhodišča pa so podrobneje pojasnjena v prispevku de Marneffe et al. (2021). Do danes je bila shema UD prenesena že na več kot 245 korpusov v več kot 140 svetovnih jezikih (Zeman et al. 2023), med njimi tudi na drevesnico pisne (SSJ; Dobrovoljc et al. 2017, Dobrovoljc et al. 2023) in govorjene slovenščine (SST; Dobrovoljc in Nivre 2016).

Čeprav se shema UD večinoma uporablja za razčlenjevanje pisnih besedil, se danes vse pogosteje uporablja tudi za slovnično označevanje transkripcij govorjenega jezika. Poleg že izpostavljenih prednosti sheme, kot sta široka uveljavljenost ter dobro razmerje med mednarodno standardizacijo na eni in jezikovnospecifično fleksibilnostjo na drugi strani, je shema z vidika raziskav govorjenega jezika zanimiva zlasti zaradi svoje visoke stopnje interoperabilnosti, saj omogoča neposredne kontrastivne raziskave med drevesnicami različnih jezikov ali jezikovnih zvrsti (npr. primerjave med govornimi drevesnicami v različnih jezikih ali primerjave med drevesnicami pisnega in govorjenega jezika). Širok nabor univerzalnih slovničnih

kategorij, kot so denimo skladijske relacije za zvalnike, diskurzne členke in samopopravke, obenem omogoča celosten, enonivojski pristop k slovnični analizi jezika, v skladu s katerim se lahko oblikoslovne oz. skladijske oznake pripišejo vsem izgovorjenim pojavom, brez kakršnegakoli predhodnega izključevanja netekočnosti in drugih strukturnih posebnosti govora, kot je bilo to pogosto praksa v preteklosti.

Shema UD je bila za razčlenjevanje govornjenega jezika prvič preizkušena prav na slovenski drevesnici SST, odtlej pa je temu vzoru sledilo že več kot 20 drugih drevesnic govornjenega jezika po vsem svetu, ki tako kot SST vključujejo zgolj transkripcije govora, ter 40 drevesnic, ki vsebujejo tako pisna kot govornjena besedila. Ta trend potrjuje, da je shema dovolj fleksibilna, da jo je mogoče uporabiti tudi za razčlenjevanje govornnih korpusov.

3.2 Nabor oznak sheme UD

Shema UD obsega 17 splošnih, 'univerzalnih' oznak za besedne vrste (npr. *ADJ* za pridevnike), 24 univerzalnih oznak za oblikoslovne lastnosti (npr. *Gender* za spol) z več kot 200 različnimi vrednostmi (npr. *Fem* za ženski spol) ter 37 odvisnostnih skladijskih relacij (npr. *obj* za predmet), pri čemer univerzalnost ne pomeni, da se ti slovnični pojavi pojavljajo v vseh jezikih, temveč da se pojavljajo v dovolj velikem številu jezikov, da so jezikoslovno relevantni. Medtem ko je nabor besednih vrst nespremenljiv, lahko avtorji drevesnic za posamezne jezike predlagajo tudi dodatne oblikoslovne lastnosti in/ali njihove vrednosti (kot npr. *Gender[psor]* za označevanje spola svojine v slovanskih jezikih) ter izpeljave posamičnih skladijskih relacij v obliki z dvopičjem ločenih podoznak (kot npr. *cc:preconj* za prvi del dvodelnih veznikov).

Za slovenščino so podrobnejše smernice, ki s podrobnejšimi pojasnili in številnimi primeri opisujejo prenos oznak vseh treh tipov na konkretne jezikovne pojave v slovenščini, dokumentirane tako na krovni spletni strani projekta UD³ (v angleščini) kot v obliki samostojnega priročnika v slovenščini (Dobrovoljc in Terčon 2023a).⁴ Za lažjo predstavbo Tabela 2 prikazuje nabor vseh univerzalnih relacij s splošnimi

³ <https://universaldependencies.org/>

⁴ <https://wiki.cjvt.si/books/07-universal-dependencies/page/oznacevalne-smernice>

opisi v slovenščini, podrobnejše pregledne tabele za druge ravni pa so poleg omenjenih priročnikov na voljo tudi na spletišču CJVT.⁵

Tabela 2: Seznam jedrnih odvisnostnih skladenjskih relacij sheme UD

Relacija	Kratek opis
acl	stavčni prilastki
advcl	prislovni odvisniki
advmod	prislovna določila (v širšem smislu)
amod	pridevniški prilastki
appos	pristavčna določila
aux	pomožni glagoli
case	predlogi
cc	priredni vezniki
ccomp	stavčna dopolnila (predmetni odvisniki)
conj	priredno zloženi elementi
cop	vezni glagoli
csubj	osebkovni odvisniki
dep	nedoločena povezava
det	določilniki
discourse	diskurzni členki
dislocated	dislocirani elementi
expl	ekspletivne besede
fixed	funkcijske zveze
flat	eksocentrične zveze
goeswith	razdruženi deli besed
iobj	nepremi predmeti
list	sezname
mark	podredni vezniki
nmod	samostalniški prilastki
nsubj	samostalniški osebki
nummod	številčna določila
obj	premi predmeti
obl	odvisne samostalniške zveze
orphan	elementi v eliptičnih strukturah
parataxis	stavčna soredja
punct	ločila
reparandum	samopopravljanja
root	koren povedi
vocative	ogovori
xcomp	odprta stavčna dopolnila

Vir: <https://wiki.cjvt.si/books/07-universal-dependencies/page/predstavitev-oznak>

Smernice UD se tako osredotočajo predvsem na raven oblikoslovnega in skladenjskega označevanja, manj specifične pa so glede načel segmentiranja, tokenizacije in lematizacije, kjer splošne smernice UD podajajo zgolj nekaj okvirnih

⁵ <https://wiki.cjvt.si/books/07-universal-dependencies/page/predstavitev-oznak>

priporočil. Kot smo že omenili, drevesnica SST sledi zapisovalnim odločitvam izvirnega korpusa Gos, leme pa so ji bile pripisane v skladu s smernicami sheme MULTEXT-East (Holozan et al. 2023),⁶ na katerih temeljijo tudi drugi referenčni korpusi slovenskega jezika.

Tudi nasploh je shema MULTEXT-East neposredno vključena v obe slovenski drevesnici UD, saj uradni format sheme, CONLL-U (gl. razdelek 4.2), v enem izmed stolpcev omogoča ohranjanje lokalnih oz. jezikovnospecifičnih oznak, kar smo glede na razmeroma visoko stopnjo podobnosti med obema shema na ravni oblikoslovne analize v samem procesu označevanja tudi izkoristili.

4 Označevanje in objava

4.1 Ročno pregledovanje

Po izdelavi vzorčenega korpusa, opisanega v 2. razdelku, so bila besedilom v prvi fazi označevanja tako najprej pripisane leme in oblikoskladenjske oznake po shemi MULTEXT-East (Erjavec 2012, Holozan et al. 2023), ki smo jih nato avtomatsko pretvorili še v besedne vrste in oblikoslovne oznake UD. Za to pretvorbo je bil namreč že ob nastanku drevesnice pisne slovenščine SSJ (Dobrovoljc et al. 2017) izdelan računalniški program,⁷ ki temelji na številnih ročno zasnovanih pravilih za preslikavo med tema podobnima označevalnima sistemoma. Oblikoslovno označeni korpus je bil nato v drugi fazi strojno skladdenjsko razčlenjen z orodjem MaltParser in naložen na spletno platformo WebAnno CLARIN.SI,⁸ na kateri so bile vse relacije tudi ročno pregledane. Ta proces podrobneje popisuje prispevek Dobrovoljc in Nivre (2016), ki omenja tudi identificirane posebnosti govornega jezika, kakršne predstavimo v razdelku 5.3.

4.2 Primer razčlenjene izjave in format CONLL-U

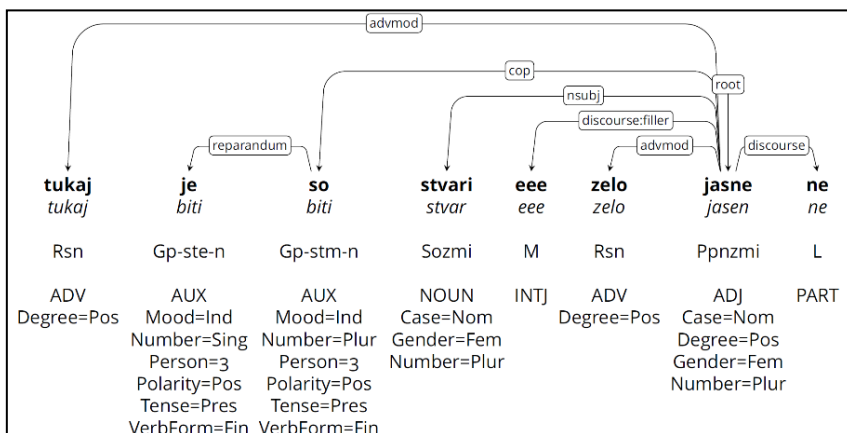
Korpus SST torej za vsako zapisano besedo prinaša ročno pripisani podatek o osnovni obliki leme, oznako MTE ter oblikoslovne in skladdenjske lastnosti (odvisnostne relacije) po shemi UD. Primer tako označene izjave v drevesnici SST je prikazan na sliki 2, na kateri oznake pod transkribiranimi besednimi oblikami

⁶ <https://wiki.cjvt.si/books/04-oblikoskladnja-multext-east/page/oznacevalne-smernice>

⁷ <https://github.com/clarinsi/jos2ud>

⁸ <https://clarin.si/webanno/>

predstavljajo ročno pripisane leme (prva vrstica v poševnem), oznake MTE (druga vrstica) in oblikoslovne lastnosti po shemi UD (tretja in vse nadaljnje vrstice), puščice nad njimi pa odvisnostna razmerja med posameznimi besedami.



Slika 2: Primer označene izjave v drevesnici SST

Vir: lasten

Drevesnica SST je objavljena v standardnem formatu sheme UD, tj. tabelaričnem formatu CONLL-U, v vrsticah zapisane besede oz. pojavnice, v stolpcih pa njihove ročno pripisane lastnosti, kot prikazuje primer izjave v formatu CONLL-U na sliki 3. Konkretno format CONLL-U sestavlja 10 stolpcev:

1. ID: identifikator pojavnice
2. FORM: besedna oblika pojavnice
3. LEMMA: osnovna oblika pojavnice
4. UPOS: besedna vrsta po shemi UD
5. XPOS: oznaka po lokalni označevalni shemi (v primeru SST je to MTE)
6. FEATS: oblikoslovne lastnosti po shemi UD
7. HEAD: identifikator nadrejene pojavnice
8. DEPREL: vrsta skladenjskega razmerja do nadrejene pojavnice
9. DEPS: nadgrajeni odvisnostni graf (v primeru SST ne pripisujemo)
10. MISC: poljubna oznaka (v primeru SST se tukaj beleži podatek o obliki pojavnice v pogovornem zapisu)

Mejo med posameznimi povedmi oz. izjavami označuje ena prazna vrstica in vrstice z metapodatki, ki se začnejo z znakom #. Med slednjimi sta obvezni vrstici z unikatnim identifikatorjem povedi (# sent_id) in izpisanim golim besedilom (# text), dodajanje drugih metapodatkov pa je poljubno. Kot prikazuje primer na sliki 3, drevesnica SST trenutno vsebuje še povezavo do zvočnega posnetka izjave (# sound_url) in identifikacijsko številko govorca (# speaker_id). Tako identifikator povedi kot identifikator govorca sledita nomenklaturi izvornega korpusa Gos, v katerem je na ta način mogoče poiskati tudi podrobnejše podatke o dogodkih in govornikih, npr. da dogodek Gos119 označuje intervju v informativni TV oddaji na temo poslovanja Slovenskih železnic in da je govorec Bm-gost-07155 moškega spola.

# sent_id = Gos119.s72									
# speaker_id = Bm-gost-07155									
# sound_url = https://nl.ijs.si/project/gos20/Gos119/Gos119.s72.mp3									
1	tukaj	tukaj	ADV	Rgp	Degree=Pos	8	advmod	_	pron=tuki
2	je	biti	VERB	Va-r3s-n	Mood=Ind...	4	reparandum	_	pron=je
3	so	biti	AUX	Va-r3p-n	Mood=Ind...	8	cop	_	pron=so
4	stvari	stvar	NOUN	Ncfpn	Case=Nom...	8	nsubj	_	pron=stvari
5	eee	eee	INTJ	I	_	8	discourse:filler	_	pron=eee
6	zelo	zelo	ADV	Rgp	Degree=Pos	8	advmod	_	pron=zelo
7	jasne	jasen	ADJ	Agpfpn	Case=Nom...	0	root	_	pron=jasne
8	ne	ne	PART	Q	Polarity=Neg	8	discourse	_	pron=ne
9	tukaj	tukaj	ADV	Rgp	Degree=Pos	8	advmod	_	pron=tuki

Slika 3: Primer zapisa označene izjave v formatu CONLL-U⁹

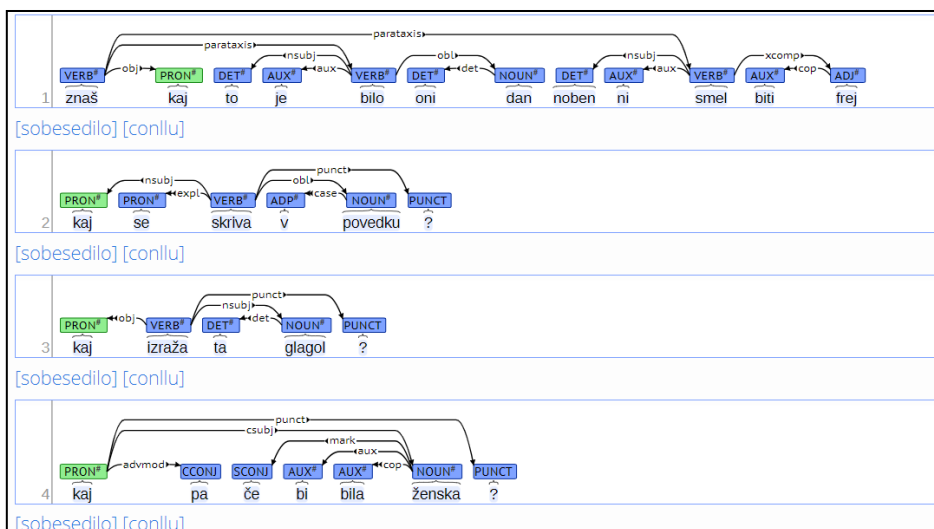
Vir: lasten

4.3 Dostopnost

Drevesnica SST je bila kot odprtodostopna podatkovna množica prvič objavljena leta 2015 pod enako licenco kot transkripcije izvornega korpusa (CC BY-NC-SA), in sicer kot del različice 1.3 uradne zbirke drevesnic UD. Drevesnice UD so namreč distribuirane kot del enotne, skupne korpusne zbirke, ki izhaja dvakrat letno (maja in novembra) in na ta način omogoča kontinuirano priključevanje novih drevesnic in redne izboljšave že obstoječih. V okviru teh polletnih posodobitev se je od prve objave redno izboljševala tudi vsebina drevesnice SST, denimo z odpravljanjem posamičnih napak, posodabljanjem glede na spremenjene označevalne smernice ali

⁹ Zaradi omejenosti s prostorom je v šestem Stolpcu (FEATS) navedena zgolj prva oblikoslovna lastnost v nizu, v zadnjem stolpcu (MISC) pa je atribut 'pronunciation' okrajšan na 'pron'.

z dodajanjem novih metapodatkov. Zadnja različica drevesnice SST je tako izšla kot del zbirke UD v2.12 (Zeman et al. 2023).



Slika 4: Primer rezultatov iskanja po drevesnici SST na portalu Drevesnik

Vir: lasten

5 Nadaljnji razvoj

Da bi zagotovili še kvalitetnejšo gradivno osnovo za nadaljnje raziskave, je kot eden izmed ciljev nacionalnega projekta *Na drevesnici temelječ pristop k raziskavam govornjene slovenščine* (SPOT),¹⁰ ki se v letih 2022–2024 izvaja na Filozofski fakulteti v Ljubljani, v teku temeljita nadgradnja drevesnice SST, ki jo skupaj z drugimi smernicami nadaljnega razvoja predstavimo v nadaljevanju.

5.1 Povečanje korpusa

Obstoječa različica drevesnice SST obsega 30.000 pojavnic, kar jo umešča v zgornjo polovico drugih sorodnih govornih drevesnic za druge jezike, a obenem to predstavlja zgolj eno devetino drevesnice pisne slovenščine SSJ, ki trenutno obsega nekaj več kot 267.000 pojavnic oz. 13.000 razčlenjenih povedi. Poleg že omenjene potrebe po vzpostavitvi statistično močnejših empiričnih temeljev za bodoče

¹⁰ <https://spot.ff.uni-lj.si/>

kvalitativne in kvantitativne raziskave te podatkovne množice je razširitev smiselna tudi zaradi dveh drugih aktualnih pomanjkljivosti.

Prva izvira iz dejstva, da je bila pred kratkim kot rezultat projekta RSDO objavljena nova, skoraj enkrat obsežnejša različica referenčnega korpusa govorne slovenščine, korpus Gos 2 (Verdonik et al. 2023), ki v primerjavi s prvotno različico vsebuje nekoliko drugačno strukturo besedil, saj je bil korpus razširjen z novimi tipi govornih dogodkov, kot so posnetki znanstvenih srečanj, javnih dogodkov in parlamentarnih sej.

Druga pomembna omejitev obstoječe drevesnice SST pa je dejstvo, da je z besedilnega vidika zelo fragmentirana, saj glede na zasnovo vzorčenja, opisanega v 2. razdelku, vsebuje precej kratke segmente zelo širokega nabora govornih dogodkov. To ima seveda številne prednosti za raziskave, pri katerih sta pomembni raznolikost govorcev in govornih situacij, kot sta na primer leksikologija in dialektologija, omejuje pa uporabo v raziskavah jezikovnih pojavov nad nivojem izjave, kot so denimo raziskave strukturiranja diskurza in pragmatike.

5.1.1 Vzorčenje novih besedil

Ob upoštevanju vseh treh naštetih dejavnikov, tj. potrebe po povečanju obsega referenčnega slovnico označenega korpusa govorne slovenščine, zagotavljanju njegove reprezentativnosti glede na Gos 2 in izboljšanju njegove uporabnosti za diskurznoanalitične raziskave, je bil v sodelovanju s projektom Mezzanine, v okviru katerega je prav tako predvidena izdelava ročno označenih korpusov za dialoška dejanja, netekočnosti in prozodično segmentacijo, pred kratkim izdelan vzorec korpusa Gos 2.0 v napovedanem obsegu 50.000 pojavnic, ki bo služil kot gradivna osnova za vse navedene kampanje.

Ker opis postopka vzorčenja presega namen in obseg tega prispevka, na tem mestu zgolj povzamemo, da je vzorčenje potekalo na podlagi ročnega predizbora specifičnih govornih dogodkov iz korpusa Gos 2 v dveh korakih. V prvem so bili povečani oz. podaljšani vzorci izbranih 22 dogodkov v obstoječi drevesnici SST (pribl. 450 novih besed na dogodek oz. skupno pribl. 10.000 pojavnic z delovnim imenom SPOG), v drugem pa so bili izdelani vzorci 57 povsem novih govornih dogodkov iz baze Artur (pribl. 800 novih besed na dogodek oz. skupno pribl. 40.000 besednih pojavnic z delovnim imenom IRISS).

5.1.2 Sestava nove različice drevesnice SST

Kot prikazuje povzetek vsebine vseh treh podkorpusov razširjene različice drevesnice SST (tj. prvotne različice drevesnice SST ter njene razširitve s podkorpusoma SPOG in IRISS), bo nova različica drevesnice bistveno večja (+199 % pojavnic), vsebovala bo še bolj raznolik nabor govorcev (+13 %) in dogodkov (+20 %), vzorčni segmenti dogodkov pa bodo v povprečju tudi daljši. V primerjavi s prvotno različico, v kateri posamezni izsek povprečno obsega 103 besede, 11 izjav in 9 izmenjanih vlog, izseki v novi različici namreč v povprečju obsegajo 257 besed, 19 izjav oz. 12 izmenjanih vlog.

Tabela 3: Velikost in sestava nove, razširjene različice drevesnice SST (v označevanju)

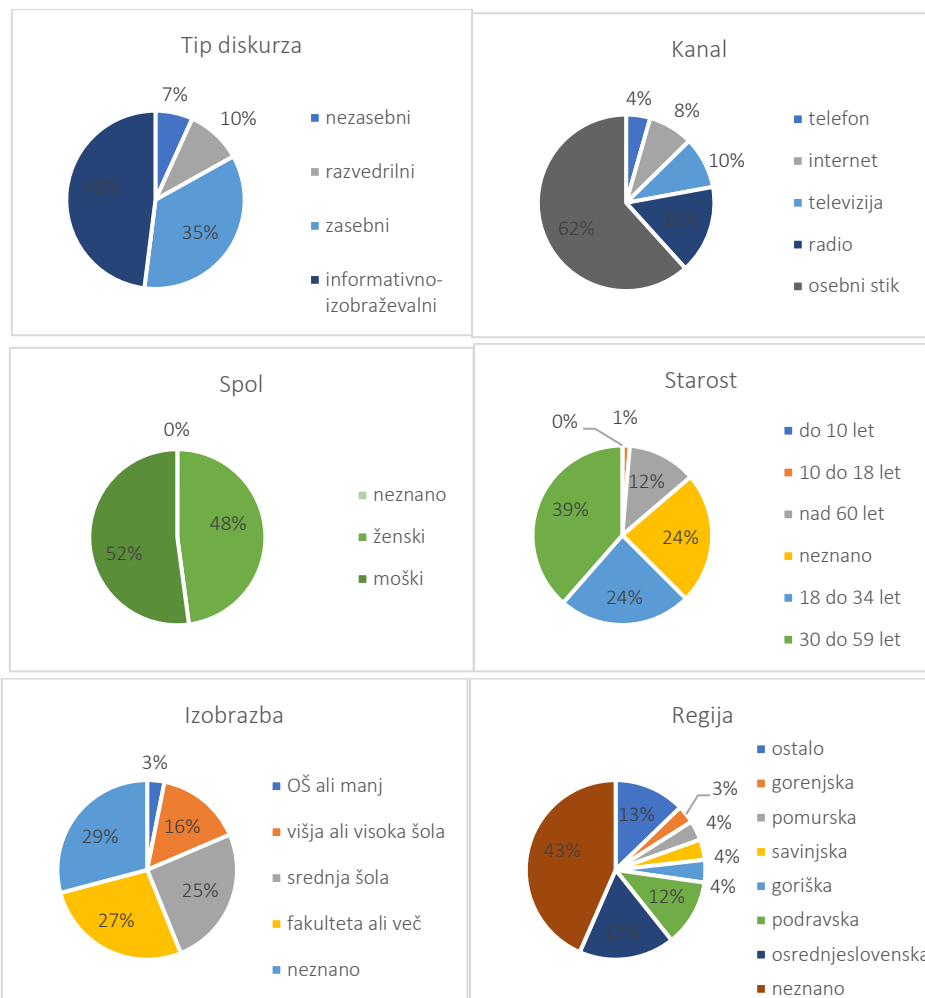
Tip diskurza	Besedila	Govorci	Vloge	Izjave	Pojavnice ¹¹
SST-izvorna	287	606	2.460	3.188	29.488
SPOG	22	63	1.224	1.374	10.184
IRISS	57	72	612	1.968 ¹²	48.624
SST-novi	344	687	4.296	6.530	88.296

Vir: lasten

Kot prikazujejo grafi na sliki 3, pa razlike v načinu vzorčenja prvotne različice drevesnice (tj. manjše število pojavnic velikega nabora dogodkov) in načinu vzorčenja njene razširitve (tj. večje število pojavnic manjšega nabora dogodkov) ne vplivajo na samo reprezentativnost nove različice drevesnice SST, saj tudi ta vsebuje raznolik in uravnotežen nabor govornih situacij ter demografskih lastnosti govorcev.

¹¹ Štetje pojavnic v tabeli 1 je bilo izvedeno na podlagi korpusov v formatu CONLL-U, v katerih so v nasprotju z izvornim zapisom korpusa v XML trenutno kot pojavnice obravnavana tudi anonimizirana imena (npr. [ime], [priimek]), in oznake za premore in nerazumljivi govor, ki se v transkripcijah pojavljajo tudi kot samostojne izjave. Štetje pojavnic v korpusu IRISS vključuje tudi ločila (ki jih v drugih dveh korpusih ni). Morebitno poenotenje podkorpusov z vidika vključevanja ločil, nebesednih pojavnic in uporabe velikih začetnic je predmet širše diskusije o poenotenju podkorpusov Gos 2 (Verdonik et al. 2022) in bo implementirano ob koncu označevalne kampanje.

¹² Štetje izjav korpusa IRISS temelji na resegmentirani različici, v kateri so bile meje med izjavami postavljene na podlagi končnih ločil, kot so pika, vejica in vprašaj. Izvorna segmentacija, podedovana iz baze Artur, je bila namreč bistveno bolj fragmentirana in ni upoštevala skladijsko-pomenske zaključenosti segmentov. Tako je bil na primer niz štirih izjav v bazi Artur oz. Gos 2 (1) /*Drage prijateljice, dragi prijatelji!*, (2) /*govorjene slovenščine.*, (3) /*razmišljal sem, kako naj začnem ta/* in (4) /*svoj nastop/* avtomatsko resegmentiran v niz dveh izjav: (1) /*Drage prijateljice, dragi prijatelji govorne slovenščine.* / in (2) /*razmišljal sem, kako naj začnem ta svoj nastop.* /



n = 88.296

Slika 5: Sestava nove različice drevesnice SST (v označevanju) z vidika deleža pojavnic glede na tip govornega dogodka, komunikacijski kanal in demografske lastnosti govorcev¹³

Vir: lasten

¹³ Graf regionalne pripadnosti govorcev prikazuje delež govorcev glede na statistično regijo prebivališča, pri čemer 'ostalo' označuje regije z manjšo zastopanostjo, tj. jugovzhodna Slovenija (3,0 %), koroška (2,5 %), posavska (2,2 %), obalno-kraška (2,1 %), primorsko-notranjska (1,2 %), tujina (0,8 %), Italija (0,4 %), Avstrija (0,3 %) in Madžarska (0,3 %). Pri zbiranju posnetkov za korpus Gos 1, iz katerega izhajajo besedila korpusa SST in SPOG, je bilo mogoče navesti več regionalnih pripadnosti. V primeru tovrstnih govorcev (pribl. 10 % vseh pojavnic) graf prikazuje zgolj prvo navedeno regijo.

5.2 Sistematična označevalna kampanja

Na podlagi zgoraj opisanih novih korpusnih podatkov (tj. korpusov SPOG in IRISS) že poteka ročno pripisovanje slovničnih oznak v obliki dveh vzporednih označevalnih kampanj. V prvi se besedilom pripisujejo leme in oblikoskladenjske oznake po shemi MTE, ki bodo nato po že preizkušenem pretvorbenem postopku strojno preslikane v besedne vrste in oblikoslovne oznake sheme UD. Konkretno označevalci pregledujejo leme in oznake, ki so bile besedilom strojno pripisane z označevalnikom CLASSLA-Stanza,¹⁴ pri čemer se pregledovanje po izkušnjah nedavne sorodne kampanje na podatkih pisne slovenščine (Pori et al. 2022) osredotoča zgolj na besedne oblike, pri katerih je glede na oblikoslovni leksikon Sloleks (Čibej et al. 2023) možnih več različnih interpretacij (npr. *škarje* kot samostalnik v imenovalniku ali tožilniku), ne pa oblikoslovno nedvoumne oblike (npr. *srajca* kot samostalnik v imenovalniku).

Vzporedno s to aktivnostjo trenutno poteka tudi kampanja skladenjskega razčlenjevanja po shemi UD, v kateri označevalci pregledujejo in popravljajo odvisnostne relacije, ki so bile tem besedilom pripisane s strojnim označevalnikom Trankit,¹⁵ ki se je na skriti tesni množici platforme SloBench izkazal kot najuspešnejši pri nalogi skladenjskega razčlenjevanja.¹⁶ Konkretno sta bila korpusa SPOG oz. IRISS razčlenjena z lokalno razvitim modelom (Krsnik in Dobrovoljc 2024), ki je bil glede na že znane prednosti združevanja različnih vrst učnih podatkov pri razčlenjevanju govorjene slovenščine (Dobrovoljc in Martinc 2018) naučen na kombinaciji obeh referenčnih drevesnic UD za slovenščino, drevesnice SSJ in SST.

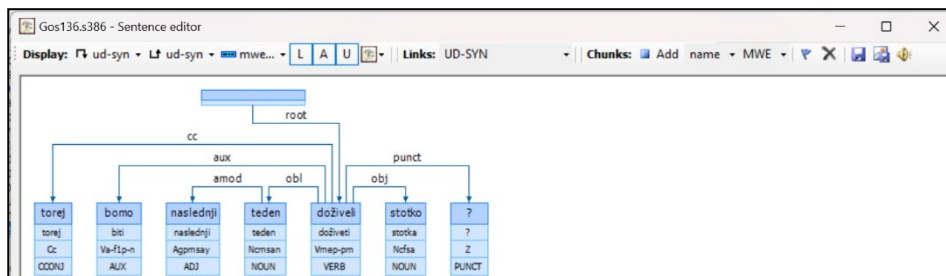
V nasprotju s prvotno drevesnico SST, ki je bila označena s strani ene same označevalke, bodo v tokratni kampanji besedila pregledali 2–3 neodvisni označevalci, kar po eni strani zagotavlja večjo zanesljivost pripisanih relacij, po drugi strani pa tudi lažjo identifikacijo težavnejših mest tovrstne jezikoslovne analize, ki opozarjajo na pomanjkljivosti izhodiščnih jezikoslovnih kategorij oz. smernic za njihovo pripisovanje. Označevanje poteka v orodju Q-CAT (Brank 2023), ki je bilo za potrebe te specifične kampanje nadgrajeno tudi z možnostjo hkratnega poslušanja zvočnih posnetkov izjav za korpusne v formatu CONLL-U, če so povezave do

¹⁴ <https://pypi.org/project/classla/>

¹⁵ <https://github.com/nlp-uoregon/trankit>

¹⁶ Primerjava orodij Trankit, Stanza in CLASSLA-Stanza na skriti tesni množici platforme SloBench: <https://slobench.cjvt.si/leaderboard/view/11>.

posnetkov podane v vrsticah z atributom # sound_url (slika 6). Ker Q-CAT ne podpira primerjave oznak različnih označevalcev oz. njihovega poenotenja v obliki končnih odločitev, za ta korak (t. i. kuriranje) uporabljamo spletno označevalno platformo WebAnno (slika 7).



Slika 6: Razčlenjevanje izjav v orodju Q-CAT (gumb za poslušanje posnetka je v desnem zgornjem kotu)

Vir: lasten

Slika 7: Primerjava oznak dveh označevalcev na platformi WebAnno

Vir: lasten

5.3 Nadgradnja označevalnih smernic s posebnostmi govora

Kot smo omenili že v 3. razdelku, so splošne, jezikovno neodvisne smernice sheme UD objavljene na spletni strani projekta, podrobnejše smernice s pojasnili in primeri prenosa sheme na slovenske podatke pa so bile popisane v obliki priročnika (Dobrovoljc in Terčon 2023a), ki se je v svoji prvi različici osredotočal predvsem na izčrpen opis razčlenjevanja besedil pisne slovenščine, kakršna se pojavljajo v učnem korpusu pisne slovenščine SUK/SSJ.

Za potrebe razčlenjevanja transkripcij govorjene slovenščine smo nedavno ta priročnik dopolnili še z opisi označevanja skladenjskih posebnosti govorjenega jezika, ki služijo kot izhodišče za razčlenjevanje v okviru zgoraj opisane kampanje (Dobrovoljc in Terčon 2023b). Te nadgrajene smernice poleg prenosa (zelo ohlapnih) splošnih smernic UD, ki govor omenjajo zgolj pri najbolj izstopajočih skladenjskih pojavih, kot so samopopravljanja in diskurzni členki, popisujejo doslej nedokumentirano podrobnejšo obravnavo teh in številnih drugih pojavov v okviru označevanja prvotne drevesnice SST (Dobrovoljc in Nivre 2016), obenem pa upoštevajo tudi priporočila, ki so se oblikovala v poznejših diskusijah (Kahane et al. 2021, Dobrovoljc 2022) in skoti primere dobre prakse drugih sorodnih drevesnic govorjenega jezika.

Na oblikoslovni ravni so bili tako dopoljnjeni ponazoritveni sezname nepregibnih in zaprtih besednih vrst (npr. prislovi *plus, kao, komot, direkt, tukajle, ene, prvo*; priredni vezniki *aber, ar*; podredni veznik *ka*; določilnik *ovi*) ter številni medmeti in členki, kot so *eee, eem, žinjo, porkaš, vav, opala, alora, arki, evo, tipo*. Popisane so bile tudi odločitve glede dogovorne besednovrstne kategorizacije nedokončanih besed, nebesednih pojavnic in anonimiziranih imen.

Bistveno obsežnejše pa so dopolnitve na skladenjski ravni, kjer smernice prinašajo podrobnejše opise in ponazoritve relacije *discourse*, s katero se označujejo medmeti, diskurzni označevalci in druge oblike ustaljenih, skladenjsko manj vpetih izrazov, kot so *oh, ja, (a) ne, tako, hvala, škoda* in (s podoznako *discourse:filler*) tudi zapolnjeni premori tipa *eee*. Prav tako je bilo obsežno dopolnjeno poglavje s predstavitvijo relacije *reparandum*, s katero se označujejo samopopravljanja različnih tipov, od popravljanja napačno začeti besed (npr. *kako orožje- orožje pa to*) ali napačnih besednih oblik (npr. *da so te eee ti stroški čim manjši*) do ponavljanj in popravljanj znotraj stavka (npr. *nekega dne sem se eee sem se skregal*).

Med drugimi izpostavljenimi posebnostmi govora smernice denimo naslavljajo tudi označevanje navideznih odvisnikov znotraj relacije *advcl* (npr. *če smem vprašati, kot rečeno*), nadaljevalnikov znotraj relacije *conj* (npr. *in tako naprej, ali pa kaj takega*), ekspletivne vloge kazalnega zaimka *to* znotraj relacije *expl* (npr. *tako da to pol nekega posebnega izobraževanja pa verjetno ni ne*), izpustov povedka znotraj relacije *orphan* (npr. *pri nas pa občasno, kam pa?, tudi Francozinja v težavah*), ponovne začetke izjav znotraj relacije *parataxis* (npr. *kaj si zdaj pravkar katero črto boš narisala*) in obravnavo netipičnega besednega reda (npr. *imamo pa tudi debelo ono uro jekleno*), če jih naštejemo le nekaj.

Nenazadnje smernice prinašajo tudi razširjeni seznam stalnih besednih zvez, pri katerih se notranja struktura členi z relacijo *fixed* (npr. *a la, a ne, hvala bogu, ker da, se pravi da*), ter znotraj relacije *flat* opisujejo skladenjsko členjenje zapisov besednih zvez, kakršne bi v izvorno pisnih besedilih pričakovali zapisane drugače, npr. izgovorjenih decimalnih števil (*dva cela pet*) ali naslovov spletnih strani (npr. *trikrat dvojni v pika radio capris pika si poševnica kikirik*).

Kot smo že omenili v razdelku 5.2, nameravamo smernice pred uradno objavo dodatno dopolniti še na podlagi analize najpogostejših nestrinjanj med označevalci, ki utegnejo opozoriti na jezikoslovna vprašanja, ki so bila v smernicah ali literaturi nasploh naslovljena pomanjkljivo, kot je denimo določanje nadrejenega elementa diskurznih členkov in drugih 'šibko' vpetih struktur. Z namenom zagotavljanja dosledne in izčrpno dokumentirane označenosti nove, razširjene drevesnice SST bodo morebitne novosti ob koncu prenesene tudi na besedila prvotne različice drevesnice SST.

6 Zaključek

V prispevku smo predstavili zasnovno, vsebino, dostopnost in aktualno nadgradnjo drevesnice SST, odprto dostopnega skladenjsko razčlenjenega korpusa govorne slovenščine, v katerem so vsaki transkribirani pojavnici ročno pripisane informacije o besednih vrsti, oblikoslovnih lastnostih in odvisnostnih skladenjskih relacijah po mednarodno uveljavljenih shemah MULTEXT-East in Universal Dependencies.

Kot taka drevesnica predstavlja pomembno podatkovno množico za nadaljnji razvoj in evalvacijo tehnologij za obdelavo slovenskega govora, kot so na primer na govor prilagojeni slovnični označevalniki, ter izjemno dragoceni gradivni vir za kvalitativne

in kvantitativne jezikoslovne analize slovničnih značilnosti govorjene slovenščine, tudi preko primerjave z drugimi sorodnimi drevesnicami pisnih in govorjenih besedil.

Za polni izkoristek tega potenciala v raziskavah slovenskega govora je seveda poleg nadaljnjega razvoja drevesnice z vidika obsega, zanesljivosti oznak in transparentne dokumentiranosti, ki smo ga nakazali v tem prispevku, smiselno tudi njeno kontinuirano dopolnjevanje z drugimi ravnmi jezikoslovnega opisa, ki bi omogočale celostne slovnične analize govora s hkratnim upoštevanjem prozodičnih, slovničnih in pragmatičnih vidikov govornega sporazumevanja.

Literatura

- Špela ANTLOGA, 2022: Identifikacija metafore in metonimije v jezikovnih korpusih: Poskus kategorizacije označenih metonimičnih prenosov v korpusu g-KOMET. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, 11/1, 91–117.
<https://doi.org/10.4312/slo2.0.2023.1.91-117>.
- Špela ARHAR HOLDT et al., 2022: Training corpus SUK 1.0, *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1747>.
- Janez BRANK, 2023: Q-CAT Corpus Annotation Tool 1.5, *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1844>.
- Brian MacWHINNEY, 2000: The CHILDES Project: Tools for Analyzing Talk, 3. izdaja. Psychology Press.
- Marie-Catherine de MARNEFFE, Christopher D. MANNING, Joakim NIVRE, Daniel ZEMAN, 2021: Universal Dependencies. *Computational Linguistics*, 47/2, 255–308.
- Jaka ČIBEJ et al., 2022: Morphological lexicon Sloleks 3.0, *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1745>.
- Kaja DOBROVOLJC, 2018: *Leksikalne prvine govorjenega jezika v uporabniških spletnih vsebinah: primer večbesednih diskurzivnih označevalcev*. Doktorska disertacija. Ljubljana: Filozofska fakulteta UL.
- Kaja DOBROVOLJC, 2022: Spoken Language Treebanks in Universal Dependencies: an Overview. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 1798–1806.
- Kaja DOBROVOLJC, Tomaž ERJAVEC, Simon KREK, 2017: The Universal Dependencies Treebank for Slovenian. *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing, BSNLP@EACL 2017*. 33–38.
- Kaja DOBROVOLJC, Joakim NIVRE, 2016: The Universal Dependencies Treebank of Spoken Slovenian. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. 1566–1573.
- Kaja DOBROVOLJC, Matej MARTINC, 2019: Er ... Well, it matters, right? On the Role of Data Representations in Spoken Language Dependency Parsing. *Proceedings of the Second Workshop on Universal Dependencies (UDW 2018)*. 37–46.
- Kaja DOBROVOLJC, Luka TERČON, 2023a: *Universal Dependencies: Smernice za označevanje besedil v slovenščini. Različica 1.0*.
- Kaja DOBROVOLJC, Luka TERČON, 2023b: *Universal Dependencies: Smernice za označevanje besedil v slovenščini. Različica 1.3*. Ljubljana: Center za jezikovne vire in tehnologije Univerze v Ljubljani.
- Kaja DOBROVOLJC, Luka TERČON, Nikola LJUBEŠIČ, 2023: Universal Dependencies za slovenščino: nove smernice, ročno označeni podatki in razčlenjevalni model. *Slovenščina 2.0: empirične, aplikativne in interdisciplinarne raziskave*, 11/1, 218–246.

- Sašo DŽEROSKI, Tomaž ERJAVEC, Nina LEDINEK, Petr PAJAS, Zdenek ŽABOKRTSKY, Andreja ŽELE, 2006: Towards a Slovene Dependency Treebank. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. 1388–1391
- Tomaž ERJAVEC, 2012: MULTEXT-East: morphosyntactic resources for Central and Eastern European languages. *Language Resources and Evaluation* 46, 131–142.
<https://doi.org/10.1007/s10579-011-9174-8>
- Tomaž ERJAVEC, Darja FIŠER, Simon KREK, Nina LEDINEK, 2010: The JOS Linguistically Tagged Corpus of Slovene. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. 1806–1809.
- John J. GODFREY, Edward C. HOLLIMAN, Jane McDANIEL, 1992: SWITCHBOARD: Telephone speech corpus for research and development. *Acoustics, Speech, and Signal Processing, IEEE International Conference*. 517–520.
- Erhard W. HINRICHS, Julia BARTELS, Yasuhiro KAWATA, Valia KORDONI, Heike TELLJOHANN, 2000: The Tübingen treebanks for spoken German, English, and Japanese. *VerbMobil: Foundations of Speech-to-Speech Translation*. Ur. Wolfgang Wahlster. Springer Berlin Heidelberg. 550–574.
- Peter HOLOZAN et al., 2023: *Specifikacije za učni korpus: lematizacija in MSD. Različica 2.0*.
- Nany IDE, James PUSTEJOVSKY, 2017: *Handbook of linguistic annotation*. Berlin: Springer.
- Sylvain KAHANE, Bernard CARON, Emmett STRICKLAND, Kim GERDES, 2021: Annotation guidelines of UD and SUD treebanks for spoken corpora: A proposal. *Proceedings of the 20th International Workshop on Treebanks and Linguistic Theories (ILT, Syntaxfest 2021)*. 35–47.
- Andre KÅSEN, Kristin HAGEN, Anders NØKLESTAD, Joel PRIESTLY, Per Erik SOLBERG, Dag Trygve Truslew HAUG, 2022: The Norwegian Dialect Corpus Treebank. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 4827–4832.
- Simon KREK et al., 2020: The sss500k Training Corpus for Slovene Language Processing. *Zbornik Konferenca Jezikovne tehnologije in digitalna humanistika 2020*. 24–33.
- Luka KRŠNIK, Kaja DOBROVOLJC, 2024: Trankit model for linguistic processing of spoken Slovenian, *Slovenian language resource repository CLARIN.SI*, ISSN 2820-4042, <http://hdl.handle.net/11356/1909>.
- Sandra KÜBLER, Ryan MCDONALD, Joakim NIVRE, 2009: *Dependency Parsing*. Morgan and Claypool Publishers.
- Anne LACHERET-DUJOUR, Sylvain KAHANE, Paola PIETRANDREA, 2019: *Rhapsodie: A Prosodic and Syntactic Treebank for Spoken French*. John Benjamins Publishing Company.
- Nina LEDINEK, 2014: *Slovenska skladnja v oblikoskladensko in skladenjsko označenih korpusih slovenščine*. Ljubljana: Založba ZRC.
- Igor A. MELČUK, 1988: *Dependency Syntax: Theory and Practice*. State University Press of New York.
- Lilja ØVRELID, Andre KÅSEN, Kristin HAGEN, Anders NØKLESTAD, Per Erik SOLBERG, Janne Bondi JOHANNESSEN, 2018: The LIA Treebank of Spoken Norwegian Dialects. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 4482–4488.
- Eva PORI, Jaka ČIBEJ, Tina MUNDA, Luka TERČON and Špela ARHAR HOLDT, 2022: Lematizacija in oblikoskladensko označevanje korpusa SentiCoref. *Zbornik konferenca Jezikovne tehnologije in digitalna humanistika 2022*. Ur. Darja Fišer, Tomaž Erjavec. Ljubljana: Inštitut za novejšo zgodovino.
- Miha ŠTRAVS, Kaja DOBROVOLJC, 2022: Service for querying dependency treebanks Drevesnik 1.0, *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1715>.
- Lucien TESNIÈRE, 1959: *Éléments de Syntaxe Structurale*. Paris: Klincksieck.
- Ton VAN DER WOUDE et al., 2002: Harvesting Dutch trees: Syntactic properties of spoken Dutch. In *Computational Linguistics in the Netherlands: Selected Papers from the Thirteenth CLIN Meeting*. Ur. Tanja Gaustad. Brill. 129–141.
- Darinka VERDONIK, 2020: Dialogue act annotated spoken corpus GORDAN 1.0 (transcription), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1291>
- Darinka VERDONIK, Ana ZWITTER VITEZ, 2011: *Slovenski govorni korpus GOS*. Ljubljana: Trojina, zavod za uporabno slovenistiko.

- Darinka VERDONIK, Andreja BIZJAK, Andrej ŽGANK, Simon DOBRIŠEK, 2022: Metapodatki o posnetkih in govoricah v govornih virih: primer baze Artur. *Zbornik konference Jezikovne tehnologije in digitalna humanistika 2022*. Ur. Darja Fišer, Tomaž Erjavec. Ljubljana: Inštitut za novejšo zgodovino.
- Darinka VERDONIK et al., 2023: Spoken corpus Gos 2.1 (transcriptions), *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1863>.
- Daniel ZEMAN et al., 2023: Universal Dependencies 2.12, *LINDAT/CLARLAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University*, <http://hdl.handle.net/11234/1-5150>.
- Ana ZWITTER VITEZ et al., 2013: Spoken corpus Gos 1.0, *Slovenian language resource repository CLARIN.SI*, <http://hdl.handle.net/11356/1040>.

PREDNOSTI IN SLABOSTI DVOTIRNEGA ZAPISOVANJA GOVORA V SLOVENSКИH GOVORNIH VIRIH

DARINKA VERDONIK,¹ MITJA TROJAR,² ANDREJA BIZJAK¹

¹ Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,
Maribor, Slovenija

darinka.verdonik@um.si, andreja.bizjak1@um.si

² ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Ljubljana, Slovenija
mitja.trojar@zrc-sazu.si

Zapisovanje govora v govornih korpusih je nedvomno časovno največji vložek v procesu izdelave govornega korpusa in pomemben razlog, da so govorni korpusi neprimerno manjši od pisnih. Zapis govora je prevod iz izvorno večmodalnega kanala komunikacije, v katerem verbalno izraženi pomen sooblikujejo glas in način govora, govorica telesa in situacija, v kateri poteka komunikacija, v eno, pisno modalnost. Zaradi variabilnosti govora na vseh jezikovnih ravneh se zapisovalec pri tem nenehno sooča z vprašanjem, kako naj to, kar sliši, zapiše. Da bi bil zapis čim bolj verodostojen, hkrati pa časovno vseeno izvedljiv za velik obseg gradiv, se je tako poleg standardiziranega zapisa vzpostavil tudi tako imenovani pogovorni zapis, ki sledi načelu zapiši, kakor je izgovorjeno. Toda dvojni zapis zahteva dodaten trud, zato v tem prispevku kritično preverjamo njegovo utemeljenost glede na prakse drugod, zahtevan dodaten trud in njegove prednosti ter kritično analiziramo še druga težavnejša vprašanja zapisovanja.

DOI

[https://doi.org/
10.18690/um.ff.4.2024.4](https://doi.org/10.18690/um.ff.4.2024.4)

ISBN

978-961-286-882-6

Ključne besede:

transkribiranje,
standardizirani zapis,
ortografska transkripcija,
pogovorni zapis,
fonetična transkripcija



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.4](https://doi.org/10.18690/um.ff.4.2024.4)

ISBN

978-961-286-882-6

Keywords:

transcribing, standardized transcription, orthographic transcription, literal transcription, phonetic transcription

ADVANTAGES AND DISADVANTAGES OF TWO-TIER SPEECH TRANSCRIPTION IN SLOVENIAN SPEECH RESOURCES

DARINKA VERDONIK,¹ MITJA TROJAR,² ANDREJA BIZJAK¹

¹ University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia

darinka.verdonik@um.si, andreja.bizjak1@um.si

² ZRC SAZU, Fran Ramovš Institute of the Slovenian Language, Ljubljana, Slovenia
mitja.trojar@zrc-sazu.si

Transcribing speech in speech corpora is undoubtedly the largest time investment in the process of creating a speech corpus and an important reason that speech corpora are considerably smaller than written ones. Speech transcription is a translation from an originally multimodal channel of communication, in which verbally expressed meaning is shaped by the voice and manner of speaking, body language, etc., and converted into a single, written modality. Due to the variability of speech at all linguistic levels, the transcriber constantly faces the question of how to transcribe what s/he hears. In order to make the transcription as exact as possible, but at the same time feasible when working with large amounts of data, a pronunciation-based transcription was introduced in Slovenian speech corpora along with the standardized transcription. However, two-tier transcription requires additional effort. For this reason, this paper critically assesses its rationale, comparing practices used elsewhere, estimates of the additional effort and its advantages. Additionally, we assess other challenging aspects of speech transcription.



1 Uvod¹

Časovno in finančno najzahtevnejši korak izdelave govornih korpusov je t. i. transkribiranje posnetkov oziroma natančneje zapisovanje in označevanje posnetkov. Ne gre namreč samo za zapis govora, ampak je treba popisati tudi podatke o govornikih in posnetih govornih dogodkih, segmentirati govor na osnovne enote – segmente, označiti menjavanje govorcev in kdaj kdo govori, označiti akustično ozadje (npr. prisotnost šuma ali glasbe) in akustične dogodke (nenadni zvoki od zunaj ali nastali z govorili, kot so kašljanje, glasni vdih ipd.) ter osnovne prozodične značilnosti (smeh, premori ipd.). Zaradi časovne in finančne zahtevnosti zapisovanja govora in označevanja posnetkov se ob izdelavi govornih korpusov vedno iščejo načini, kako izvedbo čim bolj ekonomizirati. Predvsem za javno govorjeno rabo, na primer v medijih ali parlamentu, je tako mogoče dobiti arhive posnetkov in včasih tudi zapisov, pri čemer pa se pogosto soočamo s pravnimi in drugimi omejitvami ter posledično njihovo nedostopnostjo (Verdonik 2023: 32). Vse bolj se v proces uvaja uporaba avtomatskega razpoznavanja govora za pripravo zapisa. Toda za posnetke nejavne, zasebne rabe govora lahko še naprej pričakujemo, da bo zaradi zahtevnih akustičnih pogojev in variabilnosti govora na vseh jezikovnih ravneh še dolgo obstajala potreba po ročnem zapisovanju in označevanju.

Na podlagi priporočil EAGLES (Gibbon et al. 1997) in praks v govornih korpusih ločujemo več ravni zapisa govora: bodisi gre za ortografski zapis bodisi za avtomatsko pripravljen fonetični zapis z algoritmi grafemsko-fonemske pretvorbe bodisi za natančen fonemski zapis v fonetični abecedi. Pri ortografskih zapisih lahko nadalje ločujemo:

1. standardni ortografski zapis, to je povsem standardiziran zapis, v katerem se izgovorjene besede in besedne oblike zapiše z ustreznimi standardnimi besedami in besednimi oblikami,
2. razširjeni ortografski zapis (angl. *expanded orthographic transcription*) z navodili, pravili in/ali seznami za dodatno zapisovanje posebnosti izgovorjenih besed in besednih oblik, ki so drugačne od standardnih.

¹ Prispevek je nastal v okviru temeljnega raziskovalnega projekta Temeljne raziskave za razvoj govornih virov in tehnologij za slovenščino (J7-4642) in raziskovalnega programa Slovenski jezik v sinhronem in diahronem razvoju (P6-0038), ki ju financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije.

V slovenskih govornih korpusih je uveljavljena praksa zapisovanja tako v standardnem ortografskem kot v razširjenem ortografskem zapisu. Prvi način je poimenovan standardizirani, drugi pogovorni zapis. To pomeni, da je govor zapisan dvakrat. Ta praksa je bila vzpostavljena s korpusom Gos (Verdonik et al. 2013) ter delno prilagojena in posodobljena ob izdelavi govorne baze in korpusa Artur² (Verdonik et al. 2023b). Kot je razvidno iz poglavja 2, najdemo podobno prakso tudi v drugih jezikih. Potem ko je bilo na ta način v korpusu Artur zapisanih več kot 300 ur javnega, nejavnega in parlamentarnega govora, je smiselno vprašanje, ali je takšen dvotirni sistem zapisovanja govora potreben tudi vnaprej: kaj so njegove prednosti in kaj slabosti glede na dodaten zahtevani trud ob že tako ali tako obsežnem delu zapisovanja govora in označevanja posnetkov?

2 Zapisovanje govora v govornih korpusih

Temeljni problem zapisovanja govora za namen izdelave govornega korpusa je izredna variabilnost in inovativnost oblik in besed v govorjeni rabi, ki bistveno presega standardni slovarski nabor besed in njihovih oblik, zato zanje ne obstaja noben uveljavljen način zapisovanja in se je o njem treba šele dogovoriti. V pregled praks smo zajeli nekaj najbolj vplivnih in z vidika zapisovanja govora primerljivih korpusov za različne evropske jezike s posebnim poudarkom na slovanskih jezikih.

2.1 Korpusa Spoken BNC2014 in FOLK

V govorni komponenti enega od bolj vplivnih govornih korpusov, British National Corpus, oz. v njenem najnovjšem dodatku Spoken BNC2014 (Love et al. 2017), avtorji navajajo izbiro med ortografskim in dodatnim fonetičnim zapisom, vendar slednjega zaradi zahtevnosti v Spoken BNC2014 ne vključijo. Toda njihov ortografski zapis vsebuje nabor sprejemljivih oblik za zapisovanje dialektalnih in nestandardnih besed, torej je standardni ortografski zapis razširjen z dodatnimi oblikami. Fonetične zapise običajno najdemo v dialektalnih korpusih, v splošnih govornih korpusih pa redko, in če že, v zelo omejenem obsegu. Pogosto pa se išče kompromis, kako ob čim manjšem vložku vseeno zadovoljivo opisati posebnosti izgovorjave, kot bomo videli v nadaljevanju tega poglavja.

² Za razlikovanje med govorno bazo in korpusom gl. Campbell 2005: 114–115.

Najbolj neposredno primerljiv način dvotirnega zapisovanja govora, kot se je uveljavil v slovenskih govornih korpusih, najdemo v nemškem prostoru, med drugim v govornem korpusu FOLK (Schmidt 2016). Ta ima začetke v letu 2009 in je zastavljen kot dolgoročni projekt sistematičnega zbiranja raznolikih govornih interakcij med govorcami v Nemčiji. Po številu uporabnikov je eden najbolj uporabljenih govornih korpusov nemščine. Do leta 2019 je obsegal 1,6 mio. pojavnic, letno pa se poveča za okrog 300.000 pojavnic. Sistem zapisovanja in označevanja v korpusu FOLK temelji na smernicah GAT (Selting et al. 2009), ki veljajo po navedbah Schmidta (2016) za enega najbolj uveljavljenih sistemov zapisovanja govora v konverzacijski analizi na Nemškem. Skladno s temi smernicami se uporablja modificiran zapis, poimenovan kot »literarni zapis« (angl. *literary transcription*) ali »očesni dialekt« (angl. *eye dialect*). Besede, ki v izreki odstopajo od standardne, imajo ustrezno prilagojen zapis, npr. *zwo* kot pogovorna različica števnikar *zwei*. Tak sistem zapisovanja je tako rekoč identičen kot pogovorni zapis v slovenskih govornih korpusih. Z namenom optimiziranja korpusa za korpusno jezikoslovje in računalniške jezikoslovne metode je v korpusu FOLK nato dodan še zapis v standardni ortografiji, torej podobno kot standardizirani zapis v slovenskih govornih korpusih.

2.2 Korpusi s-hovor-6.0, ORTOFON in HrAL

Primerljivi načini zapisovanja so prisotni tudi v slovanskem prostoru, kjer najdemo večje primerljive govorne korpuse za češki in slovaški jezik. Slovaški govorni korpus v različici s-hovor-6.0 obsega 6,6 mio. pojavnic in vključuje vsakdanje pogovore v najrazličnejših govornih situacijah. Zapis posnetkov je narejen v ortografskem zapisu, ki sledi pravilom standardne slovaške ortografije, s tem da v nekaterih primerih sledijo standardni slovaški izgovorjavi v nasprotju s predpisano uradno – gre za izgovorjavo posameznih glasov, npr. palataliziranega l (Garabík, Rusko 2007: 233). Poleg standardnega zapisa je dodan še delno fonemski zapis oz. zapis izgovorjave, ki pa je narejen s slovaško ortografsko abecedo, ne s fonemsko abecedo – enako kot pogovorni zapis v slovenskih govornih korpusih. Kot navaja Garabík (2023), tak način fonemskega zapisa bistveno pohitri in olajša zapisovanje govora.

Češki korpus ORTOFON (Komrsková et al. 2017) vsebuje spontane pogovore v vsakdanjih situacijah med ljudmi, ki se med seboj poznajo. Vključuje gradivo iz obdobja 2012 do 2017 in obsega 1,2 mio. pojavnic. Tudi v tem korpusu je zapis

govora dvotirni. Ortografski zapis se kljub temu, da je ortografski, v nekaterih elementih razlikuje od zapisa v standardnem pisnem jeziku. Tako na primer vključuje dialektalne značilnosti, kot so različice končnic za vse vrste sklanjatev in spregatev, regionalne različice vokalnih sprememb ali sklanjatev ipd. Seznam vseh izjem skrbno beležijo. Ortografski zapis pa ne označuje na primer različnih dolžin glasov, reduciranih oblik ali soglasniških premen. Po ortografskem zapisu je dodan še poenostavljen fonetični zapis, ki je narejen z ortografsko abecedo, razširjeno z manjšim naborom posebnih simbolov, in ne s fonetično abecedo. Fonetični zapis med drugim beleži različne asimilacijske procese, izpuste v izgovorjavi ipd., nima pa na primer označenega naglašenege zloga.

Neke vrste dvotirni način zapisovanja oz. razširjeni ortografski zapis najdemo tudi v hrvaškem korpusu govorenega jezika odraslih HrAL (Kuvač Kraljević, Hržica 2016). Ta korpus vsebuje spontane vsakdanje pogovore v obsegu 250.000 pojavnic in je bil posnet v letih od 2012 do 2016. Zapisan in označen je skladno z načeli zapisovanja in označevanja govora v konverzijski analizi ter je izdan v seriji govornih virov TalkBank,³ ki sledijo navodilom CHAT (MacWhinney 2000). Ta glede zapisovanja dialektalnih različic v izgovorjavi omogočajo, da se zapiše dialektalni izgovor, v oglatih oklepajih pa se doda standardni zapis. Primer 1 prikazuje vzorec iz korpusa HrAL.

Primer 1: Zapis govora v hrvaškem korpusu HrAL

pa ne zato što se ja njoj nisan [: nisam] niti upucavo [: upucavao] nego san ja nju namješto [: namještao]

Navodila CHAT sicer poleg načina zapisovanja, kot je bil izbran v HrAL, omogočajo še, da izberemo ali fonemski zapis ali pa dialektalne različice ignoriramo, pri čemer je treba to informacijo vključiti v korpus.

2.3 Korpus C-ORAL-ROM

V romanskih jezikih je znan korpus C-ORAL-ROM. Nastajal je od 1999 naprej (Cresti, Moneglia 2005) in vsebuje govorene vsakdanje pogovore v štirih romanskih jezikih, francoskem, italijanskem, španskem in portugalskem, v skupnem obsegu 121 ur oz. skupno 300.000 besed po jeziku. C-ORAL-ROM je transkribiran podobno

³ <https://talkbank.org>

kot hrvaški korpus HrAL, skladno s formatom CHAT. Francoski del korpusa C-ORAL-ROM poleg ortografskega zapisa vključuje tudi fonetični zapis v abecedi SAMPA v primerih, ko izgovorjave odstopajo od predvidene oz. so kakorkoli posebne (Cresti, Moneglia 2005: 114). Tudi španski del korpusa C-ORAL-ROM pri nestandardnih besedah dodatno popisuje izgovorjavo, vendar z ortografsko abecedo (Cresti, Moneglia 2005: 142).

3 Zapisovanje govora v korpusu Artur

V korpusnem delu govorne baze Artur je govor zapisan dvotirno, s pogovornim in standardiziranim zapisom hkrati, skladno z načinom, vzpostavljenim s prvo izdajo govornega korpusa Gos (Verdonik, Zwitter Vitez 2020). Ker pa je korpus Artur prvotno namenjen razvoju avtomatskega razpoznavanja govora za slovenski jezik, so bile vključene nekatere zaželenne prilagoditve, nekaj sprememb zlasti v standardiziranem zapisu pa je izhajalo iz izkušenj in analiz zapisov v prvi različici korpusa Gos.

3.1 Priprava pogovornega zapisa

Pogovorni zapis je prva raven zapisovanja govora v slovenskih govornih korpusih. Primer 2 prikazuje izjavo iz korpusa Artur, zapisano v pogovornem in v standardiziranem zapisu.

Primer 2: Pogovorni in standardizirani zapis v korpusu Artur

Pogovorni zapis	<i>Ja no, s temu mojmu p@rjatlom Jušom midva tud d@rgač velike športava, tud tenis igrava.</i>
Standardni zapis	<i>Ja no, s tem mojim prijateljem Jušem midva tudi drugače veliko športava, tudi tenis igrava.</i>

Cilj pogovornega zapisa je, da »čim bolj olajša avtomatsko fonemsko-grafemsko pretvorbo in silabizacijo. V kombinaciji s standardiziranim zapisom je zasnovan tako, da omogoča čim boljše ekstrakcijo novih kandidatov za oblikoslovno-fonetični leksikon, ki tako ali drugače odstopajo od normirane rabe.« (Verdonik, Bizjak 2023: 28) Osrednje načelo je, da »/g/ovor zapisujemo v veljavnem slovenskem črkopisu, z dodatnim posebnim znakom za polglasnik (@). Upoštevamo veljavne strategije predstavljanja posameznih glasov z določenimi črkami. Upošteva je omejitve, ki izhajajo predvsem iz omejenega nabora črk, pri tem kolikor mogoče natančno

predstavimo glasovno podobo govora.« (Verdonik, Bizjak 2023: 28). Iz pogovornega zapisa so tako še naprej vidne redukcije glasov.

Spremembe v pogovornem zapisu v primerjavi s prvo različico korpusa Gos so predvsem tri. Prvič, uvedena je uporaba ločil in velikih začetnic skladno s pravopisno normo. Drugič, delno je spremenjen način segmentiranja govora, ki zahteva, »da segmenti niso predolge enote in da je tam, kjer naredimo mejo segmenta, dovolj premora v govoru, da lahko določimo mejo segmenta, ne da odrežemo del predhodnega ali del naslednjega fonema. Glavni vodili za meje med segmenti sta zato: (a) kratek premor v govoru in (b) dolžina segmenta, ki naj ne bo predolga, tj. več kot okoli 10 sekund,« (Verdonik, Bizjak, 2023: 9) zaradi česar segmenti ne ustrezajo vedno pojmu izjave. Tretjič, uveden je manjši nabor dodatnih znakov za foneme, po pogostosti izstopa znak @ za polglasnik, pojavljata pa se še \$g za zveneči h in \$r za mehkonebni r (Trojar, Bizjak 2023: 45).

Premen po zvenečnosti (v nasprotju na primer s češkim sistemom) ne zapisujemo: razlog je, da to zahteva visoko stopnjo pozornosti in se v zapisu, ki ga sicer pripravljajo najeti zunanji izvajalci, pojavlja veliko nedoslednosti. V parlamentarnem delu korpusa Artur, za katerega je bil sistem zapisovanja govora vzpostavljen ločeno že pred uvedbo skupnih smernic za korpus Artur, so premene po zvenečnosti ostale v zapisu, vendar se je potrdilo, da je pri tem doslednost slaba. Po drugi strani so premene po zvenečnosti za slovenščino zelo predvidljive in jih je mogoče precej natančno določiti po pravilih. Podobno velja za zapisovanje dvoustničnega *v*, kjer je navodilo skupnih smernic za korpus Artur podobno, kot je veljalo v prvi različici korpusa Gos:

Zvočnik dvoustnični v (ni nosilec zloga) zapisujemo s črko 'v', če se pojavi v besednih oblikah, ki niso knjižne (prov, nav, navm, odpravit, davn, gledave, pov@n ...). Posebej smo pozorni na primere: laufati, šlavf, genav, mav (malo), šov (šel), dov (dol), prov (prav), dav (da bo), nov (ne bo), tudi medmet av. Če dvoustnični v nastopa v besedni obliki, ki je knjižna in tudi izgovorjena skladno s standardom, ohranimo knjižni zapis (bil, gledal, siv). Če je glas u samoglasniški, tj. je nosilec zloga, ga pišemo s črko 'u' (pršu, vidu, u tem delu...). Tudi predlog v, izgovorjen kot samoglasniški u, pišemo kot u. (Verdonik, Bizjak 2023: 32)

Tudi tukaj se je predhodno v parlamentarnem delu korpusa Artur vzpostavila praksa, da so dvoustnični *v* zapisovali kot 'u', pri čemer ga zapisovalci zlasti v zbornih pozicijah niso vedno prepoznali in dosledno zapisali.

Seznam neverbalnih in polverbalnih izrazov (npr. *eee, hm, uh, ššš*) se je skozi korpus Artur dopolnjeval in je od prve različice korpusa Gos na podlagi analiz prvotnih zapisov dobil tudi vzpostavljen sistem načel, ki med drugim določajo, da izraze raje zapisujemo z eno kot z več besedami, da ne uporabljamo odvečnih različic za zelo podoben izraz, da jih prednostno zapisujemo s tremi črkami in tako dosežemo razlikovalni zapis ipd. (Verdonik, Bizjak 2023).

Pogovorni zapis v korpusu Artur so izvajali najeti zunanji izvajalci in študenti, nato je sledil pregled naključnih segmentov s strani koordinatorja zapisovanja in označevanja govora. Ob tem smo beležili, kje v pogovornem zapisu se pojavlja največ napak. Pogosto je bila prisotna neustrezna segmentacija govora na osnovne enote, zlasti v primeru zelo strnjene govora brez premorov. Zapisovalci so občasno pozabljali ustrezno označevati dolge premore ali zvočna ozadja in zvočne dogodke. Pri hkratnem govoru se je dogajalo, da je bilo napačno označeno menjavanje govorcev. Zelo nezaželena napaka je bila izpust besed, ki so bile izgovorjene, zapisane pa ne, kar je izredno zahtevno odkriti. Občasno so imeli zapisovalci težavo razumeti, kaj je bilo izgovorjeno, zlasti če posnetek oz. govor ni bil povsem razločen. Nedoslednosti so se dogajale pri zapisovanju polglasnika, dvoustničnega *v*, neverbalnih in polverbalnih izrazov ter tujih besed, kjer so zapisovalci uporabljali črki *q* in *y*, ki nista del nabora znakov za pogovorni zapis. Precej korekcij je bilo potrebnih tudi pri ločilih in velikih začetnicah, kar je bilo nazadnje izvedeno s prenosom ločil in velikih začetnic iz standardiziranega v pogovorni zapis, da smo dosegli usklajenost med obema zapisoma.

3.2 Priprava standardiziranega zapisa

Standardizirani zapis kot druga raven zapisa govora služi predvsem uspešnejšemu avtomatskemu označevanju besedil ter podpora pri slovarski in slovnični analizi govorne rabe jezika, tako da razkriva značilno govorno besedje in slovnične vzorce. Kot je navedeno v dokumentaciji korpusa Artur, pa so nekatera določila dodana tudi z namenom, »da (1) se podpira avtomatizirana pretvorba pogovornega v standardizirani zapis, (2) omogoča enoznačno ujemanje pojavnih v pogovornem in standardiziranem zapisu, (3) preprečujejo težave s kodiranjem, (4) zagotavlja anonimizacija podatkov o govornikih« (Verdonik et. al. 2023a: 6).

V standardiziranem zapisu so besede, pri katerih ni bistvene razlike v primerjavi s predvideno pravorečno izreko, zapisane standardno. Posebnega znaka za polglasnik ali druge glasove tukaj ni več. Tudi ko so v govoru prisotne bodisi glasovne premene, bodisi površnost, nedoslednost, motnje v govoru ali posebnosti izgovorjave, bodisi lapsusi, so besede zapisane standardno. Izziv pa so oblikoslovne, skladijske in besedne značilnosti pogovornih in narečnih zvrsti, za katere je v korpusu Artur veljalo navodilo:

Kadar prepoznamo oblikoslovne (npr. skladijski vzorci, ne/ določna oblika, pregibanje ipd., npr. fižola, mala namesto majhna, večim), skladijske (besedni red, vezljivost ipd., npr. ena bolj od okuženih občin; nimam se kaj za pritoževati) ali besedne značilnosti (npr. pasoš, orenk, leder) pogovornega/narečnega jezika, (1) ohranimo izvorno obliko (fora) ali (2) določimo krovno standardizirano obliko te pogovorne/narečne besede oz. njene oblike, če hkrati s slovničnimi ali besednimi značilnostmi govorjenega jezika prepoznamo tudi glasovne premene (npr. žribtov -> žribtal). (Verdonik et al. 2023a: 9)

Vendar prepoznavanje tega ni vedno enostavno in enoznačno, saj

se vedno znova odpirajo primeri, za katere ni vnaprej znane najprimernejše standardizirane oblike ali pa se za določene primere pokaže, da niso tako enoznačni, kot se zdi, ko prvič naletimo nanje (npr. poleg besede gučati se naknadno razkrije še varianta gučiti in je treba za nazaj raziskati, katera od teh oblik naj velja kot standardizirana oz. ali naj se vodita dve različni obliki). (Verdonik et al. 2023a: 9)

Zato se je pri standardiziranem zapisu vodil seznam takih težavnih oblik, ki je objavljen kot del dokumentacije baze.

Enako kot pri pogovornem je tudi pri standardiziranem zapisu novost v primerjavi s prvo različico uvedba ločil in velikih začetnic pri izjavah. Razlog za to je bila predvsem potreba avtomatskega razpoznavanja govora po podatkih za učenje orodja za postavljanje ločil v razpoznan govoro (puntuatorja). Vsebinsko pa ocenjujemo, da je v primerjavi s prvo različico korpusa Gos zaznaven premik k večjemu obsegu ohranjanja pogovornih oblik in leksemov ter manjšemu pretvarjanju/prevajanju v vzpostavljene standardne oblike in lekseme. Tak pogost primer je osrednjeslovenski veznik *ke*, ki je bil v prvi različici Gos interpretiran v različne veznike (*ker, ko, ki, kot, kjer, kar, kaj*), v korpusu Artur pa je bil uveden unikaten zapis *ke*, ki omogoča enostavnejši avtomatski dostop do primerov teh rab.

Standardizirani zapis v korpusu Artur je bil v prvem koraku avtomatsko pripravljen s pomočjo prevajalnega modela, učenega na bazi Gos Videolectures (Verdonik et al. 2021), in zatem ročno popravljen. Popravki so se izvajali samo v zapisu govora, segmentacija na osnovne enote, označevanje menjavanja govorcev, popisani podatki o govornikih in posnetkih ter označevanje zvočnih ozadij in dogodkov so se pri standardiziranem zapisu popravljali le izjemoma, če je bila opažena očitna napaka oziroma pri označevanju značilnosti izgovorjave. Standardizacija zapisa govora vključuje zaradi zgoraj omenjenih negotovih primerov, ki jih ne moremo vnaprej predvideti, zahtevne odločitve, za katere je zaželeno, da so med seboj čim bolj skladne. Te odločitve imajo hote ali nehoti vpliv na rezultate analiz korpusnih podatkov, saj lahko z zapisom govora kakšne značilnosti govorjene rabe nehoti zakrijemo. Pri izvajanju standardiziranega zapisa z zunanjimi, laičnimi zapisovalci tega ne moremo zagotoviti, zato je vse standardizirane zapise v korpusu Artur popravljali en sam ustrezno izurjen sodelavec projekta s poglobljenim jezikoslovnim znanjem, ki se je o težavnejših primerih, popisanih v dokumentaciji korpusa (Verdonik et al. 2023a), periodično posvetoval z ožjo skupino sodelavcev. Kljub temu ostaja eden osrednjih problemov standardiziranega zapisa konsistentnost zapisovanja, po eni strani skozi čas in nadgradnje (odločitve so se od prve različice korpusa Gos do Arturja delno prilagodile, kot opisano zgoraj), po drugi strani pa tudi skozi projekt. Usklajevanje odločitev za nazaj ovira predvsem pomanjkanje ustreznega orodja in okolja, v katerem bi se to lahko izvedlo, to bi bil idealno konkordančnik, ki bi omogočal iskanje, poslušanje in hkrati tudi neposredno popravljane zapisov v jezikovnem viru – zadnji korak pa v obstoječih orodjih ni podprt.

3.3 Naknadni popravki pogovornega zapisa

Hkrati s popravljanjem standardiziranega zapisa se žal vedno znova pokaže, da so določeni popravki potrebni tudi v pogovornem zapisu. V korpusu Artur so bili ti popravki večinoma: (1) ločila in velike začetnice, (2) napačni zapisi skupaj ali narazen ali (3) manjkajoča beseda v zapisu. Napake iz točke 1 so bile pogoste in utrujajoče za usklajevanje med obema zapisoma, zato so bile na koncu avtomatsko prenesene iz standardiziranega v pogovorni zapis. Napake iz točk 2 in 3 pa zahtevajo izredno visoko pozornost zapisovalca, saj so v primeru neusklajenega števila besed v izjavi med pogovornim in standardiziranim zapisom težave pri uporabi korpusa in pretvorbi v formate za javno izdajo, odkrivati in popravljati pa jih je zelo težavno.

Časovno učinkovitost ročnega popravljajanja standardiziranega zapisa je ovirala velika količina kratkih datotek v parlamentarnem delu korpusa Artur, saj je pri več tisoč datotekah veliko časa potrebnega samo za rokovanje z datotekami. Paralelna primerjava obeh zapisov, standardiziranega in pogovornega, bi bila bolj učinkovita, če bi bila oba zapisa odprta v enem oknu eden pod drugim. Z uporabljenim orodjem, Transcriber 1.5.1 (Barras et al. 2000), je bilo mogoče odpreti oba zapisa samo v ločenih oknih in ju shranjevati v ločenih datotekah.

4 Kritična analiza zapisovanja govora v korpusu Artur

4.1 Dvotirni način zapisovanja

Dobrushina in Sokur (2022) sta kritična do praks, ko se uporablja dvotirni zapis, češ da je časovno celo bolj zahteven kot fonetični zapis, ker zahteva dve ravni zapisa namesto ene. Časovna zahtevnost zapisovanja in označevanja govora zagotovo zahteva osrednjo pozornost, ko razmišljamo o strategijah za naprej.

Koliko dodatnega časa torej zahteva dvotirni sistem? Ocene po izdelavi korpusa Artur so sledeče: za segmentacijo posnetka, označevanje govorcev in menjavanja govorcev ter ročni zapis povedanega (bodisi v pogovornem bodisi v standardiziranem načinu) potrebujemo okrog 20 ur dela za 1 uro posnetka, pri čemer je treba upoštevati, da je lahko v primeru zelo zahtevnih terenskih posnetkov z veliko hkratnega govora ali zelo narečnim govorom trajanje dela tudi bistveno daljše. Redakcija zapisa in oznak zahteva približno 1 uro dela za 1 uro posnetka, ob predpostavki, da pregledamo samo naključne segmente posnetka, ne celotnega posnetka. Koordiniranje zapisovanja in označevanja govora vključuje organiziranje dela, pripravo navodil, vzpostavitev delotoka, izbor in pripravo orodij in okolij za delo z datotekami na daljavo, iskanje in angažiranje sodelavcev transkriptorjev ter administrativno projektno delo. Običajno zahteva četrtinski do polovični delovni čas za celotno obdobje trajanja, odvisno od intenzivnosti dela.

Izvedba dodatnega nivoja zapisa (v našem primeru standardiziranega) zahteva skupno okvirno 100 ur za razvoj, preverjanje in zaganjanje algoritma za avtomatsko pretvorbo iz pogovornega v standardizirani zapis, pri čemer je prvi pogoj primeren učni korpus, za kar se lahko v prihodnje uporabi korpus Artur ali vsaj del Arturja. Ročno popravljajanje avtomatsko predpripravljenega standardiziranega zapisa zahteva okvirno 4 ure dela za 1 uro posnetka. Končna validacija in usklajevanje obeh ravni zapisa besedo na besedo lahko predstavlja vse do dodatne ure dela za uro posnetka,

lahko pa tudi veliko manj, odvisno od obsega napak. V skupnem seštevku delo nikakor ni podvojeno, ampak bi lahko bila groba ocena okrog 25 % dodatnega dela, da izvedemo dvotirni namesto enotirnega zapisa govora.

Če se odločimo samo za enotirni zapis govora, je najverjetnejša izbira standardizirani zapis, saj omogoča primerno nadaljnjo avtomatsko obravnavo podatkov. Izgubimo torej pogovorni zapis, ki pa je v primerjavi s standardiziranim (1) bolj usklajen z govorjeno rabo in bolj natančno odraža dejansko podobo besed in besednih oblik, (2) omogoča bolj podrobno luščenje za govor značilnih besed in besednih oblik iz korpusa, (3) omogoča bolj natančen avtomatsko pripravljen fonemski zapis podatkov, (4) je za zunanje izvajalce transkriptorje lažje usvojljiv in manj zahteven kot standardizirani zapis.

4.2 Težavnejša vprašanja zapisovanja

Kritični pogled na vzpostavljene smernice poleg vprašanja dvotirnega zapisa odpira vsaj še štiri vprašanja.

Prvo se nanaša na segmentiranje govora na osnovne enote in uvedbo ločil (gl. 3.1). Kot že navedeno, v korpusu Artur označeni segmenti ne ustrezajo vedno temu, kar bi lahko interpretirali kot ena izjava ali osnovna enota govora, ampak se bolj opirajo na premore, torej na specifično prozodično lastnost. Tak način segmentiranja odpira tehnične probleme za višje ravni označevanja, zlasti skladenjsko in pragmatično, zaradi prelomov znotraj osnovne enote po eni strani in prehodov prek mej osnovnih enot po drugi. Uvedba ločil lahko pomaga premoščati te probleme, saj omogoča alternativne osnovne enote označevanja, ki izhajajo iz skladenjskih značilnosti. Z vidika čim širše uporabnosti govornih virov za različne vede je obstoječa rešitev v korpusu Artur vseeno lahko dobra, čeprav zahteva nekaj več naknadnega procesiranja korpusnih zapisov pred nadaljnjim označevanjem. Pomembno pa je, da so ločila dodana s premislekom in strokovno, saj dodatno interpretirajo govor. Način, da se dodajo v standardiziranem zapisu, ki ga izvaja strokovno visoko usposobljena oseba, ter nato avtomatsko prenesejo v pogovorni zapis, se zato zdi dobra praksa.

Drugič, vedno znova se odpira problematika zapisovanja opornih signalov, to so običajno besedice *ja, mhm, aba, aja* ipd., ki jih sogovornik izreče, medtem ko drugi govorec govori. V korpusu Artur je natančnost zapisovanja opornih signalov zmanjšana v primerjavi s prvo različico korpusa Gos. Razlog je velika časovna zahtevnost in dodatno segmentiranje govora za povsem natančen zapis. Za potrebe tehnologij je manjša natančnost zapisovanja opornih signalov sprejemljiva, za določene jezikoslovne, sociolingvistične in druge raziskave pa je lahko slabost. Tudi tukaj vidimo rešitev v zamenjavi orodja za zapisovanje in označevanje govora s takim, ki bo omogočalo večtirn način zapisovanja govora v enem oknu.

Tretjič, uvedba posebnega znaka za polglasnik v pogovornem zapisu je po eni strani res omogočila nekoliko bolj enoznačno morebitno avtomatsko pretvorbo v fonetični zapis, a je praksa pokazala, da zapis polglasnika ni dosleden. Kolikor bolj se pogovorni zapis približuje fonemskemu, toliko več nedoslednosti vključuje in več je nejasnih vmesnih primerov, ko se je težko odločiti, kako interpretirati izgovorjeni glas.

Četrtič, potrebna bi bila analiza zapisovanja mejnih primerov standardiziranega zapisa in razširjena sistematična razlaga, po katerih načelih določamo »oblikoslovne, skladijske ali besedne značilnosti pogovornega/narečnega jezika« (Verdonik et al. 2023a: 7), ki jim ohranimo izvorno obliko, oziroma po kakšnih kriterijih določimo standardizirano obliko, pod katero vodimo tako besedje. Samo za ilustracijo – z dodatno razlago lahko na primer naslovimo primere, kot so: (1) splošnopogovorni ali regionalni pregibni vzorci (npr. *bolan – bolana*) (2) in besedje (npr. *probavati*), (3) poznani jezikovni procesi (npr. maskulinizacija; *mleko – mleček*), (4) prevzete besede (npr. *šoping, dugi*), (5) kratice, kot sta *S. P.* ali *D. O. O.*, ipd.

4.3 Orodje za zapisovanje

Za zapisovanje in označevanje govora smo v slovenskih govornih korpusih do zdaj uporabljali orodje Transcriber 1.5.1. Razlogi za njegov izbor so bili vedno znova v preprostosti uporabe za zunanje uporabnike, zanesljivosti, hitri natančni segmentaciji govora in že vzpostavljenih orodjih za uporabo in pretvorbo izhodnega formata datotek Transcriberja v formate za javno uporabo. Čeprav je Transcriber eno od pogosto uporabljenih orodij za zapisovanje in označevanje govora, pa so

pogosto v uporabi vsaj še Praat,⁴ ELAN⁵ in EXMARaLDA.⁶ Za popravljanje dodatnega nivoja zapisa bi bilo idealno, da bi lahko odprli oba zapisa paralelno v enem oknu, kar omogočajo vsa tri navedena orodja, Transcriber pa ne. Razmislek o najučinkovitejšem orodju bo tako potreben tudi v prihodnje in tudi z upoštevanjem zaledne podpore posameznega orodja.

5 Zaključek

V prispevku smo se osredotočili na vprašanje, ali je dvotirni način zapisovanja govora v govornih korpusih, pri katerem se najprej pripravi pogovorni zapis, nato pa še standardizirani zapis, ki podpira nadaljnje avtomatsko označevanje zapisanega govornjenega besedila, v prihodnje še smiselno ali pa je vložek prevelik v primerjavi s koristmi. Zaključek je, da zlasti za jezikoslovne raziskave dvotirni način zapisovanja še vedno prinaša prednosti, zaradi katerih ga je smiselno nadaljevati. Izkušnje kažejo, da potrebno delo ni podvojeno, ampak povečano morda za četrtno, pri čemer vidimo še nekaj možnosti za večjo učinkovitost, ki bi jo lahko dosegli z novim učenjem algoritma za avtomatsko predpripravo standardiziranega zapisa na podatkih korpusa Artur in z uporabo orodja za zapisovanje in označevanje govora, v katerem bi lahko oba zapisa odprli v enem oknu, enega pod drugim.

Izpostavili smo tudi štiri točke v obstoječih standardih zapisovanja govora, na katere je treba biti v prihodnje še posebej pozoren in se nanašajo na navodila za segmentiranje in s tem povezano rabo ločil, na zapisovanje opornih signalov, natančnost pogovornega zapisa z dodatnimi znaki za foneme ter na nadgradnjo smernic za standardizirani zapis.

Ključna težava ostaja neredno financiranje s pretiranimi viški v (pre)kratkem časovnem obdobju, kar zahteva uporabo manj zaželenih bližnjic za zbiranje gradiv in onemogoča temeljite priprave orodij in avtomatskih algoritmov, s katerimi lahko delo pohitrimo, ter dolgimi vmesnimi obdobji brez kakršnegakoli financiranja, ko delo na govornih korpusih popolnoma zamre.

⁴ <https://www.fon.hum.uva.nl/praat/>

⁵ <https://archive.mpi.nl/ta/elan>

⁶ <https://exmaralda.org/en/>

Literatura

- Claude BARRAS, Edouard GEOFFROIS, Zhibiao WU, Mark LIBERMAN, 2000: Transcriber: Development and use of a tool for assisting speech corpora production. *Speech Communication* 33/1–2, 5–22.
- Nick CAMPBELL, 2005: Getting to the Heart of the Matter: Speech as the Expression of Affect; Rather than Just Text or Language. *Language Resources and Evaluation* 39, 109–118. Dostop 11. 4. 2024 na <https://doi.org/10.1007/s10579-005-2699-y>.
- Emanuela CRESTI, Massimo MONEGLIA, 2005: *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Nina DOBRUSHINA, Elena SOKUR, 2022: Spoken Corpora of Slavic Languages. *Russian Linguistics* 46, 77–93. Dostop 25. 8. 2023 na <https://doi.org/10.1007/s11185-022-09254-9>.
- Radovan GARABÍK, 2023: Slovenský hovoreny korpus. *Infrastruktura za raziskave govora v humanistiki in jezikovnih tehnologijah: Zbornik povzetkov*. Ur. Mira Krajnc Ivič. Maribor: Univerza v Mariboru, Filozofska fakulteta. Dostop 25. 8. 2023 na <https://doi.org/10.18690/um.ff.5.2023>.
- Radovan GARABÍK, Milan RUSKO, 2007: Corpus of Spoken Slovak Language. *Computer Treatment of Slavic and East European Languages*. Zbornik konference Slovko 2007. Ur. J. Levická, R. Garabík. Brno: Tribun. 222–236.
- Dafydd GIBBON, Roger MOORE, Richard WINSKI (ur.), 1997: *Handbook of Standards and Resources for Spoken Language Systems*. Berlin, New York: Walter de Gruyter Publishers. Dostop 25. 8. 2023 na http://www.homes.unibielefeld.de/gibbon/Handbooks/gibbon_handbook_1997/index.html
- Zuzana KOMRSKOVÁ, Marie KOPŘIVOVÁ, David LUKEŠ, Petra POUKAROVÁ, Hana GOLÁNOVÁ, 2017: New Spoken Corpora of Czech: ORTOFON and DIALEKT. *Journal of Linguistics/Jazykovedný časopis* 68/2, 219–228. Dostop 25. 8. 2023 na <https://doi.org/10.1515/jazcas-2017-0031>.
- Jelena KUVAC Kraljević, Gordana HRŽICA, 2016: Croatian Adult Spoken Language Corpus (HrAL). *FLUMINENSLA* 28/2, 87–102.
- Robbie LOVE, Claire DEMBRY, Andrew HARDIE, Vaclav BREZINA, Tony MCENERY, 2017: The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22/3, 319–344. Dostop 25. 8. 2023 na <https://doi.org/10.1075/ijcl.22.3.02lov>
- Brian MACWHINNEY, 2000: *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, New York: Lawrence Erlbaum Associates.
- Thomas SCHMIDT, 2016: Construction and dissemination of a corpus of spoken interaction – tools and workflows in the FOLK project. *Journal for language technology and computational linguistics* 31/1, 127–154.
- Margret SELTING, Peter AUER, Dagmar BARTH-WEINGARTEN, Jörg BERGMANN, Pia BERGMANN, Karin BIRKNER, Elizabeth COUPER-KUHLEN, Arnulf DEPPERMANN, Peter GILLES, Susanne GÜNTNER, Martin HARTUNG, Friederike KERN, Christine MERTZLUFFT, Christian MEYER, Miriam MOREK, Frank OBERZAUCHER, Jörg PETERS, Uta QUASTHOFF, Wilfried SCHÜTTE, Anja STUKENBROCK, Susanne UHMANN, et al., 2009: Gesprächsanalytisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10, 353–402.
- Mitja TROJAR, Andreja BIZJAK, 2023: Transkribiranje govora pri izdelavi govorne baze Artur: od pogovornih k standardiziranim zapisom. *Razvoj slovenščine v digitalnem okolju*, 39–59. Ljubljana: Založba Univerze v Ljubljani. Dostop 10. 4. 2024 na <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/522/852/9445>.
- Darinka VERDONIK, 2023: Zbiranje gradiv za govorne korpuse med Scilo in Karibdo. *Razvoj slovenščine v digitalnem okolju*, 15–37. Ljubljana: Založba Univerze v Ljubljani. Dostop 10. 4. 2024 na <https://ebooks.uni-lj.si/ZalozbaUL/catalog/view/522/852/9447>.

- Darinka VERDONIK, Iztok KOSEM, Ana ZWITTER VITEZ, Simon KREK, Marko STABEJ, 2013: Compilation, transcription and usage of a reference speech corpus: the case of the Slovene corpus GOS. *Language Resources and Evaluation* 47/4, 1031–1048
- Darinka VERDONIK, Ana ZWITTER VITEZ, 2020: *Slovenski govorni korpus Gos*. 1. e-izd. Ljubljana: Znanstvena založba Filozofske fakultete. (Zbirka Sporazumevanje). Dostop 25. 8. 2023 na <https://e-knjige.ff.uni-lj.si/>, <http://www.dlib.si/details/URN:NBN:SI:DOC-X9DAJU5X>.
- Darinka VERDONIK, Tomaž POTOČNIK, Mirjam SEPEŠY MAUČEC, Tomaž ERJAVEC, Simona MAJHENIČ, Andrej ŽGANK, 2021: *Spoken corpus Gos VideoLectures 4.2 (transcription)*. CLARIN.SI Data & Tools. Maribor: Faculty of Electrical Engineering and Computer Science, University of Maribor. Dostop 25. 8. 2023 na <http://hdl.handle.net/11356/1444>.
- Darinka VERDONIK, Andreja BIZJAK, 2023: *Pogovorni zapis in označevanje govora v govorni bazi Artur projekta RSDO*. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Dostop 25. 8. 2023 na <http://hdl.handle.net/11356/1772>.
- Darinka VERDONIK, Andreja BIZJAK, Mitja TROJAR, 2023a: *Standardizirani zapis v govorni bazi Artur projekta RSDO*. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru; Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU. Dostop 25. 8. 2023 na <http://hdl.handle.net/11356/1772>.
- Darinka VERDONIK, Andreja BIZJAK, Mirjam SEPEŠY MAUČEC, Lucija GRIL, Simon DOBRIŠEK, Janez KRIŽAJ, Gregor STRLE, Marko BAJEC, Iztok LEBAR BAJEC, Tjaša ŠOLTES, Jure LOKOVŠEK, Mitja TROJAR, Tomaž ERJAVEC, Mitja BERNJAK, Jerneja ŽGANEC GROS, Peter ČAKŠ, Matevž PUCER, Mitja CVETKO, Jani PAVLIČ, Marijana ZELENIK, Marija IVANOVSKA, Klemen GRM, Jure LONGYKA, Aleš MIHELIČ, Boštjan VESNICER, Naum DRETNIK, 2023b: *ASR database ARTUR 1.0 (transcriptions)*. Maribor: Faculty of Electrical Engineering and Computer Science, University. CLARIN.SI Data & Tools. Dostop 25. 8. 2023 na <https://www.clarin.si/repository/xmlui/handle/11356/1772>.
- Darinka VERDONIK, Andreja BIZJAK, Andrej ŽGANK, Mitja BERNJAK, Špela ANTLOGA, Simona MAJHENIČ, Peter ČAKŠ, Matevž PUCER, Mitja CVETKO, Jani PAVLIČ, Marijana ZELENIK, Simon DOBRIŠEK, Janez KRIŽAJ, Gregor STRLE, Marija IVANOVSKA, Klemen GRM, Marko BAJEC, Iztok LEBAR BAJEC, Tjaša ŠOLTES, Jure LOKOVŠEK, Jure LONGYKA, Mitja TROJAR, Jerneja ŽGANEC GROS, Aleš MIHELIČ, Boštjan VESNICER, Naum DRETNIK, David BORDON, 2023c: *ASR database ARTUR 1.0 (audio)*. CLARIN.SI Data & Tools. Maribor: Faculty of Electrical Engineering and Computer Science, University of Maribor. Dostop 25. 8. 2023 na <https://www.clarin.si/repository/xmlui/handle/11356/1776>.

POSLUŠATI MED VRSTICAMI: PARLAMENTARNI GOVOR IN NJEGOVI ZAPISI

INA POTEKO,¹ MARKO STABEJ,¹ KAJA JOŠT²

¹ Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
ina.poteko@ff.uni-lj.si, marko.stabej@ff.uni-lj.si

² Državni zbor Republike Slovenije, Ljubljana, Slovenija
kaja.jost@dz-rs.si

Parlamentarni govor je kot dostopen in odprt vir dragocen za raziskovanje v različnih strokah. Na podlagi izrečenega v parlamentu nastajajo zapisi, ki so vključeni v parlamentarne korpus, pri čemer pa zapisovalne prakse parlamentarnih služb običajno niso razvidne. V prispevku smo predstavili prakso zapisovanja plenarnih sej v Državnem zboru Republike Slovenije med letoma 2010 in 2022 z namenom dokumentiranja procesa. V slovenskem parlamentu nastajata dve verziji: prvi, hitri zapis, ki je objavljen najpozneje pol ure po izrečenem z namenom hitrega dostopa do informacij, a je v njem tudi več nepravilnosti, in drugi, verificiran zapis, ki nastane ob poslušanju posnetka govora z namenom trajne objave, dostopnosti in zagotavljanja celovitih informacij. Primerjalno predstavljamo tudi zapisovalne prakse nekaterih drugih evropskih parlamentov.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.5](https://doi.org/10.18690/um.ff.4.2024.5)

ISBN
978-961-286-882-6

Ključne besede:
parlamentarni govor,
parlament,
sejni zapisi,
korpus,
transkripcija

DOI
[https://doi.org/
10.18690/um.ff.4.2024.5](https://doi.org/10.18690/um.ff.4.2024.5)

ISBN
978-961-286-882-6

Keywords:
parliamentary speech,
parliament,
verbatim records,
corpus,
transcription

LISTENING BETWEEN THE LINES: THE PARLIAMENTARY SPEECH AND ITS TRANSCRIPTS

INA POTEKO,¹ MARKO STABEJ,¹ KAJA JOŠT²

¹ University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia,
ina.poteko@ff.uni-lj.si, marko.stabej@ff.uni-lj.si

² National Assembly of the Republic of Slovenia, Ljubljana, Slovenia
kaja.jost@dz-rs.si

As it is accessible and open source, parliamentary speech is valuable for research in various disciplines. Speeches delivered in parliament form the basis for the records contained in parliamentary corpora, but the recording practices of parliamentary services are usually not evident. In this paper, we present the practice of recording plenary sessions in the National Assembly of the Republic of Slovenia between 2010 and 2022 to document the process. In the Slovene Parliament, two versions are created: the first, quick transcript, which is published no later than half an hour after the speech to provide quick access to information, but which also contains some inaccuracies, and the second, verified transcript, which is created by listening to a recording of a speech and which is intended to be permanently published, accessible, and provides comprehensive information. We also present the recording practices of some other European parliaments for comparison.



1 Uvod¹

Parlamentarni govor je dragocen in privlačen vir informacij za različne discipline (mdr. zgodovino, politične znanosti, sociologijo in tudi jezikoslovje) in raziskave, še posebej, ker je dostopen tudi v obliki zapisov. A če je naše glavno raziskovalno vprašanje, ali se da zapise parlamentarnega govor(jenj)a uporabiti kot zanesljiv vir za raziskovanje govornega jezika in diskurza, lahko hitro ponudimo odgovor: ne.

Raziskovanje prakse zapisovanja in redigiranja parlamentarnega govora pa je smiselno z več vidikov. Razkriva lahko tudi odnos institucije (in posredno družbe) do specifik govornega diskurza, po drugi strani pa institucionalno (in posredno tudi družbeno) predstavo o nujnih minimalnih standardih pisnega diskurza.

Zanimivejše je raziskovati sam proces jezikovne dejavnosti kot pa njene rezultate. Ti so – pisna besedila, posnetki govornih akcij in interakcij, njihove transkripcije – v materialnem pogledu precej statični. A ukvarjanje s statičnimi, nespremenljivimi viri je udobnejše, varnejše, zanesljivejše. Načeloma tovrstni pristopi omogočajo lažjo preverljivost in ponovljivost ugotovitev kot eno od ključnih postavk sodobnega empiričnega raziskovanja. Raziskovanje procesov je metodološko zahtevnejše in teoretsko problematično, saj so jezikovni procesi izmuzljivi, kompleksni, v marsičem nepredvidljivi in tudi ne zares ponovljivi.

Pri zgornji dilemi smo imeli v mislih predvsem jezikoslovne prakse – a se spoznanje dotika marsikaterega področja humanistike in družboslovja. Zgodovino pisje recimo zapise parlamentarnih sej pogosto obravnava kot primaren raziskovalni vir. A sama narava zapisov nekega govornega dejanja (konec koncev pa tudi prepisov) je takšna, da gre zmeraj – čeprav v zelo različnih merah – za interpretacijo in intervencijo, v zelo različnih oblikah in z zelo različnimi izrecnimi ali prikritimi nameni.

V tem prispevku smo želeli osvetliti sam proces nastajanja sejnih zapisov plenarnih zasedanj² v državnem zboru, v katerega je vključenih več ljudi in tudi služb, ter predstaviti njihovo končno podobo. Vedeti, kakšno gradivo imamo pred sabo, je

¹ Marko Stabej je prispevek napisal v okviru raziskovalnega projekta ARIS Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS).

² Nastajajo tudi zapisi sej delovnih teles, ki pa niso redigirani.

namreč ključno za vsako raziskavo parlamentarnega govora na podlagi njegovih zapisov.

Zapisovalna praksa, predstavljena v prispevku, velja za obdobje med letoma 2010 in 2022. Informacije o sejnih zapisih in zapisih govora smo poleg objavljenih prispevkov in podatkov pridobili iz izkušenj z delom v obeh službah državnega zbora, ki skrbita za zapisovanje izrečenega na sejah, iz internih uredniških navodil (Moličnik et al. 2022) in na podlagi primerjave izrečenega z zapisoma obeh služb.

1.1 Terminologija

Zapis seje je zapis posameznega plenarnega sejnega dne. Je dokument, ki povzema zasedanje, in sicer vključuje protokol seje, določen s poslovnikom, glasovanja, prekinitve, obravnavane točke, poimenovanja govorcev ter zapis govorjene besede in dogajanja v dvorani, ki vpliva na razumevanje izrečenega ter je hkrati v skladu z določili poslovnika. Oblikuje se kot samostojen dokument in se objavi na spletu v sejni mapi.

Zapise posameznih dni redne ali izredne seje se združi v en celovit dokument, ki vsebuje tudi dnevni red seje, kazalo vsebine, morebitni kulturni program, indeks govornikov in legendo. Dokument se izda kot posamezni izvod v obliki serijske e-publikacije **Sejni zapisi Državnega zbora**.

Zapis govora, na katerega se sicer osredotočamo v tem prispevku, je torej samo eden od sestavnih delov zapisa seje.

V predstavljenih zapisovalnih praksah drugih parlamentov (gl. poglavje 4), v izvorniku zapisanih v angleščini, so poimenovanja za zaposlene, ki so vključeni v proces zapisovanja izrečenega na parlamentarnih zasedanjih, in njihove produkte raznovrstna (na primer ang.: *parliamentary reporters, stenographers, typists, verbatim records, verbatim transcripts*). V prispevku smo zaradi poenotenja in za lažjo primerjavo pri predstavitvi uporabili naslednje slovenske ustreznice: končni (sejni) zapis, parlamentarni poročevalec, strojepisec, urednik.

2 Parlamentarni govor

Parlamentarni govor je eden redkih odprtih³ govornih virov v slovenskem prostoru, ki se stalno dopolnjuje, ni podvržen avtorskim pravicam ali varovanju zasebnosti ter posledično omogoča »poglobljeno raziskovanje jezikovnih in družbenih sprememb skozi čas« (Fišer, Pahor de Maiti 2021). Vendar pa se je pri njegovem raziskovanju najprej treba zavedati specifik tega diskurza, parlament je namreč »institucija, za katero veljajo specifična pravila in konvencije, kar vpliva tudi na jezikovno rabo« (Fišer, Pahor de Maiti 2021).

Procedura ni enaka v vseh parlamentih. V slovenskem veljajo relativno stroga pravila: potek plenarnih zasedanj državnega zbora je natančno določen s poslovnikom, kar pomeni, da je predvideno, kdaj posameznik dobi besedo in koliko časa ima na voljo, zato do spontanega govorenega diskurza ne more priti. Naenkrat je lahko vključen samo en mikrofonski, nenapovedana pa sta le postopkovno vprašanje in replika.⁴ Čeprav gre torej za govor, je izrečeno marsikdaj vnaprej pripravljena pisna predloga (npr. pri obrazložitvah glasovanja poslanske skupine), poslanci pa lahko preklaplajo med predložnim besedilom in prostim govorom (Moličnik et al. 2022: 10, 139). Na potek diskurza vplivajo tudi javnost in formalnost govornega položaja ter množičen naslovnik. Poslanci in poslanke se med svojim izvajanjem namreč zavedajo, da jih ne poslušajo le prisotni v dvorani, temveč tudi zainteresirana javnost preko televizijskih prenosov (Jošt, Vranjek Ošlak 2021: 39).

3 Sejni zapisi

Na podlagi izrečenega v parlamentu nastajajo zapisi govornih besed, ki so del sejnih zapisov. Ti so »primarni vir pri razpravah v Državnem zboru, so politični in zgodovinski dokument, ki ponazarja proces nastajanja zakonodaje, ter jezikovni dokument (korpus besed, raziskave živega jezika)« (Moličnik et al. 2022: 6). Prav zapisi parlamentarnega govora so dragocen zgodovinski vir in omogočajo raziskovanje daljših časovnih obdobij (Gašparič, Šorn 2016: 438–440). Danes so dostopni v obliki elektronskega dokumenta, kar raziskovalcem in raziskovalkam olajša delo, predvsem pa omogoča oblikovanje različnih korpusov.

³ Razen nekaterih delov, ki so lahko zaprti za javnost zaradi tajnosti podatkov.

⁴ Postopkovno vprašanje ali predlog se nanaša na uporabo poslovnika, replika pa je odgovor na napačno razumevanje ali interpretacijo izvajanja razpravljavca; replika na repliko je dovoljena le izjemoma; postopkovni predlog in replika lahko trajata največ tri minute (Poslovnik državnega zbora).

3.1 Parlamentarni korpusi in Parlameter

Sejni zapisi slovenskega parlamenta so vključeni v korpusa siParl in ParlaMint ter v orodje Parlameter. Korpus siParl 3.0 vključuje sejne zapise več kot 11 tisoč sej, vanj pa so vključene seje od leta 1990 do 2022 (Pančur et al. 2022). ParlaMint 2.1 je večjezični korpus, ki vsebuje 17 primerljivih korpusov parlamentarnih razprav različnih držav, predvsem med letoma 2015 in 2020 (Erjavec et al. 2021). Parlameter pa je »orodje, ki z analizo glasovanj in transkriptov nastopov lajša spremljanje dela v Državnem zboru« (Parlameter).

3.2 Prosto dostopne informacije o sejnih zapisih

Fišer in Pahor de Maiti (2021) navajata, da »[u]radno objavljeni sejni zapisi niso dobesedne transkripcije parlamentarnih razprav, ampak zapisi v skladu z uveljavljenimi zapisovalnimi praksami institucije«, pri čemer pa omenjata tudi, da »[s]mernice za urejanje sejnih zapisov običajno žal niso javno objavljene, kar lahko bistveno oteži raziskovalno delo«.

Tudi Gašparič in Šorn (2016: 438) zapišeta, da »so dobesedni zapisi parlamentarnih sej vendarle rezultat dela človeka, zapisovalca, ki se je pogosto znašel pred številnimi problemi«. Med drugim se sprašujeta o situacijah, v katerih se govor ni dobro slišal ali pa so bili v ozadju dvorane vzkliki, pa tudi o jezikovnih dilemah, kako na primer zapisati narečne besede ali napačne izraze.

Brez posebnega poizvedovanja pri Službah Državnega zbora je dostopnih le malo podatkov o nastanku zapisov sej.⁵ V Poslovniku državnega zbora je v 94. členu, ki se nanaša na dobesedne zapise seje, opredeljeno sledeče:

- (1) O delu na seji državnega zbora se vodijo dobesedni zapisi (zvočni zapis in njegov prepis).
- (2) Govornik ima najkasneje v treh dneh po svojem nastopu na seji državnega zbora pravico do redakcijskih popravkov v prepisu svojega govora. Popravki ne smejo spremeniti smisla in bistva njegovega izvajanja. V dvomu odloči o dopustnosti popravka predsednik državnega zbora. Popravke prepisa govornik potrdi s podpisom.

⁵ Službe Državnega zbora so pripravljene deliti informacije in interno gradivo, do katerega raziskovalci in raziskovalke lahko dostopajo v njihovih prostorih.

Pomisleki se pojavijo pri dikciji dobesedni zapisi, ki je lahko predvsem za laično javnost zavajajoča, saj to ne pomeni natančnega zapisa od besede do besede, še manj pa od črke do črke, temveč urejen in pregledan zapis govora.

Tudi v drugih parlamentih se je poimenovanje »dobesedni zapis« oz. angleško *verbatim records* izkazalo za problematično, saj so istovrstni zapisi v poročilih Evropskega centra za parlamentarne raziskave in dokumentacijo (ECPRD) večkrat poimenovani različno, včasih celo v okviru istega odgovora (na primer ang.: *verbatim reports, verbatim transcripts, verbatim records*).⁶

Poleg člena v poslovniku informacije o načinu zapisovanja sej najdemo še na prvi strani vsake e-publikacije Sejni zapisi Državnega zbora, in sicer:

Državni zbor vsako sejo zvočno posname. Simultano ob zvočnem zajemanju nastaja besedilo, ki je na spletu dostopno s približno polurnim zamikom. V uredništvu sejnih zapisov se ob poslušanju zvočnega posnetka preveri avtentičnost zapisanega, besedilo pa se uredi v skladu s strokovnimi merili prenosa govorjene besede v zapisano. Takšno preverjeno in jezikovno urejeno besedilo na spletnem naslovu zamenja prvi zapis.

Besedilo celotne seje se izda tudi v publikaciji Sejni zapisi Državnega zbora. Sejni zapisi vsebuje dnevni red, sprejet na seji Državnega zbora, kazalo, iz katerega je razviden potek seje in v katerem so točke dnevnega reda in govorniki, osrednji del je besedilo seje, zapisano v prvi osebi, na koncu pa je dodan še indeks govornikov.

Iz teh informacij ni mogoče razbrati, do kakšnih sprememb pri zapisovanju izrečenega pride, niti kakšna so »strokovna merila za prenos govorjene besede v zapisano«.

3.3 Raziskave na podlagi zapisov parlamentarnega govora

Za smiseln pristop k raziskavam na podlagi besedil v parlamentarnih korpusih je torej nujno treba vedeti, kakšno sploh je gradivo, vključeno v korpus, dejstvo namreč je, da je dejansko izrečeno v sejnih zapisih vsaj do neke mere urejeno in tudi spremenjeno v postopku redakcije. Čeprav se zdijo nekatere jeziko(slo)vne raziskave smiselne, na primer opazovanje rabe tujejezičnih prvin (Jošt, Vranjek Ošlak 2021) ali paremioloških uvajalnih sredstev (Meterc 2019) v parlamentarnem diskurzu ter

⁶ Poleg tega so v britanskem parlamentu v enem izmed svojih odgovorov posebej poudarili, da so sejni zapisi urejeni tako, da so smiselni in slovnično pravilni, kar ni popolnoma dobesedni zapis.

naslavljanje pri predajanju, pridobivanju in ohranjanju vloge govorca na parlamentarnih razpravah (Modrijan 2007), lahko tudi pri teh podvomimo v rezultate, sploh če so ti kvantitativne narave, saj so določeni izrečeni elementi lahko v končnem besedilu tudi izpuščeni.

Manj primerno je na primer opazovanje prvin spontanega govora (Jošt 2019), predvsem diskurznihih označevalcev, ponavljanj in popravljanj. Tudi avtorica sama je izpostavila, da je gradivo, ki ga je pridobila za obravnavo, sicer zapis operaterskega servisa, ki je vmesna stopnja med izrečenim in končnim zapisom, ki gre skozi redakcijo, »zato sicer izrečeni napačni začetki, samopopravki, nekatera ponavljanja ipd. pogosto sploh niso zapisani« (Jošt 2019: 122).

4 Zapisovalne prakse plenarnih zasedanj nekaterih evropskih parlamentov

Primerjalno predstavljamo kratek pregled zapisovalnih praks govora in nastajanja sejnih zapisov plenarnih zasedanj nekaterih drugih evropskih parlamentov: avstrijskega, nemškega, Združenega kraljestva, francoskega, češkega, nizozemskega, finskega, Evropskega parlamenta.⁷ V dvodomnih parlamentih smo povzeli prakse spodnjih domov.

Podatke o zapisovalnih praksah smo pridobili iz mreže ECPRD. V prispevek so vključeni zadnji razpoložljivi podatki (december 2022).

V Avstriji so t. i. parlamentarni poročevalci v času plenarne seje prisotni v dvorani in poleg glavne vsebine zabeležijo še vse drugo dogajanje. V dvorani so po 10 ali 20 minut naenkrat v intervalih na od tri do pet ur, njihovi zapiski so stenografski. Medtem strojepisci pripravijo prepise desetminutnih digitalnih zvočnih posnetkov zasedanja. Poročevalci v času, ko niso prisotni na seji, urejajo prepise zvočnih posnetkov, ki so jih pripravili strojepisci. Nastane zapis, ki je že prestal prvo fazo urejanja. Ta je na parlamentarnem intranetu dostopen približno tri ure po govoru v dvorani, na parlamentarni spletni strani pa okrog 15. ure istega dne. Zadnji deli (zasedanja, ki se končajo npr. ob polnoči) so na spletu objavljeni okrog poldneva

⁷ Te parlamente smo v pregled vključili, ker so kulturno-zgodovinsko ali jezikovno povezani s slovensko zapisovalno prakso, se izvajajo v državah, ki so bodisi pomembneje povezane z začetki Evropske unije bodisi imajo v njej posebno vlogo, ali so vključeni v posamezne znanstvene raziskave. Prakso britanskega parlamenta predstavljamo, ker gre za parlament države z najstarejšo parlamentarno demokracijo, prakso Evropskega parlamenta pa, ker gre za prakso nadnacionalnega predstavniškega telesa.

naslednji dan. Sledijo revizija, avtorizacija ter po plenarnem zasedanju v dnevih, ko ni nobene seje, tretji, končni del v procesu urejanja, v katerem se v zapis vnesejo uredniške korekcije ter popravki, ki so jih zahtevali govorci, ki jim je bilo besedilo poslano v avtorizacijo. Končno urejanje in priprava sejnega zapisa (v enem celovitem dokumentu) se običajno opravita v tednih po posamezni plenarni seji. Rok za objavo končnega sejnega zapisa ni uradno določen, splošno pa je razumljeno, da se mora začeti objavljati čim prej na dan plenarnega zasedanja.

Tudi v Nemčiji zapisi nastajajo na podlagi stenografskih zapiskov, s katerimi se zajame celoten potek seje (vključno z vzkliki, aplavzi, nasprotovanji in drugimi dogajaji), v kombinaciji z digitalnim zvočnim zapisom. Zapis plenarnega zasedanja je ne glede na trajanje seje na voljo naslednji delovni dan. Parlamentarni poročevalci spremljajo in zapisujejo potek plenarnih zasedanj po pet ali 10 minut naenkrat. V roku 50 ali 100 minut pripravijo osnutek zapisa. Končen elektronski sejni zapis je objavljen naslednje jutro, tiskana različica pa naslednji dan opoldne. Dopoldanski deli sej so v elektronski obliki objavljeni približno pet ur po govoru.

V Združenem kraljestvu zapisi nastajajo na podlagi prepisov zvočnih posnetkov zasedanja. Urejeni so tako, da so smiselni in slovnično pravilni. Revizija zapisanega poteka na dveh ravneh, najprej jo opravijo parlamentarni poročevalci, nato še uredniki. Objavljen je samo končni sejni zapis, in sicer v treh urah po izrečenem na seji, natisnjen mora biti do 6. ure zjutraj naslednjega dne. Rok za objavo zapisov zasedanj v »Westminster Hall« je pet ur po govoru.

V Franciji imajo sejni zapisi dandanes obliko celovitega dokumenta, pred leti so se objavljala tudi krajša poročila (zbirna poročila). Čeprav imajo parlamentarni poročevalci dostop do zvočnega in slikovnega zapisa zasedanja, so v času seje prisotni v dvorani, da opazijo elemente, ki jih je na posnetkih težko zaznati (na primer neformalizirane odzive poslancev). Na javni seji si poročevalec zabeleži približno 15 minut sejnega dogajanja naenkrat. Odgovorna oseba, ki je bila na seji prisotna več ur, poskrbi za urejanje zapisanega besedila, da se zagotovi skladnost med posameznimi zapisi. Digitalizirana različica končnega sejnega zapisa je objavljena na spletu nekaj ur po koncu seje oz. po urejanju.

Na Finskem se parlamentarna razprava najprej digitalno posname z zvočnim snemalnim sistemom, ki je izdelan po meri pisarne, ki skrbi za sejne zapise, nato na podlagi zvočnega posnetka strojepisci zapišejo povedano. Parlamentarni poročevalci ob poslušanju posnetka uredijo besedilo, ki so ga pripravili strojepisci. Po urejanju je zapis seje objavljen na internetu. Poleg zapisa v finščini je izdelan tudi povzetek v švedščini, ki vsebuje podatke o obravnavanih temah in odločitvah na posamezni seji. Zapis plenarnega zasedanja se postopoma objavlja med zasedanjem. Zapis, ki je prestal prvo fazo urejanja, je objavljen v roku ene do dveh ur po koncu seje. Druga faza urejanja, v kateri se pred ponovno objavo zapis v celoti uskladi, sledi pozneje, v naslednjih nekaj dneh ali mesecih, odvisno od razpoložljivih virov in drugih obveznosti zaposlenih. Uradnega roka za objavo končnega sejnega zapisa ni. V skladu z zakonodajo je treba alternativna besedila objaviti v 14 dneh po objavi video ali zvočnega posnetka na spletni strani javnega organa.

V češkem parlamentu se za zapisovanje povedanega na zasedanju uporablja stenografija, hkrati pa se uporabljajo tudi zvočni posnetki. Parlamentarni poročevalci so v dvorani, kjer poteka seja, prisotni po 10 minut naenkrat. Po 10 minutah zapisovanja v dvorani v pisarni zabeleženo prepíšejo na računalnik in oblikujejo celovit zapis. Poročevalec zapiše izrečeno in posluša digitalni zvočni posnetek, da se prepriča, da nič ne manjka, nato besedilo še enkrat prebere, da popravi pravopisne in slovnične napake. Zatem se vrne v dvorano in postopek ponavlja do konca seje. Tako pripravljen zapis se zagotovi čim prej, javnosti je na spletni strani z opozorilom, da ni bil lektoriran in ni bil avtoriziran, dostopen v času od ene ure do ure in pol po izgovorjenem na seji. Ko se del seje konča (običajno po tedenskem zasedanju), urednik zapise pošlje v avtorizacijo govorcem, ki so govorili na dotični seji. Govorci imajo pravico predlagati odpravo napak ali pomanjkljivosti v 10 dneh. Urednik presodi, ali so predlogi ustrezni, in objavi končni sejni zapis, običajno v nekaj tednih po koncu seje. Za objavo končnega sejnega zapisa ni obveznega roka, ki bi bil določen s poslovníkom.

Na Nizozemskem parlamentarni poročevalci izrečeno na sejah zapišejo tako, da poslušajo zvočne posnetke. Poročevalci izmenično opravljajo vlogo zapisovalcev in recenzentov. Nekateri izmed njih tudi pripravljajo ter objavljajo sejne zapise. Zapis, ki nastane po prvi fazi urejanja, je objavljen dve uri po končani točki dnevnega reda in dve uri po zaključku zasedanja. Pred objavo končnega sejnega zapisa poteka revizija na dveh ravneh, na ravni jezika in na ravni procedure (potek seje). Uradnega roka za objavo končnega sejnega zapisa ni.

V Evropskem parlamentu sta za plenarno zasedanje na voljo dve vrsti zapisov, zapisnik (ang. the Minutes) in sejni zapis. Zapisnik v obliki kratkega povzetka razprav, glasovanj, izjav, sporočil in obvestil predsedujočega itd., napisan v francoščini, nastane med zasedanjem. Preveden je v vse uradne jezike in je na voljo na spletni strani Parlamenta okrog poldneva dan po vsaki seji. Sejni zapisi (ang. the Verbatim) predstavljajo prepis govorov na plenarnem zasedanju v izvirnem jeziku, skupaj s pisnimi prispevki, ki jih lahko predložijo poslanci (pisne obrazložitve glasovanja in pisne izjave). Za zapis izrečenega se za 22 od 24 jezikov uporablja razpoznavnik. Samodejni prepisi se nato urejajo med poslušanjem zvočnega posnetka, da se popravijo morebitne napake in upoštevajo pravila parlamentarnega zapisa. Za preostala dva jezika se izvede ročno prepisovanje zvočnih posnetkov. Revizija poteka v štirih fazah. Najprej se pregledajo in uredijo zapisi, ki se v času od 30 minut do nekaj ur pogovoru na seji objavijo na spletni strani Parlamenta. V naslednjem tednu se dodajo pisni prispevki članov parlamenta (pisne izjave in pisne obrazložitve glasovanja), sledi končna revizija oz. urejanje. Na koncu se besedilo še dodatno uskladi in lektorira pred objavo v uradnem listu. Celoten tako pripravljen večjezični dokument je objavljen konec naslednjega dne. Rok za objavo končnega sejnega zapisa plenarnega zasedanja je 10 dni (rok za oddajo pisnih prispevkov članov).

Pregled zapisovalnih praks plenarnih parlamentarnih zasedanj je pokazal, da so pomembna podlaga za nastanek sejnih zapisov digitalni zvočni posnetki zasedanj, saj se uporabljajo tudi, kjer so redaktorji oz. stenografi prisotni na samih sejah (Nemčija, Avstrija, Francija, Češka). V večini med seboj primerjanih zapisovalnih praks prva objava sejnih zapisov ni končna (razen v Združenem kraljestvu in Franciji). Roki za objavo končnih sejnih zapisov večinoma niso uradno določeni, čas objave je odvisen od drugih nalog in števila zaposlenih, ki skrbijo za sejne zapise, prvi, začasni zapisi pa so javnosti dostopni najpozneje v nekaj urah po zasedanju.

4.1 Uporaba razpoznavnikov govora pri nastajanju sejnih zapisov

V treh od osmih primerjanih parlamentarnih praks v procesu nastajanja sejnih zapisov uporabljajo tudi razpoznavnike govora, in sicer v Nemčiji, na Finskem in v Evropskem parlamentu. Programe za razpoznavanje in avtomatično zapisovanje govora so testirali tudi v Avstriji, Združenem kraljestvu ter na Nizozemskem. V Združenem kraljestvu in na Nizozemskem so se po testnem obdobju odločili, da avtomatičnega zapisovanja za sejne zapise ne bodo uporabljali, testirani programi

namreč še ne dosegajo dovolj visoke kakovosti za vsakodnevno uporabo. V Avstriji prehod na avtomatično zapisovanje načrtujejo takoj, ko bo program dovolj razvit za uporabo v parlamentarni praksi.

Tudi v parlamentih, v katerih je v proces nastajanja sejnih zapisov vključeno avtomatično zapisovanje govora, ocenjujejo, da avtomatično zapisano besedilo ni primerno za objavo v surovi, neurejeni obliki. Pred prvo objavo ga je treba urediti tako, da doseže zahtevane standarde, in sicer pravopisno, slovnično, oblikovno, z vidika razumljivosti ipd. Avtomatični zapis je torej osnova za nadaljnje delo, javno pa je objavljen zapis, ki je že šel skozi prvo fazo urejanja. V britanskem parlamentu so po testnem obdobju poudarili, da za pripravo razumljivega zapisa prepis, ki je stoddosten posnetek govora, običajno ni uporaben, zlasti kadar govorec ni izurjen, je zmeden, ne dokončuje stavkov ali so v govoru pogosti napačni začetki.

5 Zapisovalna praksa v slovenskem parlamentu

Zapisovalna praksa, kot jo službe državnega zbora poznajo danes, ima svoje korenine v 60. letih 20. stoletja, ko je delo stenografov zamenjal magnetofon. Seje se je začelo snemati na magnetofonski trak, strojepisci in strojepiske pa so posnete seje zapisovali (Gašparič, Šorn 2016: 449). Danes za celoten proces zapisovanja v glavnem skrbita dva oddelka, Oddelek operaterski servis in Dokumentacijsko-knjižnični oddelek.

Za »naloge pisanja, urejanja in objave dobesednih [prvih, op. p.] zapisov sej Državnega zbora in njegovih delovnih teles ter drugih dogodkov« (Službe Državnega zbora) skrbi Oddelek operaterski servis, ki spada pod Informacijski sektor služb Državnega zbora. Dokumentacijsko-knjižnični oddelek, ki je del Raziskovalno-dokumentacijskega sektorja, pa med drugim »opravlja redakcijo in ureja zapise sej Državnega zbora, lektorira zahtevnejša besedila Državnega zbora ter svetuje na jezikovnem področju« (Službe Državnega zbora). V celoten proces so vključene tudi druge službe Državnega zbora, saj ne gre le za zapis in redakcijo. Poleg omenjenih služb sodelujejo še snemalnica, informacijska podpora, službe sektorja za pripravo sej, tiskarna in arhiv (Moličnik et al. 2022: 182).

Od vključno VI. mandata dalje (od leta 2010) sejni zapisi in e-publikacije nastajajo povsem brezpapirno in so sproti objavljeni na spletni strani Državnega zbora (Moličnik et al. 2022: 6–7). Javnosti so torej e-publikacije sejnih zapisov Državnega zbora dostopne od V. mandata dalje (15. 10. 2008 dalje), Državni zbor pa je skupaj z Inštitutom za novejšo zgodovino poskrbel tudi za projekt digitalizacije zakonodajnega gradiva starejših mandatov in pretvorbe sejnih zapisov v Word. Na spletnem portalu Državnega zbora so torej zapisi celotnih sej od I. mandata, tj. od 23. 12. 1992, na portalu Sistory pa so dostopni tudi zapisi pred osamosvojitvijo Slovenije (Moličnik et al. 2022: 7).

5.1 Zapisovalna praksa v slovenskem parlamentu v obdobju 2010–2022

Med sejo se sproti izvaja zajem govornih besed (snemanje). V operaterskem servisu sledi hiter zapis izgovorjenega; prepisuje se petminutne zvočne posnetke, pri čemer je okvirna interna norma prepis 15 minut govora oziroma treh posnetkov v eni uri. Pol ure po izgovorjenem je besedilo javno objavljeno na spletni strani Državnega zbora. Po končani seji je celoten zapis poslan v redakcijo, kjer preverjajo istovetnost izrečenega in zapisanega ter besedilo uredijo v skladu z internimi pravili prenosa govornih besed v zapisano. Pri tem hkrati poslušajo zvočni posnetek ter pregledujejo in urejajo zapis. Njihova norma je 2,55 strani na uro (ena stran v Wordu, pisava Courier new, velikost 12, dvojni razmak). Preverjeno in urejeno besedilo je potem objavljeno na spletu, kjer nadomesti neverificiran zapis. Besedilo se oblikuje, izda in arhivira v e-publikaciji Sejni zapisi Državnega zbora (Moličnik et al. 2022: 9–11).

Tako torej sejni zapis nastaja v več fazah. Hiter zapis operaterskega servisa je začasen in ima ob objavi na spletni strani oznako v pregledu, ko ga nadomesti verificiran zapis, pa na spletu to oznako izbrišejo. Ker na spletni strani ni sledljivosti dokumentov (nov zapis nadomesti prejšnjega), lahko pri zajemu zapisov za korpuse pride do težave, če se vanj ne vstavljajo istovrstni zapisi. V uredništvu sicer hranijo vse, tudi delovne verzije, ki omogočajo sledljivost in transparentnost popravkov (Moličnik et al. 2022: 11).

6 Analiza zapisa govora na plenarnih zasedanjih državnega zbora

Da bi natančno predstavili jezikovno podobo zapisov govora in spremembe, ki nastanejo v procesu, smo se odločili za analizo zapisov in njihovo primerjavo z dejansko izrečenim. Najprej predstavljamo primerjavo med i) izrečenim, ii) hitrim zapisom (prvi zapis) in iii) redigiranim zapisom (končni zapis), v nadaljevanju pa podrobneje tudi delovanje obeh oddelkov, ki skrbita za sejne zapise, in značilnosti posamezne objave.

6.1 Metodologija

Za ponazarjalne primere smo analizirali govore s 16. redne seje Državnega zbora Republike Slovenije, ki je trajala od 2. do 5. marca 2020. Izbrali smo tri govorce in tri govorke, ki so v 8. mandatu državnega zbora govorili največ glede na podatke Parlametra (Robert Polnar, dr. Franc Trček, Miha Kordiš, Anja Bah Žibert, Violeta Tomić in Nataša Sukič). Analizirali smo približno tri minute govora vsakega od poslancev, in sicer iz skupnega dela razprave o posameznih točkah dnevnega reda. V tabelah za ponazorilo navajamo naključno izbrane primere iz opravljene analize. Ponekod, kjer se nam je zdelo smiselno, smo primere izpisali tudi iz internega gradiva službe državnega zbora (Moličnik idr. 2022), v tem primeru so primeri zapisani ležeče.

V tabeli 1 predstavljamo tri zapise: dobesedni zapis izrečenega, hitri zapis, ki nastane v operaterskem servisu, in končni zapis po opravljeni redakciji.

Tabela 1: Primeri zapisov izrečenega na plenarnem zasedanju

Izrečeno (dobesedni prepis)	Prvi, hitri zapis (operaterski servis)	Končni zapis (po redakciji)
Narejene so na po-, na po-, te številke so narejene na podlagi obljub, ki jih je politični razred, samovoljno, brez posvetovanja z javnosti, brez družbenega konsenza, dal vrhu zveze NATO, raznim generalom in drugim predstavnikom vojaške birokracije (in) vojnih krokarjev. In potem, k je te obljube dal, jih je pretočil v doktrinarne dokumente, ki jih	Te številke so narejene na podlagi obljub, ki jih je politični razred, samovoljno, brez posvetovanja z javnostjo, brez družbenega konsenza, dal vrhu Zveze Nato, raznim generalom in drugim predstavnikom vojaške birokracije in vojnih krokarjev in potem, ko je te obljube dal, jih je pretočil v doktrinarne dokumente, ki jih v državi imamo, v teh dokumentih	Te številke so narejene na podlagi obljub, ki jih je politični razred samovoljno, brez posvetovanja z javnostjo, brez družbenega konsenza dal vrhu zveze Nato, raznim generalom in drugim predstavnikom vojaške birokracije in vojnih krokarjev. In potem ko je te obljube dal, jih je pretočil v doktrinarne dokumente, ki jih imamo v državi. V teh dokumentih

Izrečeno (dobesedni prepis)	Prvi, hitri zapis (operaterski servis)	Končni zapis (po redakciji)
v državi imamo. V tej, teh dokumentih piše, da mormo met am sedn tisoč pripadnikov stalne am sestave. In ker te čist administrativno birokratsko določene številke niso izpolnjene, sedaj zaganja paniko. Nimamo dovolj vojakov. Kaj se dogaja s Slovensko vojsko, to ne gre.	piše, da moramo imet 7 tisoč pripadnikov stalne sestave in ker te čisto administrativno birokratsko določene številke niso izpolnjene, sedaj zaganja paniko. »Nimamo dovolj vojakov, kaj se dogaja s Slovensko vojsko, to ne gre...«	piše, da moramo imeti 7 tisoč pripadnikov stalne sestave in ker te čisto administrativno birokratsko določene številke niso izpolnjene, sedaj zaganja paniko. »Nimamo dovolj vojakov. Kaj se dogaja s Slovensko vojsko. To ne gre.«
Zdi se mi, da je tud to mau del, jaz to imenujem sla-slovensko katolištvo, pa naj mi verujoči o- oprostijo, da smo mi v tej družbi obsedeni z nekim iskanjem krivic, ne.	Zdi se mi, da je tudi to mal del, jaz to imenujem slovensko katolištvo, pa naj mi verujoči oprostijo, da smo mi v tej družbi obsedeni z nekim iskanjem krivic.	Zdi se mi, da je tudi to malo del tega, kar jaz imenujem slovensko katolištvo, pa naj mi verujoči oprostijo, da smo mi v tej družbi obsedeni z nekim iskanjem krivic.
Naša naloga je razumeti, zakaj smo v buli, kot pravimo v Mariboru, in znotraj naših pristojnosti iskati rešitve, in tega v zadnjih pet letih in sedem mesecev sem bl mau vidu v tej stavbi.	Naša naloga je razumeti, zakaj smo obuli, kot pravimo v Mariboru in znotraj naših pristojnosti iskati rešitve, in tega v zadnjih pet letih in sedem mesecev sem bolj malo videl v tej stavbi	Naša naloga je razumeti, zakaj smo v buli, kot pravimo v Mariboru, in znotraj naših pristojnosti iskati rešitve. In tega sem v zadnjih petih letih in sedmih mesecih bolj malo videl v tej stavbi.

Vir: lasten

6.2 Zapisi operaterskega servisa

Prvi, hitri zapisi nastanejo v operaterskem servisu še med potekom plenarnega zasedanja in so objavljeni najpozneje pol ure po izgovorjenem. Treba se je zavedati, da »[n]eprerejeni prepis zvočnega zapisa seje nastane pod časovnim pritiskom, zato lahko prozodične lastnosti govora ali nepoznavanje vsebine privedejo do napačnega oziroma fonetičnega prepisa ali je del besedila celo nemogoče prepoznati« (Moličnik et al. 2022: 14).

Zapisovalke in zapisovalci sledijo internim smernicam, ki v glavnem težijo k poknjževanju pri zapisu. Če se jim zdi, da posnetek ni dovolj razumljiv, lahko to v besedilu tudi označijo z / nerazumljivo/, če o zapisanem niso prepričani, pa z (?). Poleg tega petminutne odseke zapisujejo brez konteksta, kar pomeni, da jim manjka vedenje, kaj je bilo povedano pred tem, kar lahko vpliva na (ne)pravilnost zapisa.

Zapis operaterskega servisa, ki nastane, je zasnovan kot začasen, nepregledan in v prvi vrsti namenjen tistim, ki so sejo spremljali in se zavedajo tudi dogajanja v dvorani (poslanci, sekretarji, zaposleni v poslanskih skupinah in novinarji). Hitro

nastalo besedilo jim pomaga pri iskanju ali priklicu vsebine in naj bi bilo uporabno predvsem pri nanašanju na že povedano v daljših razpravah. Ta verzija je najbolj relevantna isti dan, zato je tudi nujno, da je besedilo dostopno kmalu po izrečenem.

V tabeli 2 prikazujemo opažene neustrezne spremembe, do katerih lahko pride v procesu hitrega zapisovanja, skupaj s ponazarjalnimi primeri iz analize in internega gradiva.

Tabela 2: Primeri iz zapisa operaterskega servisa

	Izrečeno (dobesedni zapis)	Prvi, hitri zapis (operaterski servis)
pravopisne napake	podizvajalski	pod izvajalski
hiperkorekcija	da greš služiti pač ta vojaški rok	da greš služiti pač ta vojaški rok
neustrezna poknjžitev	ustvarl	ustvaril (<i>ustvaril</i>)
zatički	kajti dve tretjine jih je glasovalo proti Janši	kajti dve tretjine jih je glasovalo proti Janši
neustrezni izpusti	Padec berlinskega zidu je dokončno pokopal delitev na levo in desnico v njihovih klasičnih oblikah. Niti na evropskem zahodu in niti na evropskem vzhodu leвица in desnica nista več odločilna in določujoča politična pojma.	Padec berlinskega zidu je dokončno pokopal delitev na levo in desnico v njihovih klasičnih oblikah, niti na evropskem vzhodu, leвица in desnica nista več odločilna in določujoča politična pojma.
napačen zapis lastnih imen	<i>Tega se ne bi sramoval niti Lukašenko</i>	<i>Tega se ne bi sramoval niti Luka Šenko</i>
vsebinske spremembe	<i>prepoznavamo nosilno vlogo Državnega zbora</i>	<i>prepoznavamo nasilno vlogo Državnega zbora</i>
	Skratka, tudi argument ogroženi smo, dajmo ljudi izvleči z vojsko, tudi ta argument ne pije vode.	Skratka, tudi argument, »Ogroženi smo, dajmo ljudi izvleči v vojsko«, tudi ta argument ne pije vode. *po redakciji: dajmo ljudi zveči v vojsko
	moral napisat neki konkretni datum, ne	moral napisat neki konkretni zakon
	<i>koliko je blue collar, white collar</i>	<i>koliko je blue colour, white colour</i>
nerazumljivo	ki je s svojo retoriko zvezda stalnica pejorativnega žaljivo sovražnega diskurza	ki je s svojo retoriko zvezda stalnica / nerazumljivo/ žaljivo sovražnega diskurza
dvom	leвица in desnica prestavljata efektivni gesli	leвица in desnica prestavljata efektivni gesli(?)

Vir: lasten

6.3 Zapisi po redakciji

Redakcijo opravlja Dokumentacijsko-knjižnični oddelek. Zaposlene poročajo, da je osnova za njihovo delo posnetek govora, hitri zapis pa jim je predvsem v pomoč kot delovno gradivo, da ni treba ponovno zapisovati celotnega besedila. Prvi zapis torej popravljajo glede na to, kar slišijo.

Redakcijski posegi so potrebni zaradi spremembe vrste prenosnika in naslovnika, »da se kot posledica pretvorbe govorjenega jezika v zapisanega ne izgubijo bistvene informacije« (Moličnik et al. 2022: 138). Glavno vodilo pri prenosu govorjene besede v zapisano je poleg istovetnosti oziroma vsebinske ustreznosti razumljivost (Moličnik et al. 2022: 13).

Z uredniškim posegom se »govor zapiše tako, da se izpusti mašila, nenamerne ponovitve, samopopravke in druge jezikovne elemente, ki bralca zmedejo in mu otežujejo razumevanje bistva izrečenega« (Moličnik et al. 2022: 13). Za razumljivost so nujni slovnični in pravopisni popravki. Vodi jih načelo: govor pripada govorcju, biti pa mora dostopen in razumljiv vsem.

»Redaktor nosi odgovornost za jasnost in razumljivost zapisane govorjene besede, ohraniti mora pomensko avtentičnost, prispevati h kulturni identiteti, ohraniti jezikovne oziroma narečne značilnosti govora, hkrati pa poskrbeti za razumljivost v celotnem slovenskem govornem prostoru« (Moličnik et al. 2022: 6). Pri redakciji je torej glavno vodilo oziroma cilj »z minimalnimi posegi oblikovati celovito, vsebinsko jasno, slogovno ustrezno, razumljivo končno besedilo, ki hkrati ohranja pomen in smisel izrečenega ter značilnosti posameznega govorca« (Moličnik et al. 2022: 138).

Vsa omenjena načela so v praksi med sabo lahko tudi nezdružljiva; predvsem je v besedilu, ki teži k berljivosti, težko ohraniti narečne značilnosti govora, pa tudi vseh drugih dimenzij govora marsikdaj niti ni mogoče prenesti v pisno obliko.

Čeprav gre zapis seje skozi precej dolg proces do končne verzije, se tudi po redakciji še vedno lahko pojavijo nekatere neustreznosti; poleg spregledanih napak, ki so normalne za vsak delovni proces, so lahko posegi v povedano mestoma tudi nepotrebni ali pretirani, kar kažejo nekateri primeri v tabeli 3.

S stališča ohranjanja istovetnosti izrečenega je vsaj nekoliko problematična tudi praksa, da se nekateri izrečeni lapsusi popravijo, ne da bi to bilo označeno v opombi, ter tudi, da se citirano besedilo zapiše tako, kot je v originalu, tudi če ni bilo povedano v tej obliki. Dvom se pojavlja tudi o praksi, kot jo opišejo in utemeljijo v smernicah (Moličnik et al. 2022: 116), da »[i]zrečenega ob izklopljenem mikrofону ne zapisuje[j]o (ker ni možnosti preverjanja ob zvočnem posnetku), ne zapisuje[j]o niti govora, ki ni namerno povedan v mikrofón in namenjen zboru oziroma javnosti«.

Tabela 3: Primeri iz zapisa po redakciji

	Izrečeno (dobesedni zapis)	Prvi, hitri zapis (operaterski servis)
spregledano	Jaz nimam občutka, da smo se iz-iz- iz te problematike naših bolnišnic am kaj naučil.	Jaz nimam občutka, da smo se iz te problematike naših bolnišnic kaj naučil.
neustrezno spremenjen besedni red	Ali je to demokracija, o kateri govorite, in sodelovanje?	Ali je to demokracija in sodelovanje, o kateri govorite?
izpusti	Zdej, po vseh teh vložkih se mi zdi prov, pa da je že skrajni čas, da se začnemo resno pogovarjat o tej temi	Po vseh teh vložkih se mi pa zdi, da je že skrajni čas, da se začnemo resno pogovarjati o tej temi
	Očitno pa upam, da če smo se kaj iz tega naučil, smo se očitno naučil, da bomo pri naslednjih interventnih zakonih moral napisat neki konkretni datum [...]	Če smo pa se kaj iz tega naučil, smo se očitno naučili, da bomo pri naslednjih interventnih zakonih morali napisati nek konkretni datum [...]
	To sem nekako am še želel povedati.	To sem še želel povedati.
popravek lapsusa govorca pri redakciji	<i>Druga intermenca</i>	<i>Drugi intermezzo</i>
	Desus, ki bi ga še vedno volil Karel Erjavec, ne bi nikoli vstopil v takšno koalicijo	Desus, če bi ga še vedno vodil Karel Erjavec, ne bi nikoli vstopil v takšno koalicijo
	<i>Tukaj smo danes [...] poslušali očitke o nepristranskosti</i>	<i>Tukaj smo danes [...] danes poslušali očitke o pristranskosti</i>
popravek citiranega	gospod Breznik v svojem zadnjem vzkliku: »Bog živi Janeza Janšo in bog živi Slovenijo!«	gospod Breznik v svojem zadnjem vzkliku – Bog živi Slovenijo, bog živi mandatarja Janeza Janšo.

Vir: lasten

Zaposlene v redakcijski službi poročajo, da se lapsusi govorca brez opombe popravijo le v primerih, za katere so povsem prepričane, kaj je bilo mišljeno, saj takšna praksa sledi načelu razumljivosti zapisa kot celote. Prav tako, da zapisovanje citiranega, kot je v originalu, zagotavlja večjo sledljivost navajanj, vendar pa se to tudi opušča. Da se izrečeno ob izklopljenem mikrofónu ne zapiše, je posledica poslovnih določil procedure seje.

6.4 Razlike med zapisoma

V procesu zapisovanja govora plenarnih zasedanj zapisi povedanega nastajajo v dveh fazah. Prvi, hitri zapis nastane v operaterskem servisu, čeprav je začasno objavljen na spletni strani, pa zaradi nezanemarljive nenatančnosti pri prepisu ni primeren kot raziskovalno gradivo. Nastane z namenom hitrega dostopa do informacij in med drugim predstavlja delovno gradivo za tiste, ki so sejo spremljali in so seznanjeni z dogajanjem na njej. Glavna funkcija teh prvih zapisov je namreč omogočanje hitrega vpogleda v zakonodajne, nadzorne in volilne postopke v državnem zboru (transparentnost, javnost in dostopnost podatkov) različnim javnostim, nastane pa skoraj simultano z dogajanjem v dvorani. Analiza sicer majhnega vzorca je pokazala, da so najpogostejše spremembe med izrečenim in prvim zapisom poknjževanje besed v zapisu ter izpusti vsega tistega, kar je dojeto kot nepotrebno; mašila, nehoteni govorni zapleti ali napačni začetki.

Verificiran, končni zapis je preverjen in nastane ob poslušanju posnetka govora. Glavna funkcija končnega zapisa je trajna objava oz. dostopnost ter zagotavljanje celovitih informacij o zakonodajnih, nadzornih in volilnih postopkih v državnem zboru. Je t. i. izvorno poročilo o seji, ki ni iztrgano iz konteksta, namenjeno pa je različnim zainteresiranim javnostim. Ker se pri svojem delu v uredništvu nanašajo predvsem na zvočni posnetek, v procesu redakcije pri zapisu govorjenega dopolnjujejo morebitne izpuste, ki so nastali pri prvem prepisu, in popravljajo napačno zapisano (vsebino). Poleg tega prihaja tudi do drugih sprememb, še posebej pogosto do sprememb besednega reda (zaradi sledenja načelu istovetnosti pri prenosu govorjene besede v zapisano) in izpustov (zaradi lažje berljivosti). Ti posegi so, kot ugotavljajo tudi drugi raziskovalci, npr. Podbevšek (1990) in Voutilainen (2017), potrebni zaradi spremembe jezikovnega koda. Hkrati se med redakcijo poskrbi tudi za pravopisno ustreznost zapisanega besedila.

Z vidika opisanih različnih funkcij prvega ter končnega zapisa se zdi nejasno označevanje ob objavi na spletni strani, za kateri zapis gre, ob dejstvu, da se zapisi zajemajo za parlamentarne korpuse, še posebej problematično.

7 Zaključek

Pri pretvarjanju parlamentarnega govornega diskurza v pisni diskurz je zmeraj potreben kompromis med avtentičnostjo povedanega in berljivostjo. Vsak zapis govora je sam po sebi redukcija, saj vseh dimenzij govora ni mogoče prenesti v pisno obliko (tudi če bi zapis dopolnjevale neke vrste opisne didaskalije, ki bi povzemale prozodične in druge informacije govora, kar pa bi oteževalo samo zapisovanje in branje). Presoja o tem, katere sestavine govora je mogoče v zapisu brez škode opustiti, je kljub takim in drugačnim navodilom zadeva trenutne interpretacije zapisovalk in zapisovalcev. Odločitve o tem, kaj od izgovornega je za zapis nerelevantno, pa so tudi posledica splošnega stereotipnega prepričanja, da so t. i. nepolnopomenske oz. propozicijsko prazne sestavine diskurza, kot so npr. diskurzni označevalci, nekaj, kar v povedanem ni zares relevantno – čeprav lahko zelo jasno izraža tako ali drugačno intenco govorečih.

Prakso zapisovanja parlamentarnega govora smo predstavili predvsem z namenom dokumentiranja, da se torej vsaj v glavnih obrisih ve, za kakšno gradivo gre, kako nastaja in do katerih sprememb pride v procesu. Na podlagi tega lahko določimo, za katere raziskovalne namene je gradivo primerno, za katere pa ne, oziroma česa se moramo pri analizi tovrstnega gradiva zavedati, če se želimo izogniti neustreznim in nezanesljivim sklepom in rezultatom. Parlamentarni korpusi, sestavljeni iz zapisov parlamentarnih govorov, torej niso sami po sebi zanesljiv vir za raziskovanje govornega diskurza, temveč jih je treba analizirati z zavedanjem kompleksnosti (in marsikdaj popolne nedokumentiranosti) zapisovalnega procesa.

Tudi v slovenskem parlamentu se napoveduje uvedba avtomatskega zapisovanja govora, kar bo v precejšnji meri spremenilo doslej uveljavljen zapisovalni proces. Slovenski parlament je s svojo precej strogo proceduro, ki v veliki meri preprečuje nepredvidljivo govorno sliko (po poslovniku lahko govori le ena oseba naenkrat, z zelo omejeno možnostjo repliciranja ipd.), za tak projekt precej primerno okolje. Ob tem pa se pojavljajo novi možni zapleti. Kot lahko že zdaj opazimo pri nekaterih komercialnih aplikacijah samodejnega zapisovanja, npr. v urejevalniku Word, je v ta proces že samodejno vključena precejšnja stopnja standardizacije pri pretvorbi govora v zapis.

Samodejno zapisovanje bo sicer gotovo olajšalo in pospešilo proces dokumentacije parlamentarnega govora, a kaj lahko se zgodi, da se bo pri tem avtentičnost (iz)govorjenega ohranjala še težje, kot se pri procesu, opisanem v prispevku. Kaj vse to pomeni v sodobni (in prihodnji) slovenski jezikovni situaciji, kjer postajajo dialektalne oziroma regionalne govorne poteze govork in govorcev relevantni del njihove identitete tudi v javni komunikaciji, ostaja odprto in najbrž tudi precej pereče vprašanje.

Literatura

- ECPRD. *Request 5263*. Dostop 20. 8. 2023. <https://ecprd.secure.europarl.europa.eu/ecprd/private/request-details/762854>
- Tomaž ERJAVEC idr., 2021: *Multilingual comparable corpora of parliamentary debates ParlaMint 2.1*. Dostop 20. 8. 2023. <https://www.clarin.si/repository/xmlui/handle/11356/1432>
- Darja FIŠER, Kristina PAHOR DE MAITI, 2021: *Voices of the Parliament: A Corpus Approach to Parliamentary Discourse Research*, »Prvič, sem političarka in ne politik, drugič pa ...«, *Korpusni pristop k raziskovanju parlamentarnega diskurza*. Inštitut za novejšo zgodovino. Dostop 20. 8. 2023. <https://sidih.github.io/voices/index-sl.html>
- Jure GAŠPARIČ, Mojca ŠORN, 2016: Od žive debate do zapisane besede: dobesedni zapisi parlamentarnih sej kot zgodovinski vir. *Zgodovinski časopis* 70/3–4, 438–454.
- Kaja JOŠT, 2019: Javni govor na plenarnih zasedanjih državnega zbora glede na pojavljanje na funkcijsko različnih delih seje in stopnjo spontanosti. *Slovenski javni govor in jezikovno-kulturna (samo)zavest. Obdobja 38*. Ur. Hotimir Tivadar. Ljubljana: Znanstvena založba Filozofske fakultete. 119–126.
- Kaja JOŠT, Urška VRANJEK OŠLAK, 2021: Analiza in klasifikacija tujejezičnih prvin v korpusu slovenskega parlamentarnega diskurza siParl. *Škerbčevi dnevi 11: zbornik prispevkov s simpozija 2019*. Ur. Franc Marušič, Petra Mišmaš, Rok Žaucer. Nova Gorica: Založba Univerze. 38–51.
- Matej METERC, 2019: Analiza rabe paremioloških uvajalnih sredstev v parlamentarnih razpravah iz korpusa SiParl. *Slovenski javni govor in jezikovno-kulturna (samo)zavest. Obdobja 38*. Ur. Hotimir Tivadar. Ljubljana: Znanstvena založba Filozofske fakultete. 135–141.
- Nina MODRIJAN, 2007: Naslavljanje pri predajanju, pridobivanju in ohranjanju vloge govorca na parlamentarnih razpravah. *Jezik in slovstvo* 52/5, 3–17.
- Vesna MOLIČNIK idr., 2022: *Uredniška navodila za oblikovanje preverjenega zapisa procedure in govorjene besede na sejah Državnega zbora, njegovo javno objavo in arhiviranje*. [Gradivo Dokumentacijsko-knjižničnega oddelka Državnega zbora Republike Slovenije.]
- Andrej PANČUR idr., 2022: *Slovenian parliamentary corpus (1990-2022) siParl 3.0*. Dostop 20. 8. 2023. <https://www.clarin.si/repository/xmlui/handle/11356/1748>
- Parlamer*. Dostop 20. 8. 2023. <https://parlamer.si/>
- Poslovnik državnega zbora (PoDZ-1)*. Dostop 20. 8. 2023. <http://pisrs.si/Pis.web/pregledPredpisa?id=POSL34>
- Sejni zapisi Državnega zbora, 1. seja (22. junij 2018)*. Dostop 20. 8. 2023. https://fotogalerija.dz-rs.si/datoteke/Publikacije/Sejni_zapisi_Drzavnega_zbora/2018-2022/2018_06_22_S_1.pdf
- Službe Državnega zbora*. Dostop 20. 8. 2023. <https://www.dz-rs.si/wps/portal/Home/odz/sdz/!ut/p/z1/>

STANDARDI TRANSKRIBIRANJA IN OZNAČEVANJA NAREČNEGA KORPUSA GOKO

KLARA ŠUMENJAK

Univerza na Primorskem, Fakulteta za humanistične študije, Koper, Slovenija
klara.sumenjak@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

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 Uvod¹

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*.² Gre za prvi slovenski narečni korpus, ki odpira številna vprašanja in možnosti za nadgradnjo.³

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, IRIS 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 ravni omogoča tudi analizo na oblikoslovni⁴ in (delno) na skladenjski ravni. 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).

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

² Korpus je realiziral dr. Jernej Vičič.

³ Več o korpusu gl. Šumenjak (2012, 2013a in 2013b).

⁴ 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.⁵

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.⁶

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

⁵ 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.

⁶ 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 govorenem 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 govorenih 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='sje...?'/> 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 'viəm, <nedokon value='sje...?'/> 'anix 'sjeđəm 'metrou*

Isti primer, kot je viden uporabniku: *ne 'viəm, <sje...> 'anix 'sjeđəm '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. <sje...>, 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 zako'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'vɛzəlɪ zə ta 'mali z'ɣun jin mi 'dali an 'kɥəs <?> se'na, 'ne, in kə je zɣ'rabu, jə <ə:> zə'jɣɣu, 'ne, pe je zaz'ɣuənu, 'polej so 'Kup'riɥci š'li pəɣ'lɛdət, kə'du z'ɣuəne, 'kaj z'ɣuəne, '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:> 'Aɥbəlɕi 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'vɛzəlɪ zə ta 'mali z'ɣun jin mi 'dali an 'kɥəs se'na, 'ne, in kə je zɣ'rabu, jə zə'jɣɣu, 'ne, pe je zaz'ɣuənu, 'polej so 'Kup'riɥci š'li pəɣ'lɛdət, kə'du z'ɣuəne, 'kaj z'ɣuəne, '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 'Aɥbəlɕi 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 sobesedilu⁷.

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 količnost glasov. Čeprav pri zapisovanju s fonetično

⁷Več 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'biɛləm, 'ne, <nar> 'ma </nar> mula'rija 'təkət, ki smo blɛ, 'viɛš, kə'ku je, <ə:> in smo 'tisti 'miɛt, 'viɛš, 'ku sə jə u'liɛku 'tisti 'miɛt, 'ne ku z'deɪ, 'tu si 'miɛla u'se po 'rokəx <?> smo š'li se u'mit in 'tako, 'ne, o'joj, sə smo 'miɛl '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 oslavljeni monoftong *ɛ* (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
 ɛ → 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 'Pòst'ojni 'dost 'dʲɛlli, zə 'Tərst zə 'marsik'ej, 'ne ve'likix 'dʲɛl, 'əmpək so 'dʲɛlli, zə'stòpəš, [Aha] <ə:> in tək'ku zɣ'ljèda, da jə 'jəmu ɥ'sak 'suɣ kamno 'lòmčək in 'pač, 'kukər jə š'lo, jə š'lo, 'ne, 'sej 'vięš [Aha], 'tu jə 'taka š'torje, 'təm ɥət 'təwžənt ɥəsəms 'tu 'de'viède'sedya du 'pòl, 'kònc 'vojne, 'vięš 'təm, kə jə bla 'wəjska kònc'ana 'pərva.

Poenostavljeni zapis:

In 'pòle si je, <ə:> zglièda, də so t Poknjženo: zgodba 'təst'ojni 'dost 'dielli, zə 'Tərst zə mársik'ej, 'ne velikih dièl, 'əmpək so dièlli, zə'stòpəš, [Aha] <ə:> in tək'ku zglièda, da jə 'jəmu vsək sɥj kamnolòmčək in 'pač, 'kukər jə š'lo, jə š'lo, 'nə, 'sėj viéš [Aha], 'tu jə 'taka š'torje, 'təm wət 'təwžənt wəsəms'tu 'devièdesédga du 'pòl, 'kònc 'vojne, viéš 'təm, kə jə bla 'wəjska kònc'ana 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 ZRCola⁸

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).

⁸ 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.

in

Poisvedba:

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 wsə vesela, kakú bəm domā poviēdla, ja blo za plačet dvěj liri. Wsi so šli, sámo jas né, ni bilo dvěj liri za me dát, má so jmiēli, ki wsək drúgi dán, máti smo <rek...>, nóna bi rekli [Aha] zděj, né, <a:> má še bəl se spounem, tərko mlička ja miēla **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 dialektolog), 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 funkcionalnost 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.

STANDARDIZACIJA PREKMURSKE TRANSKRIPCIJE SAMOGLASNIKOV: ŠTUDIJA PRIMERA

MIHAELA KOLETNIK, MELITA ZEMLJAK JONTES

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija

mihaela.koletnik@um.si, melita.zemljak@um.si

V pričujočem prispevku bomo v okviru pilotne obravnave prekmurskega narečja preverili primernost v projektu predvidenega postopka standardizacije slovenske narečne transkripcije, pri čemer bomo (1) preverili fonetično transkripcijo prekmurskega gradiva za SLA za issln. *ǣ, issln. *ǰ- in issln. *-ǰ; (2) opravili primerjavo s fonetičnim zapisom v dialektološki literaturi in gradivu, dostopnem po posameznih regionalnih središčih; (3) na izbranih primerih opravili akustično analizo dostopnih zvočnih posnetkov (eksperimentalnofonetična analiza).

DOI
[https://doi.org/
10.18690/um.ff.4.2024.7](https://doi.org/10.18690/um.ff.4.2024.7)

ISBN
978-961-286-882-6

Ključne besede:

panonska narečna skupina,
prekmursko narečje,
gradivo za SLA,
fonetični zapis,
eksperimentalnofonetična
analiza



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.7](https://doi.org/10.18690/um.ff.4.2024.7)

ISBN
978-961-286-882-6

Keywords:

Pannonian dialectal group,
Prekmurje dialect,
material for SLA,
phonetic transcription,
experimental phonetic
analysis

STANDARDISATION OF PREKMURJE DIALECTAL TRANSCRIPTION OF VOWELS: A CASE STUDY

MIHAELA KOLETNIK, MELITA ZEMLJAK JONTES

University of Maribor, Faculty of Arts, Maribor, Slovenia
mihaela.koletnik@um.si, melita.zemljak@um.si

In the context of the pilot research of the Prekmurje dialect, this paper examines the suitability of the process of standardisation of Slovene dialectal transcription, where we will (1) analyse the phonetic transcription of the Prekmurje dialectal material for SLA for issln. * \bar{e} , issln. * \bar{a} - and issln. * \bar{a} ; (2) compare the phonetic transcription in the dialectological literature and the accessible material of individual regional centres; (3) perform an acoustic analysis of the accessible audio recordings based on selected examples (experimental phonetic analysis).



1 Uvod¹

Eden izmed ključnih ciljev sodobnih raziskav spontane govornje rabe jezika je pregled stanja in opredelitev potreb po govornih podatkih ter pripadajoči raziskovalni infrastrukturi. Pri tem je pomemben tudi socialnozvrstni vidik, pri čemer predstavljajo narečja velike raziskovalne izzive. V projektu Mezzanine se bomo predstavniki vseh slovenskih dialektoloških središč v okviru delovnega sklopa *Narečna variabilnost* ukvarjali z (1) zanesljivostjo veljavne slovenske narečne transkripcije, (2) določitvijo prostorske razširjenosti nestandardnih fonemov, (3) izdelavo prostorskega modela za pripravo diasistemskih narečno-knjižnih kontrastivnih tabel fonemov in (4) opredelitvijo optimalnega nabora slovenskih fonemov, uravnoteženega med standardnimi in narečnimi različicami fonemov.

V pričujočem prispevku bomo v okviru pilotne obravnave prekmurskega narečja panonske narečne skupine preverili primernost v projektu predvidenega postopka standardizacije slovenske narečne transkripcije samoglasnikov. Pregledali in ovrednotili bomo (1) fonetično transkripcijo prekmurskega gradiva za SLA za naslednje izhodiščnoslovenske glasove: stalno dolgi polglasnik (**ǣ*),² staroakutirani polglasnik v nezadnjem besednem zlogu (**ǣ-*) in staroakutirani polglasnik v zadnjem besednem zlogu (**-ǣ*); (2) opravili bomo primerjavo s fonetičnim zapisom v dialektološki literaturi in gradivu, dostopnem po posameznih regionalnih središčih; (3) na izbranih primerih bomo opravili akustično analizo dostopnih zvočnih posnetkov (eksperimentalnofonetična analiza).

2 Narečna podoba Prekmurja³

Prekmurje je nižinska, poljedelska pokrajina na skrajnem severovzhodnem delu Slovenije ob meji z Avstrijo na zahodu, Madžarsko na severu in vzhodu ter Hrvaško na jugu. Zemljepisno, kulturno in zgodovinsko je vpeta med tri v več pogledih zelo različne narode. Zaradi številnih in ostrih razdelitev Prekmurja vse do 18. stoletja,⁴

¹ Prispevek je nastal v okviru raziskovalnega projekta ARRS Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije (ARIS) iz državnega proračuna.

² Besedilo je bilo delno 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.

³ Prirejeno po Koletnik (2008: 9–16).

⁴ Prekmurje je bilo od 11. stoletja sestavni del ogrskega kraljestva, po trianonski mirovni pogodbi leta 1920 pa je bilo slovensko etnično ozemlje med Muro in Rabo (t. i. Slovenska krajina) razdeljeno tako, da je večji del pripadel

ko so na ta del slovenskega etničnega prostora dotekali različno močni kulturni ter posledično tudi jezikovni vplivi,⁵ so se na tem območju oblikovala tri temeljna podnarečja: severno – goričko (ob zgornji Ledavi in v Porabju), osrednje – ravensko (od Cankove prek Murske Sobote do Filovcev in Kobilja) in južno – dolinsko (vzdolž Mure). Glede na prvotno delitev slovenskega jezikovnega prostora, ki je odraz dveh naselitenih valov, pripada prekmursko narečje jugovzhodni skupini, torej tistemu delu, ki so ga naselili Slovani z območja južnoslovsanske jezikovne skupine ob Savi in Dravi navzgor. Kot najvzhodnejše slovensko narečje ohranja prekmurščina prvine, ki so v drugih narečjih že izgubljene, tu pa so se ohranile zaradi tisočletne izoliranosti od matičnega naroda.

Prekmurska podnarečja se med seboj razlikujejo po mlajših prekmurskih razvojih, ki se po Zorkovi (1998a: 42) kažejo zlasti v (1) različnih odrazih za dolgi in kratki *a* – v goričkem in ravenskem podnarečju se kratki in nenaglašeni *a* zaokrožujeta, v dolinskem pa se pod vplivom prleškega narečja zaokrožuje dolgi *a* v *o:/â*; kratki naglašeni in nenaglašeni *a* pa ostajata odprta vokala; (2) v razvoju končnega *-l* v *-o* v goričkem in ravenskem podnarečju ter *-u* v dolinskem; (3) v izgovoru zvočnika *j*, ki se v dolinskem in delu vzhodnega ravenskega podnarečja izgovarja kot *j*, sicer kot *dj*, *g*, *k* ali *dž*. Starejši glasoslovni in oblikoslovni razvoji so dokaj enotni.

Prekmursko narečje nima tonemskih nasprotij, ima pa kolikostna; naglašeni samoglasniki so dolgi ali kratki, nenaglašeni samo kratki. Dolgi in kratki naglašeni zlogi so možni v vseh besednih zlogih večzložnih besed. Izvedena sta bila oba splošnoslovsenska naglasna premika (*zlâto* → *zlatô*, *ôko* → *okô*; *dûšâ* → *dúša*), naglašeni so *e*, *o* in *æ* pred nekdanjimi kratkimi zadnjimi naglašeni zlogi (*l'žena*, *'noga*, *'megla*) ter samoglasniki (tudi v predponah in predlogih) po umiku naglasa s kratkega odprtega zloga (*'bogāt*, *'otpren*, *p'rinās*). Ohranitev dolgega tematskega samoglasnika *-e* v sedanjiku *ne'se:n*, *be'rę:n*, *zo'vę:n*; *koren'lje:*, *pošten'lje:* kaže na zveze z zahodnoslovsanskimi jeziki, posebno s slovaščino (Zorko 2003: 74).

Kraljevini SHS (kasneje Jugoslaviji, danes Sloveniji), 9 slovenskih vasi ob reki Rabi (t. i. slovensko Porabje) pa je ostalo na Madžarskem. Po črti, ki danes loči južni in osrednji del Prekmurja, je od leta 1176 pa vse do leta 1777 potekala meja med györsko in zagrebško škofijo.

⁵ Jezik prekmurskih Slovencev je bil ob močnih madžarskih pritiskih že od 11. stoletja naprej izpostavljen tudi hrvaško-kajkavskim vplivom, saj je bilo ozemlje med Muro in Rabo l. 1094 cerkvenoupravno priključeno zagrebški škofiji. Zaradi narodnozavednih in tudi jezikovnih vzrokov, prekmurski Slovenci namreč kajkavščine niso nikoli sprejemali za svoj jezik, ampak so jo imeli za »neprijubljen, vendar potreben pripomoček pri posredovanju grških, latinskih in madžarskih besedil v njim najbližji jezik« (Jesenšek 1991/92: 178), se je prekmurski jezik ohranil kljub zapletenim zgodovinskim okoliščinam, njegovo pisno tradicijo pa danes ohranja tudi sodobna prekmurska narečna književnost.

Prekmurski samoglasniški vokalni sistem je monoftongično-diftongičen. Dolgi široki praslovanski samoglasnik *ja* se je zožil in diftongiral v dvoglasnik *ej* ('be:žli, 'le:žs, 'se:žkan), z njim pa sta se kot fonološki par zožila tudi praslovanska dolgi *o* in nosni *o* ter se diftongirala v dvoglasnik *ou* ('mo:ust, 'no:uč, 'go:uš, k'lo:up). Posebnost prekmurščine so še sprednja zaokrožena samoglasnika *ü* za *u* ('dii:ša, 'kii:p) ter *ö* za *e* in *u* v položaju ob zvočnikih *v* in *r* ('vöter 'veter', 'kō:jati 'kuhati'), *u* za samoglasniški *ĭ* ('ku:čti 'tolči', 'pun 'poln', 'vu:k 'volk'), dolgi ozki *e* za praslovanski stalno dolgi polglasnik ('de:n, 'le:n 'lan', 've:s 'vas') ter onemitev ali prehod velarnega pripornika *x* v *j*, odvisno od glasovne soseščine ('la:pec < xlapec, streja < strexa, p'ra:j/p'ra: < prax). Končni *-m* se zamenjuje z *-n* ('de:lan 'delam', 'tām 'tam'), zvočnik *v* pa ima položajno različico *f* ('kii:p 'skupaj', 'ke:f 'kri', 're:kef 'redkev'). Zvočnik *j* se v dolinskem in delu vzhodnega ravenskega podnarečja izgovarja kot *j* ('jes 'jaz', 'jabuko 'jabolko'), sicer kot *dj/dž*, *g* ali *k* ('dja:rek 'jarek', 'džo:ukati 'jokati', pi'ge:n 'pijem', 'li:stke 'listje'). Posebnosti v prekmurskem oblikoslovju so končnica *-i* v dajalniku in mestniku ednine moške sklanjatve ('si:ni (daj. ed.), pri 'si:m), v množini je ohranjeno podaljševanje osnove z *-je* (vlas'je: 'lasje'), pred končnico *-je* pa se osnova pogosto podaljšuje še z *-ov-* (zo'bo:u:ve 'zobje'). Edninska orodniška ženskospolska končnica se glasi *-ov* (z 'mäterjo:f). Dvojina je trdno ohranjena pri vseh spolih. V sedanjiški glagolski spregatvi se v prvi osebi dvojine ohranja osebilo *-va* za moški ('de:lava) in *-ve* ('de:lave) za ženski spol, pogosta pa je še tvorba ponavljalnih glagolov s sedanjiško pripono *-je* (pla'čii:n, 'le:čen).

3 Nabor narečnega gradiva

Narečno gradivo z leksemi iz prekmurskih krajevnih govorov, vključenih v mrežo krajev za SLA, nam je priskrbel vodja Dialektološke sekcije Inštituta za slovenski jezik Frana Ramovša dr. Karmen Kenda Jež. V mrežo krajev za SLA so vključene naslednje krajevne točke: T387 Cankova, T388 Gorica, T389 Martjanci, T390 Strehovci, T391 Beltinci, T392 Gomilica, T393 Nedelica, T394 Velika Polana, T395 Gornja Bistrica, T396 Hotiza, T397 Večeslavci, T398 Grad, T399 Gornji Petrovci, T400 Križevci, T401 Kančevci, T402 Šalovci, T403 Markovci, T404 Gornji Senik – Felsőszölnök, T405 Slovenska ves – Rábatótfalu, T406 Ženavci – Jennersdorf in 414 Žetinci – Sieldorf. Iz gramatičnega dela vprašalnice za SLA pridobljeno gradivo za fonetični inventar posameznih krajevnih govorov, ki je bilo ob prejemu že

vneseno v excelove tabele,⁶ smo dopolnili z zapisi iz novejših znanstvenih razprav in diplomskih del, katerih avtorji in avtorice so: Marc Greenberg (1993), Tine Logar (1981; 1996), Zinka Zorko (1998; 2003; 2008), Francek Mukič (2005), Marija Bajzek Lukač (2009), Jožica Lukač (1997), Darinka Mulaosmanović (2001), Martina Koroša (2005), Alojz Tuškei (2000), Milena Salaj (2003), Renata Špilak (2002), Janja Adanič (2009) in Suzana Panker (1997).

3.1 Preglednice z gradivom

Za prikaz glasovne variabilnosti je bil v Dialektološki sekciji Inštituta za slovenski jezik oblikovan besedni seznam s 3–10 vzorčnimi leksemi za posamezne izhodiščnoslovenske glasove. Gradivo v tabelah je zapisano citatno, kot je to v listkovni in zvezkovni kartoteki,⁷ in to v dveh transkripcijah: starejši Ramovševi in novejši Logarjevi oz. novi nacionalni transkripciji.⁸ S starejšo transkripcijo so zapisani odgovori v točkah, označenih s številkami: 388 (iz let 1960, 1973), 389, 390, 391, 392 (iz let 1957, 1963), 393, 394, 396, 399, 404 (iz leta 1976). Z novejšo transkripcijo so zapisane točke 387 (iz let 2003, 2008), 392 (iz leta 1986–1988), 395, 397, 401, 402, 404 (iz let 1978, 1983, 1986) in 414 (iz leta 1989), gradivo iz Bakovcev (iz let 1997 in 2001), Moravskih Toplic, Tišine, Črenšovcev,⁹ Turnišča (iz let 2002 in 2009)¹⁰ in Števanovcev. Greenbergovo gradivo za Cankovo, Polano in Martinje iz leta 1993 je zapisano v mednarodni fonetični transkripciji IPA. Slovarska gesla v Mukičevem Porabsko-knjižnoslovensko-madžarskem slovarju (2005) so onaglašene tako: ostrivec zaznamuje dolge samoglasnike: *í, ú, ú*, dolgi in kratki *é*, dolgi ozki *ó*, dolgi *á* in dolgi *ř*; krativec označuje kratke naglašene *ì, ù, ù*, široki *è*, na kratkem *à* ni naglasa; kratki naglas se zapisuje tudi na dvoglasniku *èj*, dvoglasnik *áu* pa nima naglasa (Zorko 2005: I).

⁶ V desetih excelovih tabelah, razdeljenih po področnih središčih, je 53 zavihkov s po 3–10 vzorčnimi leksemi za posamezne izhodiščnoslovenske glasove (s povprečno 3 leksemi), ki prinašajo fonetično gradivo najmanj 160 leksemov za okrog 400 raziskovalnih točk.

⁷ Digitalizacija listkovnega in zvezkovnega gradiva za SLA se je v Dialektološki sekciji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU pričela leta 2023. Več o tem glej pri Weiss, Žejn (2004: 179–188).

⁸ Ramovševa transkripcija, ki je bila v uporabi v prvem obdobju zbiranja gradiva, je zlasti v letih 1946–1958 izpopolnil Tine Logar. Spremembo načina zapisovanja je spodbudilo prizadevanje delovne skupine za OLA za sestavo enotne transkripcije za vse slovanske jezike. Logar je temeljnim načelom transkripcije OLA sledil le z uvedbo posebnih diakritičnih znamenj za označevanje naglašenosti (´), kolikosti (˙) in tona (˘), medtem ko je zapisovanje kolikosti ostalo nespremenjeno (Kenda Jež 2011: 27). Prvi zapis v novi, Logarjevi transkripciji je iz leta 1975 (T369), zadnji zapis v stari, Ramovševi transkripciji pa iz leta 1986 (T261) (Kenda Jež 2011: 27).

⁹ Ta točka s staro številko 305 v Ramovševi mreži krajev po preureditvi ni bila vključena v mrežo krajev za SLA. Številka točke je zato v preglednicah zapisana v oklepaju.

¹⁰ Ta točka s staro številko 306a v Ramovševi mreži krajev po preureditvi ni bila vključena v mrežo krajev za SLA. Številka točke je zato v preglednicah zapisana v oklepaju.

Gradivo predstavljamo tako, da je v prvem stolpcu zapisana številka točke iz mreže krajev za SLA,¹¹ v drugem kraj zapisa, v tretjem čas zapisa odgovora, v naslednjih stolpcih so zabeleženi odgovori na različna vprašanja iz gramatičnega dela Vprašalnice za SLA, v zadnjem pa je podana abstrahirana transkripcija (AT). Znak # pomeni, da točka ni vključena v mrežo krajev za SLA.

3.2 Izsledki na podlagi narečnega gradiva

Tabela 1: Odrazi za issln. *⁵

Točka SLA	Kraj zapisa	Leto zapisa	V700 vas	V 700 dan	V700 lan	V700 laž	V700 čast	AT
387	Cankova	1993	v'e:s	d'e:n	l'e:n			e:
387	Cankova	2003, 2008		'dɛ:n				ɛ:
387	Cankova	2013						
388	Gorica	1960	vêš	dêŋ	lêŋ			ɛ:
388	Gorica	1973	vêš	dêŋ	lêŋ	lâš	čâst	ɛ:, a:, e:i
389	Martjanci	1958	vêš		lêŋ	lâš	čêst	ɛ:, a:
390	Strehovci	1983	vêš	dêŋ	lêŋ	lâš	čêst	ɛ:, a:
#	Bakovci	1997	'vɛ:s	'dɛ:n	'lɛ:n	'la:š	'čɛ:st	ɛ:, ɛ:, a:
#	Bakovci	2001	'vɛ:s	'dɛ:n	'lɛ:n	'la:š	'čɛ:st	ɛ:, a:, â
#	Moravske Toplice	2005	'vɛ:s	'dɛ:n	'lɛ:n			ɛ:, ɛ
#	Tišina	2000	'vɛ:s	'dɛ:n	'lɛ:n			ɛ:
(305)	Črenšovci	2003	'vi:ɛš	'di:ɛŋ	'li:ɛŋ			i:ɛ
(306a)	Turnišče	2002	'vi:ɛš	'di:ɛŋ	'li:ɛŋ			i:ɛ
(306a)	Turnišče	2009		'di:ɛŋ				i:ɛ
391	Beltinci	1972	vêš	dêŋ	lân	lâš	čâst	ɛ:, a:
392	Gomilica	1957	vêš	dêŋ			čêst	ɛ:
392	Gomilica	1963	vêš	dêŋ	lêŋ	lâš	čêst	ɛ:, a:
392	Gomilica	1986–1988	'vɛ:s	'dɛ:n	'lɛ:n	'la:š	'čɛ:st	ɛ:, a:
393	Nedelica	1959	vêš		lêŋ		čêst	ɛ:
394	Velika Polana	1957	vêš	dêŋ			čâst	ɛ:, â:
395	Gornja Bistrica	1983	'viɛš	'diɛŋ	'liɛŋ	'lâ:aš	'čâ:ast	iɛ, â:a
396	Hotiza	1957	vêš	dêŋ				ɛ:

¹¹ Ramovševo mrežo krajev za SLA, ki je prvotno obsegala 312 točk in se sčasoma povečala na 686, sta v letih 1982–1984 preuredila Tine Logar in Jakob Rigler. Številni kraji so dobili nove številke, nekaj pa jih je tudi izpadlo (Benedik 1999: 18).

Točka SLA	Kraj zapisa	Leto zapisa	V700 vas	V 700 dan	V700 lan	V700 laž	V700 čast	AT
#	Polana	1993		<i>d'ien</i>	<i>l'ien</i>			<i>ie</i>
397	Večeslavci	1983	<i>'vɛ:ɨs</i>	<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>'lɥá:š</i>	<i>'ča:st</i>	<i>'ɛ:, a:, e:i, uá:</i>
398	Grad	1958						<i>ie</i>
399	Gornji Petrovci	1967	<i>vɛs</i>	<i>dɛ'n</i>	<i>lɛ'n</i>	<i>láš</i>	<i>čɛst</i>	<i>ɛ:, ɛj, e:'i, a:</i>
400	Križevci	1958						<i>ɛj</i>
401	Kančevci	1997	<i>'vɛ:s</i>	<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>la:š</i>	<i>'čɛ:st</i>	<i>ɛ:, a:</i>
401	Kančevci	2013						
402	Šalovci	1985	<i>'vɛ:s</i>	<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>'la:š</i>	<i>'čɛ:st</i>	<i>ɛ:, a:</i>
403	Markovci	1958						<i>ɛj</i>
404	Gornji Senik – Felsőszölnök	1976	<i>vɛs</i>	<i>dɛn</i>	<i>lɛn</i>			<i>ɛ:</i>
404	Gornji Senik – Felsőszölnök	1978	<i>'vɛ:s</i>	<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>'la:š</i>		<i>ɛ:, a:</i>
404	Gornji Senik – Felsőszölnök	1981						<i>ɛ:</i>
404	<i>Gornji Senik – Felsőszölnök</i>	1983	<i>'vɛ:s</i>	<i>'dɛ:n</i>	<i>'lä:n</i>	<i>'la:š</i>	<i>'ča:st</i>	<i>ɛ:, ä:, a:</i>
404	Gornji Senik – Felsőszölnök	1986	<i>'vɛ:s</i>		<i>'lɛ:n</i>	<i>'la:š</i>		<i>ɛ:, a:</i>
404	Gornji Senik – Felsőszölnök	2005	<i>vɛs</i>	<i>dɛn</i>	<i>lɛn</i>	<i>láž</i>		
404	<i>Gornji Senik – Felsőszölnök</i>	2009		<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>'la:ž</i>	<i>'čɛ:st</i>	<i>ɛ:, a:</i>
405	Slovenska ves – Rábatótfalu	2013						
406	Ženavci – Jennersdorf							
#	Števanovci – Apátistvánfalva	1998		<i>'dɛ:n</i>	<i>'lɛ:n</i>		<i>'čɛ:st</i>	<i>ɛ:</i>
#	Martinje	1993	<i>v'ɛis</i>	<i>d'ɛin</i>	<i>l'ɛin</i>			<i>ɛi</i>
414	Žetinci – Sieldorf	1989	<i>'vɛ:s</i>	<i>'dɛ:n</i>	<i>'lɛ:n</i>	<i>'la:š</i>	<i>'ča:st</i>	<i>ɛ:, a:</i>
414	Žetinci – Sieldorf	2014						

Vir: Dialektološka sekcija ISJFR

V tabeli 1 so predstavljeni odrazi za issln. * $\bar{\epsilon}$ v besedah V700 *vas*, *dan*, *lan*, *laš* in *čast*.

Razvoj stalno dolgega polglasnika v ozki dolgi ϵ : je v vseh ravenskih govorih enak, le v položaju pred *n* se v Moravskih Toplicah zanj govori kratki ozki ϵ (*'dɛn*, *'lɛn*).

V točki 388 Gorica sta bila za issln. * $\bar{\epsilon}$ prepoznana dva odraza, dolgi ozki ϵ : v zapisu iz leta 1960 in dvoglasnik *e:ɨ* v trinajst let mlajši zbirki podatkov. Zaradi kratkega časovnega obdobja med zaporednima zapisoma menimo, da v Gorici ne gre za potekajočo jezikovno spremembo, dvoglasnik *eɨ* < issln. * $\bar{\epsilon}$ je bil v prekmurskih govorih pri Logarju zabeležen že konec petdesetih leti prejšnjega stoletja, pač pa za različno dojetje fonetične vrednosti glasov obeh zapisovalcev, pri čemer bi kot problematično izpostavili še fonetično vrednost dvoglasnika. Zapis pridevnika *pejski* istega zapisovalca v gradivu iz leta 1973 kaže na to, da *e:ɨ* ni razširjen kot pri *e:ɨ* < * $\bar{\epsilon}$, ampak da je ožji,¹² zato v točki 388 Gorica (1973) predlagamo transkripcijski popravek pri V700 *vas* iz *vɛjs* v *'vɛ:js*.

V gradivu bakovskega govora iz leta 1997 se za issln. * $\bar{\epsilon}$ v položaju pred *n* pojavlja kolikostna dvojica: *'dɛn* ~ *'lɛ:n*. Ker se v zaporednem zapisu iz Bakovcev (2001) issln. * $\bar{\epsilon}$ v leksemu *dan* odraža kot ozki dolgi ϵ : (*'dɛ:n*), menimo, da gre pri starejšem zapisu za drugačen slušni vtis zapisovalke. Zapis *'lɛn* v časovno mlajšem bakovskem gradivu razlagamo kot vpliv knjižnega jezika in ne kot položajni refleks stalno dolgega polglasnika pred *n*.

Odras *a*: < issln. * $\bar{\epsilon}$ v besedah *laš* in *čast* razlagamo kot vpliv knjižnega jezika. Medtem ko je razvoj stalno dolgega polglasnika > *a*: kontinuirano izkazan v besedi *laš*, se v besedi *čast* v ravenskih govorih pojavlja več primerov z ozkim dolgim ϵ : zanj.

Odras za issln. * $\bar{\epsilon}$ se od splošnega prekmurskega razvoja v dolgi ozki ϵ : razlikuje v južnih dolinskih govorih, saj se je ϵ : diftongiral v *i:ɛ*. Odstopanje od regularnega razvoja pri besedi *lan* v zapisu iz Beltincev (1972), kjer je za stalno dolgi polglasnik odraz *a*: (*'lɛn*), je mogoče razložiti z vplivom knjižnega jezika.

¹² V pridevniku *pasji* se namreč zaradi zgodnje podaljšave nezadnjega akutiranega zloga * $\bar{\epsilon}$ odraža kot * $\bar{\epsilon}$.

Polansko gradivo iz dveh časovnih obdobj¹³ v razvoju stalno dolgega polglasnika kaže na jezikovno spremembo. Medtem ko se issln. **ǣ* v zapisu iz Velike Polane (1957) še odraža kot monoftong *e*;, se šestintrideset let pozneje zanj govori dvoglasnik *i:ɛ*.

Tudi na Dolinskem se v besedi *laž* stalno dolgi polglasnik pod vplivom knjižnega jezika kontinuirano odraža kot *a*-jevski fonem, v besedi *čast* pa sta bila prepoznana dva odraza – *e*: in *a:/ǎ*;¹⁴ pri čemer je prvi pogostejši. V teh primerih se v gornjebistriškem gradivu iz leta 1983 pojavlja odraz *ǎ:a*, ki ga je mogoče pojasniti z drugačnim slušnim vtisom zapisovalke, zato v točki 395 Gornja Bistrica (1983) predlagamo transkripcijski popravek pri V700 *laž* in *čast* iz *ˈlǎ:aš*, *ˈčǎ:ast* v *ˈlǎ:š*, *ˈčǎ:st*.

Na Goričkem se je stalno dolgi polglasnik v vseh krajevnih točkah in v vseh položajih največkrat razvil v dolgi *e*:. Kot dvoglasnik se odraža v točki 397 Večeslavci (*ˈvɛ:i*) in v Martinju (*ˈvʲɛ:i*), pri čemer je opaziti razliko v fonetični vrednosti dvoglasnika. V severnem martinjskem govoru je namreč razširjen kot pri *e:i* < **ǣ*, medtem ko je v zahodnem večeslavskem govoru ožji.

Kot različica issln. **ǣ* v položaju pred *n* se v točki 397 pojavlja diftongični odraz *ie*: (*ˈdʲe:n*, *ˈlʲe:n*),¹⁵ v točki 399 Gornji Petrovci pa sta v tem glasovnem okolju zabeležena dvoglasnika *eʲ* in *eʲi*, ki se v zapisu razlikujeta tako po kvantiteti kot tudi kvaliteti (*dě̃n*, *lě̃in*). Ali res gre za dva različna fonema ali le za drugačen slušni vtis zapisovalca zaradi pomanjkanja gradiva, ni mogoče z gotovostjo trditi. Za to bi potrebovali natančnejšo raziskavo z več zbranimi primeri.

V razvoju stalno dolgega polglasnika v položaju pred *n* je opazna še sprememba med gradivom v točki 414 iz leta 1983 ter poznejšo zbirko podatkov iste zapisovalke iz leta 2009. V starejšem zapisu se na mestu issln. **ǣ* kot dvojnica k *e*: pojavlja tudi zelo široki dolgi *ǎ*: (*ˈdɛ:n* ~ *ˈlǎ:n*), v mlajšem gradivu pa je izpričan le dolgi ozki *e*: (*ˈdɛ:n*,

13 Zapis za točko 394 Velika Polana je iz leta 1957, mlajša, Greenbergova raziskava pa je iz leta 1993. Greenberg je terensko raziskavo izpeljal v zaselku Bükovje v Mali Polani. V pogovorih s prebivalci Velike Polane se je avtorju potrdilo, da je govor obeh delov (Male in Velike Polane) enak, kakor trdijo domačini (prim. Greenberg 1993: 468).

14 Nizki sredinski samoglasnik *a*-jevske barve se govori v severnih, zelo široki sredinski zadnji samoglasnik pa v južnih dolinskih govorih.

15 Dvoglasnik *ie* (fonologiziran zapis) za točko 398 Grad beleži tudi Logar, a brez gradiva, zato ni mogoče vedeti, ali gre za regularen razvoj **ǣ* ali le za njegovo položajno različico (pred zvočnikom *n*).

'lɛ:n). Sklepamo, da gre v novejšem gradivu za natančnejši fonetični zapis, zato predlagamo transkripcijski popravek iz 'lä:n v 'lɛ:n.

V leksemu *laž* se issln. *ǫ pod vplivom knjižnega jezika odraža kot *a*; le v Večeslavcih kot *uā*; pri čemer gre verjetno le za idiolekt informatorke.¹⁶ Pri leksemu *čast* gradivo izkazuje več primerov z regularnim refleksom za stalno dolgi polglasnik, torej z dolgim ozkim *e*; manj pogost *a*: v tem primeru pa razlagamo z vplivom knjižnega jezika.

Tabela 2: Odrazi za issln. *ǫ-

Točka SLA	Kraj zapisa	Leto zapisa	V702 maša	V702 pasji	V702 vaški	V702 premakne	V702 gane	AT
387	Cankova	1993						
387	Cankova	2003 in 2008	'meša					e
387	Cankova	2013						
388	Gorica	1960	měšā	pěškij	věškij			e:, e:
388	Gorica	1973	měša	pěšikij	věškij		gěne	e, e:i
389	Martjanci	1958	māša		věški		gěne	ā, e:, e:
390	Strehovci	1983	měša	pěšk'i	věški	preměkne	gěne	e, e:
#	Bakovci	1997	'mäšā	'pe:ske, 'pe:šče	'väške	prä'mäknä		ä, e:
#	Bakovci	2001	'mäšā	'pe:šči	'väški	prä'mäknä		ä, e:
#	Moravske Toplice	2005	'mäšā	'pe:stj	'väškij	prä'mäknä	'gəne	ä, e:, e
#	Tišina	2000	'mäšā					ä
(305)	Črenšovci	2003	'mäša			prä'mäknä	'gänä	ä
(306a)	Turnišče	2002	'mäša		'väški		'gänä	ä
(306a)	Turnišče	2009	'mäša					ä
391	Beltinci	1972	měša		věškij	preměkne	gěne	e
392	Gomilica	1957	měša		věškij			e
392	Gomilica	1963	māša	pěšij	väški		gäne	ä, e:
392	Gomilica	1986–1988	'mäša	'pe:sje	'väške		'gäne	ä, e:
393	Nedelica	1959	māša					ä
394	Velika Polana	1957	měša				gěneņ	e

¹⁶ V leksemu *laž* se issln. *ǫ v vseh prekmurskih govorih razvija kot issln. *ā, zato bi tudi v Večeslavcih na tem mestu pričakovali odraz *a*. Issln. *ā se kot *uā*, ki je sicer iz issln. -ā- in *ā-, odraža samo v položaju za *r* (Gosteničnik 2022: 134).

Točka SLA	Kraj zapisa	Leto zapisa	V702 maša	V702 pasji	V702 vaški	V702 premakne	V702 gane	AT
395	Gornja Bistrica	1983	'mäša	'pięsji	'väški	pre'mäkne	'gäne	ä, ə, ię
396	Hotiza	1957	męša				gęnęn	e
#	Polana	1993	m'äša					ä
397	Večeslavci	1983	'm ^ε ä:ša	'p ^ε :įske	'v ^ε :įške		'g'ę:ne	^ε ä:, e:į, 'ę:
398	Grad	1958						e:, ä:
399	Gornji Petrovci	1967	męša	pęskę	vęškę	premękne	gęne	ę, e
400	Križevci	1958						e:, ä:
401	Kančevci	1997	'meša	'p ^ε :skį	'v ^ε :škį		g'je:ne	e, ę:, e:
401	Kančevci	2013						
402	Šalovci	1985	'mä:ša	'p ^ε :stę	'vä:štę		'dä:ne	ä:, ę:
403	Markovci	1958	'mjä:ša					ä, ä:, jä:
404	Gornji Senik – Felsőszölnök	1976	męša		vęškę			e
404	Gornji Senik – Felsőszölnök	1978	'meša	'p ^ε :nskę		'mekne	'gene	e, ę:
404	Gornji Senik – Felsőszölnök	1981						e (zemen)
404	Gornji Senik – Felsőszölnök	1983	'ma:ša	pe'so:uvde	'vä:skę		'gä:ne	a:, ä:
404	Gornji Senik – Felsőszölnök	1986	'me:ša					e:
404	Gornji Senik – Felsőszölnök	2005	męša		vęški		g ^ę niti se	è
404	Gornji Senik – Felsőszölnök	2009					'g ^ę n ^ę -tęn	e
405	Slovenska ves – Rábatótfalu	2013						
406	Ženavci – Jennersdorf							
414	Žetinci – Sichelendorf	1989	'meša		'veški	pre'mekne	'gene	e
414	Žetinci – Sichelendorf	2014						
#	Števanovci – Apátistvárfalva	1998	'mieša	'piesja	'vieški	pr'miekne		ie
#	Martinje	1993	m'ieša					ie

Vir: Dialektološka sekcija ISJFR

V tabeli 2 so predstavljeni odrazi issln. *ǣ- v besedah V702 *maša, pasji, vaški, premakne* in *gane*.

V ravenskih govorih se za issln. *ǣ- praviloma pojavljata dva odraza, zelo široki kratki ä, ki je pogostejši, in široki kratki e. Njuno dejansko fonetično vrednost bi bilo smiselno določiti z eksperimentalnofonetično analizo.

V točki 388 Gorica je v fonetičnem zapisu gradiva za novoakutirani polglasnik v nezadnjem besednem zlogu na podlagi slušnega vtisa zapisovalcev opaziti razhajanje. Starejši zapis iz leta 1960 kaže na podaljšavo nezadnjih akutiranih kračin (*měša, věškej*), mlajši iz leta 1973 (*měša, věškej*) pa potrjuje Ramovšev misel (1936: 157), da se akutirane kračine sredi besede niso podaljšale. Ker kolikost odraza za novoakutirani polglasnik v nezadnjem besednem zlogu ni jasna, bo potrebno za potrditev razvoja issln. *ǣ- v tem krajevnem govoru opraviti nadaljnje raziskave.

V točki 389 Martjanci sta bila v Logarjevem gradivu iz leta 1958 za issln. *ǣ- prepoznana dva odraza: zelo široki kratki ä (*măša*) in dolgi široki e: (*věškej*). Prvi fonem kaže na ohranitev akutiranih kračin sredi besede, drugi pa na njihovo podaljšanje. Ramovš (1936: 157) to podaljšavo razlaga kot mlado analogično podaljšavo, kjer dolžina še ni mogla vplivati na vokalno kvaliteto, torej ima *'veški* široki e še iz prejšnjega *'veški*.

Z zgodnjo podaljšavo nezadnjega akutiranega zloga pa lahko pojasnimo ravenski odraz e: z diftongično dvojnico e:j v drugem (zaporednem) zapisu v točki 388 Gorica za issln. *ǣ- v besedi *pasji*. Novoakutirani polglasnik v nezadnjem besednem zlogu se v tej besedi odraža kot issln. *ǣ, kar pomeni, da je podaljšava dokaj stara, saj je dolžina že vplivala na kvaliteto vokala (Ramovš 1936: 157).

Odras za issln. *ǣ- se od splošnega razvoja v ä ali e največkrat razlikuje le v položaju ob zvočniku n. V točki 388 Gorica (1973) in v Moravskih Toplicah je v tem glasovnem okolju zapisan kratki ozki e (*gěne; 'gene*), Logarjev zapis *gěne* v Martjancih pa kaže na podaljšavo nekdanjih akutiranih kračin sredi besede.

Na Dolinskem se v vseh krajevnih govorih in v vseh položajih na mestu issln. *ǣ- poleg širokega kratkega e pojavlja zelo široki kratki ä, ki je pogostejši. Njuno dejansko fonetično vrednost bomo določili z eksperimentalnofonetično analizo. Odstopanje od regularnega razvoja je opaziti samo v besedi *gane* v točki 395

Gornja Bistrica, kjer polglasniški odraz zanj (^l*gəne*) razlagamo z idiolektom informatorja ali drugačnim slušnim vtisom zapisovalke.

V točki 392 Gomilica je v gradivu iz treh časovnih obdobj (1957, 1963, 1986–1988) opaziti razliko v fonetični vrednosti enoglasnika (*mĕša ~ mäša ~ 'mäša*). Medtem ko se issln. *ǣ v najstarejšem gradivu odraža kot široki kratki *e*, je v drugem in tretjem zaporednem zapisu prepoznan kot zelo široki kratki *ä*. Sklepamo, da so mlajši zapisi natančnejši, zato v zapisu iz leta 1957 predlagamo transkripcijski popravek v *ä*. Enako razhajanje je opaziti tudi v zapisih za Polano. V točki 394 Velika Polana se za *ǣ govori *e*, v mlajši Greenbergovi zbirki podatkov pa *ä*. Zaradi natančnejšega mlajšega zapisa predlagamo v starejšem gradivu transkripcijski popravek v *ä*, čeprav dopuščamo tudi možnost, da je na tem mestu v daljšem časovnem obdobju (36 let) prišlo do glasoslovne spremembe, in sicer razširitve starejšega kratkega širokega *e*.

V leksemu *paŷi* odraza ozki dolgi *e*: v severnih in diftongični *iɛ* v južnih dolinskih govorih kažeta na zgodnjo podaljšavo nezadnjega akutiranega zloga, zato se issln. *ǣ se odraža kot *ǣ̄.

V krajevnih govorih na Goričkem so bili za issln. *ǣ prepoznani naslednji odrazi: dvoglasniški *ɛa*: v Večeslavcih, zelo široki dolgi *ä*: pri Gradu (brez gradiva), v Križevcih (brez gradiva), Šalovcih, Markovcih (tu s hkratnim odrazom *ä* (brez gradiva) ter sporadičnim diftongom *ja*:¹⁷), široki dolgi *e*: s hkratnim odrazom *e* v Kančevcih, dvoglasnik *ie* v Martinju, kratki dvoglasnik *ie* v Števanovcih in široki kratki *e* na Gornjem Seniku in v Gornjih Petrovcih.

Dolžino na mestu issln. *ǣ na Goričkem (^l*m^ɛü:ša, 'mä:ša, 'mjä:ša, m'ieša*) razlagamo s podaljšavo nekdanjih akutiranih kračin v nezadnjem besednem zlogu. Težnjo podaljševanja akutiranih kračin sredi besede je pri zbiranju gradiva za SLA na Goričkem opazil že Tine Logar (1996: 200), na izgubo kolikosti v prekmurskih govorih pa opozarjata tudi Vera Smole (2002: 55) in Marc Greenberg (1993: 483). Greenberg ugotavlja, da so prekmurski govori brez kolikostnih nasprotij v posameznih vaseh in ne na strnjenem območju,¹⁸ kar kaže, da gre pri tej izgubi za strukturni vzrok (Greenberg 1993: op. 17). Da ne gre za pozno podaljšavo akutiranih zlogov, temveč za krajevno sporadično podaljšavo kratkih zlogov, neodvisno od

¹⁷ Zanj glej Logarjev zapis ^l*mjä:ša* (1996: 200).

¹⁸ Vasi, ki so v Greenbergovem gradivu zabeležene kot sistemi brez kolikosti, so Števanovci, Sakalovci, Martinje, Večeslavci in Gančani. Zinka Zorko (1998: 104–105) izgube kolikosti za Števanovce ne navaja.

nekdanjega tonema, piše tudi Januška Gostenčnik (2022: 126). Da podaljšanje akutiranih kračin na Goričkem ni bilo dosledno, potrjujejo ohranjene kračine na mestu issln. *ǰ v Gornjih Petrovcih (*měša*), Kančevcih (*'meša*) in na Gornjem Seniku (*měša*, *'meša*). Opozoriti velja na izstopajoče gornjeseniško gradivo iz leta 1983, saj se na mestu issln. *ǰ pojavljata *a*: in *ä*:. Ker gornjeseniški govor ohranja kolikostna nasprotja¹⁹ in ker se v novejšem gradivu Marije Bajzek Lukač (2009) za novoakutirani polglasnik v nezadnjem besednem zlogu pojavlja kratki široki *e*,²⁰ sklepamo, da gre v gradivu iz leta 1983 ter gradivu iz leta 1986, kjer je prav tako opazno neupoštevanje samoglasniške kolikosti, za napako v zapisu, zato predlagamo transkripcijski popravek v široki kratki *e*. Dolgi *a*: v besedi *maša* v zapisu iz leta 1983 razlagamo z vplivom knjižnega jezika.

V točki 401 Kančevci sta bila v gradivu iz leta 1997 za issln.*ǰ prepoznana dva odraza: široki kratki *e* (*'meša*) in dolgi široki *e*: (*'ve:ški*, *g'je:ne*). Prvi fonem kaže na ohranitev akutiranih kračin sredi besede, drugi pa na njihovo podaljšanje. Ramovš (1936: 157) to podaljšavo razlaga kot mlado analogično podaljšavo, kjer dolžina še ni mogla vplivati na vokalno kvaliteto, torej sklepamo, da imata *'ve:ški* in *g'je:ne* široki *e* še iz prejšnjega *'veški* in *g'jene*.

Nepojasnjeno ostaja dejansko stanje v števanovskem govoru. Medtem ko se je po Zorkovi (1998: 105) issln. *ǰ tu razvil v diftongični kratki *ie*,²¹ je ta porabski govor v Greenbergovem gradivu zabeležen kot sistem brez kolikosti (1993: 483, op. 17). Za potrditev razvoja bi potrebovali natančnejšo raziskavo z več zbranimi primeri. Odras za issln. *ǰ- se od že opisanih razvojev razlikuje v položaju ob nosnih zvočnikih. V točki 397 Večeslavci ima issln. *ǰ položajno različico *ie*: (*'g'e:ne*), v Gornjih Petrovcih pa *e* (*g'ene*, *premešene*).

V besedi *pasji* se tako kot v ravenskem in dolinskem podnarečju novoakutirani polglasnik v nezadnjem besednem zlogu odraža kot stalno dolgi polglasnik. To kaže na zgodnjo podaljšavo nezadnjega akutiranega zloga,²² opazno še v zapisu *'ve:jške* 'vaški' v Večeslavcih. Odstopanje od regularnega razvoja je opaziti samo v Gornjih Petrovcih, kjer se na tem mestu pojavlja ozki kratki *e* (*pěšēē*), ki ga je mogoče pojasniti z idiolektom informatorja ali drugačnim slušnim vtisom zapisovalca.

¹⁹ Glej Logarjev fonološki opis govora na Gornjem Seniku (1981: 213–218) in primerjalno gornjeseniško gradivo.

²⁰ Glej iztočnici *geniti* in *geniti se* (Bajzek Lukač 2009: 140).

²¹ In to tudi v leksemu *pasji*, kjer se v vseh preostalih prekmurskih govorih issln. *ǰ odraža kot *ǰ.

²² Več o podaljšanju nezadnjih akutiranih zlogov v prekmurščini glej pri Ramovš (1936: 157).

Tabela 3: Odrazi za issln. *ě

Točka SLA	Kraj zapisa	Leto zapisa	V703 pes	V 703 tešč	V 527 dež	AT
387	Cankova	1993			'dež	ε
387	Cankova	2003, 2008	'pes			e
387	Cankova	2013				
388	Gorica	1960	pěs		děšč	e:
388	Gorica	1973	pěš	těšč		e
389	Martjanci	1958	pās			ä:
390	Strehovci	1983	pěš	natěšče		e
#	Bakovci	1997	'pās	'tāščā		ä
#	Bakovci	2001	'pās	'tāšč		ä
#	Moravske Toplice	2005	'pās	'tāščā	'dāšč	ä
#	Tišina	2000	'pās		'dāš	ä
(305)	Črenšovci	2003	'pās	'tāšč		ä
(306a)	Turnišče	2002	'pās			ä
(306a)	Turnišče	2009				
391	Beltinci	1972	pěš	těšč		e
392	Gomilica	1957	pěš		děšč	e
392	Gomilica	1963	pās	na tāšče		ä
392	Gomilica	1986-1988	'pās	na 'tāšče		ä
393	Nedelica	1959	pās		dāšč	ä
394	Velika Polana	1957	pěš			e
395	Gornja Bistrica	1983	'pās	na 'tāšče		ä
396	Hotiza	1957	pěš			e
#	Polana	1993	p'ās			ä
397	Večeslavci	1983	p'ä:s	na'tä:šče	'd'ä:š	ä:
398	Grad	1958				ä
399	Gornji Petrovci	1967	pěš	na těšče		e
400	Križevci	1958				ä:
401	Kančevci	1997	'pes	na 'tešče	'deš	e
401	Kančevci	2013				
402	Šalovci	1985	'pä:s	'tāšte		ä:, ä

Točka SLA	Kraj zapisa	Leto zapisa	V703 pes	V 703 tešč	V 527 dež	AT
403	Markovci	1958				ä:, ä:
404	Gornji Senik – Felsőszölnök	1976	pěš	na těščē		e
404	Gornji Senik – Felsőszölnök	1978	'pes	na 'tešče, 'teške		e
404	Gornji Senik – Felsőszölnök	1981				e (g'nes)
404	Gornji Senik – Felsőszölnök	1983	'pä:s	na'tä:šče		ä:
404	Gornji Senik – Felsőszölnök	1986	'pe:s	na 'te:šče		e:
404	Gornji Senik – Felsőszölnök	2005	pěš	těšč	děž	(e)
404	Gornji Senik – Felsőszölnök	2009				e (g'nes)
414	Žetinci – Sieldorf	1989	'pes	'teše, na 'teše		e
414	Žetinci – Sieldorf	2014				
#	Števanovci – Apátistvánfalva	1998	'pies	na 'tiešt'e		ie
#	Martinje	1993			d'iež	ie

Vir: Dialektološka sekcija ISJFR

V tabeli 3 so predstavljeni odrazi issln. *-ə v besedah V703 *pes*, *tešč* in *dež*.

V gradivu za issln. *-ə se v ravenskih krajevnih govorih poleg širokega kratkega *e* pojavlja tudi zelo široki kratki *ä*, ki je nekoliko pogostejši. Njuno dejansko fonetično vrednost bi bilo smiselno določiti z eksperimentalnofonetično analizo.

V točki 388 Gorica je v zaporednem fonetičnem zapisu gradiva za novoakutirani polglasnik v zadnjem besednem zlogu opaziti razhajanje. Starejši zapis iz leta 1960 kaže na podaljšavo zadnjih akutiranih kračin (*pěš*, *děšč*), mlajši zapis iz leta 1973 (*pěš*, *těšč*) pa potrjuje, da se akutirane kračine v zadnjem ali edinem besednem zlogu niso podaljšale. Ker enoznačnega odgovora o razvoju novoakutiranega polglasnika v zadnjem besednem zlogu v Gorici ni mogoče podati, bo potrebno za potrditev njegovega razvoja opraviti nadaljnje raziskave.

Izstopajoče je še gradivo v točki 389 Martjanci, kjer se na mestu issln. *-ə pojavlja zelo široki dolgi *ä*: (*pää*). Ker je gradivo zelo skopo, razvoja issln. *-ə ne moremo potrditi. Za razjasnitev kolikosti odraza novoakutiranega polglasnika v zadnjem besednem zlogu bo treba v prihodnjih zapisih načrtno poiskati dodatno gradivo.

Tudi na Dolinskem se issln. *-ə odraža kot *e*-jevski samoglasnik, katerega dejansko fonetično vrednost bo potrebno določiti z nadaljnjimi raziskavami, saj v zapisih na podlagi slušnega vtisa zapisovalcev prihaja do razhajanj. V Beltincih in Hotizi je zapisan kot širok sredinski sprednji kratki samoglasnik (*pěs*, *těšć*, *pěš*), v preostalih dolinskih točkah pa kot zelo širok sprednji samoglasnik iste količnosti (*ä*).

V točki 392 Gomilica je pri razvoju *-ə opaziti spremembo med najstarejšim gradivom iz leta 1957 ter mlajšimi zbirkami podatkov. V najstarejšem zapisu se novoakutirani samoglasnik v zadnjem besednem zlogu odraža kot nekoliko ožji kratki *e*-jevski glas, mlajši informatorji pa na tem mestu izgovarjajo zelo široki *ä* iste kvantitete, ki ga potrjuje tudi primerjalno Logarjevo gradivo (1993: 98), zato v točki 392 Gomilica (1957) predlagamo transkripcijski popravek pri V703 *pes* in *dešć* iz *pěš*, *děšć* v *päs*, *däšć*.

V točki 394 Velika Polana se v prvem zapisu iz leta 1957 za *-ə govori *e* (*pěš*), v mlajši Greenbergovi zbirki podatkov (1993) pa *ä* (*p'äš*). Zaradi natančnejšega mlajšega zapisa predlagamo v starejšem gradivu transkripcijski popravek v *ä*, čeprav dopuščamo tudi možnost, da je na tem mestu v daljšem časovnem obdobju (36 let) prišlo do glasoslovne spremembe, in sicer razširitve starejšega kratkega širokega *e*.

Na Goričkem so bili za issln. *-ə prepoznani naslednji odrazi: diftongični *ɛä*: v Večeslavcih, zelo široki dolgi *ä*: v Križevcih (brez gradiva), Šalovcih (tu s hkratnim odrazom *ä* – *'üšćte*) in Markovcih²³ (brez gradiva), dvoglasnik *ie* v Martinju, široki kratki *e* v Gornjih Petrovcih Kančevcih in na Gornjem Seniku ter kratki dvoglasnik *ie* v Števanovcih (*'pies*).

Dolžina na mestu issln. *-ə (*'p'ä:s*, *'pä:s*, *d'ieš*) kaže na podaljšavo nekdanjih akutiranih kračin v zadnjem besednem zlogu. Po Logarju (1996: 200) si namreč dvoglasnikov na tem mestu ob ohranjenih kračinah (*pěš*, *'peš*), zabeleženih v točkah 398 Grad (brez gradiva), 399 Gornji Petrovci, 401 Kančevci in 404 Gornji Senik, sicer ne bi mogli razložiti.

V točki 402 Šalovci issln. *-ə izkazuje različna odraza – zelo široki dolgi *ä*: in zelo široki kratki *ä*. Zaradi skopega gradiva in različnosti odrazov v pridobljenem gradivu kvantiteta odraza za issln. *-ə ni jasna, zato bi bilo smiselno zbrati več gradiva in dopolniti zapis.

²³ Tu s sporadičnim diftongom *ja*; o čemer glej več pri Logar (1996: 200).

Opozoriti velja še na izstopajoče gornjeseniško gradivo iz leta 1983, saj se na mestu issln. *-ə pojavlja ä: Ker gornjeseniški govor akutiranih kračin v zadnjem ali edinem besednem zlogu ni podaljšal²⁴ in ker se v novejšem gradivu Marije Bajzek Lukač (2009) za novoakutirani polglasnik v zadnjem besednem zlogu pojavlja kratki široki e,²⁵ sklepamo, da gre v gradivu iz leta 1983 ter gradivu iz leta 1986, kjer je prav tako opazno neupoštevanje samoglasniške kolicosti, za napako v zapisu, zato predlagamo transkripcijski popravek v široki kratki e.

Nepojasnjeno ostaja dejansko stanje v števanovskem govoru. Medtem ko se je po Zorkovi (1998: 105) issln. *-ə tu razvil v diftongični kratki ie, je ta porabski govor v Greenbergovem gradivu zabeležen kot sistem brez kolicosti (1993: 483, op. 17). Za potrditev razvoja bi potrebovali natančnejšo raziskavo z več zbranimi primeri.

4 Eksperimentalnofonetična analiza²⁶

Fonetika ali glasoslovje kot jezikoslovna veda preučuje zvočno stran jezika, tj. od glasu, naglasa, glasovnih zvez, besed, besednih zvez, stavkov in povedi, tudi do odstavkov in celotnega besedila. Preučuje jo s stališča izgovarjave (artikulacijska ali izgovorna fonetika) in slušnosti (akustična ali slušna fonetika), slednja seveda z organi človeškega telesa (ušesna fonetika) ali z različnimi orodji (eksperimentalna fonetika). Pri izrazni podobi jezika raziskuje fonetika poleg tvarnosti (načina tvorbe glasov z organi človeškega telesa ali sintetizirano) tudi razločevalno oblikovnost (s tem se ukvarja fonologija) (povzeto po Toporišič 1992: 42 in Toporišič 2000: 41). Eksperimentalnofonetična analiza katerega koli zvočnega posnetka vključuje poleg zaznavanja slušnosti z različnimi orodji obvezno tudi uporabo zaznavanja glasov z organi človeškega telesa, tj. z ušesom kot glavnim slušnim organom. Šele uporaba obojega namreč omogoča natančnejšo instrumentalno-slušno raziskavo bodisi kvalitete bodisi kvantitete posameznega glasu. Pri tem je izrednega pomena natančna določitev kriterijev razmejevanja glasov povezanega akustičnega govornega signala in konstantna uporaba istih kriterijev, saj sta le na ta način omogočena glasovna primerljivost ter pojasnjevanje vzrokov raziskovanim narečnim pojavom.

²⁴ Glej Logarjev fonološki opis govora na Gornjem Seniku (1981: 213–218) in primerjalno gornjeseniško gradivo.

²⁵ Glej iztočnico *gues* (Bajzek Lukač 2009: 144).

²⁶ Delno prirejeno po Zemljak (2004: 12–22).

Osnovne (segmentne) lastnosti samoglasnikov, ki jih določa položaj govornil v ustni votlini, in prozodične (nadsegmentne) lastnosti (trajanje, jakost in ton), lahko opazujemo s sledenjem sinusnega gibanja valovanja akustičnega signala iz ničelne točke na preseku osi x in y do naslednje ničelne točke, konec samoglasnika pa je običajno označen na mestu, kjer je še vidna zaključena celota sinusnega vala, in ne v zadnji točki tvorjenja glasu oz. fonacije, ki ima običajno obliko šuma. Drugi način opazovanja samoglasniškega segmenta je za natančnejšo zamejitev opazovan še s sledenjem značilnosti, vidnimi na spektrogramski sliki (formanti od 1 do 4 (F1–F4), pri čemer sta za opazovanje kvalitete pomembna predvsem F1 in F2 (Toporišič 2000: 49)). Najbolj smiselno se izkazuje časovnoporedno opazovanje in primerjanje obojega, sploh ko gre za problematiko prepoznavanja dvoglasniških posameznih delov – značilnosti, ki se kažejo v kvaliteti in/ali kvantiteti.

Za slovenski jezik je značilna povezanost samoglasnikov in njihove osnovne frekvence: čim višja je lega jezika, tem krajši je samoglasnik in tem višja je njegova frekvenca; čim nižja je lega jezika, tem daljši je samoglasnik in tem nižja je njegova frekvenca. Večina raziskovalcev slovenskega jezika ugotavlja daljšanje zadnjega zloga ali vsaj zadnjega glasu pred premorom, krajšanje trajanja glasov z večanjem števila glasov v zlogu, daljše trajanje glasov v odprtih zlogih, najkrajše trajanje nenaglašanih samoglasnikov v prednaglasni legi in najdaljše trajanje nenaglašanih samoglasnikov v govornem izglasju (oz. vsaj v zadnjem zlogu). Slednje naj bi bilo lahko celo daljše od trajanja naglašanih samoglasnikov. Jakostno izrazitejši samoglasniki so v primerjavi z manj izrazitimi običajno daljši.

Pri opazovanju kvantitete glasov se pojavlja vprašanje, kaj je pravzaprav dolgo in kaj kratko oz. kakšna je razlika med trenutnim in trajnim. Zanimivo je, da se jakostno močnejši in frekvenčno višji govorni signal zdita daljša. Ob večji jakosti (in višji frekvenci) govornega signala od tistega, ki omejuje jakostno šibkejšega (oz. z nižjo frekvenco), se zdi, da je slednji krajši. Z daljšanjem omejitvenega intervala daje omejeni interval vtis daljšega trajanja. Nekatere študije kažejo, da je za jasno prepoznavo določenega samoglasnika potrebno trajanje od vsaj 30 ms do 40 ms, je pa dejstvo, da so glasovi zaradi spontanega narečnega govora večinoma dosti krajši od studijsko posnetih izoliranih besed, zato pogosto tudi težje jasno prepoznavni, razen če so izgovorjeni skupaj z drugimi glasovi, torej vsaj v izgovorni enoti.

Razvoj tehnike omogoča uporabo čedalje boljših orodij, s katerimi je povezanost akustičnih enot laže vidna, njihova različnost pa bolj prepoznavna. Kljub temu pa osnovni problem določitve kriterijev razmejevanja glasov ostaja aktualen. Kriteriji so v veliki meri odvisni od načina in vrste analize akustičnega signala in niso razumljivi sami po sebi, zato jih je treba definirati vnaprej, pred željeno analizo. Raziskovalci (npr. Peterson & Lehiste (1960), Ladefoged (2001a, 2001b), Neweklowsky (1973), Srebot – Rejec (1988), Ozbič (1998), Bakran (1984), Gros (2000), Jurgec (2005, 2011), Tivadar (2009, 2017, 2019)) se lahko kljub uporabi iste instrumentalne podpore v kriterijih segmentiranja močno razlikujejo. Kljub temu so si enotni v prepričanju, da je za relevantnost dobljenih podatkov treba pri analizi govornega gradiva vseskozi uporabljati iste kriterije, saj je le na ta način omogočena primerljivost glasovnih značilnosti.

4.1 Primeri dejanskih fonetičnih vrednosti izbranih odrazov narečnega polglasnika

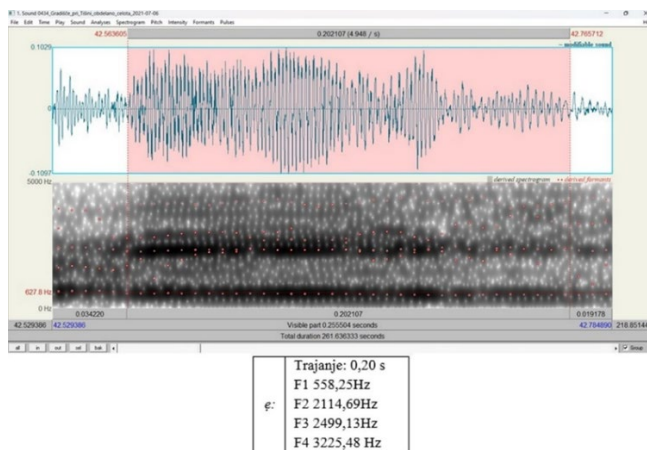
Narečno zvočno gradivo z leksemi iz prekmurskih krajevnih govorov, vključenih v mrežo krajev za SLA Dialektološke sekcije Inštituta za slovenski jezik Frana Ramovša (dostopnih v mapicah arhiv Dialektološke sekcije, arhiv Zinke Zorko, trakovi), smo s pomočjo orodja za prenos večjih datotek, *Dropboxa*, kot datoteke .wav prenesli na računalnik in jih kolikostno ter kakovostno analizirali z orodjem *Praat*, različica 6.3.10. Za potrebe raziskave smo v dialektološki literaturi in zvočnem gradivu najprej poiskali tiste fonetične zapise polglasnika, ki v transkripcijah izkazujejo znatna razhajanja, nanašajoč se tako na kvantiteto kot kvaliteto. Zaradi obsežnosti gradiva v prispevku prikazujemo po en izbrani primer analize kvalitete in kvantitete narečnega odraza polglasnika za izbrane točke posameznega prekmurskega podnarečja, ki bodisi predstavljajo točko po SLA bodisi so evidentirane na podlagi gradiva, ki je trenutno na voljo v digitalni obliki: (a) Gradišče pri Tišini (ni točka za SLA; arhiv Zinke Zorko (oznaka: 0434 Gradišče pri Tišini_obdelano_celota_2021_07_06; ravensko podnarečje)), (b) Gornji Petrovci (točka za SLA 399; arhiv Zinke Zorko (oznaka: 0021_T399 Gornji Petrovci_obdelano_2021_07_09; goričko podnarečje)) in (c) Gomilica (točka za SLA 392; trakovi (oznaka: T30 Gomilice v Prekmurju, Jože Horvat, 1967; dolinsko podnarečje)).

Ob prikazu zvočne datoteke v formatu .wav in poslušanju zvočnega gradiva smo analizirani samoglasnik najprej čim bolj natančno segmentirali, nato pa segmentu dodali še spektrogramski prikaz formantov (F1–F4) z njihovo oznako. Zaradi čim bolj enostavne in hitre, a vendar še vedno dovolj natančne analize smo uporabili avtomatski izračun formantov glede na osrednji del samoglasnika, ki formantnostrukturno predstavlja njegovo stabilno stanje.²⁷

4.1.1 Problematika odrazov za stalno dolgi polglasnik v besedi *vas*: eno- oz. dvoglasniškost in trajanje ter kakovost zapisa prvega dela dvoglasnika – ravensko podnarečje

V tabeli 1²⁸ so predstavljeni odrazi za issln. **ɤ* v besedah V700 *vas*, *dan*, *lan*, *laž* in *čast*. Za instrumentalno-slušno fonetično analizo sta bili izbrani problematika eno- oz. dvoglasniškosti odraza (enoglasnik *ɛ*: proti dvoglasniku *e:i/e:ɨ*) in trajanje ter kakovost zapisa prvega dela dvoglasnika (*e:i* proti *e:ɨ*) v besedi *vas*.

Iz zvočnega gradiva Gradišča pri Tišini smo primerjali dolgi ozki *ɛ*: v zadnjem oz. edinem besednem zlogu (v im. in tož.) in v nezadnjem besednem zlogu (v mestn.).

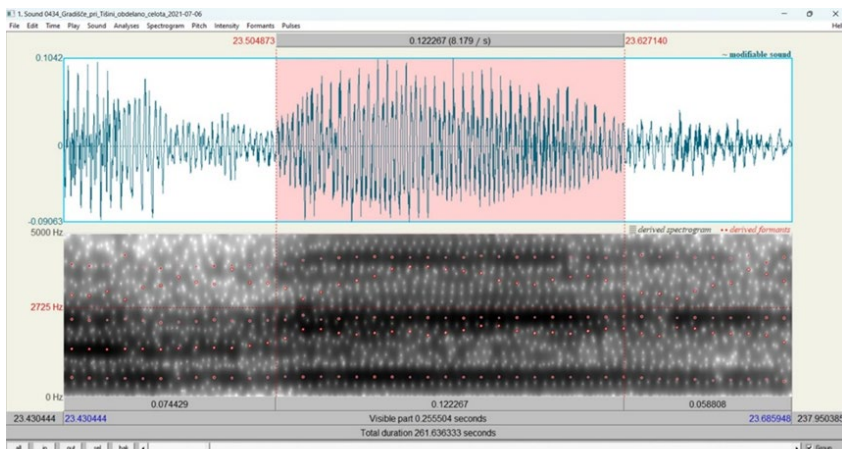


Slika 1: (*Na:ša*) '*vɛ:s* (zadnji oz. edini besedni zlog: im. (Gradišče pri Tišini))

Vir: lasten

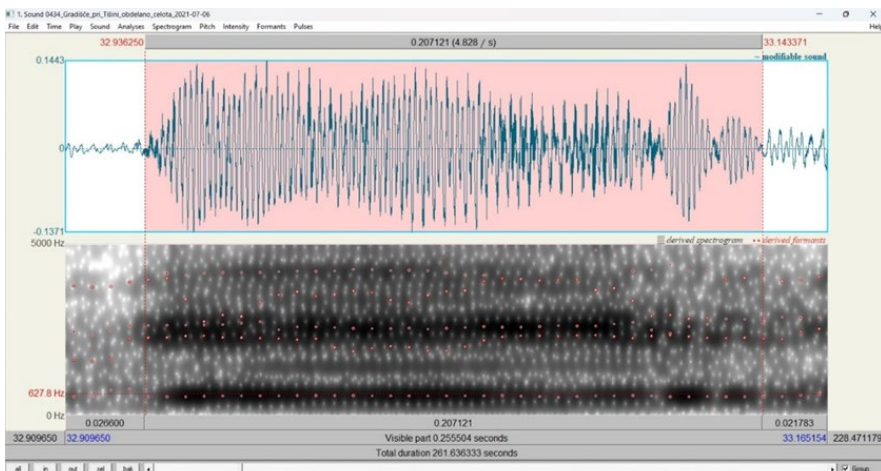
²⁷ Na vrednosti formantne strukture posameznih formantov vplivajo številni dejavniki (npr. spol, dialektološko ozadje vključno z (ne)tonemskostjo informanta, informantke, kot tudi glasovno okolje, (ne)naglašenost, položaj v besedi in stavku, stavčna (ne)poudarjenost), ki v začetni raziskavi zaradi vzpostavitve metodologije nadaljnega projektnega dela niso bili upoštevani. Predstavljeni primeri v prispevku so študija primera.

²⁸ Gl. poglavje Izsledki na podlagi narečnega gradiva.



Slika 2: (*'Na:šo*) *'və:s* (zadnji oz. edini besedni zlog; tož. (Gradišče pri Tišini))

Vir: lasten

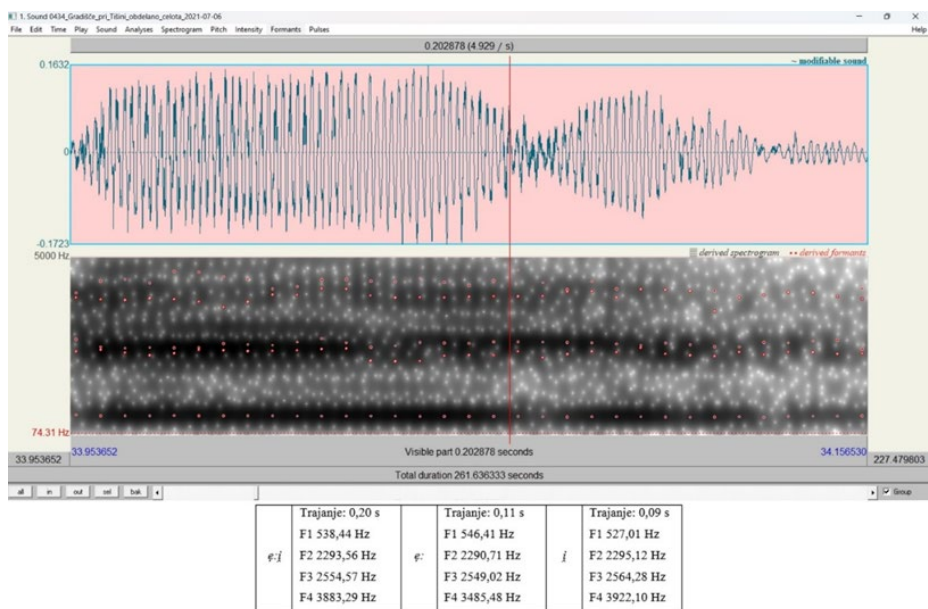


Slika 3: (*U'na:ši*) *'və:si* (nezadnji besedni zlog: mestn. (Gradišče pri Tišini))

Vir: lasten

Formantna struktura dolgega ozkega *e*: v imenovalniku, tožilniku in mestniku izkazuje njihovo primerljivost, še posebej F1 (med 554 Hz in 598 Hz) in F2 (1987 Hz do 2189 Hz), pri čemer sta najbolj primerljivi imenovalniška in tožilniška oblika, morda tudi zaradi krajše izgovarjave tožilniške oblike. Položaj samoglasnika v zadnjem ali nezadnjem (predzadnjem) besednem zlogu pri tem ni razlikovalen.

V vseh treh primerih se dolgi ozki *e*: nahaja v uvodnem delu povedi v drugi polnopomenski besedi besedne zveze, v mestniku obvezno predložne, zaradi česar je primerljivo trajanje imenovalnika in tožilnika (pridevnik + samostalnik), mestniško pa je zaradi nezadnjega besednega zloga in trodelnosti besedne zveze (predlog + pridevnik + samostalnik) predvidljivo krajše.



Slika 4: *Teče:ji* (*'po:tok*) (trajanje delov dvoglasnika in kvaliteta prvega dela dvoglasnika v zadnjem besednem zlogu (Gradišče pri Tišini))

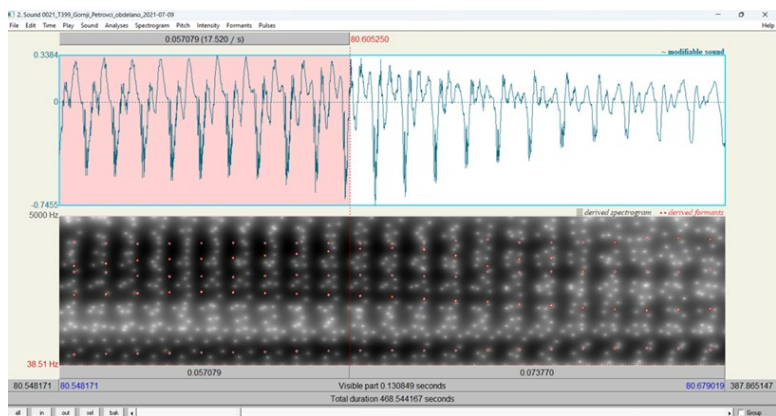
Vir: lasten

Kot problematična sta v gradivu ravenskega podnarečja izpostavljena tudi kakovost, ali gre torej za *e:i* ali *e:ĩ*, in kvantiteta prvega dela dvoglasnika.

Primerjava trajanja $e:i$ ($te'ʔe:i$) in $e:$ ($'ne:s$) v začetnem stavčnem izgovornem položaju potrjuje primerljivo trajanje fonemov ne glede na dvo- oz. enoglasniškost. Notranja razmejitev dvoglasnika na daljši začetni e (11 ms) in krajši jezični i (9 ms), kljub splošno znanim zadregam pri segmentaciji jezičnega dela dvoglasnika (gl. npr. Zemljak 2004: 20) in kljub za slušno zaznavanje nepomembni kvantitetni razliki, potrjuje način zapisovanja, torej z dvopičjem za prvim, vendarle daljšim delom fonema. Delitev na dva dela istega fonema potrjuje tudi formantna podoba obeh delov dvoglasnika, pri čemer je F1 dolgega naglašena i nižji od F1 dolgega naglašena e , F2 dolgega naglašena i pa višji od F2 dolgega naglašena e .²⁹

4.1.2 Problematika odrazov za novoakutirani polglasnik: kakovost in trajanje naglašena samoglasnika – goričko proti dolinsko podnarečje

V gradivu goriškega podnarečja se za novoakutirani polglasnik v zadnjem besednem zlogu pojavljata dve transkripciji, ki nakazujeta kakovostno razliko, in sicer običajno široki e ter zelo široki $ä$, v zapisu je prisotna tudi dvoglasniškost ($e'ü$, $i'ü$).



$i'ü$	Trajanje: 0,13 s	i	$e'ü$
		Trajanje: 0,06 s	Trajanje: 0,07 s
		F1 434,81 Hz	F1 538,40 Hz
		F2 2402,73 Hz	F2 2043,72 Hz
		F3 3012,57 Hz	F3 3085,90 Hz
		F4 3574,49 Hz	F4 3806,23 Hz

Slika 5: *Gnjäs* (trajanje delov dvoglasnika in kvaliteta drugega dela dvoglasnika v edinem besednem zlogu (Gornji Petrovci))

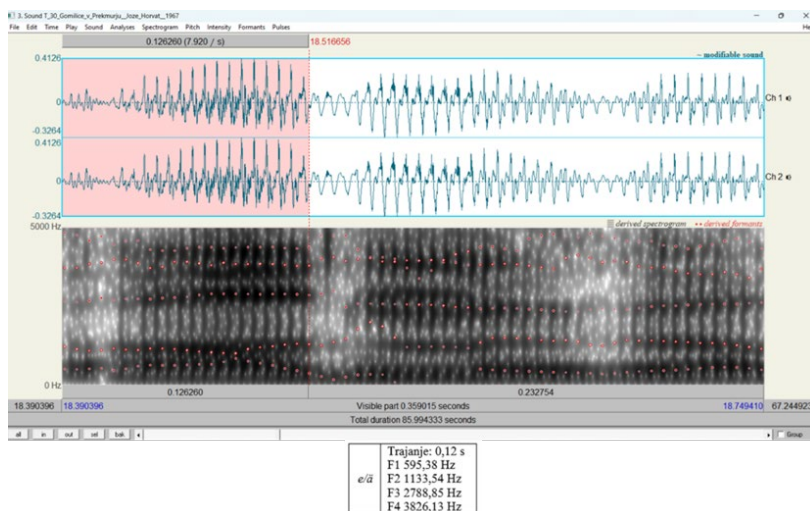
Vir: lasten

²⁹ Takšne vrednosti izpostavlja že Toporišič (1975: 153–196).

V zvočnem gradivu smo tako poiskali primer slišane dvoglasnika *iü* z jezičnim prehodom in zelo širokim *e*-jem (Gornji Petrovci) ter ga primerjali s formantno strukturo širokega in zelo širokega *e*. Zanimalo nas je tudi, ali je dvoglasnik kratek ali dolg in ali je zaznati pomembno kolikostno razliko med prvim in drugim delom dvoglasnika.

Izmerjena kvantiteta fonema v zadnjem oz. edinem zlogu v začetnem stavčnem izgovornem položaju potrjuje kvalitativno delitev na dva dela, tj. na *i* in *ä*, kvantitativno pa potrjuje tudi njegovo kračino (prim. npr. s trajanjem dvoglasnika *e:i* v ravenskem podnarečju). Trajanje posameznih delov dvoglasnika izkazuje njuno enakovrednost, kljub temu da je drugi del dvoglasnika tisti, ki definira tako instrumentalno izmerjeno kot tudi slušno zaznano kvaliteto novoakutiranega polglasnika v zadnjem besednem zlogu kot širok *e*-jvski glas (razvidno s slike 5).

Zbrano narečno gradivo izkazuje tudi nesistematičnost zapisa kvalitete širokega *e*, saj je ta narečno lahko izgovorjen kot *e* (tj. kot knjižni široki *e*) ali kot zelo široki *ä*, ki se v izgovarjavi lahko bolj ali manj približuje *a*. Tako smo za razpoznavo širine primerjali formantno strukturo jasno vidnega zelo širokega *e* v Gornjih Petrovcih goriškega podnarečja (*ghjäs*) s formantno strukturo naglašene *e* v Gomilici dolinskega podnarečja (*ednoga/ädnoga*), ki je v zbranem gradivu po slušnem vtisu zapisovan različno.



Slika 6: *'Ednoga/ädnoga* (kvaliteta vzglasnega samoglasnika (Gomilica))

Vir: lasten

Analiza je pokazala, da sta široka *e* glede na vrednosti F1 pravzaprav zelo primerljiva (*g'njäs*: F1 538 Hz; *'ednoga/'ädnoga*: F1 595 Hz), razliko pa izkazujeta vrednosti za F2 (*g'njäs*: F2 2042 Hz; *'ednoga/'ädnoga*: F2 1133 Hz). Pri tem je moč razbrati, da sta vrednosti F1 in F2 pri zelo širokem *e*-jevskem samoglasniku, kot je to v analiziranem primeru dolinskega podnarečja, bliže drug drugemu, kot to še posebej izrazito velja tudi za (knjižni) *a*, kot v analiziranem primeru goričkega podnarečja. Ustrezen zapis kvalitete naglašene *e* je tako v dolinskem podnarečju *'ädnoga*.

Glede na formantno strukturo goričkega *e* (*g'njäs*) pa se ravno zaradi vrednosti F2, ki je bliže tisti, ki jo lahko označujemo za vrednost knjižnega širokega *é*, zastavlja vprašanje, ali je zapis kot zelo široka fonemska različica pravzaprav ustrezen. Za razrešitev te problematike bo potrebna dodatna analiza več primerov izgovarjave.

5 Sklep

Zbiranje narečnega gradiva in njegovo zapisovanje, še posebej zaradi daljšega časovnega razpona in različno usposobljenih zapisovalcev, je v slovenskem prostoru prineslo številne izzive v težnji po poenoteni transkripciji glede na posamezna narečja, podnarečja, krajevne govore. Prispevek predstavlja poskus standardizacije prekmurske transkripcije samoglasnikov na podlagi zbranega narečnega gradiva, tj. mreže krajev za SLA Dialektološke sekcije Inštituta za slovenski jezik Frana Ramovša ZRC SAZU in novejših znanstvenih razprav ter diplomskih del. Študija primera primerjave odrazov izhodiščnoslovenskega stalno dolgega polglasnika, staroakutiranega polglasnika v nezadnjem besednem zlogu in staroakutiranega polglasnika v zadnjem besednem zlogu je pokazala na gradivo, za katero je potrebna dodatna akustična analiza zvočnih posnetkov. Opravljeni analizi (1) odrazov za stalno dolgi polglasnik v besedi *vas* glede na eno- oz. dvoglasniškost in trajanje ter kakovost zapisa prvega dela dvoglasnika v ravenskem podnarečju in (2) odrazov za novoakutirani polglasnik glede na kakovost in trajanje naglašene samoglasnika, primerjalno med goričkim in dolinskim podnarečjem, sta pokazali vso kompleksnost tovrstnih raziskav, hkrati pa njeno nadaljnjo pomembnost.

Literatura

Janja ADANIČ, 2009: Ledinska in hišna imena v občini Turnišče. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.

- Marija BAJZEK LUKAČ, 2009: *Slovar Gornjega Senika. A–L*. Maribor: Filozofska fakulteta, Mednarodna založba Oddelka za slovanske jezike in književnosti. (Zora, 66).
- Juraj BAKRAN, 1984: *Model vremenske organizacije hrvatskoga standardnog govora*. Zagreb: Sveučilište u Zagrebu.
- Francka BENEDIK, 1996: *Vodnik po zbirki narečnega gradiva za Slovenski lingvistični atlas (SLA)*. Ljubljana: Založba ZRC (ZRC SAZU).
- Aleks BIRSA JOGAN, 2022: Glasovne spremembe v krajevnem govoru Komna (SLA 107, OLA 5). *Jezikoslovni zapiski* 28/1, 111–122.
- Januška GOSTENIČNIK, 2022: Fonološki opis govora kraja Večeslavci (SLA T397). *Jezikoslovni zapiski* 29/2, 125–149.
- Marc GREENBERG, 1993: Glasoslovni opis treh prekmurskih govorov in komentar k zgodovinskemu glasoslovju in oblikoslovju prekmurskega narečja. *Slavistična revija* 41/4, 465–487.
- Jerneja GROS, 2000: *Samodejno tvorjenje govora iz besedil*. Ljubljana: Založba ZRC, ZRC SAZU.
- Marko JESENŠEK, 1991/92: Jezikovni sistemi v slovenskem (alpskem in panonskem) govornem območju. *Jezik in slovstvo* 37/7, 173–181.
- Peter JURGEC, 2005: Položaj v besedi in formantne frekvence samoglasnikov (standardne slovenščine): I. Naglašeni samoglasniki. *Jezikoslovni zapiski* 11/1, 87–95.
- Peter JURGEC, 2011: Slovenščina ima 9 samoglasnikov. *Slavistična revija* 59/3, 465–487.
- Karmen KENDA-JEŽ, 2011: Fonetična transkripcija. *Slovenski lingvistični atlas 1, Človek, (telo, bolezn, družina), 1. Atlas (SLA 1.1), 2. Komentarji (SLA 1.2)*. Ur. Jožica Škofic. Ljubljana: Založba ZRC, ZRC SAZU. 27–30.
- Mihaela KOLETNIK, 2008: *Panonsko lončarsko in kmetijsko izraze ter druge dialektološke razprave*. Maribor: Mednarodna založba Oddelka za slovanske jezike in književnosti, Filozofska fakulteta. (Zora, 60).
- Martina KOROŠA, 2005: Prekmurski govor v Moravskih Toplicah. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Mojca KUMIN HORVAT, 2022: Fonološki opis kraja Žetinci – Sichelendorf (SLA T414). *Jezikoslovni zapiski* 29/2. 151–168.
- Peter LADEFOGED, 2001a: *An Introduction to fieldwork and instrumental phonetics*. <https://linguistics.ucla.edu/people/ladefoge/> (21. 8. 2023).
- Peter LADEFOGED, 2001b: *Vowels and Consonants. An Introduction to the Sounds of Language*. Malden: Blacwell Publishers Inc.
- Tine LOGAR, 1996: *Dialektološke in jezikovnozgodovinske razprave*. Ur. Karmen Kenda Jež. Ljubljana: ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša.
- Tine LOGAR, 1981: Gornji Senik (Felsőszölnök; OLA 149). *Fonološki opisi srpskohrvatskih/ hrvatskosrpskih, slovenačkih i makedonskih govora obuhvaćenih opšteslovenskim lingvističkim atlasom*. Ur. Nedim Filipović. Sarajevo: Akademija nauka i umjetnosti Bosne i Hercegovine. 213–218.
- Tine LOGAR, 1993: *Slovenska narečja*. Ljubljana: Založba Mladinska knjiga.
- Jožica LUKAČ, 1997: Prekmurski govor vasi Bakovci. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Franček MUKIČ, 2005: *Porabsko-knjižnoslovensko-madžarski slovar*. Szombathely: Zveza Slovencev na Madžarskem.
- Darinka MULAOSMANOVIĆ, 2001: Govor vasi Bakovci. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Gerhard NEWEKLOWSKY, 1973: *Slovenische Akzentstudien. Akustische und Linguistische Untersuchungen am Material slowenischer Mundarten aus Kärnten mit 46 Abbildungen und 76 Figuren im Text*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.
- Martina OZBIČ, 1998: Akustična spektralna FFT analiza samoglasniškega sistema slovenskega jezika. Ljubljana: *Jezikovne tehnologije za slovenski jezik/ Mednarodna konferenca Informacijska družba – IS '98*, 55–59.

- Suzana PANKER, 1997: Govor vasi Kančevci. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Gordon E. PETERSON, Ilse LEHISTE, 1960: Duration of Syllable Nuclei in English. *The Journal of the Acoustical Society of America* 32/6, 693–703.
- Fran RAMOVŠ, 1936: *Kratka zgodovina slovenskega jezika*. Ljubljana: Akademski založba.
- Milena SALAJ, 2003: Poimenovanja za vrt, sadovnjak in polje v črenšovskem govoru. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za razredni pouk.
- Vera SMOLE, 2002: Kolikostna nasprotja v slovenskih narečjih. Med dialektologijo in zgodovino slovenskega jezika: ob življenjskem in strokovnem jubileju prof. dr. Martine Orožen. Ur. Marko Jesenšek, Bernard Rajh, Zinka Zorko. Maribor: Slavistično društvo.
- Marko SNOJ, Janoš JEŽOVNIK, 2001: Kolikostna nasprotja v knjižni izreki. 1. *slovenski pravorečni posvet*. Ur. Tanja Mirtič in Marko Snoj. Ljubljana: Slovenska akademija znanosti in umetnosti. (Razprave/Slovenska akademija znanosti in umetnosti, Razred za filološke in literarne vede = Dissertationes/Academia scientiarum et artium Slovenica, Classis II: Philologia et litterae; 25). 217–225.
- Jožica ŠKOFIC, Karmen KENDA-JEŽ, 2021: Tonemska in kolikostna nasprotja v slovenskih narečjih. 1. *slovenski pravorečni posvet*. Ur. Tanja Mirtič in Marko Snoj. Ljubljana: Slovenska akademija znanosti in umetnosti. (Razprave/Slovenska akademija znanosti in umetnosti, Razred za filološke in literarne vede = Dissertationes/Academia scientiarum et artium Slovenica, Classis II: Philologia et litterae; 25). 194–216.
- Renata ŠPILAK, 2002: Čevljarsko, kovaško in kmetijsko izraze v turniškem govoru. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Hotimir TIVADAR, 2009: Fonetika in realno gradivo – izbira(nje), pridobivanje, urejanje in analiza govornjenih medijskih besedil. *Infrastruktura slovenščine in slovenistike*. 1. natis. Ur. Marko Stabej. Ljubljana: Znanstvena založba Filozofske fakultete. Obdobja. 37–56. <https://centerslo.si/wp-content/uploads/2015/10/28-Tivadar.pdf> (21. 8. 2023).
- Hotimir TIVADAR, 2017: Speech rate in phonetic-phonological analysis of public speech (using the example of political and media speech). *Jazykovedný časopis* 68/1, 37–56. <https://www.degruyter.com/downloadpdf/j/jzcas.2017.68.issue-1/jzcas-2017-0016/jzcas-2017-0016.pdf>, DOI: 10.1515/jzcas-2017-0016 (21. 8. 2023).
- Hotimir TIVADAR, 2019: Opis samoglasnikov slovenskega knjižnega jezika z vidika spola govorečega. *Slavistična revija* 67/2, 233–242. <https://srl.si/ojs/srl/article/view/2019-2-1-11> (21. 8. 2023).
- Jože TOPORIŠIČ, 1975: Formanti slovenskega knjižnega jezika. *Slavistična revija*, 23/2, 153–196.
- Jože TOPORIŠIČ, 1992: *Enciklopedija slovenskega jezika*. Ljubljana: Cankarjeva založba.
- Jože TOPORIŠIČ, 2000: *Slovenska slovnica*. Maribor: Založba Obzorja.
- Alojz TUŠKEJ, 2000: Besedje v tišinskem govoru. *Diplomsko delo*. Maribor: Univerza v Mariboru, Pedagoška fakulteta, Oddelek za slovanske jezike in književnosti.
- Peter WEISS, Andrejka ŽEJN, 2004: Digitalizacija pisnega narečnega gradiva v dialektološki sekciji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU v Ljubljani. *Jezikoslovni zapiski* 10/2, 179–188.
- Melita Zemljak Jontes, 2004: *Trajanje glasov štajerskega zabukovskega govora. Instrumentalno-slušna analiza*. Maribor: Slavistično društvo Maribor. (Zora, 30).
- Zinka ZORKO, 1998: *Haloško narečje in druge dialektološke študije*. Maribor: Slavistično društvo. (Zora, 4).
- Zinka ZORKO, 1998a: Narečne prvine v Ivanocyjevem Kalendarju Najsvetjesege srzca Jezusovega (1904–1906). *Slovensko-madžarski jezikovni in književni stiki od Košiča do danes*. Ur. István Nyomárkay, Stjepan Lukač. Budimpešta: Košičev sklad. 39–59.
- Zinka ZORKO, 2003: Oblikoslovje in leksika v govoru Cankove. *Avugst Pavel*. Ur. Zinka Zorko, Miha Pauko. Maribor: Slavistično društvo. 73–94. (Zora, 23).

- Zinka ZORKO, 2005: Slovarju na pot. *Porabsko-knjižnoslovensko-madžarski slovar*. Szombathely: Zveza Slovencev na Madžarskem. I–II.
- Zinka ZORKO, 2008: Prekmursko ravensko podnarečje na Cankovi. *Življenje in delo Jožefa Borovnjaka*. Ur. Marko Jesenšek. Maribor: Filozofska fakulteta, Oddelek za slovanske jezike in književnosti. 257–269. (Zora, 55).

TRANSKRIBIRANJE V SOCIOLINGVISTIČNIH RAZISKAVAH IN KORPUSIH GOVORJENE SLOVENŠČINE

MAJA BITENC

Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
Maja.Bitenc@ff.uni-lj.si

Prispevek obravnava transkribiranje kot kompleksen, interpretativen in selektiven proces, pri katerem pretvarjanje iz govornega v zapisano besedilo odpira številna temeljna in praktična vprašanja. V sociolingvističnih raziskavah ni standardnih transkripcijskih načel, temveč so ta vedno odvisna od teoretske usmeritve raziskovalca, predmeta in namena raziskave. Uvodni del predstavlja splošne teoretske pomisleke in načelne smernice, osrednji del pa se osredotoča na različne prakse transkribiranja v korpusih in raziskavah govorne slovenščine. Posebna pozornost je posvečena problematiki zapisovanja variant fonema /v/, polglasnika, specifičnih narečnih glasov in premen po zvonečnosti. Med drugim so opisane različne prakse transkribiranja v avtoričinih sociolingvističnih raziskavah variantnosti govorne slovenščine in relevantnih socialnopsiholoških tem, od začetnih natančnih fonetičnih transkripcij do različnih stopenj poenostavljanja in prehajanja k ortografskemu zapisu s posebnimi znaki za posamezne glasove. Prispevek izpostavlja tudi izzive glede razločevanja posameznih glasov in nedoslednosti pri njihovem zapisovanju.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.8](https://doi.org/10.18690/um.ff.4.2024.8)

ISBN
978-961-286-882-6

Ključne besede:
transkribiranje,
fonetična in ortografska
načela,
sociolingvistika,
jezikovna variantnost,
fonem /v/,
polglasnik



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.8](https://doi.org/10.18690/um.ff.4.2024.8)

ISBN
978-961-286-882-6

Keywords:
transcription,
phonetic and orthographic
conventions,
sociolinguistics,
language variation,
phoneme /v/,
semi-vowel

TRANSCRIPTION IN SOCIOLINGUISTIC RESEARCH AND CORPORA OF SPOKEN SLOVENE

MAJA BITENC

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
Maja.Bitenc@ff.uni-lj.si

The article discusses transcription as a complex, interpretive, and selective process in which the transformation of spoken into written text raises numerous fundamental and practical questions. There are no standard transcription principles in sociolinguistic research; they always depend on the researcher's theoretical orientation, subject matter, and research purpose. The introductory part of the article presents theoretical considerations and general guidelines, while the central part focuses on various transcription practices in corpora and studies of spoken Slovene. Particular attention is paid to the notation of the variants of the /v/ phoneme, the semi-vowel, specific dialectal sounds and changes in sonority. Among other things, various transcription practices in the author's sociolinguistic studies of variation in spoken Slovene and relevant socio-psychological topics are described, ranging from initial precise phonetic transcriptions through various levels of simplification to the transition to orthographic representation with special symbols for individual sounds. The article also highlights challenges in distinguishing individual sounds and inconsistencies in their representation.



1 Uvod¹

Transkribiranje oz. pretvarjanje iz govornega v zapisano besedilo je ključen del obdelave gradiva pri vsaki raziskavi govornega jezika, glede katerega je treba odgovoriti na številna temeljna in praktična vprašanja. Pomembno je, da so odločitve, ki jih sprejme raziskovalec, konsistentne, zapisovanje pa natančno, saj transkripcije predstavljajo osnovo tako za nadaljnjo analizo kot interpretacijo podatkov (Tagliamonte 2007; Kvale in Brinkmann 2009: 177–187). Analiza je lahko osnovana tudi na prvotnih virih, torej avdio posnetkih, h katerim se ob transkripcijah vračamo med celotnim analitičnim procesom; v tem primeru da so transkripcije predvsem način predstavitve podatkov bralcem (Lampropoulou 2012: 76–77).



Slika 1 Shema o izgubi informacij pri pretvorbi žive pripovedi v zapis
(graf Barbara Ivančič Kutin, ilustracije Ciril Horjak)

Vir: Ivančič Kutin 2011: 64

¹ Prispevek je nastal v okviru projekta Sociolingvistična variantnost govorne slovenščine na primeru mobilnih govorcev: Viri, metode in analiza (Z6-9371). Avtorica se Javni agenciji za raziskovalno dejavnost Republike Slovenije zahvaljuje za financiranje, Darinki Verdonik, Heleni Dobrovoljc, Karmen Kenda-Jež, Luku Horjaku in Marku Stabeju pa za razprave glede transkripcijskih načel.

Zavedati se je treba, da gre pri transkripciji vedno za interpretacijo, saj nikoli ne more biti povsem zanesljiva predstavitev originalne interakcije. Že z avdio snemanjem se izgubi vrsta informacij, kot so telesna govorica, mimika in gestikulacija, ob zapisu govora pa je nemogoče v celoti opisati ton glasu, intonacijo, dihanje ipd. Obenem je vsaka transkripcija vzeta iz konteksta, saj ne vključuje časovnih in prostorskih dimenzij žive komunikacije (Kvale in Brinkmann 2009: 177–178). Slika 1 nazorno prikazuje izgubo informacij pri pretvorbi žive pripovedi v zapis.

Dejstvo je torej, da pri pretvarjanju besedila iz posnetka v zapis ni mogoče vsega istočasno, izčrpno in neprekinjeno zabeležiti in da se mora raziskovalec vedno odločiti, koliko in katere vrste informacij zapisovati (Brinker in Sager 1996: 35). Ne obstajajo standardna transkripcijska načela, pač pa so ta odvisna od teoretske usmeritve raziskovalca, predmeta in namena raziskave, torej kaj in zakaj se proučuje. Selektivnost je tako tesno povezana s subjektivnostjo. Pomembna dejavnika pri odločanju glede natančnosti zapisa sta tudi čas in finančna sredstva, ki jih imamo na razpolago, saj je transkribiranje zelo zamudno. Okvirna ocena je, da so za prvo transkripcijo enournega posnetka potrebne štiri ure, vendar je to zelo odvisno od kakovosti zvoka, števila govorcev in transkriptorjevega poznavanja varietet(e), ki jih (jo) zapisuje; tako lahko enourni intervju terja delo od enega dneva do enega tedna ali celo več.²

V sociolingvističnih raziskavah se večinoma uporabljajo standardna ortografija in ločila, saj je to ključno za karseda enostavno transkribiranje in berljivost transkripcije. Načeloma se torej besede zapisuje tako, kot so običajno zapisane, razen če obstaja upravičen razlog za drugačno odločitev. Navadno se uporablja tudi standardna ločila, torej pike, vprašaje, vejice itd., kot običajno pri pisanju. Obenem je pomembno, da se transkribira, kar je oseba dejansko rekla, čeprav to ni v skladu s predpisanimi slovničnimi pravili standardnega jezika (Preston 2000; Tagliamonte 2006: 55).

Osnovni vodili pri transkripcijah za potrebe variantnostne analize sta, da naj bodo transkripcije dovolj natančne, da učinkovito ohranijo dovolj podatkov za jezikoslovno analizo, in da naj bodo dovolj preproste, da jih je mogoče sorazmerno lahko brati in zapisovati. Ob upoštevanju časovnih in finančnih omejitev je

² Na zamudnost transkribiranja opozarja veliko avtorjev; Jana Zemljarič Miklavčič (2008: 180–181) npr. piše, da je za transkripcijo dveh minut posnetka povprečno porabila eno uro, pri posnetkih spontanega multiloga pa pogosto ni dosegla niti transkripcijske hitrosti ene minute posnetka v eni uri, pa čeprav je šlo za ortografsko transkripcijo. V tuji raziskavi staranja je izkušena tajnica za transkribiranje ene ure posnetka intervjuja potrebovala približno 5 ur (Kvale in Brinkmann 2009: 180).

ohranjanje ravnotežja med tema dvema ciljema ključno pri oblikovanju korpusa (Tagliamonte 2006: 54). Zapisujemo torej tisto, kar potrebujemo in kar je relevantno za namen sedanje oz. morebitne kasnejše raziskave (Wray in Bloomer 2006: 185).

Glavni izziv pri izdelavi korpusa je zagotoviti karseda avtentično, natančno in dosledno predstavitev posnetkov govornega jezika. Za enotno zapisovanje je potreben transkripcijski protokol, ki kot referenčni dokument za transkripcijsko prakso zagotavlja doslednost znotraj korpusa (Poplack 1989: 434; Tagliamonte 2006: 55).

Za urejanje posnetkov, transkribiranje in obdelavo transkripcij so na voljo različni računalniški programi in orodja. Pogosto uporabljana in prosto dostopna orodja za urejanje posnetkov so Transcriber, WinPitch in Exmaralda, od katerih ima vsak svoje posebnosti. Poslušanje in sorazmerno enostavno urejanje posnetkov omogočata npr. programa Adobe Audition ali brezplačen Audacity. Za poslušanje in natančnejšo analizo krajših izsekov in posameznih glasov je na voljo program Praat. Gradnjo lastnega korpusa s številnimi funkcijami omogoča program Sketch Engine, za korpusno obdelavo transkripcij sta uporabna Notepad++ in PSpad.

V slovenski praksi, tako v dialektologiji kot drugih raziskavah, je še vedno pogosto zapisovanje v programu Word, ki omogoča zapis narečnih glasov z znaki slovenske narečne fonetične oz. Logarjeve transkripcije v vnašalnem sistemu ZRCola, posebej če ni potrebno označevanje prekrivnega govora, obenem pa prav pridejo različne možnosti za oblikovanje besedila in zapisovanje komentarjev. Pri korpusu GOS³ in projektu RSDO⁴ so za transkribiranje uporabili program Transcriber.

2 Transkribiranje v korpusih in raziskavah govornje slovenščine

V okviru dialektologije v slovenskem prostoru je za zapis narečnih govornih besedil od 70. let prejšnjega stoletja v rabi t. i. nova nacionalna oz. Logarjeva transkripcija (Logar 1993: 101–103; Kenda-Jež 1996: VII–XIV in 2011: 27–30). Font z naborom transkripcijskih znakov različnih fonetičnih transkripcij iz slovenske in slovanske zapisovalske tradicije ter mednarodne fonetične abecede IPA (ang.

³ <http://www.korpus-gos.net/Support/About>, dostop 23. 6. 2023.

⁴ <https://rsdo.slovenscina.eu/>, dostop 23. 6. 2023.

International Phonetic Alphabet) je skupaj z vnašalnim sistemom ZRCola⁵ na ZRC SAZU v Ljubljani razvil Peter Weiss. Glede načina zaznamovanja prozodičnih in paralingvističnih prvin ter glasovnih refleksov v dialektoloških študijah prim. Weiss 1990: 181; Kenda-Jež 2002: 166–167.

Pri korpusu GOS so se odločili za t. i. pogovorni zapis, katerega cilj je zvesta predstavitev glasovne podobe govora v karseda berljivi obliki. Uporabljali so veljavni slovenski črkopis in upoštevali odstopa od pravopisne norme, kjer izgovor odstopa od standardne izreke. Transkripcijska načela sta določili Darinka Verdonik in Ana Zwitter Vitez (2011: 58–68).

Tudi pri projektu Razvoj slovenščine v digitalnem okolju (RSDO) govor zapisujejo v skladu s knjižno normo, če izgovor ne odstopa bistveno od zbornega, le polglasnik vedno, tudi v primeru zborne izreke, pišejo s posebnim znakom – @. Pri zapisu besed, ki v izreki odstopajo od zborne ali niso del knjižne leksike, sledijo smernicam, ki sta jih opredelili Darinka Verdonik in Andrejka Bizjak (2022: 25–26).

Za namen proučevanja oblikovanja neknjižnega pogovornega oz. pogovarjalnega jezika na primeru govork iz Mengša in Beltincev, ki obiskujejo srednjo šolo v Ljubljani, je Jožica Škofic posnetke šolskega spraševanja, obnov besedil, medsebojnih pogovorov in intervjujev o jezikovni zavesti uporabljala fonetično transkripcijo po zgledu zapisov v *Slovenskih narečjih* Tineta Logarja (1975 oz. 1993), ni pa označevala tonemov v gorenjskem narečju, saj bi, kot piše, glede na njeno poznavanje lahko prišlo do napak (Škofic 1991: 225–227, transkripcije na str. 227–377).

Pri zapisovanju spontano govornih besedil ljubljanskih govorcev za analizo besedilnih vrst in zgradbe ter sociolingvističnih intervjujev za raziskavo variantnosti govorne slovenščine na primeru šestih mobilnih mladih Korošcev sta se Mojca Smolej in Katja Kotnik odločili za knjižno oz. ortografsko transkripcijo z opredeljenimi izjemami, ki niso skladne s standardnim slovarjem (Smolej 2012: 14–15, zapis besedil na str. 145–236; Kotnik 2022: 48).

⁵ <http://zrcola.zrc-sazu.si/>, dostop 23. 6. 2023.

V raziskavi variantnosti govorne slovenščine pri petih mobilnih govornicah z Idrijskega, ki se šolajo ali delajo v Ljubljani, sem Maja Bitenc (2016a, 2016b) oblikovala več različnih transkripcijskih protokolov. V začetni fazi sem se predvsem zaradi posebnosti in neraziskanosti slovenskih medzvrstnih razlik in razmerij odločila za natančno slovensko narečno fonetično transkripcijo. Čeprav sem že takrat okvirno določila nabor variabel, sem želela preveriti, ali so morda za variantnostno analizo relevantne še druge. V pilotnem poskusu se je izkazalo, da so izbrane variable ustrezne, da je na različnih jezikovnih ravneh sicer še nekaj drugih, vrednih pozornosti, da pa bi bilo transkripcijo vendarle smiselno poenostaviti in pospešiti, saj je za zanesljivost rezultatov potrebna zadostna količina podatkov. V doktorski raziskavi sem tako izbrane izseke iz celodnevnega samosnemanja prvih dveh informantk zapisala v ozki fonetični transkripciji s pisavo ZRCola v skladu z določenimi načeli, pri preostalih treh informantih pa sem se odločila za ortografsko transkripcijo z nekaterimi posebnimi znaki. Pri transkribiranju sociolingvističnih intervjujev so bila transkripcijska načela še nekoliko poenostavljena (Bitenc 2016b: 182–183, transkripcije izsekov iz celodnevnega samosnemanja na 72 straneh so v prilogi 8.4.1, transkripcije relevantnih delov intervjujev pa na 74 straneh v prilogi 8.4.4 na CD-ju v Bitenc 2016a).

Več je torej primerov uporabe ortografske transkripcije z različnimi načeli zapisovanja posameznih jezikovnih elementov, večinoma so tudi ločila zapisana v skladu s standardom. Pri J. Škofic (1991) in M. Bitenc (2016a, 2016b) ni vejic in pik, le vprašaji pri vprašalni intonaciji, pri J. Škofic (1991.) tudi klicaji pri vzklični intonaciji, premori pa so označeni s pikami, od ene do treh pri M. Bitenc (n. d.) in tudi več pri J. Škofic (n. d.), odvisno od trajanja. M. Bitenc (n. d.) sem velike začetnice uporabljala le pri lastnih imenih. Pri K. Kotnik (2022) in M. Bitenc (n. d.) je označeno mesto naglasa, če je nenavadno ali razlikovalno.

3 Transkripcijska načela za specifične pojave govornega jezika

V nadaljevanju izpostavljam primere, glede katerih se pri odločanju za transkripcijska načela znotraj ortografske transkripcije pojavlja največ dilem, predvsem so to načela zapisovanja redukcije, variant fonema /v/, polglasnika, specifičnih narečnih glasov in premen po zvonečnosti.

3.1 Redukcije

V vseh omenjenih korpusih in raziskavah so upoštevane redukcije – glasov, ki niso izgovorjeni, torej raziskovalci oz. transkriptorji ne zapisujejo. Pri RSDO oz. D. Verdonik in A. Bizjak (2022: 25) so navedeni primeri *tud, neki, tko, mam, čevli*; pri M. Smolej (2012: 14) po alinejah glede na obliko in reduciran glas *mlek, okn, mel, igral, spal, dogajal, spal, pripravt, prijatu, drugač, men, clo, zlo, tko*.

3.2 Variante fonema /v/

Slovenski pravopis v poglavju o glasoslovju (Toporišič idr. 2003: 73) glede variant fonema /v/ navaja, da ustničnoustnični šumni *w* (zveneči) govorimo pred zvenečim soglasnikom in na koncu za zvočnikom, *m* (nezveneči) pa pred nezvenečim soglasnikom: *vnuke, barv, odvzeti – vsake, v tebi, predvsem*. Med posebnostmi je omenjeno, da namesto *w* in *m* zlasti pri počasnem govorjenju govorimo tudi *u*, da se pri besedah, ki se začenjajo s pisanim predponskim *u*-, ta izgovarja kot *u*, *w* oz. *m*; pred samoglasniki pa se predlog *v*, če se ne veže na končni samoglasnik prejšnje besede, izgovarja kot *w* ali *u*, odvisno od hitrosti govora.

V pravopisnih pravilih *Pravopis 8.0* in slovarju *ePravopis* je navedeno, da fonem /v/ izgovarjamo zobnoustnično ali dvoustnično: **zobnoustnično** pred samoglasniki iste besede in v nekaterih zvočniških sklopih; **dvoustnično** v dvofonemskih zvezah oz. fonetičnih dvoglasnikih (za samoglasnikom pred soglasniki in v izglasju) ter v vzglasju pred soglasniki in na morfemski meji (Glasoslovni oris, odstavek 39, ostala določila v zvezi s fonemom /v/ so v odstavkih do 45⁶). Po novem predlogu vse tri dvoustnične variante zapisujemo v oglatem oklepaju s črko *u*, po sistemu IPA pa bi bile zapisane s črko *w*. Če je dvoustnična varianta v izglasju, jo zapisujemo tudi s črko *v* in vezajem: očetov [očetov-]. Poenostavitev je pojasnjena s sledečim: »Razlik med tradicionalno **zapisanimi dvoustničnimi variantami fonema /v/** ([w], [ɱ], [u]) večina govorcev slušno ne zaznava, zato so jih jezikoslovci v kodifikacijskih in slovničnih opisih razvrščali in poimenovali glede na glasovno bližino kot **zvenečo** [w], **nezvenečo** [ɱ] in **dvoglasniško** varianto [u] (za samoglasniki). Dvoustnični izgovor se po slušnem vtisu približuje ujevskemu izgovoru, razlikujemo pa ju po zložnosti« (odstavek 40 v navedenem poglavju).

⁶ <https://www.fran.si/pravopis8/Podpoglavje/7-2>, dostop 23. 6. 2023.

V korpusu GOS zvočnik dvoustnični *v*, ki ni nosilec zloga, zapisujejo kot »v« (*prov, nav, navm, odpravi, dawn, pršov* – deležnik stanja na -l z glasovno premeno) oz. z »k«, če tako izhaja iz knjižne norme (*kosil, mel*). Če je u samoglasniški, tj. nosilec zloga, ga pišejo kot »u« (*pršu, vidu, u tem delu*) (Verdonik in Zwitter Vitez 2011: 59). Tudi v projektu RSDO zvočnik dvoustnični *v*, ki ni nosilec zloga, zapisujejo s črko »v«, če se pojavi v besednih oblikah, ki niso knjižne: *prov, nav, navm, odpravi, dawn, gledave, pov@n*. Posebej opozarjajo na primere *lavfati, slavf, genav, mav (malo), šov (šel), dov (dol), prov (prav), dav (da bo), nov (ne bo)*, tudi medmet *av*. V poglavju o redukcijah navajajo tudi premene oblik za prihodnjik, izpostavljen je zapis primerov *čev (če bo), navm (ne bom), nav (ne bo)*. Če dvoustnični *v* nastopa v besedni obliki, ki je knjižna in tudi izgovorjena skladno s standardom, ohranijo knjižni zapis: *bil, gledal, siv*. Če je glas u samoglasniški, tj. je nosilec zloga, pa ga pišejo s črko »u«: *pršu, vidu, u tem delu*. Tudi predlog *v*, izgovorjen kot samoglasniški *u*, pišejo kot *u* (Verdonik in Bizjak 2022: 26).

M. Smolej (2012: 14) med izjemami od standardnega zapisa navaja, da deležnik na -l za moško obliko ednine zapisuje po izgovorjavi z *u*, po drugi strani pa med nestandardnimi oblikami, ki jih zapisuje skladno s standardnim zapisom, navaja predlog *v*, ki ga zapisuje dosledno s črko *v* tudi v primerih, ko se izgovarja kot *u*. Pri K. Kotnik (2022) so deležniki na -l zapisani z -*u*, izgovor zvočnika *v* je zaradi poenostavitve zapisa in nerelevantnosti razlikovanja za namen raziskave ohranjen kot *v*, le pri (narečnem) dvoustničnem izgovoru je zapisala *w*. Pri J. Škofic (1991) v transkripcijah najdemo zapis *u* in predvsem v vlogi predloga *v*, čeprav je verjetno izgovorjen dvoustnično, npr. *V šo:li pa 'use govo'rim*.

M. Bitenc (2016a, 2016b) v fonetični transkripciji uporabljam *v*, *u* in *w* za narečni dvoustnični glas v položaju pred samoglasniki, ki je tudi varianta proučevane variable, v ortografski transkripciji z nekaterimi posebnimi znaki za variantnostno analizo treh informantov in v dodatno poenostavljeni transkripciji za sociolingvistične intervjuje pa *u* tako za *u* kot dvoustnični *u*, ki se izgovarja pred soglasnikom in na koncu besede, ter *w* za narečni dvoustnični glas pred samoglasniki. Zapisovanje dvoustničnih glasov z *u*-jem je bilo deležno kritike, saj gre za soglasniški glas, zapisan s samoglasniškim. Tako sem se pri podoktorskem projektu pri zapisu dvoustničnih variant fonema /v/ odločala med dosedanjim *u*, ki se kljub kritičnim pomislekom zdi izgovoru in splošnemu slušnemu vtisu najbližji ter se tudi v zapisu pogovornega jezika v spletnih in telefonskih sporočilih kaže kot najpogostejša

izbira,⁷ *v* in *l* v skladu s standardnim zapisom, *u*, kot je v predlogu novega Pravopisa 8.0, in *m*, kot je v sistemu IPA.

Po eni stani bi bilo najbolj korektno za poseben glas uporabljati poseben znak, vendar pa se zdi, da je, če ta glas ni predmet analize in mu ne posvečamo posebne pozornosti, lažje zapisovati in brati *v* in *l*. Tudi pri pretvorbi v poknjžžen zapis za vsebinsko analizo bi zapis s posebnim znakom terjal več sprememb. Posebej delikatni so primeri, ki se lahko izgovorijo z *l* ali *u*, kot npr. v besedi *mal* (npr. *guvurim maŭ dargac̣*). Da bi ohranili razlikovanje med obema načinoma, je v primeru izgovorjave *maŭ* vsekakor na mestu zapis z *v*, kar je skladno tudi s smernicami pri projektu RSDO, saj gre za nekñjižno pozicijo dvoustničnega *v*. Dilema se pojavlja tudi pri zapisovanju nekaterih deležnikov na *-l*. Pri pogovornih oblikah pride do primerov, kot sta npr. *nadzorovou* ali *šov*. Čeprav se po navodilih deležniki načeloma pišejo z *-l*, je pri pogovornih oblikah predviden zapis z *v*, ki se zdi manj nenavaden, torej *nadzorovov* ali *šov*.

3.3 Polglasnik

V pravilih *Pravopis 8.0* in slovarju *ePravopis* pri zapisu izgovora v oglatih oklepajih za polglasnik uporabljajo znak iz IPE, in sicer obrnjeno črko e tj. ⟨ə⟩, kot je navedeno v odstavku {4} poglavja Slovnični oris za pravopis. Zapis polglasnika je sicer opredeljen v odstavkih 76–96 poglavja Glasoslovni oris,⁸ v Slovenskem pravopisu pa na kratko v odstavku 620 s povezavami na druge relevantne odstavke (Toporišič 2003: 70).

V korpusu govornjene slovenščine GOS polglasnik ni zapisan pri zvočnikih l, r, m, n (*złoml*, *pr*, *misl**m*, *sn*); enoglasovnih predlogih, členkih, čeprav so izgovorjeni s polglasnikom (*s*, *z*, *d*), enozložnih besedah (*nč*, *jž*, *jsi*); lahko je zapisan z e v dvo- ali večzložnih besedah (*kešm*), razen pred l, r, m, n (Verdonik in Zwitter Vitez 2011: 58). V projektu RSDO polglasnik vedno zapisujejo z znakom @: pri zvočnikih l, r, m, n: *s@n*, *p@r*, *mis@l@m*, *bit@r*, *złom@l*, *prjat@lci*; pri enoglasovnih predlogih, členkih ipd., izgovorjenih vokalizirano, s polglasnikom: *s@*, *z@*, *d@*; pri enozložnih

⁷ Pomenljiv je tudi komentar sina, ki je ob prvih vajah branja negodoval nad zapisanim *l*-jem, ki se ne izgovarja kot *l*: »To je mal zoprno, ta glupi *l*. Zakaj se ne piše *u*?«

⁸ https://fran.si/pravopis8/Poglavje/7/slovnici_oris in <https://www.fran.si/pravopis8/Podpoglavje/7-2>, dostop 23. 6. 2023.

besedah: *j@ɤ, n@ɛ*; v dvo- ali večzložnih besedah: *k@šni (kakešni)*; v zborni izreki: *b@z@g, p@s* (Verdonik in Bizjak 2022: 25)

Pri J. Škofic (1991), K. Kotnik (2022) in M. Bitenc (2016a, 2016b) v vseh treh oblikah transkripcije je polglasnik zapisan kot *ə*. M. Smolej (2012) polglasnika ne zapisuje, med nestandardnimi oblikami, ki jih zapisuje skladno s standardnim zapisom, pa omenja kratke samoglasnike, ki so v izgovoru reducirani v polglasnik in pri katerih ohranja zapis npr. s črko *a* oz. *i*, npr. *jaɤ, brət, kear, niš* (Smolej 2012: 14).

3.4 Specifični narečni glasovi

V korpusu GOS so pokrajinsko specifični glasovi zapisani z najbližjimi ustreznimi črkami – *ej* (rejs – res), *aj* (lajtus – letos), *uj* (tujdi – tudi), *ov* (prov – prav), *av* (tav – to), *uo* (puole – pol), *je* (rjekla – rekla), *ju* (tjuk – toliko), *ue* (zmuetlu – zmotilo), *ea* (nea – ne), *ua* (uaknu – okno), *uo* (duobru – dobro); *u* ali *i* tudi za *u* s preglasom, *b* ali *g* za zveneči primorski *b*, *r* za mehkonobni koroški *r* (Verdonik in Zwitter Vitez 2011: 59). Tudi v projektu RSDO diftonge in druge pokrajinsko specifične foneme, ki jih ni v knjižnem jeziku, pišejo z najbližjimi ustreznimi črkami, odvisno tudi od izgovorjave v konkretnih primerih: v navodilih navajajo *ej, ov, je*; »u« za *u* s preglasom; »h« ali »\$g« za zveneči primorski *h*; »r« ali »\$r« za mehkonobni koroški *r* (Verdonik in Bizjak 2022: 26).

M. Smolej (2012: 14) med izjemami, ki niso skladne s standardnim slovarjem, navaja zapisovanje ukanja v prednaglasnem zlogu (npr. *utrok, pupadki, purudnišnica*). K. Kotnik (2022) za zapis narečnih širokih samoglasnikov uporablja strešico (*ožéroma, kéda, dóma, Primórske*), za oznako narečnih naglasov tudi krativec in ostrivec (*zriðčér, òtrok, dèrvač, dóma, babíca, v Máriaibórn, óbadvá*).

Tudi v *Pravopisu 8.0* in slovarju *ePravopis* so pri zapisu v oglatih oklepajih naglasna znamenja na črkah *e* in *o* uporabljena tudi za označevanje kakovosti, tj. širine oziroma ožine; za široka *e* in *o* torej ⟨ê⟩ in ⟨è⟩ oz. ⟨ô⟩ in ⟨ò⟩ (v IPI ⟨ɛ⟩ oz. ⟨ɔ⟩) ter ⟨é⟩ in ⟨ó⟩ za ozka *e* in *o* (v IPI ⟨e⟩ oz. ⟨o⟩) (odstavek {3} v poglavju Slovnični oris za pravopis).

M. Bitenc (2016a, 2016b) v ortografski transkripciji uporablja znake za narečne glasove, ki predstavljajo variante proučevanih variabel, in sicer γ za narečni pripornik γ ; γ , že omenjen w za narečni dvoustnični glas pred samoglasniki, i : za dolgi i (iz izhodiščno-slovenskih $\acute{e}(:)$, $\acute{e}(:)$ in $e(:)$; z izjemo kratkih v zadnjih zlogih); u : za dolgi u (iz izhodiščno-slovenskih $o(:)$ in $o(:)$; z izjemo kratkih v zadnjih zlogih); e : za dolgi ozki e (kjer bi bil v narečju navadno i : ali ie v neidrijskih varietetah; npr. *be'se:d*); o : za dolgi ozki o (kjer bi bila v narečju navadno u : ali uo v neidrijskih varietetah; npr. *'o:sem*); wo za idrijski dvoglasnik uo in je za idrijski dvoglasnik je .

3.5 Premene po zvonečnosti

V GOS in RSDO premen po zvonečnosti v pisavi ne upoštevajo – zapišejo *tud dobr*, *tud tak*, čeprav se izgovori *tut dobr*, *tut tak*, kot navajata D. Verdonik in A. Bizjak (2022: 26). Pri RSDO kot izjemo izpostavljajo predloge s/z in k/h , ki jih pišejo tako, kot so izgovorjeni (Verdonik in Bizjak 2022: 26).

Tudi M. Smolej (2012) v skladu s standardnim zapisom v vseh primerih ohranja zapis zvonečega nezvočnika, K. Kotnik (2022: 48) glede nezapisovanja utemeljuje, da v obravnavanih narečjih gre za takšne premene, kot jih pozna tudi knjižni jezik.

M. Bitenc (2016a, 2016b) v transkripcijah za variantnostno analizo premene po zvonečnosti upoštevam in zapisujem, pri transkribiranju intervjujev, kjer je poudarek na vsebini, pa ne, razen pri zapisu variant proučevane variable (g): te pišem kot $/\gamma/$ oz. $/g/$ skupaj z alofoni.

Zapisovanje in proučevanje premen po zvonečnosti in njihove variantnosti bi bilo smiselno pri proučevanju škofjeloškega, horjulskega in poljanskega narečja rovtarske narečne skupine, v katerih zvoneči nezvočniki pred pavzo in nezvonečimi soglasniki ohranjajo zvonečnost.

4 Dileme pri razločevanju in nedoslednosti pri zapisovanju glasov

V tem poglavju izpostavljam nekatere izkušnje in izzive, s katerimi sem se srečevala v sociolingvističnih raziskavah. Pri poslušanju in zapisovanju se je namreč večkrat pojavil dvom o tem, za kateri glas gre, tudi pri določenih pojavitvah variant izbranih variabel, predvsem pri hitrem govornem tempu, tišjem ali manj artikuliranem govoru

in pri posnetkih z več udeleženci oziroma drugimi spremljajočimi šumi. Izseke s takimi primeri sem večkrat poslušala in nekatere dodatno analizirala. Tako sem npr. za določitev dolžine *a*-ja pogosto porabila precej časa: po večkratnem poslušanju je sledila analiza v programu Praat, pa nazadnje večkrat še ni bilo povsem jasno, ali gre za kratek ali dolg glas. V drugih primerih sem isti glas lahko zaznala kot *ə*, *a* ali *o* (npr. pri *kət/kot/kaɪ*) ali pa je bilo glas v sredinski legi težko določiti. Če se tudi po večkratnem poslušanju, tudi pri upočasnjem predvajanju, nisem mogla odločiti, sem za presojo prosila še koga. Pogosto se je zgodilo, da so drugi – vsi jezikoslovci slovenisti – glas opredelili drugače, kot sem ga opredelila sama.

Izkazalo se je, da poslušalčev oz. transkriptorjev lastni jezikovni sistem oz. sistemi, ki jih ima uzaveščene, znatno vplivajo na zaznavo. Pri jezikoslovcih je ta pogojena tudi s tem, da so zaradi poznavanja sistemskih lastnosti govora nagnjeni k temu, da slišijo, kar želijo slišati, oz. kar mislijo, da bi morali slišati. Naj kot primer navedem akanje. Narečno bi govorec rekel *ɣawa'rim 'ku:kər ɣawa'rim*, vendar se na posnetku namesto /a/-jev kot vmesna varianta pojavlja težko opredeljiv glas, ki smo ga različni poslušalci različno slišali in zapisali. Tako sem se kot narečna govorka pri vmesnem glasu na tem mestu odločila za zapis *a*-jev, slovenistka iz Vipavske, ki ima v svojem narečju na mestu kratkih nenaglašanih samoglasnikov polglasnik, polglasnike, dialektologinja iz Dolenjske, ki že dolgo živi v Ljubljani, *o*-je, jezikoslovec Ljubljančan pa *a*-je in polglasnike. Vmesni glas bi se dalo bolj natančno zapisati tudi z ožjo transkripcijo kot *a*-jevski polglasnik *ɔ*, vendar se za to nisem odločila, saj sem presodila, da bi vnašanje dodatnih kategorij presojo lahko še otežilo in jo naredilo manj zanesljivo. Tako sem se nazadnje v vseh primerih odločila po svoji presoji ob upoštevanju zaznav drugih jezikoslovcev, kadar so ti podali svoje mnenje. Podobne dileme so bile pogoste pri variablah (v) in (e:) ter pri zlogotvornih zvočnikih, ki predstavljajo potencialne variable. Najbolj sporne primere, ki so jih različni poslušalci, včasih celo en sam, različno zaznavali, sem posebej označila in so v eni izmed različic zapisa obarvani.

J. Škofic (1991) je v primerih, ko se natančnega izgovora besed s posnetka ni dalo razbrati, obliko, ki se ji je zdela verjetnejša, napisala najprej, drugo možno obliko pa v poševnem oklepaju, npr. *do'ma: /du'ma:/* (Škofic 1991: 226). Tudi izkušnje s transkribiranjem pri projektu RSDO kažejo na težave v razločevanju glasov: »v določenih kontekstih se je res težko odločiti med *e* in *@*, pa v nekaterih tudi med kratkim *a* in *@*« (Verdonik v osebni korespondenci jeseni 2022).

Različne vmesne oz. medvarietetne⁹ oblike pri jezikovnem prilagajanju imamo lahko za rezultat številnih strategij kodnega preklapljanja in kodnega mešanja, pri čemer lahko govorimo o t. i. dimenzijah nevtralnosti: govorci kombinirajo elemente različnih varietet in s tem kažejo na družbeno bližino ali identifikacijo z več kodi in njihovimi govorci hkrati. Pri rabi značilnosti različnih kodov se družbena opozicija med njimi nevtralizira. Ta sinteza različnih kodov ali menjavanje med njimi govorcu omogoča izražanje dvojne identitete – tako lahko za različne poslušalce zveni bodisi kot narečni govorec in pripadnik lokalne skupnosti bodisi kot standardni govorec (Mæhlum 2000: 104–105; Bitenc 2016b: 115).

Nedoslednosti pri zapisovanju posameznih glasov glede na določene smernice so opazne pri vseh omenjenih raziskavah, čeprav so transkripcije delo enega transkriptorja. Zagotovo je problematika bolj izrazita v primeru več zapisovalcev v večjih kampanijah. D. Verdonik v osebni korespondenci (jeseni 2022) omenja gradivo, ki so ga transkribirali na Fakulteti za računalništvo in informatiko, kjer so uvedli črko u za u-jevski alofon /v/-ja, pa so to zapisovalci pogosto spregledali in zapisali v ali l. Podobno izkušnjo imam sama s transkriptorko, ki je pomagala transkribirati intervjuje – kljub navodilom za zapisovanje u-ja je v več primerih zapisan v ali l. Pri K. Kotnik (2022) in M. Smolej (2012) se, obratno, kljub načelu zapisovanja z v ali l pojavi tudi zapis z u. D. Verdonik (osebna korespondenca jeseni 2022) obenem izraža nezadovoljstvo z odločitvijo za dosledno zapisovanje polglasnika (z znakom @) v projektu RSDO, ker v tako velikih kampanjah ni mogoče pričakovati, da bo dosledno zapisan.

5 Zaključek

Tudi v korpusih in raziskavah govorne slovenščine so transkripcijski protokoli, načela in prakse različni glede na teoretsko usmeritev raziskovalcev, predmet in namen raziskav, pa tudi obseg projekta, razpoložljivi čas in finančna sredstva. Iskanje kar najbolj optimalnega ravnotežja med čim večjo natančnostjo in doslednostjo ter obenem preprostostjo zapisovanja in branja je vsakič znova izziv. Natančna fonetična transkripcija se uporablja v dialektologiji, sicer pa večinoma ortografske transkripcije z različnimi načini in stopnjami poenostavljanja, navadno s standardno ortografijo in po potrebi posebnimi oznakami za specifične narečne glasove ter ločili. V skladu s splošnimi smernicami se je v izogib (pre)dolgim in (pre)zapletenim

⁹ Gre za oblike, ki jih ne moremo pripisati ne narečju ne standardu (ang. *interdialectal*).

procesom smiselno odredi transkribiranju podrobnosti, za katere kaže, da pri načrtovani študiji ne bodo potrebne, in se odločiti zapisovati za analizo relevantne podatke. Ob morebitni želji po proučevanju drugih variabel se je vedno mogoče vrniti k posnetkom, jih z osredotočenostjo na njihove variante spet poslušati in prilagoditi transkripcije. Zdi se tudi smiselno, da so za morebitne nadaljnje primerjave ali vključitve v podatkovne zbirke (npr. CLARIN.SI) načela zapisovanja čim bližje tem, ki so v veljavi v dosedanjih projektih in vključenih gradivih.

Izkušnje pričajo, da je pri določenih vmesnih glasovih, tudi pri pojavitvah posameznih izbranih variabel, težko presoditi, za katero varianto gre. Pri jezikovnem prilagajanju namreč prihaja do različnih oblik kodnega mešanja in preklapljanja, s katerimi se družbena opozicija med različnimi varietetami nevtralizira, vmesne oblike pa govorcem omogočajo izražanje dvojne identitete. Pri zaznavanju in zapisovanju glasovja se kot pomemben kaže vpliv jezikovnega sistema in vnaprejšnjih pričakovanj transkriptorjev. Za namen transkribiranja predvsem v večjih projektih bi bilo tako smiselno in potrebno usposabljanje in z njim povezana praksa, ki bi te vplive postopoma zmanjšala. Dobrodošle bi bile tudi natančnejše meritve in zato sodelovanje fonetika. Predvsem v primeru večjega števila transkriptorjev pri zapisovanju neizogibno prihaja do določenih nedoslednosti, predvsem pri bolj delikatnih primerih, kot so npr. variante fonema /v/ ali polglasnik.

V okviru projekta *Mezzanine oz. Temeljne raziskave za razvoj govornih virov in tehnologij za slovenščino*¹⁰ so v sklopu Narečna variabilnost med drugim v načrtu naloge, ki bodo lahko pomembno prispevale tudi k enostavnejšemu odločanju in razreševanju dilem glede zapisovanja glasov v nadaljnjih obravnavah jezikovne variantnosti, npr. zasnova sintetične jezikovnogeografske predstavitev fonetičnega/fonemskega inventarja slovenskih narečij, razvoj robustnega modela za prostorski prikaz fonetične variantnosti na podlagi starejših jezikovnogeografskih podatkov, optimizacija široke fonetične transkripcije za govorne korpusi slovenskega jezika in opredelitev razširjenega nabora slovenskih fonemov, ki vključuje na novo opredeljene narečne različice.

¹⁰ <https://mezzanine.um.si/>, dostop 23. 6. 2023.

Literatura

- Maja BITENC, 2016a: *Varietnost govorne slovenščine pri geografsko mobilnih osebah: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- Maja BITENC, 2016b: *Z jezikom na poti med Idrijskim in Ljubljano*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Klaus BRINKER, Sven F. SAGER, 1996: *Linguistische Gesprächsanalyse: Eine Einführung*. Berlin: Erich Schmidt.
- Barbara IVANČIČ KUTIN, 2011: *Živa pripoved v zapisu: kontekst, tekstura in prekodiranje pripovedi Tine Kravanja iz Baršice*. Ljubljana: Založba ZRC, ZRC SAZU. (Ethnologica – Dissertationes, 3).
- Karmen KENDA-JEŽ, 1996: Uvodna pojasnila. *Dialektološke in zgodovinske razprave*. Tine LOGAR, ur. Karmen Kenda-Jež. Ljubljana: ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša. VII–XVIII.
- Karmen KENDA-JEŽ, 2002: *Cerkljansko narečje: Teoretični model dialektološkega raziskovanja na zgledu besedišča in glasoslovja: Doktorska disertacija*. Ljubljana: Filozofska fakulteta.
- Karmen KENDA-JEŽ, 2011: Fonetična transkripcija. *Slovenski lingvistični atlas 1: Človek (telo, bolezni, družina)*. Ur. Jožica Škofic. Ljubljana: Založba ZRC, ZRC SAZU.
- Katja KOTNIK, 2022: *Varietnost govorne slovenščine pri geografsko mobilnih mladih govorcih, govorkah izbranega korškega narečja: Magistrsko delo*. Maribor: Filozofska fakulteta Univerze v Mariboru.
- Steinar KVALE, Svend BRINKMANN, 2009: *InterViews: Learning the Craft of Qualitative Research Interviewing*. Los Angeles: SAGE Publications.
- Sofia LAMPROPOULOU, 2012: *Direct speech, Self-Presentation and Communities of Practice*. London: Continuum International Publishing Group.
- Tine LOGAR, 1993: *Slovenska narečja*. Ljubljana: Založba Mladinska knjiga.
- Brit MÆHLUM, 2000: Strategies of Neutrality in the Arctic. *Dialect and Migration in a Changing Europe*. Ur. Klaus Mattheier. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang. 99–108.
- Shana POPLACK, 1989: The Care and Handling of a Megacorpus: The Ottawa-Hull French Project. *Language Change and Variation*. Ur. Ralph Fasold, Deborah Schiffrin. Amsterdam, Philadelphia: John Benjamins. 411–444.
- Pravopis 8.0: Pravila novega slovenskega pravopisa za javno razpravo*. Dostop 23. 6. 2023 na www.fran.si/pravopis8.
- Dennis PRESTON, 2000: Mowr and Mowr Bayud Spellin': Confessions of a Sociolinguist. *Journal of Sociolinguistics* 4/4, 614–621.
- Mojca SMOLEJ, 2012: *Besedilne vrste v spontanem govoru*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Jožica ŠKOFIC, 1991: *Problemi slovenskega pogovornega jezika: Magistrsko delo*. Ljubljana: Filozofska fakulteta.
- Sali A. TAGLIAMONTE, 2006: *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Sali A. TAGLIAMONTE, 2007: Representing Real Language: Consistency, Trade-Offs and Thinking Ahead! *Using unconventional digital language corpora 1: Synchronic corpora*. Ur. Joan Beal, Karen Corrigan, Hermann Moisl. Basingstoke, Hampshire: Palgrave Macmillan. 205–240.
- Jože TOPORIŠIČ, Franc JAKOPIN, Janko MODER, Janez DULAR, Stane SUHADOLNIK, Janez MENART, Breda POGORELEC, Kajetan GANTAR, Martin AHLIN (ur.), 2003: *Slovenski pravopis*. Ljubljana: Založba ZRC, ZRC SAZU.
- Darinka VERDONIK, Ana ZWITTER VITEZ, 2011: *Slovenski govorni korpus GOS*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Darinka VERDONIK, Andreja BIZJAK, 2022: *Navodila za zapisanje in označevanje govora v projektu RSDO: Verzija 15*. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko Univerze v Mariboru.
- Peter WEISS, 1990: *Govori Zadrebke doline med Gornjim Gradom in Nazarjami: Glasoslovje, oblikoslovje in skladnja: Magistrsko delo*. Ljubljana: Filozofska fakulteta.

- Alison WRAY, Aileen BLOOMER, 2006: *Projects in Linguistics: A Practical Guide to Researching Language*. London: Hodder Arnold.
- Jana ZEMljARIČ MIKLAVČIČ, 2008: *Govorni korpusi*. Ljubljana: Znanstvena založba Filozofske fakultete, Oddelek za prevajalstvo. (Zbirka Prevodoslovje in uporabno jezikoslovje).

JEZIKOVNI MODELI ZA PRIPRAVO GOVORNEGA KORPUSA: PROGRAMI ZA PREPOZNAVANJE GOVORA

TEODOR PETRIČ

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
teodor.petric@um.si

V preteklem desetletju, še posebej v zadnjih petih letih po uveljavljanju velikih jezikovnih modelov, ki temeljijo na arhitekturi transformerjev (pretvorbenih modelov), smo dobili vrsto programskih orodij, ki pospešujejo ustvarjanje večplastnih jezikovnih gradiv. Preizkušali smo programska orodja za prepoznavanje in pretvorbo govora v pisno obliko (tj. orodja *Razpoznavalnik*, *Microsoft Word Prepisi*, *Vosk/Kaldi* in *OpenAI Whisper*), ki so ključni za pospešeno ustvarjanje govornih korpusov. Uporabljali smo vrsto meril, ki zadevajo preprostost uporabe, časovni prihranek, morebitne stroške, zagotavljanje anonimnosti govorcev in različne vidike kakovosti pretvorbe (delež besednih napak, število zamenjav, vstavitev in izpustov). Orodja za pretvorbo govora v pisno obliko so vidno napredovala, vendar bi si vsekakor želeli, da bi lahko izhodne formate programov prilagajali posameznim raziskovalnim potrebam, npr. vključitev diskurznihih označevalcev (npr. tako imenovanih »mašik») ali dejansko izgovorjenih skrčenih besednih oblik v zapis.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.9](https://doi.org/10.18690/um.ff.4.2024.9)

ISBN
978-961-286-882-6

Ključne besede:
pretvorbeni modeli,
govorni korpus,
kakovost pretvorbe,
programska orodja,
raziskovalne potrebe

DOI
[https://doi.org/
10.18690/um.ff.4.2024.9](https://doi.org/10.18690/um.ff.4.2024.9)

ISBN
978-961-286-882-6

Keywords:
transformer models,
spoken corpora,
conversion quality,
software tools,
research needs

LANGUAGE MODELS FOR SPOKEN CORPUS PREPARATION: SPEECH RECOGNITION SOFTWARE

TEODOR PETRIČ

University of Maribor, Faculty of Arts, Maribor, Slovenia
teodor.petric@um.si

In the last decade, particularly in the last five years after the emergence of large language models based on transformer architectures, we have seen the development of a number of software tools that accelerate the creation of multi-layered corpora. We have tested software tools for speech recognition and conversion to written form (i.e. tools such as Razpoznavnik, Microsoft Word Dictate, Vosk/Kaldi and OpenAI Whisper), which are crucial for accelerating the creation of spoken corpora. We have employed various criteria concerning ease of use, time-saving features, potential costs, ensuring speaker anonymity and various aspects of conversion quality (e.g. word error rates, number of substitutions, insertions and deletions). While the tools for converting speech to written form have made considerable progress, we would certainly wish for the ability to customize the output formats of these programmes to meet individual research needs, e.g. including discourse markers (such as the so-called ‘fillers’) or the actual spoken contracted word forms in the transcription.



1 Uvod

1.1 Pretvorbeni modeli

V preteklem desetletju, še posebej v zadnjih petih letih po uveljavljanju velikih jezikovnih modelov, ki temeljijo na arhitekturi *transformerjev* (v nadaljevanju tudi v poslovenjeni obliki: *pretvorbeni modeli*), smo dobili vrsto programskih orodij, ki pospešujejo ustvarjanje večplastnih jezikovnih gradiv. Pretvorbeni modeli imajo potencial za velike spremembe v družbi, med drugim tudi v načinu raziskovalnega dela in poučevanja.

Pretvorbeni modeli (Vaswani et al. 2017: 2–6) so vrsta arhitekture nevronske mreže, temeljijo pa na mehanizmi pozornosti (*attention mechanism*) in samopozornosti (*self-attention mechanism*). *Mehanizem pozornosti* modelu omogoča osredinjanje na najpomembnejše dele vzorca in zanemarjanje manj pomembnih delov vzorca: npr. pri prevajanju stavka mehanizem pozornosti določi, katera beseda v izvornem jeziku je najbolj povezana z besedo v ciljnem jeziku. Mehanizem pozornosti pomaga izboljšati kakovost prevoda in razumevanje jezika.

Mehanizem *samopozornosti* je posebna vrsta mehanizma pozornosti, ki povezuje različne položaje ene same sekvence in nato izračuna vektorsko predstavitev (reprezentacijo) sekvence. Modelu omogoča, da oceni pomembnost različnih besed v stavku in dinamično prilagodi njihov vpliv na izhod. Pri razumevanju naravnih jezikov je to pomembna lastnost, saj se pomen besede lahko spremeni glede na sobesedilo znotraj stavka ali besedila. Zaradi opisanih lastnosti so pretvorbeni modeli primernejši za vzporedne računske operacije in analizo odvisnosti medsebojno oddaljenih elementov kot druge nedavno razvite programske arhitekture (npr. rekurentni ali konvolucijski modeli).

1.2 Veliki jezikovni modeli

Veliki jezikovni modeli (*Large language models, LLM*) so pretvorbeni modeli, ki so izurjeni na osnovi velikih količin besedilnih podatkov (npr. celotne Wikipedije ali več). Naučijo se lahko slovnice, pravopisa, besedišča in drugih znanj, ki jih potrebujejo za uporabo jezika. Prilagodimo jih lahko za opravljanje posebnih nalog, kot so povzemanje besedila, strojno prevajanje, odgovarjanje na vprašanja, pisanje besedila, iskanje podatkov in idej, za izdelovanje aplikacij in orodij za učenje in

poučevanje jezikov, za razpravljanje o različnih družbeno relevantnih temah in drugo. LLM lahko ustvari koherentno in naravno zveneče besedilo, ki ga pogosto ni mogoče razlikovati od človeškega besedila (npr. GPT-3.5/4, BERT, T5). Zaradi vseh teh lastnosti so izredno zanimivo orodje pri pripravi in ustvarjanju jezikovnih gradiv ter pri preučevanju in poučevanju naravnih jezikov.

Aktualni jezikovni modeli omogočajo različne postopke za analizo stavčnih elementov ali celotnih besedil v jezikovnih gradivih: npr.

- glasoslovje, pravopis: samodejno prepoznavanje (ASR) in zapisovanje govora (STT - speech to text),
- oblikoslovje: označevanje besednih vrst (POS),
- skladnja: odvisnostna razmerja, besedni vrstni red,
- NER: prepoznavanje poimenovanih entitet in razmerja med entitetami,
- Q&A: vprašanja uporabnika in odgovori programa o jezikovni tematiki, npr. članka ali knjige,
- povzemanje: ustvarjanje izvlečkov in povzetkov člankov ali knjig,
- koreferenčnost: ugotavljanje soodnostnosti in sopomenskosti stavčnih prvin,
- semantika: odkrivanje semantičnih sprememb skozi čas,
- pragmalingvistika: analiza čustev (sentimenta) in čustvenosti govora,
- avtomatizacija: programski agenti GPT (GPT agents) uporabljajo LLM za avtomatizacijo različnih nalog na podlagi cilja, ki ga je v naravnem jeziku določil človeški uporabnik.

V nadaljevanju sestavka se osredinjamo na preizkus in oceno programskih orodij za samodejno prepoznavanje in pretvorbo govora v pisno obliko (*Automatic Speech recognition*, ASR).

2 Preizkus orodij ASR

2.1 Motivacija preizkusa

Če želimo z govornimi viri izvesti podobne raziskave kot s pisnimi (gl. prejšnji odsek), je ključna hitra in uspešna pretvorba govora v pisno obliko. Modeli za samodejno razpoznavanje govora nam obetajo precejšen časovni prihranek pri zapisovanju govornih prispevkov.

Modeli ASR lahko zajamejo tako globalne kot lokalne značilnosti govornih virov, predhodna priprava na velikih količinah neoznačenih govornih podatkov pa jim omogoča usvajanje splošnih akustičnih vzorcev in znanj.

Prilagodimo jih lahko že na osnovi majhne količine označenih podatkov za določeno nalogo (*fine-tuning*), kot je npr. prepisovanje telefonskih klicev. Cilj, ki smo si ga zastavili, tj. ustvarjanje multimodalnega govornega korpusa, je kompleksnejši. Prav nam bi prišla programska oprema za pospeševanje zamudnega zapisovanja in urejanja govornih prispevkov.

2.2 Izbor orodij ASR

Programskih orodij za samodejno prepoznavanje govora je kar nekaj, večinoma pa imajo določene omejitve ali pa imamo določene zadržke do njihovih zahtev (npr. previsoki stroški, posnetek mora biti kratek, anonimnost ni zagotovljena, med jeziki ni slovenščine ...).¹

Preizkusili smo nekaj izmed razpoložljivih orodij za pretvorbo govora v pisno obliko, ki imajo razmeroma preprost vmesnik za delo z njimi:²

- *Razpoznavalnik*, različica *e2e*,
- *Razpoznavalnik*, različica *Kaldi*,
- *Microsoft Word* z vstavkom *Prepiši (Transcribe)*,
- *Vosk/Kaldi*, vtičnik v programu *Subtitle Edit*,
- *Whisper* in različica *Whisperx*.

Razpoznavalnik (tj. splošni razpoznavalnik *e2e*) je program za prepoznavanje govora v slovenščini. Razvil in trži ga konzorcij, ki ga sestavljata Zemanta in Univerza v Ljubljani (Lebar Bajec et al. 2022). Gre za model ASR, ki temelji na tehnologiji NVIDIA NeMo, ki je okvir za razvoj modelov AI za konverzacijske aplikacije. Model je bil treniran na naboru podatkov *Artur* (Verdonik et al. 2023), ki vsebuje 630 ur transkribiranega govora v standardni slovenščini. Program je dostopen na

¹ Na spletni strani *paperswithcode* (<https://paperswithcode.com/datasets?task=speech-recognition>) so tudi povezave do podatkovnih nizov za učenje modelov. (31. 8. 2023).

² Zelo obetaven projekt je *FAIRSEQ* podjetja *Meta* (prej: *Facebook*), ki pa je trenutno dostopen predvsem programerjem, zato ga v našem izboru programskih orodij ni.

spletni strani slovenscina.eu³, kjer lahko uporabniki naložijo zvočne posnetke in dobijo njihovo prepisano besedilo. Program lahko prepozna govor do dolžine 300 sekund. Repozitorij programa je na GitHubu⁴, kjer je mogoče najti več informacij o modelu in njegovem delovanju.

Kaldi (Povey et al. 2011), program za prepoznavanje govora v slovenščini, je dostopen na spletni strani slovenscina.eu⁵, kjer lahko uporabniki izberejo med dvema modeloma ASR: splošnim razpoznavalnikom *Kaldi* in razpoznavalnikom *e2e*. Splošni razpoznavalnik *Kaldi* je model ASR, ki temelji na odprtokodnem orodju *Kaldi*, ki je namenjeno raziskovalcem s področja prepoznavanja govora. Program lahko prepozna govor do dolžine 300 sekund. Več informacij o projektu *Kaldi* na spletnih straneh *Kaldi*⁶ in na *GitHubu*⁷.

Microsoft Office 365 in Word Prepisi (Transcribe): program za prepoznavanje govora je del storitve Microsoft Office 365, ki jo je razvil in trži Microsoft. Gre za funkcijo *Word Prepisi (Transcribe)*, ki omogoča pretvorbo govora v besedilo. Program uporablja model za prepoznavanje govora, ki temelji na storitvi Azure Cognitive Services AI, ki je platforma za razvoj modelov umetne inteligence za konverzacijske aplikacije. Model je bil naučen na velikih količinah transkribiranega govora v več kot 100 jezikih. Program lahko prepozna govor iz različnih virov, kot so mikrofoni, zvočne datoteke ali shramba blob. Spletna različica programa lahko prepozna govor do dolžine 300 sekund, omejitev za nameščeno različico programa pa je velikost naložene zvočne datoteke (manj kot 300 MB). Več informacij o programu za prepoznavanje govora lahko je na spletnih straneh podjetja *Microsoft* (npr. *Microsoft support*⁸ in *Azure*⁹).

Subtitle Edit je brezplačen in priljubljen urejevalnik podnapisov za video posnetke, ki v različici 3.6.13 podpira uporabo modelov ASR (*Vosk/Kaldi* in *Whisper*) kot vtičnikov.

³ <https://slovenscina.eu/razpoznavalnik> (31. 8. 2023).

⁴ https://github.com/clarinsi/Slovene_ASR_e2e (31. 8. 2023).

⁵ Prav tam.

⁶ <https://www.kaldi-asr.org/doc/about.html> (31. 8. 2023).

⁷ <https://github.com/kaldi-asr> (31. 8. 2023).

⁸ <https://support.microsoft.com/en-us/office/transcribe-your-recordings-7fc2efec-245e-45f0-b053-2a97531ecf57> (31. 8. 2023).

⁹ <https://azure.microsoft.com/en-us/products/ai-services/speech-to-text> (31. 8. 2023).

Vtičnik *Vosk/Kaldi* v programu *Subtitle Edit*: program za prepoznavanje govora je vtičnik *Vosk/Kaldi*. Vtičnik *Vosk/Kaldi* je razvil in trži *Alpha Cephei*, ki je podjetje, specializirano za rešitve na področju prepoznavanja govora. Vtičnik *Vosk/Kaldi* uporablja model ASR, ki temelji na odprtokodnem orodju *Kaldi*, ki je namenjeno raziskovalcem s področja prepoznavanja govora. Model je bil treniran na različnih naborih podatkov, odvisno od podprtega jezika. Vtičnik *Vosk/Kaldi* lahko prepozna govor v več kot 30 jezikih in variantah, med njimi slovenščine še ni. Program lahko prepozna govor iz različnih virov, kot so mikrofoni, zvočne datoteke ali video posnetki. Program nima časovnih omejitev za dolžino posnetkov za pretvorbo. Več informacij o programu za prepoznavanje govora in jezikovnih modelih lahko je na *Alphacephei*¹⁰ in na *GitHubu*¹¹.

Whisper je razvil in trži *OpenAI*, raziskovalna organizacija, ki se ukvarja z umetno inteligenco. *Whisper* je brezplačen model ASR, ki temelji na tehnologiji transformerjev, ki so vrsta nevronske mreže, ki uporabljajo mehanizem samopozornosti za učenje iz sekvenc podatkov. Model je bil treniran na velikih količinah transkribiranega govora v več kot 50 jezikih in variantah. Program lahko prepozna govor v angleščini, francoščini, nemščini, kitajščini, španščini, slovenščini in drugih jezikih. Sistem za prepoznavanje govora je bil naučen na 680.000 urah večjezičnega in večopravilnega gradiva, zbranega v medmrežju. Po mnenju podjetja vodi uporaba tako velikega in raznolikega nabora podatkov do izboljšane robustnosti pri prepoznavanju različnih naglasov, ozadnega šuma in tehničnega jezika. Poleg tega omogoča prepisovanje v več jezikih, pa tudi prevajanje iz teh jezikov v angleščino. Program lahko prepozna govor iz različnih virov, kot so mikrofoni, zvočne datoteke ali video posnetki. Omejitve glede dolžine posnetka program nima. Modele *Whisper* podjetja *OpenAI* lahko uporabljamo preko različnih vmesnikov ali programov: kot vtičnik programa *Subtitle Edit* ali v *ukazni vrstici* ali v skriptu računalniškega jezika *Python* na lokalnem računalniku ali v oblčnih storitvah kot npr. *Google Colab*. *Whisper* podpira tudi uporabo grafičnega pospeševalnika, kar omogoča bistveno hitrejšo prepoznavanje govora. Več informacij o programu za prepoznavanje govora lahko je na spletnih straneh *Whisper*¹² in na *GitHubu*.¹³

¹⁰ <https://alphacephei.com/vosk/> (31. 8. 2023).

¹¹ <https://github.com/alphacep/vosk-api> (31. 8. 2023).

¹² <https://openai.com/research/whisper> (31. 8. 2023).

¹³ <https://github.com/openai/whisper> (31. 8. 2023).

2.3 Merila ocenjevanja

Uporaba orodja ASR naj jezikoslovcu čimbolj olajša pretvorbo govornih virov in njihovo implementacijo v multimodalni korpus, ki ga želi sestaviti s programsko opremo (npr. s programom *Elan* ali drugimi korpusnimi orodji). Merila, po katerih smo ocenjevali dosežke programskih orodij:

- cenovno ugodna ali celo brezplačna uporaba,
- preprost uporabniški vmesnik,
- primerna hitrost (pretvorbe, shrambe, nalaganje in povezanih opravil),
- primerne strojne zahteve za širši krog uporabnikov,
- zagotovljeno varstvo osebnih podatkov govorcev na posnetkih,
- uporabnost za več jezikov (poleg slovenščine),
- prilagodljivost govornim virom slabše kakovosti (brez dodatnega programskega izboljševanja zvočnega posnetka),
- prilagodljivost neknjižnemu govoru,
- različne vhodne oblike (oblike zvočnih in video datotek),
- izpis pretvorjenega gradiva v različne oblike,
- časovni žigi govornih prispevkov (timestamps),
- sposobnost razlikovanja govorcev (speaker diarization, separation, identification),
- natančnost prepoznavanja govornih besed (knjižna in neknjižna izreka), kar preverjamo z metriko WER (word error recognition) in preizkusi urejanja ustvarjenega pisnega gradiva,
- samodejno postavljanje ločil,
- uporabnost za dalj časa trajajoče zvočno ali filmsko gradivo,
- programska razširljivost ali prilagodljivost orodja.

2.4 Zvočno in filmsko gradivo

Zvočno gradivo v preizkusih za ocenjevanje uporabnosti programa za zgoraj predstavljen namen je bilo raznoliko sestavljeno iz:

- slovenska zborna izreka (Šeruga-Prek et al. 2004, 5 minut),
- vsakdanji pogovor (smalltalk) v slovenščini (*Sosedska neljubezzen*, 2 min.),

- pogovor z otrokom v slovenščini (5 in 90 min.),
- pogovori z otrokom v nemščini (90 in 120 min.),
- razprava v nemški televizijski oddaji *13 Fragen* (tema: Instagram, 35 min.),
- Fantom iz opere v nemščini (pesem, 4 min.),
- Mešanje jezikov - angleščine in španščine (nadaljevanka *Kraljica juga*, > 5 min.).

Večinoma je bilo vsako gradivo preizkušeno z vsakim orodjem dvakrat, izjemoma celo več kot dvakrat ali samo enkrat.

2.5 Preizkus programskih orodij

WER (*Word error rate*, delež besednih napak) in CER (*Character error rate*, delež črkovnih napak) sta kazalnika uspešnosti sistema za samodejno prepoznavanje govora. Merita razliko med referenčnim besedilom in samodejno prepoznanim besedilom s štetjem števila popravkov (zamenjav, izpustov ali vstavitev), ki so potrebni za pretvorbo enega besedila v drugo. WER deluje na besedni ravni, CER pa na črkovni ravni. V primerjavi je treba upoštevati vrsto in dolžino besedilnih zaporedij. Če so kratka in sestavljena iz določenih zaporedij (kot so telefonske številke, številke socialnega zavarovanja itd.), je CER po navadi ustrežnejši od WER, če pa so dolga in sestavljena iz povedi (kot npr. knjige, časopisi itd.), je WER po navadi uporabnejši kot CER. Praviloma je WER tri- do štirikrat višji od CER in je premo sorazmeren z njim. Če se izboljšuje CER, se bo izboljšal tudi WER. Vendar ta zveza ni vedno linearna ali dosledna, zlasti pri velikem številu vstavitev.¹⁴

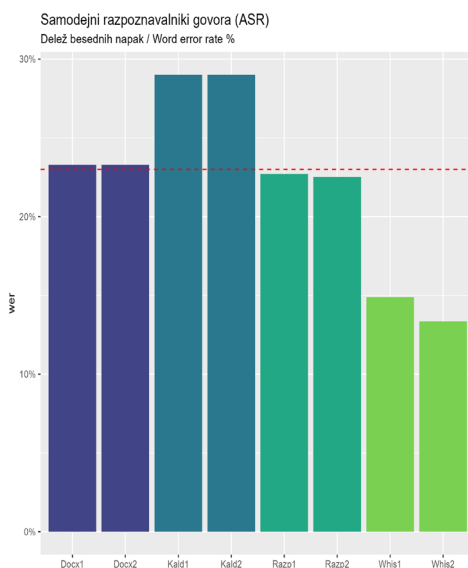
Pri ocenjevanju kakovosti pretvorbe govora smo uporabljali le količino *WER*, ki je po navadi primeren za primerjavo napak na besedilni ravni. Za izračun vrednosti *WER* smo uporabljali knjižnico *wersim* (Proksch, Wratil, and Wäckerle 2018) v računalniškem jeziku R (R Core Team 2023). Sestavili smo programski skript za izračun količin, sestavo tabel in izpis grafikonov.

¹⁴ Evaluate OCR Output Quality with Character Error Rate (CER) and Word. Towards Science (<https://towardsdatascience.com/evaluating-ocr-output-quality-with-character-error-rate-cer-and-word-error-rate-wer-853175297510>), (31. 8. 2023).

2.5.1 Slovenska knjižna izreka

Najprej bomo predstavili izide preizkusa orodij z zvočnim posnetkom, ki predstavlja vzorec slovenske zborne izreke (Šeruga-Prek et al. 2004). Izbrali smo peti posnetek na zgoščenci, v katerem slišimo *Ajdo Kalan* izgovarjati posamezne besede, besedne zveze ali kratke povedi o različnih, vsebinsko nepovezanih temah. Zvočni vzorec je bil posnet v studiu RTV Slovenije in zato tudi akustično zelo kakovosten. Za preizkus smo posnetek omejili na prvih pet minut.

Preizkušali smo *Razpoznavalnik* (oba modela: *e2e* in *Kaldi*), OpenAI *Whisper* in Microsoft Word *Prepiši* (*Transcribe*). Izide prikazujeta preglednica 1 in diagram 1.¹⁵ Deleži besednih napak (*wer*) so pri vseh orodjih nizki. Najnižje vrednosti je dosegel *Whisper*, čeprav slovenščina ni med jeziki, ki jih najbolje obvlada. Razmeroma tesno mu sledita *Razpoznavalnik* (model *e2e*) in *Word Prepiši*. Najslabše se je odrezal model *Kaldi*.



Slika 1: Slovenska knjižna izreka (deleži besednih napak, WER)

Vir: lasten

¹⁵ Uporabljene kratice: *wer* = delež besednih napak v odstotkih ($wer = 100 * (ins + del + sub) / word.ref$); *sub* = število besednih zamenjav; *ins* = število vstavljenih besed, *del* = število zbranih besed; *word.ref* = število pojavnic v referenčnem besedilu; *words.hyp* = število pojavnic v samodejno prepoznanim besedilu; rdeča črtkana črta v diagramu je *mediana*; *Docx* = Microsoft Word Prepiši, *Razp* = Razpoznavalnik, model *e2e*; *Kald* = Razpoznavalnik, model Kaldi; *Whis* = Openai Whisper; številka za imenom orodja je številka preizkusa

Preglednica 1 prikazuje podrobnejšo sliko in nam približuje vzorce programskega vedenja v procesu prepoznavanja govora, kar bo moč videti tudi v rezultatih sledečih preizkusov. Oba modela *razpoznavalnika* (*e2e* in *Kaldi*) izkazujeta majhno število vstavljenih izrazov (vrednosti *ins* so nizke), *Whisper* pa vstavlja več izrazov kot ostala orodja, teži torej k vstavljanju sobesedilno verjetnih besed, če izgovorjene besede ne prepozna. *Word Prepiši* podobno kot *Razpoznavalnik* (model *e2e*) malokdaj vstavlja izraze, zelo rad pa jih izpusti ali v določenih primerih zamenja.¹⁶

Tabela 1: Slovenska knjižna izreka (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Kald1	29	86	6	58	524	470
Kald2	29	86	6	58	524	470
Razp1	23	64	1	54	524	471
Razp2	23	64	1	53	524	472
Whis1	15	47	10	18	524	516
Whis2	13	47	4	19	524	509
Docx1	23	36	1	85	524	440
Docx2	23	36	1	85	524	440

Vir: lasten

Po pregledu podrobnih rezultatov se je pri prepoznavanju vzorca slovenske zborne izreke najbolje obnesel *Whisper* (najnižji delež besednih napak, malo izpuščenih besed), sledi *Razpoznavalnik* (model *e2e*). *Word Prepiši* pa je v preizkusu prepoznanih besed pristal na tretjem mestu, saj je izpustil več besed kot *Razpoznavalnik* (*e2e*). Četrty model (*Kaldi*) pa zaostaja z razmeroma velikim številom zamenjav in izpustov. Tudi z ozirom na ostala zgoraj navedena merila bomo prihranili največ časa z orodjem *Whisper*, saj omogoča pretvorbo dolgih posnetkov, uporabo grafičnega procesorja za pospeševanje prepoznavanja govora in več izhodnih oblik (npr. tudi podnapise, ki jih zlahka uvozimo v *Elan* za ustvarjanje večsteznega jezikovnega gradiva), idr. Razlika med *Razpoznavalnikom* in *Wordom Prepiši* pa se zmanjšuje v korist Microsoftovemu orodju zaradi (opcionalnega) razlikovanja govorcev in izpisa v različnih izhodnih oblikah (s časovnimi žigi ali brez). Oboje lahko prihrani čas pri ustvarjanju jezikovnega gradiva.

Orodja za prepoznavanje govora spremljajo programi za *vstavljanje ločil*. Ustreznosti vstavljenih ločil nismo preverjali, vendar smo dobili vtis, da *Whisperjeva* programska sestavina za vstavljanje ločil ne zaostaja za *Razpoznavalnikom* (*e2e*). *Microsoftov* program

¹⁶ Microsoft Word Prepiši po navadi zamenja vulgarne izraze z zvezdicami, kar se v evidenci šteje kot zamenjava.

in *Kaldi* sta bila v tem pogledu pogosteje manj dosledna kot *Whisper* ali *Razpoznavalnik (e2e)*.

Precejšen delež *Whisperjevih* napak je šlo na račun neupoštevanih pravopisnih pravil. Program je torej pravilno “slišal”, ampak izbral napačen grafem. V zvočnem vzorcu je bilo več besed, kjer je bil izgovorjen nezvenciči pripornik [s], zapisati pa ga je bilo treba z grafemom <z> (npr. v besedah *razprava*, *razplet*, *privež*). *Whisper* je take besede pogosto zapisal z grafemom <s>, kar ustreza nezvencičemu priporniku [s] (npr. kot v hrvaški besedi *rasprava*). Podobno se je *Whisper* tudi pri dvoumnih besedah (npr. *poseg* ali *posek*, *obseg* ali *obsek*) odločil za grafem <k> in ne za <g>. Nekatere zamenjave se nanašajo na oblike, kjer pišemo grafem <l>, izgovarjamo pa [w] (kot npr. v glagolskih oblikah *preletel*, *obvestil*). Pri glagolskih oblikah je *Whisper* glagole nekajkrat spregal s tematskim samoglasnikom /a/ namesto z /e/ (npr. *vlečajo* namesto *vlečejo*, *preletal* namesto *preletel*, *privedal* namesto *privedel*, *pripomogal* namesto *pripomogel*).

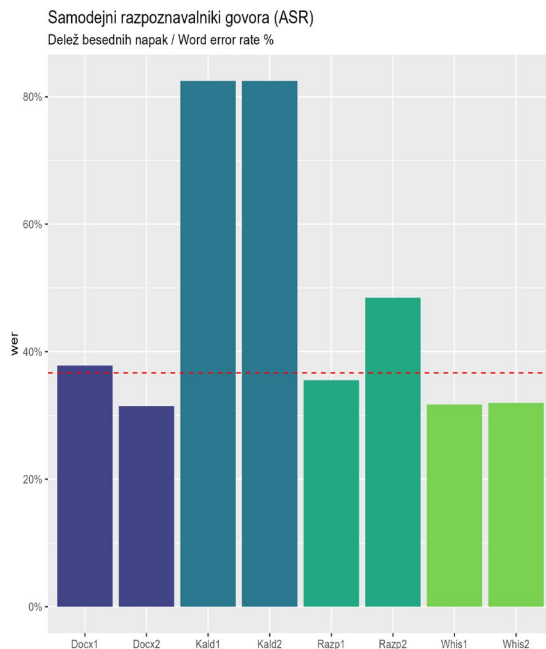
Občasno je prišlo tudi do težave, ali vstaviti črko <j> ali ne (npr. *pomenjska razlika* namesto *pomenska*, *dalnovid* namesto *daljnovid*). Te besedne oblike pričajo o tem, da je program sicer akustično prepoznal besedne oblike, da pa pravopisno ni dovolj podkovan ali da meša pravopisne oblike v slovenščini s tistimi v podobnih slovanskih jezikih.

2.5.2 Slovenski pogovorni jezik

Dvominutni odlomek iz slovenskega filma *Sosedska neljubezen*¹⁷ je vzorec, ki se od prejšnjega razlikuje po več lastnostih: ne govori samo ena oseba, temveč tri (mati, oče, hči najstnica), osebe ne govorijo v zbornem jeziku in zvočna kakovost zaostaja za tistim v prejšnjem posnetku, saj se pogovor odvija v naravnem prostoru v hiši. V ozadju pogovora ni izrazitih ali motečih šumov. Izidi o deležu besednih napak (*wer*) so podobni tistim za zborni jezik: *Whisper* ima nekoliko nižji delež besednih napak kot *Razpoznavalnik (e2e)* in *Word Prepisi*, močneje pa zaostaja *Kaldi* (slika 2).

¹⁷ DD Studio produkcija, 2012

(https://www.youtube.com/watch?app=desktop&v=aRot1XH3RDE&embeds_referring_curi=https%3A%2F%2Fmichalekskolky.cz%2F&feature=emb_imp_woyt) (31. 8. 2023)



Slika 2: Slovenski pogovorni jezik (deleži besednih napak, WER)

Vir: lasten

Podrobnejši izidi v preglednici 2 kažejo podobne težnje kot pri razpoznavanju vzorca zborne izreke: *Whisper* dodaja besede, tako da je število vseh besed celo večje kot v referenčnem besedilu. Druga orodja (*Razpoznavalnik e2e*, *Word Prepiši* in *Kaldi*) pogosteje izpuščajo besede.

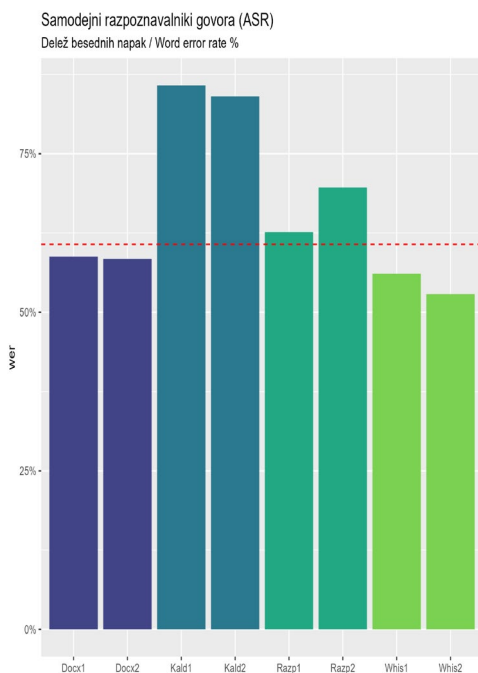
Tabela 2: Slovenski pogovorni jezik (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Kald1	82	18	0	307	394	87
Kald2	82	18	0	307	394	87
Razp1	36	73	35	25	394	405
Razp2	48	88	6	91	394	308
Whis1	32	61	35	19	394	414
Whis2	32	62	35	19	394	414
Docx1	38	58	5	82	394	314
Docx2	31	44	8	66	394	330

Vir: lasten

2.5.3 Pogovor s slovenskim otrokom

Glede na to, da sem pred leti sestavil manjši govorni korpus o otroškem govoru in o izsledkih poročal v več znanstvenih prispevkih (npr. Petrič 2016, 2021), me je še posebej zanimalo, kako uspešni so novi modeli za pretvorbo govora v težavnem naravnem pogovornem okolju. Izbrani petminutni zvočni vzorec je v dveh pogledih še težji za razpoznavanje govora: posnetki so slabše kakovosti kot filmski dialog ali posnetek zborne izreke, govornici v ozadju se ne slišijo tako dobro kot sogovornik v vlogi snemalca blizu mikrofona in eden izmed sogovornikov je trileten otrok. Dobili smo vtis, da je otroški glas orodjem glasovno manj domač kot odrasli glasovi. Med učnimi vzorci orodij za razpoznavanje govora najbrž ni takih (ali le malo takih) z otroškimi glasovi. Rezultati kažejo, da je mediana deleža besednih napak (WER) pri vseh preučevanih orodjih precej slabša, z vrednostjo preko 60 %. Po deležu besednih napak rahlo prednjači *Whisper* pred *Wordom Prepiši* in *Razpoznavalnikom (e2e)*, *Kaldi* pa tudi tu močno zaostaja (slika 3).



Slika 3: Pogovor s slovenskim otrokom (deleži besednih napak, WER)

Vir: lasten

Vedenjski vzorec orodij je podoben zgoraj opisanemu (tabela 3): *Whisper* pogosteje dodaja besede kot ostala orodja (čprav ne tako močno kot v prejšnjih vzorcih), slednja orodja pa pogosteje izpuščajo besedno gradivo (*Razpoznavnik e2e* in *Word Prepiši* sta izpustila več kot 300 besed, *Kaldi* celo več kot 600 od 808 v referenčnem besedilu). Vsa orodja (razen *Kaldi*) so v pogovornem gradivu pogosto zamenjevala besede.

Tabela 3: Pogovor s slovenskim otrokom (deleži besednih napak, WER)

comparison	wer	sub	ins	del	words.ref	words.hyp
Kald1	86	50	0	643	808	165
Kald2	84	58	0	621	808	187
Razp1	63	184	8	307	808	502
Razp2	70	177	7	372	808	436
Whis1	56	225	24	181	808	629
Whis2	53	165	8	246	808	563
Docx1	59	152	3	317	808	491
Docx2	58	156	11	297	808	515

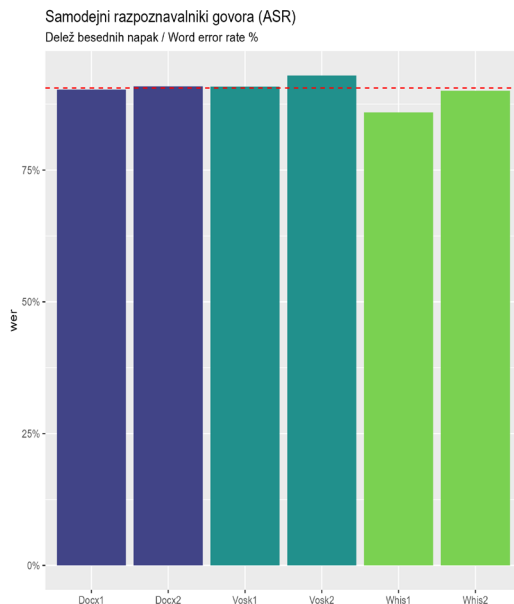
Vir: lasten

Razpoznavanje vsakdanjega spontanega govora, ki je bil posnet v naravnih prostorih, je delal vsem orodjem velike preglavice. Dodatna težava je bilo sodelovanje otroka v pogovoru. Pri zapisovanju vsakdanjih pogovorov v slabših akustičnih razmerah, še posebej, če sodelujejo majhni otroci, žal ne bi prihranili veliko časa. Najbrž primanjkuje tovrstnih zvočnih vzorcev za učenje programskih orodij, kar še posebej velja za slovenščino.

2.5.4 Pogovor z nemškima otrokoma

Preizkus s podobnima zvočnima vzorcema v nemščini, tj. pogovor z nemškima otrokoma *Emely* in *Falko*¹⁸, je prav tako prinesel dokaj slabe izide. Posnetka sta trajala dobrih 60 minut oz. dobrih 90 minut. Občasni hkratni govor dveh ali več oseb, otroški glasovi, neenakomerna glasnost človeških glasov in slabša zvočna kakovost v snemalnem prostoru (igralnici) so botrovali visokim deležem besednih napak (WER > 75 %) pri vseh uporabljenih orodjih, čprav spet rahlo manj z orodjem *Whisper* (slika 4).

¹⁸ Childes, <https://childes.talkbank.org/access/German/Szagun.html> (31. 8. 2023).



Slika 4: Pogovor z nemškima otrokoma (deleži besednih napak, WER)

Vir: lasten

Poglavitna značilnost razpoznavanja zvočnih posnetkov z otrokoma je bilo izpuščanje neprepoznanega besednega gradiva (tabela 4). V krajšem pogovoru (ok. 60 minut) je najmanj besed izpuščal *Whisper*, v daljšem pogovoru (ok. 90 minut) pa ga je v tem oziru s tesnim izidom premagal *Word Prepiši*. Podobno razmerje je vidno tudi pri številu zamenjav. Model *Vosk* za nemščino se je glede na število izpustov in zamenjav slabše odrezal kot *Whisper* in *Word Prepiši*. Vendar se to pri deležih napačnih besed skorajda ne pozna.

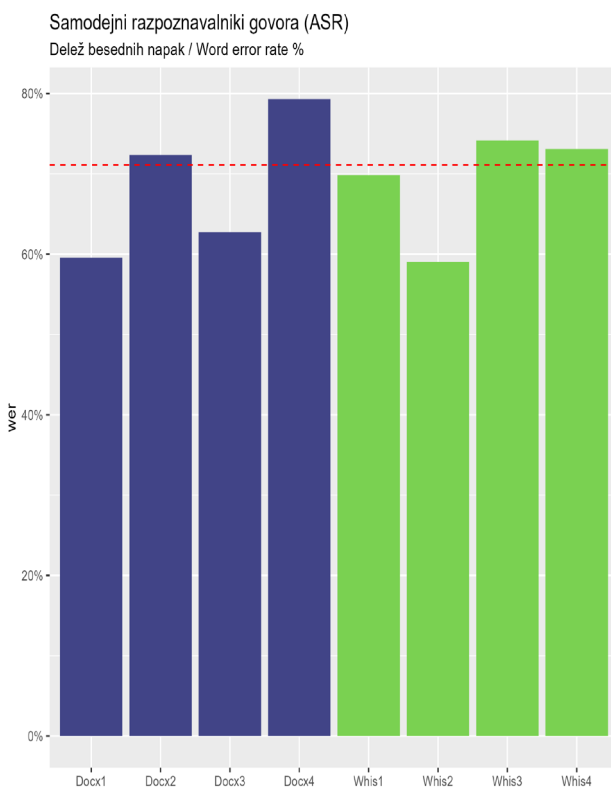
Tabela 4: Pogovor z nemškima otrokoma (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Vosk1	91	182	0	1277	1606	329
Vosk2	93	324	0	1742	2223	481
Whis1	86	331	19	1026	1606	598
Whis2	90	596	0	1406	2223	817
Docx1	90	261	0	1189	1606	417
Docx2	91	660	0	1361	2223	862

Vir: lasten

2.5.5 Petje: slovenske pesmi

Pesmi spadajo redko kdaj med učna gradiva orodij za razpoznavanje govora. Tudi za človeka predstavlja prepoznavanje besedila v opernem libretu ali pesmi v muziklu precejšen izziv. Zato lahko pričakujemo slabše izide kot pri prepoznavanju govora. V tem primeru je *Razpoznavalnik* povsem odpovedal, saj je namesto besedila izpisal samo dva vprašaja. Orodji *Whisper* in *Word Prepiši* sta sicer izkazovala višje deleže besednih napak kot pri studijskih ali filmskih posnetkih govora, vendar pa nižje (slika 5 in tabela 5) kot pri posnetkih spontanah pogovorov z otrokom v naravnih prostorih (slika 3). Izbrali smo štiri slovenske pesmi (*Nipke – Popoln lajf*, *Mi2 – Oda gudeki*, *N'toko – Seks v mestu*, *Trkaj – 1 mf 2*).¹⁹



Slika 5: Pesmi v slovenščini (deleži besednih napak, WER)

Vir: lasten

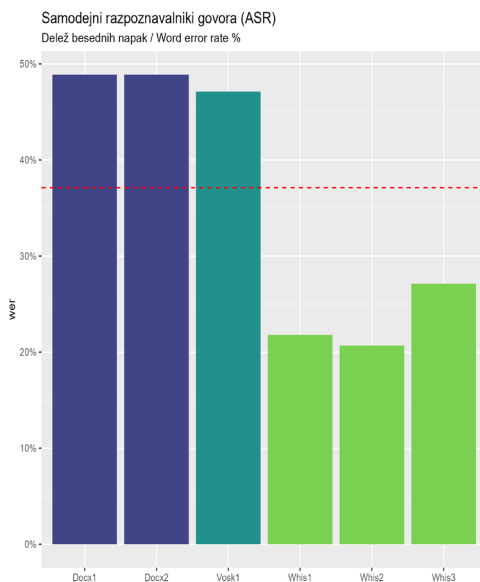
¹⁹ Uporabljena besedila so na naslovu <https://www.besedilo.si/> (31. 8. 2023).

Tabela 5: Pesmi v slovenščini (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Whis1	70	334	36	151	789	676
Whis2	59	277	26	21	586	586
Whis3	74	276	39	319	867	587
Whis4	73	677	184	178	1643	1545
Docx1	60	250	9	206	789	591
Docx2	72	193	1	229	586	359
Docx3	63	252	37	234	867	665
Docx4	79	543	22	718	1643	929

Vir: lasten

2.5.6 Nemški pogovorni jezik



Slika 6: Nemški pogovorni jezik (deleži besednih napak, WER)

Vir: lasten

Izbrali smo približno dvominutni odlomek *Sind deine Eltern Terroristen?* (Ali sta tvoja starša terorista?) iz nemškega filma *Zweihrküken*.²⁰ Pričakovali smo dobre izide kot v odlomku iz slovenskega filma (gl. zgoraj). Filmski odlomek se odvija v baru, v ozadju slišimo klavirsko glasbo, žvenketanje kozarcev in zamolke pogovore ljudi.

²⁰ *Zweihrküken* je režiral Til Schweiger (2009). Odlomek: <https://www.youtube.com/watch?v=yj4Bgij6AsI> (31. 8. 2023).

Gre za tri kratke pogovore med moškim in žensko, kot peti govorec pa sodeluje še točaj. Kljub zvočni kulisi so človeški glasovi večinoma razločni. Osebe se pogovarjajo v nemškem pogovornem jeziku brez izrazitega regiolekt. Deleži besednih napak, ki jih izkazujeta *Word Prepiši* in *Vosk*, sta opazno večja kot *Whisperjev* delež (slika 6).

Po podrobnejših podatkih v preglednici sodeč (tabela 6), sta *Vosk* in *Word Prepiši* izpustila znatno več besed kot *Whisper*. Tudi zamenjav je v *Whisperjevem* zapisu manj. Opazen je že zgoraj večkrat opisan “vedenjski” vzorec, da *Whisper* vstavlja več besed kot druga orodja. Če *Whisper* besednega gradiva ne prepozna ali če je računalniški procesor preobremenjen, si *Whisper* izmišljuje besede, besedne zveze ali cele povedi.

Lep primer za programske “halucinacije” je prav iz tega nemškega filmskega odlomka. Mlada Lana vpraša Moritza, ali bi kaj spil z njo (*Trinkst du was?*). Mlademu Moritzu pa se je zavozljal jezik in ni sposoben odgovoriti. Namesto jeclanja je *Whisper* petkrat zapored zapisal poved, da ne želi biti v razmerju z neko osebo (*Nein, ich hab keine Lust auf eine Beziehung.*). Ta poved se v tem filmskem odlomku ne pojavlja nikjer. Zdi se, da je *Whisper* vstavil poved, ki je v tej situaciji možna alternativa, da bi tako preprečil vrzel v besedilu. Glede na obsežnost programa je možno, da je bil računalniški procesor občasno preobremenjen, tako da *Whisper* del zvočnega vzorca ni mogel analizirati. *Whisper* je najbrž privzeto tako nastavljen za večjo ustvarjalnost, podobno kot ChatGPT. *Whisper* ima stikalo *temperature*. Če to stikalo ali parameter nastavimo na nižjo stopnjo ali celo na ničlo, potem začne tudi *Whisper* izpuščati več besed.

Tabela 6: Nemški pogovorni jezik (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Vosk1	47	98	0	114	450	336
Whis1	22	31	32	24	450	458
Whis2	21	24	6	60	450	393
Whis3	27	48	32	38	450	440
Docx1	49	79	21	110	450	353
Docx2	49	71	21	118	450	345

Vir: lasten

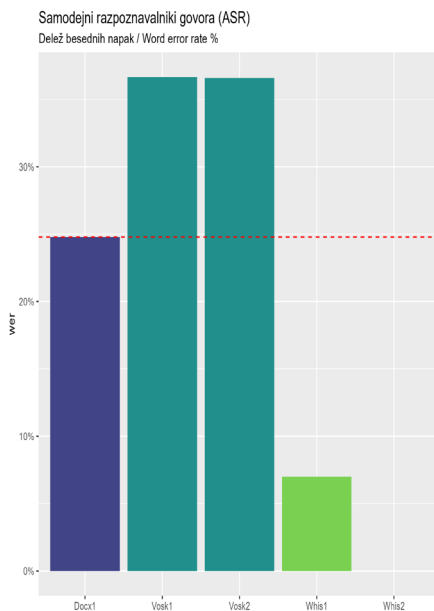
2.5.7 Nemška razprava o instagramu

Nemška pogovorna oddaja *13 Fragen* na temo *instagram*²¹ je značilen primer študijskega posnetka: odlična akustična kakovost, vsi govorniki so dobro slišni, govornicem se v glavnem ne zatika, ozadnega šuma je malo. Posnetek je trajal 35 minut. Referenčnega besedila v tem primeru nimamo. Zato smo se odločili, da zapise ostalih programov primerjamo z drugim *Whisperjevim* zapisom (*Whis2*). Tudi v tem primeru je opazno, da sta si privoščila *Voske* in *Word Prepiši* veliko število izpustov (tabela 7 in slika 7).

Tabela 7: Nemška razprava o instagramu (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Vosk1	37	1471	174	1223	7958	6890
Vosk2	37	1466	174	1223	7958	6890
Whis1	7	222	121	210	7958	7866
Whis2	0	0	0	0	7958	7958
Docx1	25	836	407	665	7958	7697

Vir: lasten



Slika 7: Nemška razprava o instagramu (deleži besednih napak, WER)

Vir: lasten

²¹ ZDF, *13 Fragen*, Filter, Fake und nackte Haut – Macht Instagram uns unglücklich?: <https://www.youtube.com/watch?v=-6smnitK4zs> (31. 8. 2023).

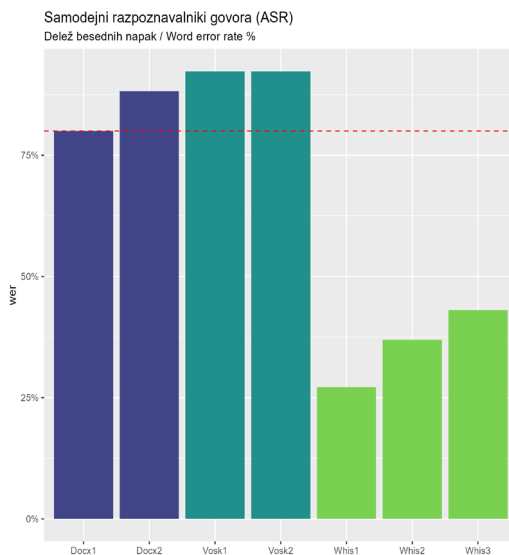
2.5.8 Petje: rock opera v nemščini

Pri zapisovanju slovenskih pesmi so se vsa orodja približno enako slabo odrezala ali celo odpovedala. Drugačno sliko vidimo na primeru nemške različice rock opere *Fantom iz opere*.²² Orodja smo preizkušali z naslovno pesmijo opere v nemščini. *Whisper* je izkazoval bistveno nižji delež besednih napak kot *Vosk* in *Word Prepisi*. Iz tabele 8 je razvidno, da je to povezano ali z večjim številom zamenjav, ali v drugih primerih z večjim številom izpustov.

Tabela 8: Petje v rock operi v nemščini (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Vosk1	92	35	0	145	195	50
Vosk2	92	97	0	83	195	112
Whis1	27	30	6	15	195	186
Whis2	37	36	2	31	195	167
Whis3	43	51	3	26	195	173
Docx1	80	51	6	86	195	122
Docx2	88	90	1	80	195	115

Vir: lasten



Slika 8: Petje v rock operi v nemščini (deleži besednih napak, WER)

Vir: lasten

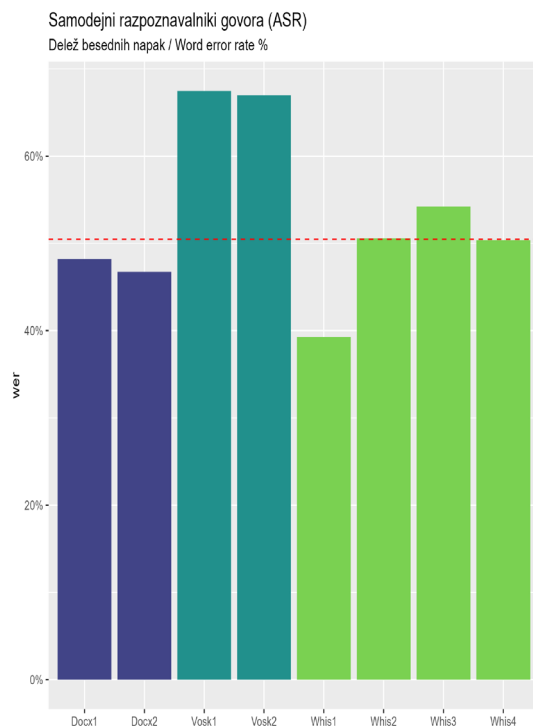
²² <https://www.youtube.com/watch?v=HJJPxwneN-wk> (31. 8. 2023).

2.5.9 Preklapljanje med jeziki

Tabela 9: Preklapljanje med jeziki (deleži besednih napak, WER)

ASR	wer	sub	ins	del	words.ref	words.hyp
Vosk1	67	74	0	206	415	209
Vosk2	67	73	0	205	415	210
Whis1	39	69	19	57	415	391
Whis2	51	93	17	83	415	360
Whis3	54	77	63	30	415	480
Whis4	50	86	21	84	415	362
Docx1	48	54	6	134	415	293
Docx2	47	54	6	134	415	287

Vir: lasten



Slika 9: Preklapljanje med jeziki (deleži besednih napak, WER)

Vir: lasten

Preizkušali smo, kako se orodja za prepoznavanje govora obnesejo, če osebe govorijo v različnih jezikih. Tovrstni izidi so še posebej zanimivi za raziskovalce, ki preučujejo preklapljanje med jeziki (code switching ipd.) in slenga. Glede na to, da so mnogi razpoznavalniki najuspešnejši v nekaterih svetovno bolj razširjenih jezikih

(v angleščini, španščini, francoščini, nemščini, ...) smo izbrali odlomek iz nadaljevanke *Kraljica Jaga* (*Queen of the South*²³), in sicer več kot sedem minut iz desete epizode pete sezone. V tej nadaljevanki je značilno mešanje angleščine in španščine. Prevladuje sicer angleščina, španščina pa prevzame pobudo, ko gre za čustveno napete trenutke in preklinjanje. Glede deleža besednih napak *Vosk* zaostaja za *Wordom Prepiši* in *Whisperjem* (tabela 9 in slika 9). *Whisper* je v enem preizkusu prednjačil pred *Wordom Prepiši*, v treh preizkusih pa je imel podobno slabe izide kot *Word Prepiši*. Deleži napak so bili pri vseh orodjih razmeroma visoki, ker smo izbrali odlomek, med katerim se je odvijal dvoboj med mehiškima akterjema in med katerim se je poleg preklinjanja in stokanja slišalo tudi pokanje strelnega orožja in drugi šumi.

2.5.10 Povzetek preizkusov

Tabela 10: Povzetek lastnosti programskih orodij ASR

Merila	Zemanta/UL Razpoznavnik	Microsoft Word Prepiši	Kaldi Vosk/Kaldi	OpenAI Whisper
Brezplačna uporaba	Da	Komercialna naročnina	Da	Da
Preprost vmesnik	Da	Da	Da (Subtitle Edit)	Da (Subtitle Edit)
Primerna hitrost	Da (strežnik)	Da (strežnik)	Da	Da (z GPU, pospeševalnikom)
Zmerne strojne zahteve	Da	Da	Da	Odvisno od velikosti modela
Varstvo osebnih podatkov	Neznano	Neznano	Da (lokalna uporaba)	Da (lokalna uporaba)
Uporabnost za več jezikov	Ne (samo slovenščina)	Da	Da (ni slovenščine)	Da
Prilagodljivost ob slabšem zvoku	Ne	Ne	Ne	Omejeno
Natančnost (knjižni jezik)	Visoka (knjižni jezik)	Visoka (knjižni jezik)	Zmerno visoka	Visoka (knjižni jezik)
Natančnost (neknjižni jezik)	Zmerno visoka	Zmerno visoka	Pod povprečjem	Zmerno visoka
Različne vhodne oblike	Mikrofon, zvočni vzorec	Mikrofon, zvočni in video vzorec	Zvočni vzorec in video	Zvočni vzorec in video
Izpis v različne oblike	Ne	Da	Omejeno (podnapisi)	Da
Časovni žigi	Ne	Da (z napakami)	Da	Da (z napakami)

²³ Ameriška nadaljevanka, ki sta jo razvijala M.A. Fortin in Joshua John Miller. Peta sezona je bila prvič predvajana 2021.

Merila	Zemanta/UL Razpoznavnik	Microsoft Word Prepiši	Kaldi Vosk/Kaldi	OpenAI Whisper
Razlikovanje govorcev	Ne	Da (z napakami)	Ne	Omejeno (Python in Whisperx)
Samodejno postavljanje ločil	Da	Da	Omejeno uporabno	Da
Uporabnost za daljše gradivo	Ne (do 300 sekund)	Omejeno (do 300 MB, 60 minut)	Da	Da (tudi več vzorcev zapored)
Programska prilagodljivost	Ne	Ne	Omejeno	Da (Python in Whisperx, Google Colab z GPU)

Vir: lasten (stanje 30. 4. 2023)

Tabela 11: Srednje vrednosti orodij ASR pri zapisovanju slovenskega govornega besedila²⁴

ASR	wer_m	sub_m	ins_m	del_m	quo_m
Docx	58	90	6	134	0.71
Kald	82	54	0	307	0.23
Razp	42	80	6	72	0.84
Whis	41	73	20	48	0.94

Vir: lasten

Tabela 12: Srednje vrednosti orodij ASR pri zapisovanju nemškega govornega besedila

ASR	wer_m	sub_m	ins_m	del_m	quo_m
Docx	58	90	6	134	0.71
Vosk	67	98	0	206	0.51
Whis	41	73	20	48	0.94

Vir: lasten

3 Sklep

Preizkušali smo več programskih orodij za prepoznavanje in pretvorbo govora v pisno obliko, ki naj bi skrajšala čas, potreben za ustvarjanje govornih korpusov. Uporabljali smo vrsto meril, ki zadevajo preprostost uporabe, časovni prihranek, morebitne stroške, zagotavljanje anonimnosti govorcev in različne vidike kakovosti pretvorbe. V sklepu strnemo nekaj izmed rezultatov, ki izpostavljajo prednosti ali slabosti pretvorbe slovenskega govora v pisno obliko:

²⁴ Kratice: wer_m, sub_m, ins_m, del_m = mediana deležev besednih napak in števil zamenjav, vstavitve in izbrisov vsakega orodja za prepoznavanje govora, quo_m = mediana količnika med številom besed prepoznane besedila in referenčnega besedila.

- natančnost prepoznavanja govora je v slovenščini slabše kot v angleščini, španščini ali nemščini, kar je mogoče povezovati z razmeroma majhnim in premalo raznovrstnim zvočnim gradivom v slovenščini za učenje modelov ASR;
- slabše prepoznavanje v slovenščini je opaznejše, kadar gre za pogovorni, otroški jezik, zapeto besedilo ali posnetke slabše zvočne kakovosti;
- postavljanje ločil je na zadovoljivi ravni (velja za: *Razpoznavalnik e2e*, *Word Prepiši*, *Whisper*);
- časovni žigi so netočni (razen pri *Vosk/Kaldi*);
- razlikovanje govorcev je sicer programsko mogoče, vendar ne deluje prav dobro (*Word Prepiši*, *Whisper*);
- najenostavnejše je razlikovanje govorcev z *Word Prepiši*;
- največ različnih izhodnih formatov ponuja *Whisper*;
- vhodna oblika je po navadi zvočna datoteka, vendar *Whisper*, *Word Prepiši* in *Vosk/Kaldi* sprejemajo tudi video posnetke;
- enostavna in hkrati vsestransko uporabna je kombinacija programov *Subtitle Edit* in *Whisper(x)*;
- največjo hitrost pretvorbe dosežemo s programom *Whisper(x)* na grafičnem procesorju (na lokalnem prenosniku ali npr. pri *Google Colab*);
- zaporedno pretvorbo več zvočnih datotek je mogoče s programom *Whisper*;
- na internetu je več predlog za ustvarjanje računalniških skriptov (npr. v Pythonu) za pretvorbo govora s programom *Whisper*;
- v različnih preizkusih je *Whisper* v povprečju naredil najmanj besednih napak;
- največ jezikov (tudi slovenščino) poznata *Word Prepiši* in *Whisper*.

Novejša orodja za pretvorbo govora v pisno obliko so vidno napredovala. Kljub vsemu napredku bi si jezikoslovci vsekakor želeli, da bi lahko programe za samodejno prepoznavanje govora in pretvorbo v pisno obliko prilagajali posameznim raziskovalnim potrebam (npr. vključitev diskurzivnih označevalcev ali dejansko izgovorjenih skrčenih besednih oblik v zapis).

Literatura

- Iztok, LEBAR, Marko BAJEC, Žan BAJEC, Mitja RIZVIČ, 2022: *Slovene Conformer CTC BPE E2E Automated Speech Recognition Model RSDO-DS2-ASR-E2E 2.0*. <http://hdl.handle.net/11356/1737> (31. 8. 2023).
- Teodor PETRIČ, 2016: Dolgoročna raziskava o razvoju otroškega govora: Slovenske Samostalniške Sklanjatve. *Zbornik Prispelkov s Simpozija 2015*. Ur. Franc Marušič, Petra Mišmaš, Rok Žaucer. Nova Gorica: Založba Univerze. 91–112. http://www.ung.si/media/storage/cms/attachments/2016/10/21/13/45/01/Zbornik-%C5%A0D9_okt.16_splet.pdf (31. 8. 2023).
- Teodor PETRIČ, 2021: Razvoj slovenskih glagolskih oblik in spregatev na primeru predšolskega otroka. *Škrabčevi Dnevi 11: Zbornik Prispelkov s Simpozija 2019*. Ur. Franc Marušič, Petra Mišmaš, Rok Žaucer. Nova Gorica: Založba Univerze. 78–101. <http://www.ung.si/media/storage/cms/attachments/2021/01/27/12/29/22/Zbornik-%C5%A0D11-2021-3.pdf> (31. 8. 2023).
- Daniel POVEY, Ghoshal ARNAB, Gilles BOULIANNE, Lukas BURGET, Ondrej GLEMBEK, Nagendra GOEL, Mirko HANNEMANN, et al., 2011: The Kaldi Speech Recognition Toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US: IEEE Signal Processing Society*. DOI: <https://doi.org/10.1017/pan.2018.62> (31. 8. 2023).
- Sven-Oliver PROKSCH, Christopher WRATIL, Jens WÄCKERLE, 2018: Testing the Validity of Automatic Speech Recognition for Political Text Analysis. *Political Analysis* 27/3, 339–359.
- R Core Team, 2023: R: *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/> (31. 8. 2023).
- Cvetka ŠERUGA-PREK, Emica ANTONČIČ, Ajda KALAN, Ivan LOTRIČ, 2004: *Slovenska Zborna Izreka. Aristej*.
- Ashish VASWANI, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan N GOMEZ, Łukasz KAISER, Illia POLOSUKHIN, 2017: Attention Is All You Need. *Advances in Neural Information Processing Systems*. Ed. I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett. Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf (31. 8. 2023).
- Darinka VERDONIK, Andreja BIZJAK, Mirjam SEPESY MAUČEC, Lucija GRIL, Simon DOBRIŠEK, Janez KRŽAJ, Gregor STRLE et al., 2023: *ASR Database ARTUR 1.0 (Transcriptions)*. <http://hdl.handle.net/11356/1772> (31. 8. 2023).

PREDLOG IZDELAVE KORPUSA HUMORJA V GOVORU ZA SLOVENŠČINO

MIRA KRAJNC IVIČ,¹ ŠPELA ANTLOGA²

¹ Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija

mira.krajnc@um.si

² Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko,

Maribor, Slovenija

s.antloga@um.si

V prvem delu prispevek predstavlja teorije humorja (superiornosti, sproščanja in neskladja), kot njihovo nevtralno združitev in dopolnitev Raskinovo semantično teorijo humorja ter njeno nadgradnjo, tj. Attardovo splošno teorijo verbalnega humorja (v nadaljevanju GTVH), namenjeno analizi šal. A humor v konverzaciji nastaja tudi spontano in v večji odvisnosti od neposrednega konteksta, zato je na kratko predstavljeno teoretično ozadje konverzacijskega humorja. Humorno je po teoriji superiornosti omejeno z nacionalnimi mejami in s časom, odvisno je od zvestobe skupini, identifikacije z njo. Raziskovanje konverzacijskega humorja v slovenskem jeziku je torej nujno. To bi lahko pospešila izdelava ustreznega korpusa. V drugem delu prispevka so zato pod drobnogled vzeti korpusi humorja, in sicer: v katerih jezikih so na voljo govorni, pisni ali multimodalni korpusi, označeni kot humorni; kako obsežni so ti korpusi, katero gradivo zajemajo, kateri od njih so prosto dostopni, kako je potekalo označevanje humornih pasаж, kateri jezikovni pojavi in drugi nejezikovni kodi so bili označevani. V zaključku predstavlja rezultate ankete, ki sva jo izvedli z namenom presoje, ali je lahko tovrstna metoda učinkovita za določanje humornih pasаж in ali lahko pomaga pri presoji označevalca, katera jezikovna ali nejezikovna sredstva vplivajo na humornost izreka oz. pasажe.

DOI

[https://doi.org/
10.18690/um.ff.4.2024.10](https://doi.org/10.18690/um.ff.4.2024.10)

ISBN

978-961-286-882-6

Ključne besede:

humor,
teorije humorja,
govorjeno besedilo,
korpus,
slovenščina



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.10](https://doi.org/10.18690/um.ff.4.2024.10)

ISBN
978-961-286-882-6

Keywords:
humor,
humor theory,
spoken text,
corpus,
Slovene

SPOKEN SLOVENE CORPUS OF HUMOR: A DRAFT PROPOSAL

MIRA KRAJNC IVIČ,¹ ŠPELA ANTLOGA²

¹ University of Maribor, Faculty of Arts, Maribor, Slovenia
mira.krajnc@um.si

² University of Maribor, Faculty of Electrical Engineering and Computer Science,
Maribor, Slovenia
s.antloga@um.si

The first part of the paper presents the theories of humor (Superiority Theory, Release Theory, and Incongruity Theory), and as their neutral integration and complementation, Raskin's Semantic Theory of Humor, as well as its extension, Attard's General Theory of Verbal Humor developed for analysing jokes. However, humor in conversation also arises spontaneously and depends more on the immediate context. Therefore, we also present a brief theoretical background of conversational humor. According to the Superiority Theory, humor is limited by national boundaries and time; it depends on loyalty to a group and identification with it. Research on conversational humour in the Slovene language is therefore essential. This could be accelerated by developing a suitable corpus. The second part of the article examines corpora of humor, namely: which languages have available spoken, written or multimodal corpora characterized as humorous; their size, material they cover, accessibility, annotation procedure of humorous segments, and which linguistic phenomena and other non-linguistic instruments were annotated. In conclusion, we present the survey results to help us evaluate whether the method used can be effective for determining the humor passages and deciding about the linguistic or non-linguistic means that affect the humorousness of the utterance or passage.



1 Uvod¹

Humor je človeški in družbeni pojav (Bergson 1977: 12). Zaradi odvisnosti od konteksta zahteva ustvarjalno razumevanje ubesedenega in situacije. To pa je mogoče le, ko človek lahko vidi preko trenutne (neugodne) situacije. Humor je človekov notranji mehanizem za preraščanje težav. V nasprotju s šalo, s katero tvorec skuša pridobiti le ugodje ali pa postavlja ugodje v službo agresije, humor ne vsebuje le nečesa sproščujočega, ampak nekaj veličastnega in zmagovitega, zato je dostojanstven (Freud 1991: 94). Humor je kot družbeni pojav v smislu odziva na politično-družbena dogajanja v porastu. Veča oziroma izraža družbeno angažiranost posameznika ali skupnosti. Temu porastu raziskave ne sledijo v celoti. Razvijanje humorne zmožnosti zahteva spontano in načrtno opazovanje, samoopazovanje, urjenje in spoznavanje tudi teoretičnega ozadja o humorju. Njegovo raziskovanje je multidisciplinarno in multimodalno. Vključuje številne discipline, kot so psihologija, filozofija, sociologija, komunikacija in literarne vede, a se pogosto zgodi, da te discipline prezrejo (jezikovno) teorijo humorja ali pa so brez nje (Attardo, Raskin 2017: 51). To naj bi se spremenilo po uveljavitvi Raskinove semantične teorije humorja, tj. po letu 1985 oz. 1991 (Attardo, Raskin 2017: 51), in njene nadgradnje v splošno teorijo verbalnega humorja. Kljub temu pa je humor na področju slovenistike še vedno raziskovan le v okrnjeni pojavnosti. Nekaj raziskav humorja najdemo pri U. Jarnovič (Jarnovič 2008), M. Stritar (Stritar 1999/2000) in M. Krajnc Ivič (Krajnc Ivič 2011, 2013, 2017). Drugi avtorji slovenističnega jezikoslovja (Jesenšek 2010; Stramljič Breznik 2002) so humor analizirali z vidika neskladja, dvoumja na ravni pomena v umetnostnih besedilih, znanih kot humorna. Avtorji npr. komunikologije, etnologije, folkloristike in sorodnih ved se te tematike redko lotevajo (Babič 2018), izjema je slovenska literarna zgodovina in literarna teorija (npr. Zadavec 1991, Borovnik 1997), vendar tudi ta humorja ne raziskuje z vidika GTVH in ob upoštevanju multimodalnosti besedila. Na področju slovenistike bi bilo treba zlasti na področju govornega diskurza sistematično raziskati, katera semantična sredstva (metafora, ironija, hiperbola itd.), pragmatične strategije (odnos do vsebine, sogovorca ...), kateri žanri, katero znanje in katera neverbalna sredstva (mimike, kretnje) ter druge nejezikovne kode (barva glasu) uporabljamo oz. izkoriščamo v konverzaciji za doseganje humornega učinka, in tudi kako se vsa ta omenjena sredstva in različni kodi medsebojno prepletajo ter dopolnjujejo.

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

Namen prispevka je ob predstavitvi temeljnih teorij humorja in s predlogom za izgradnjo korpusa humorja deloma pripraviti gradivno ozadje za raziskovanje ugotovljenih vrzeli pri preučevanju konverzacijskega humorja v slovenščini.

2 Teorije humorja

Pri raziskovanju humorja se pokaže troje: 1) da beseda humor v preteklosti ni imela pomena, kot ga poznamo danes, 2) da je bilo humorno, šaljivo vezano na telesno, kar je zlasti z vidika religij predstavljalo nizko, nevredno pozornosti, in 3) da je bilo jezikoslovje z vidika raziskovanja humorja vedno obrobno, manj pomembno področje, vodilno področje pri raziskovanju humorja je bila psihologija (Attardo, Raskin 2017: 50), spregledati pa ne gre vloge filozofije (Bergson, Kant, Descartes idr.). Glede na to so se razvijale klasične teorije humorja (teorije superiornosti, teorije sproščanja in teorije neskladja). Kot ugotavlja Raskin (1985: 40–41 po Attardo, Raskin 2017: 51), te tri teorije humor opisujejo z zelo različnih vidikov in se pri tem dopolnjujejo. Teorija neskladja podaja izjave o dražljaju; teorija superiornosti označuje odnose ali stališča med tvorcem in naslovnikom; teorija sproščanja pa komentira naslovnikove občutke in njegovo psihologijo (Attardo, Raskin 2017: 51).

Za področje jezikoslovja sta ključni na scenariju temelječa Raskinova semantična teorija humorja, ki je zasnovana kot nevtralna glede na vse te teorije in je z njimi združljiva, ter njena nadgradnja, to je Attardova splošna teorija verbalnega humorja.

2.1 Klasične teorije humorja: teorija superiornosti, teorija sproščanja in teorija neskladja²

2.1.1 Teorija superiornosti

Za teorijo superiornosti, imenovano tudi teorija nadrejenosti, je pomemben *Esej o smehu* Henrija Bergsona (1977), ki enako kot Descartes vidi smeh kot družbeni dejavnik, tj. smeh je vedno smeh skupine, ki ga ne moremo uživati, kadar se čutimo osamljene, ločene od ostalega sveta, čeprav se smeh izživi na nekakšni brezčutnosti (Bergson 1977: 12–14).

² Če ni označeno drugače, je besedilo povzeto po Cristina Larkin-Galiñanes (2017).

Za to skupino teorij humorja je pomembno razumevanje smeha. Ta je sredstvo, s katerim naj bi se posameznik osvobodil negativnih, z vidika produktivnosti posameznika v skupnosti nezaželenih nagonov (agresije, spolnosti itd.). Za te nagone civilizacija in družba od posameznika pričakujeta, da jih zatre. Smeh zato zgodovinsko gledano predstavlja nekaj vulgarnega, grešnega, motečega. Lahko je sredstvo družbenega nadzora, celo glavna naloga smeha je nadzorovati ljudi. Posameznika namreč že potencialna možnost, da bi bil predmet posmeha, omejuje pri dejanjih, besedah, čustvovanjih in razmišljanjih in mu preprečuje, da bi se oddaljil od tega, kar v družbi velja za normalno, primerno in spodobno. Vendar pa človeka napake drugih spravljajo v smeh zaradi svoje nesocialnosti in ne zaradi svoje nemoralnosti. Čeprav smeh po eni strani pomeni 'ustrahovati s poniževanjem', to je nasprotnika narediti majhnega, manj vrednega, vrednega posmeha ali preprosto komičnega, to po drugi strani pomeni, da je kaj humorno le do te mere, da izboljša predmet naklonjenosti; z drugimi besedami tistega, ki se smeje, predstavi v boljši luči. Od tod izvira poimenovanje te teorije kot teorije superiornosti.

Ali je kaj humorno/smešno oz. ali je kdo humoren/smešen, je tako odvisno od zvestobe skupini, od identifikacije z določeno skupino in od tega, v kolikšni meri imajo udeleženci komunikacijskega stika skupno družbeno ozadje. Iz tega izhaja, da je humor pogosto omejeno z nacionalnimi mejami in s časom. Humor je v veliki meri vprašanje referenc skupin in afektivne naravnosti (Larkin-Galiñanes 2017: 9).

A prav dejstvo, da je smeh sredstvo, s katerim naj bi se posameznik osvobodil nagonov (ki jih družba ne tolerira), hkrati tudi pomeni, da smeh predstavlja obrambni mehanizem, kar je osnova teorij sproščanja.

2.1.2 Teorija sproščanja

Teorija sproščanja predstavlja psihološki in fiziološki pristop k razumevanju humorja. Povezavo med humorjem in zdravjem sicer omenjajo že klasični avtorji (Aristotel, Cicaro), saj humor prinese užitek in sprostitiv. Ko s smehom nasprotnika naredimo manj vrednega ali komičnega, dosežemo užitek in moč, saj smo ga premagali. S Hipokratovimi tipi osebnosti pa je bila vzpostavljena tudi povezava med smehom in humorjem. Sangvinik kot tip osebnosti je optimist, vesel in ljubitelj zabav teržitka. V tem smislu je humor posebna ali prevladujoča lastnost, ki tvori temelj človekovega značaja.

Za teorijo sproščanja so pomembna spoznanja Immanuela Kanta in Herberta Spencerja. Prvi navaja, da je smeh naklonjenost, ki nastane zaradi nenadne preobrazbe napetega pričakovanja v nič. To opredelitev povzame drugi v opisu delovanja smeha, ko navaja: »V primeru smeha se zgodi, da se razvijejo čustva, ki se nato pokažejo kot neprimerna, tako da se čustvo spremeni iz močnega v šibko, tako nastala odvečna energija pa se sprosti v vrsti mišičnih gibov (smeš), ki ne vodijo v nobeno dejanje, temveč služijo le sprostitvi živčne energije, podobno kot to počne odprtje varnostnega ventila v parni cevi.« (Larkin-Galiñanes 2017: 10). V posamezniku se tako ustvari ravnovesje ali katarza, kot bi to poimenoval Aristotel.

Poleg koncepta sproščanja je za razumevanje humorja pomemben še koncept vzburjenosti. Šala povzroči neprijetno stanje rahle vzburjenosti. To stanje je nastalo zaradi nelagodja, zmedenosti, napetosti, tudi strahu. Tako nastala vzburjenost se v varnem okolju sprosti, zmanjša. Da pa sploh nastopi sprostitvev, mora obstajati določeno stanje duha/uma, in sicer splošno veselo razpoloženje, nagnjeno k smehu, in pričakovanje komičnega.

Čeprav teorija sproščanja v prvi vrsti pojasnjuje nastanek smeha in čeprav njena smiselnost pri besednem humorju ni vedno očitna, prav ta poudarja spoznanje, da je za razumevanje humorja potreben predhodni obstoj ugodnega razpoloženja.

2.1.2 Teorija neskladja

Lahko bi rekli, da to, kako se smejimo in čemu se smejimo, velja za kazalnik izobrazbe, občutljivosti za druge in družbenega razreda, a že od Aristotela dalje si raziskovalci smeha in humorja prizadevajo analizirati in opredeliti, kateri drugi dejavniki poleg posmeha in norčevanja prispevajo k zabavi. Pojem 'zabava' je v primerjavi s teorijo superiornosti in teorijo sproščanja vpeljan »šeše« pri teoriji neskladja. Teorija neskladja je kognitivni pristop k razumevanju humorja. Ta teorija se ukvarja z 'duhovitostjo' in ne z naključnim ali priložnostnim smehom, ki ga povzročijo situacija ali osebne značilnosti; ukvarja se z nasmehom in ne s smehom. In to sprva počne s pragmatičnega vidika zabave, ki ima kot komunikacijski dogodek cilj zabavati, s čimer ustvari nepričakovano neskladje. Neskladnost pomeni zaznavanje in odkrivanje odnosov med idejami ali stvarmi, ki se na prvi pogled zdijo popolnoma nepovezane in med katerimi je um s svojo prefinjenostjo ter hitrostjo odkril razmerja (Larkin-Galiñanes 2017: 13). Tu je treba omeniti Sigmunda Freuda, ki je predvidel teorijo Victorja Raskina o zamenjavi scenarijev. Freud namreč navaja,

da svoje zaznave organiziramo v okviru tako imenovanih abstraktnih pojmov, ki pogosto združujejo različne stvari pod isto oznako, tako da zelo različne predmete pogosto poimenujemo z isto besedo. Dodati je treba še Bergsonovo ugotovitev, da je situacija komična, ko hkrati pripada dvema povsem neodvisnima nizoma dogodkov in jo je mogoče hkrati razumeti v dveh popolnoma različnih pomenih. S tem je neskladju pripisal družbeno, kontekstno dimenzijo. A samo neskladje ni dovolj, potreben je še element nepričakovanosti ali presenečenja. Zato je vrednotenje humorja zasnovano kot dvostopenjski proces, ki vključuje odkritje neskladja. Temu odkritju sledi razrešitev neskladnosti z uporabo drugačnega kognitivnega pravila. Ta razrešitev je tisto, kar ločuje humor od tistega, kar bi sicer bilo preprosto nesmisel.

Jezikovna sredstva, s katerimi lahko izzovemo smeh/humor, so enakozvočnice, sopomenke, nekonvencionalna raba jezika, nemogoče hiperbole in pretiravanje na splošno, besedne igre, ironija, dvoumnost, nepričakovani preobrati, risanje čudnih podobnosti ali različnosti, primerjave in metafore ... Jezikovna sredstva ustvarijo niz, sestavljen iz vnosa (ubeseditvev), napovedi verjetne vsebine, potrjevanja ali nepotrjevanja, iz novega vnosa in iz prilagajanja razumevanja ubesedenega novemu vnosu itd. Ta niz je enak za katero koli besedilo, a pri šali naslovnik pride do točke, običajno na koncu, ko se njegova pričakovanja nenadoma ne potrdijo, nastalega neskladja, ker je besedila konec, pa ni mogoče razrešiti (glej dalje 2.2.2 Konverzacijski humor po H. Kotthoff). Naslovnik se zato oprime reševanja nastale težave tako, da se vrne na mesto v jedru besedila/šale, ki mu neskladje pomaga razrešiti. Gre za zaznavanje entitete (situacije ali ideje) v dveh referenčnih okvirjih ali 'valovih dolžinah' oziroma v dveh asociativnih kontekstih, kar je znano kot bisociacija, tj. povezava med dvema asociativnima kontekstoma, ki sta običajno nezdružljiva.

2.2 Jezikoslovne teorije humorja

Za področje jezikoslovja sta pomembni Raskinova semantična teorija humorja in njena nadgradnja, tj. Attardova GTVH. Ker obe teoriji v ospredje postavljata semantični scenarij, sta predstavljeni skupaj. A ti teoriji sta osredotočeni na analizo kratkih šal, njuna uporaba za analizo drugega gradiva, npr. konverzacije, zahteva nadgradnjo (Attardo, Raskin 2017: 51). Zato in zato, ker je žarišče prispevka t. i. konverzacijski humor, ki v konverzaciji nastaja spontano (Kotthoff 1998) ali pa je predoblikovan (Dynel 2009) in je njegovo razumevanje še bolj vezano na neposredni

kontekst ter tudi na druge nejezikovne kode za doseganje humornega učinka, je na kratko predstavljen tudi pojem konverzijskega humorja.

2.2.1 Semantična teorija humorja in GTVH

Raskinova semantična teorija humorja je – kot rečeno – nadaljevanje in povezava vseh treh velikih skupin teorij humorja. Ta teorija pri analizi humorja v ospredje postavi semantiko in pragmatiko. Njena težava je – tako Attardo in Raskin (2017: 53) –, da si tudi jezikoslovci težko predstavljajo, kaj pomeni besedna zveza 'neskladje dveh scenarijev'. Čeprav je scenarij na najpreprostejši ravni enak leksikalnemu pomenu besede (Attardo 2001: 3), je to za razumevanje tako pojma 'scenarij' kot 'neskladje' premalo. Glavna predpostavka semantične teorije in GTVH je, da je humor mogoče analizirati z istimi orodji, ki so potrebna za predstavitev pomena/smisla katerega koli, tj. tudi nehumornega besedila (Attardo, Raskin 2017: 51).

2.2.1.1 Scenarij kot element znanja

Scenarij je kognitivna struktura (Attardo 2001: 2), element znanja o svetu, ki ga sestavlja veliko število rutin o tem, kako kaj delamo, vidimo (Attardo, Raskin 2017: 53).³ Vsebuje informacije, ki so prototipične za opisovano entiteto (Attardo 2001: 3). Po Raskinu (Attardo 2001: 3) je scenarij neposredno povezan z leksikalnimi enotami: določen leksem priključuje določen scenarij. To pomeni, da scenarij ni le izkustvena/spoznavna enota. S psihološkega vidika so prisotnost oz. resničnost scenarija ugotavljali eksperimentalno v raziskavah, ki so vključevale priklic dogodkov v zgodbi ali priklic zgodbe z dogodki v drugačnem vrstnem redu od običajnega. Ugotovljeno je bilo, da so se udeleženci eksperimenta spomnili dogodkov, ki so bili v scenariju, tudi če se v dejanski zgodbi niso zgodili, hkrati pa so ti udeleženci spremenili vrstni red dogodkov tako, da je ta bil skladen s scenarijem (p. t.). Čeprav se scenariji aktivirajo med semantično in pragmatično obdelavo besedila, se lahko bistveno razlikujejo od površinske uresničitve besedila, saj je interpretacija besedila (s strani tvorca, naslovnika ali raziskovalca) vedno nujno konstrukt interpreteta (Attardo 2001: 7).

³ Scenarij kot primer globalnega vzorca omenjata tudi Beaugrande in Dressler (1992: 69). Opredelita ga kot ustaljen načrt, ki ga aktiviramo z namenom določanja vlog in določanja pričakovanega ravnanja udeležencev komunikacijskega stika. Za scenarij je značilna vnaprej utrjena rutina (p. t.).

Attardo (2001: 8–9) pojasnjuje, da je scenarij del večjega scenarija, ki povezuje vsa semantična vozlišča oz. scenarije, značilne za določeno družbo, kulturo. To trditev pojasni na primeru razmerja med scenarijema 'mati' in 'oseba' – *mati* je podpomenka leksema *oseba*.

Scenariji so torej povezani preko medsebojnih vezi (medleksemska pomenska razmerja). Zanimivo je, da so te vezi lahko različnih dolžin, kar kaže na to, da se določeni vozli pogosteje asociativno aktivirajo kot drugi (Attardo 2001: 8). Vsi scenariji, vozli in vezi med njimi tvorijo semantično omrežje, ki vsebuje vse informacije, ki jih ima govorec nekega jezika o svoji kulturi. Nova informacija lahko scenarij predrugači, zato so scenariji odprto-zaprti (Attardo 2001: 6).

Po Raskinu naj bi semantično teorijo sestavljali seti vseh scenarijev, s katerimi razpolaga tvorec, in seti pravil njihovega kombiniranja. Naloga teh pravil je združiti vse možne pomene scenarijev in zavreči tiste kombinacije, ki ne omogočajo koherentnega razumevanja besedila (Attardo 2001: 9). Razumevanje besedila pomeni hranjenje uspešnih, koherentnih kombinacij in vključevanje drugih prav tako koherentnih kombinacij, dokler niso obdelane vse besedilne enote. Če obstaja vsaj eno koherentno, dobro oblikovano razumevanje – Attardo tu (2001: 9) uporabi celo izraz interpretacija –, je to razumevanje besedila potrjeno kot besedilni pomen, semantična teorija pa stavek/besedilo oz. besedilno enoto označi kot 'dobro oblikovan/-o'. Semantičnemu razumevanju sledi pragmatično razumevanje oziroma interpretacija, ki iz konteksta semantičnemu besedilnemu pomenu doda inference, implikacije itd. (Attardo 2001: 9 po Raskin 1985: 80). Attardo (p. t.) doda, da je ločevanje semantičnega razumevanja od pragmatičnega dodano zaradi jasnosti.⁴

2.2.1.2 GTVH in humorno besedilo

Kot omenjeno, je GTVH prvotno nastala za analiziranje šal (Attardo, Raskin 2017: 51). Šala je kratko pripovedno besedilo, ki vsebuje napoved (angl. *set up*) in poanto (angl. *punch line*) (Attardo 2001: vii, Dynel 2009). Po GTVH mora šala kot humorno besedilo izpolnjevati dva pogoja: 1) deloma ali popolnoma mora biti združljiva z dvema različnima scenarijema (npr. alkohol in marihuana oba povzročata odvisnost

⁴ O tvorjenju smisla besedila s pragmatičnega vidika glej P. Grice (1989: 31), o povezavi med semantiko in pragmatiko pa tudi G. Deleuz po T. Erzar (1997: 64–69), ki navaja, da smisel ni lociran v nobeni od treh dimenzij jezika (semantika, skladnja, pragmatika), saj te dimenzije niso ločene, ampak si skupaj s smislom, ki kroži med njimi, podajajo tudi te vidike pomena.

in sta drogi), 2) oba scenarija, čeprav združljiva (odvisnost, droga), morata hkrati v določenem segmentu biti nasprotna (dobro/slabo; prepovedano/neprepovedano). Poleg opozicije dobro/slabo Attardo (2001: 20) po Raskinu navaja še štiri za človekovo življenje bistvene opozicije: življenje/smrt, spodobno/nеспodobno, denar/ne-denar, visok/nizek položaj. Doda pa tudi še dodatna nasprotja po Chlopickem: odsotnost/prisotnost, nujnost/nenujnost oz. neizogibnost/izogibnost, veliko/malo.

Temeljna razlika med Raskinovo semantično teorijo humorja in GTVH je v uvedbi petih drugih virov znanja, ki jih je treba poleg scenarija uporabiti pri tvorjenju šale. Ti dodatni viri so: logični mehanizem (pretiravanje, analogije, zavajanje, nepričakovani preobrati ...), tarča, pripovedna strategija, jezik in situacija. GTVH tako poleg leksikalnega scenarija vključuje teorijo naratologije in pragmatično teorijo v še večjem obsegu (Attardo 2001: 22).

2.2.3 Konverzacijski humor po H. Kotthoff

Konverzacijski humor je krovni termin za obsegovno različne humorne enote, ki so predoblikovane ali v interakciji tvorjene spontano (Dynel 2009). Helga Kotthoff (1998: 43) navede 6 definicijskih prvin za opredelitev konverzacijskega humorja. Te prvine so: a) za vse šaljive dejavnosti je značilno, da jih lahko spodbudi smeh ali da se z njim lahko zaključijo; smeh ni nujni pogoj ali rezultat, temveč le možen, b) med šaljivimi dejavnosti se igra z ustaljenimi/ouveljavljenimi pomeni in vsebinami, c) sestavni deli humorističnih dejavnosti so dvoumje, razmišljanje na več nivojih hkrati (bisociacija /glej 2.1.2 Teorija neskladja/) in interpretativna raznolikost, č) humor deluje na osnovi kakršnega koli skupnega védenja, d) pri humorju imajo pomembno vlogo domišljija, ustvarjalnost in igra, f) šaljenje omogoča sklepanja o družbi, občutih in zgradbi vrednot udeležencev komunikacijskega stika in g) humor najbolje deluje v veselem oz. sproščenem vzdušju.

Omenjena avtorica (1998: 46) izraze šala, humor in šaljivo komuniciranje uporablja kot nadpomenke za vse fenomene, ki jih lahko označimo kot 'neresne' v konverzaciji. Ne izgone se niti razmejitvi med termini smešno, humor in komično. Smešno se po njenem mnenju posreči s prekrivnostjo pomena, ki ustvari zabavno sprostitev, in ni omejeno le na šalo kot tipično šaljivi žanr, ustvari pa poanto in presenetljiv pomenski zasuk. Za humor navaja, da ta praviloma označuje določeno držo in čustveno stanje, ko lahko cenimo smešno, komično in ustvarjamo veselo, vedro vzdušje. Humor se nanaša tako na proces tvorjenja kot proces interpretiranja ubesedenega, čeprav se

lahko zgodi, da je humorno lahko razumljeno tudi kot resno. Če ji je humorno lahko le namerno, za komično to ne velja. Komično se večinoma nanaša na vidik razumevanja, interpretiranja dejanj, misli idr. drugih. Prav v konverzaciji so pogosta prekrivanja komičnega, smešnega in humorja. Humor lahko vključuje komično in smešno. Smešno potrebuje nepričakovano razmišljanje vsaj na dveh nivojih o nečem, kar se zdi nekompatibilno (Kotthoff 1998: 46). Avtorica meni, da humor ne temelji tako zelo na besedni semantiki, kot to poudarjata semantična teorija humorja in GTVH, ampak bolj na igri med različnimi oblikami tipizacij, institucionalizacij in normiranji. Pravi, da za smešno niso konstitutivne temeljne semantične opozicije (glej 2.2.1.2), ampak nasprotno, konstitutivne so stranske, presenetljive in ustvarjalne bisociacije – razmišljanja na več nivojih – o različnih področjih vedenja in vedénja. Če komuniciranje sicer poteka v dobri veri (angl. *bona-fide-communication*),⁵ humor, šaljenje v komuniciranje tega principa ne vključujeta. Še več, šaljivo komuniciranje sodi med temeljne oblike implikativnega komuniciranja (Kotthoff 1998: 45). S tega vidika avtorica izpostavi prav Griceovo teorijo pogovornih sklepanj, ki je primerna za razumevanje humornih besedil. Maksime Griceovega sodelovalnega načela se tudi za doseganje humorne učinka praviloma izkoriščajo. Za humor je značilno še, da udeleženca komunikacijskega stika v komunikacijskem procesu sodelujeta, a v primerjavi s sodelovalnim načelom imajo zdaj pomembnejšo vlogo druge, družbeno pogojene determinante konteksta. Sodelovalno načelo se mora nemudoma po neupoštevanju maksim ponovno vzpostaviti. Načelo, ki to poudarja, Attardo (2001: 111) imenuje načelo najmanjše količine motenj.

Kot izkoriščanje maksime kakovosti Grice navaja tudi metaforo in ironijo. Za ironijo je bistveno, da so jezikovna sredstva 'le' ironično uporabljena. Ironija je »ena najzlahtnejših oblik človekovega intelekta. Je kazalnik ustvarjalnega duha« (Korošec 2002: 25, tudi Freud 1991: 94). V smislu Ducrotove polifonične teorije izjavljanja Ducrotu (1988: 209) govoriti ironično pomeni, da tvorec predstavlja izrečeno kot stališče, za katero ne odgovarja, saj ima tvorec to stališče za nesmiselno. Hkrati pa po Olgi Kunst Gnamuš (1984: 16 po Habermas) ironičen izrek uporabljamo takrat, kadar želimo doseči svoj namen, ne da bi se želeli zameriti naslovniku, na katerega bi neironičen izrek utegnil imeti negativen učinek. Ubесedeno mora biti glede na naslovnika tvorjeno tako, da upošteva postavke o naslovnikovih vrednotah in normah. Po Attardu (2001: 111) je moč ironije v čustveni dinamiki, ki je sprejemljiva za vse udeležence komunikacijskega stika.

⁵ O tem sicer piše tudi že Raskin (1984: 100).

3 Korpusi humorja

Ker je humor pomemben del kulture vsakega naroda, je ključno, da je preučevan sistematično in z različnih vidikov na gradivu iz lokalnega okolja. Korpusno jezikoslovje je od 80. let 20. stoletja v slovenskem prostoru pomembno sooblikovalo raziskovanje jezika prav z vidika metodičnega pristopa s pisnimi in z govornimi korpusi, kasneje pa so začeli nastajati tudi multimodalni korpusi,⁶ ki temeljijo na video posnetkih različnih govornih situacij in poleg transkripcije govora vsebujejo tudi oznake drugih modalnosti tako na ravni prozodije kot na ravni govornice telesa (Verdonik in drugi 2020). Za jezikovno (in tudi multimodalno) analizo humorja se tako zdi najprimernejša oblika preučevanja posameznih jezikovnih in neverbalnih instrumentov humornosti z le za humor označenim korpusom, ki bi predstavljal po eni strani dragocen vir za jezikoslovno analizo in razumevanje rabe jezikovnih in nejezikovnih sredstev v humornih kontekstih, po drugi strani pa prosto dostopen jezikovni vir za nadaljnjo strojno obdelavo naravnega jezika. V zadnjih nekaj letih so namreč za druge jezike označeni korpusi humorja in druge označene zbirke besedil postali tudi pomemben vir učnih množic za strojno razpoznavanje, razumevanje in generiranje humorja. Za slovenščino takega korpusa še ni.

Predlagava izdelavo korpusa humorja v konverzaciji za slovenščino po naslednjih korakih:

- 1. faza: Pregled obstoječih korpusov humorja in standardov označevanja.
- 2. faza: Definiranje označevalnega procesa.
- 3. faza: Izbira in pridobivanje gradiva.
- 4. faza: Samodejno označevanje korpusa.
- 5. faza: Ročno označevanje korpusa.

3.1 Pregled obstoječih korpusov humorja

Da bi lahko podali predlog izdelave korpusa humorja v konverzaciji za slovenščino, sva pregledali že obstoječe pisne in govorne korpusne humorja v drugih jezikih. Zanimalo naju je predvsem:

⁶ Za slovenski jezik obstaja multimodalni korpus EVA (Mlakar in drugi 2019) v obsegu 1 ure govornega posnetka, označenega na številnih ravneh, od gest in mimike na fizični ravni do njihove semiotične interpretacije, oznak prozodije, sentimenta itd.

- Ali obstajajo prosto dostopni označeni korpusi humorja v drugih jezikih?
- Kako obsežni so in v katerem kodu (pisnem, govornem, multimodalnem) so?
- Kateri elementi so označeni v korpusu (vrsta označenih elementov, kot so posamezni jezikovni elementi za doseganje humornosti, humorne pasaže, intenziteta humornosti ipd.) in kako so označeni (način in merila za označevanje)?

Pri analizi obstoječih korpusov in označenih zbirk humorja sva upoštevali: 1) korpus humorja, ki so nastali za (jezikoslovno ali drugo) analizo posameznih prvin v humorni produkciji, interakciji in pri prepoznavanju humornosti ali drugih s humorjem povezanih pojavov, ter 2) korpus in označene zbirke, ki so nastali za namen strojnega prepoznavanja, generiranja in vrednotenja humorja, ampak le, če so bili kakorkoli označeni za humor, torej so v njih označene humorne pasaže ali drugi humorni elementi, ali če je humornost v teh korpusih kakorkoli človeško presojana z vidika smešnosti/nesmešnosti oziroma stopnje/intenzitete smešnosti.

Pregledane korpus humorja, ki so navedeni v tabeli 1, sva razdelili v skupine glede na vrsto in kanal posredovanega gradiva v 1) pisne, 2) govorne in 3) multimodalne. Pisni korpusi humorja so načeloma nastajali za namene:

- sistematičnega opazovanja posameznega izoliranega elementa za doseganje humorja, npr. metafore (Attardo 2016) ali primere (Veale 2013);
- analize humornega sloga posameznega avtorja (Duguid 2009);
- združeno tako za jezikoslovno analizo humorja kot strojno prepoznavo ali generiranje humorja, npr. pisni korpus v manjšem obsegu kot prvi poskus strojnega razumevanja humorja (Mihalcea in drugi 2006), španski korpus humorja (Chiruzzo in drugi 2020), korpus angleško-hindujskih tvitov s šalami (Khandelwal in drugi 2018), ročno označena zbirka wikinavedkov (Buscaldi in Rosso 2007);
- za strojno razumevanje in ocenjevanje intenzitete humorja, npr. kitajski korpus za raziskovanje humorja (Tseng in drugi 2019), francosko-angleški korpus za analizo humornih besednih iger in strojno vrednotenje humorja (Ermakova in drugi 2022).

Med govornimi korpusi trije primarno niso nastali za namene strojne obdelave: korpus, namenjen jezikoslovni analizi metafor v pogovoru za doseganje humornega učinka (Kyrtzis 2003), in korpusa, namenjena raziskovanju humorne interakcije v spontanem pogovoru (Feyaerts in drugi 2015; Val.Es.Co 2014). Posebej opredeljujeva korpuse, ki so nastali z vzorčenjem že obstoječih večjih govornih korpusov in bili nato posebej označeni za humorne pasaže ali druge humorne elemente (Günther 2003; Nesi 2012; Pickering in drugi 2009; Archakis in drugi 2010).

Multimodalni korpusi oz. označene podatkovne zbirke vključujejo videoposnetke s transkripcijami za namen preučevanja sovpivanja različnih elementov jezikovnega in nejezikovnih kodov pri humorjem izražanju, npr. multimodalna analiza označevalcev humorja (ang. *markers of humor*) (Attardo 2011), analiza zbadanja v ameriških sitcomih (Feyaerts in drugi 2015), analiza smeha med humorno produkcijo (Priego-Valverde 2016) in analiza humorjih izrekov v TV-serijah (Singh 2023). Vsi razen prvega so hkrati nastali za namene strojne obdelave humorja.

Tabela 1: Pregledani pisni, govorni in multimodalni korpusi humorja in za humorne elemente ali/in pasaže označene govorne, pisne ali multimodalne podatkovne zbirke

Ime korpusa, zbirke	Avtorji	Leto	Jezik	Dostopnost korpusa, označene zbirke	Velikost	Gradivo	Namen
/	Attardo, S.	2016	angleščina	neobjavljen	neznana (manjši korpus za lastno raziskavo)	besedila s spletnih strani in novinarskih prispevkov	kvalitativna analiza humornih metafor
/	Kyratzis, S.	2003	grščina	neobjavljen	180 minut	transkripcije posnetega pogovora med petimi znanci o različnih temah	uporaba metafor v pogovoru za doseganje humornega učinka
/	Duguid, A.	2009	angleščina	neobjavljen	205.000 besed	ocene, kritike in druga besedila prepoznavnih komičnih britanskih kolumnistov	analiza humornega sloga avtorja, analiza humornih primer
/	Mihalcea, R., Strapparava, C.	2006	angleščina	neobjavljen	16.000 enovrstičnic (<i>oneliners</i>)	humorne enovrstičnice, izluščene s spletnih strani	strojna obdelava (prepoznavna humorja)
/	Attardo, S.	2011	angleščina	neobjavljen	10 minut	video posnetek in transkripcije pogovorov med študenti, ki drug drugemu pripovedujejo šale in se medtem prosto pogovarjajo	analiza označevalcev humorja (<i>markers of humor</i>)
/	Feyaerts, K., Bróne, G., De Ceukelai, R.	2015	angleščina	neobjavljen	402 humorna izseka z zbadljivkami	epizode štirih ameriških sitcomov	empirična analiza zbadljivosti v sekvencah dialoga

Ime korpusa, zbirke	Avtorji	Leto	Jezik	Dostopnost korpusa, označene zbirke	Velikost	Gradivo	Namen
HAHA 2019 Dataset	Chiruzzo, L., Castro, S., Rosá, A.	2019	španščina	objavljen, prosto dostopen	30.000 tvitov	tviti	analiza humorja, označena zbirka za strojno obdelavo (prepoznavna, generiranje in ocenjevanje stopnje humornosti)
Chinese humor corpus	Tseng, Y.-H., Wu W.-S., Chang, C.-Y., Chen, H.-S., Hsu, W.-L.	2020	kitajščina	objavljen, prosto dostopen	3365 šal iz več kot 40 virov	pisne šale	strojna obdelava (razumevanje humorja, ocenjevanje stopnje humornosti)
Cheese!	Priego-Valverde	2016	francoščina	objavljen, prosto dostopen	165 minut	video posnetki s transkripcijami pogovorov	multimodalna analiza smeha pri tvorjenju humorja
Corinth (Corpus of Interactional Humor)	Feyaerts, K., Oben, B., Geert, B., Dirk, S.	2011	angleščina	neobjavljen	300 minut	posnetki in transkripcije spontanega pogovora med mladimi	analiza humorne interakcije
/	Veale, T.	2013	angleščina	neobjavljen	38.294 primer	splet	analiza humoristnih primerov
/	Khandelwal, A., Swami, S., Akthar, S. S., Shrivastava, M.	2018	angleščina in hindijščina	objavljen, prosto dostopen	10.478 tvitov	tviti	strojna obdelava (prepoznavna humorja)
VALESCO. HUMOR corpus	skupina Val.Es.Co (vodja Briz, A.)	2014	španščina	objavljen, prosto dostopen	70 minut	posnetki in transkripcije spontanega pogovora	analiza humorja, zbadanja in posmeha

Ime korpusa, zbirke	Avtorji	Leto	Jezik	Dostopnost korpusa, označene zbirke	Velikost	Gradivo	Namen
The JOKER Corpus	Ermakova, L., Miller, T., Regattin, F., Bosser, A.-G., idr.	2022	francoščina in angleščina	prosto dostopen za raziskovalne namene	3506 šal + 1708 namerno pokvarjenih šal	različni spletni viri, izluščene šale, besedne igre, uganke ...	primerjalna analiza humornih besednih iger v francoščini in angleščini, strojna obdelava (vrednotenje humorja)
Multimodal Multiparty Hindi Humor (M2H2) dataset	Singh, D., Singh, V., Ekbal, A., Bhattacharyya, P.	2023	angleščina in hindijščina	prosto dostopen za raziskovalne namene	6.191 izrekov	13 epizod TV-serije	strojna obdelava (prepoznavna humorja in sentimenta v multimodalnem gradivu)
Emotocorpus	Buscaldi, D., Rosso, P.	2009	italijanščina	prosto dostopen	1966 vikinavedkov	vikinavedki, Wikipedija	strojna obdelava (klasifikacija in prepoznavna humorja)

Vir: lasten

3.2 Označevalni proces

Definiranje označevalnega procesa, ki je eden od ključnih korakov izdelave korpusa humorja v govoru, vključuje:

- definiranje in specificiranje označevanja humornih izrekov (enot konverzacije, ki praviloma na intonacijski, skladenjsko-semantični ravni predstavljajo najmanjšo enoto konverzacije in hkrati na naslovnika učinkuje humorno) oz. pasaž (več humornih izrekov, ki so smiselno vsebinsko zaključen segment konverzacije);
- definiranje in specificiranje označevanja instrumentov, ki pripomorejo k doseganju humornega učinka pri ubesedenem (npr. semantične, pragmatične, multimodalne prvine);
- izbira primernega orodja za označevanje;
- definiranje načina označevanja (število označevalcev, da bo proces označevanja čim bolj zanesljiv; kdo sploh bodo označevalci; ali bodo določeni vnaprej ali bodo neznani (označevanje z množičenjem)) ipd.).

Tabela 2: Pregled označevalnih praks v korpusih humorja

Način označevanja humorja/humornih elementov	Merila za presojanje humornosti besedila	Vrsta označenih elementov
ročno	presoja označevalca	humorna pasaža/element; metafora, primera; logični mehanizem; izražena namera
delno samodejno	smeh (dolžina), aplavz (intenziteta in dolžina); ključniki (#haha, #lol, #rofl, #funny); odziv uporabnikov (<i>user feedback</i>)	smeh, ploskanje; prozodični elementi; smešnost
množičenje	analiza vprašalnika	smešnost (stopenjsko)

Vir: lasten

Pri pregledanih korpusih humorja za druge jezike sva ugotovili (tabela 2), da ročno označeni korpusi večinoma temeljijo na presoji označevalca ali več označevalcev – to pomeni, da označevalci glede na zastavljeno označevalno shemo presojajo, ali je za označevanje izbrani element humoren ali ne, kje se začne in konča humorna

pasaža, ali ima izbrani element humorni učinek ali ne ipd. Delno samodejno označeni korpusi prevzemajo oznake iz že obstoječega označenega gradiva, ki je vključeno v korpus, kot so na primer oznake za smeh in aplavz, tudi dolžina in intenziteta smeha in ploskanja ali ključniki v tvitih, npr. #haha, #lol, #rofl, ki predstavljajo podlago za presojo o humornosti določenega tvita. Poleg tega so upoštevani tudi odzivi uporabnikov, kot je na primer število všečkov humorne objave na družbenem omrežju. Z metodo množičenja se preverja, ali je neko besedilo oziroma izrek humoren ali ne in kako močno humoren je (z različnimi lestvicami).

4 Predlog izdelave korpusa humorja v konverzaciji za slovenščino

Na podlagi pregleda obstoječih korpusov humorja podajava predlog za izdelavo korpusa humorja v konverzaciji za slovenščino:

1. Korpus je primarno namenjen jezikoslovni (in multimodalni analizi), a je primeren tudi za nadaljnjo strojno obdelavo (v formatu TEI, ustrezno tokeniziran, lematiziran, oblikoslovno in skladenjsko označen).
2. Gradivo vključuje video posnetke in pripadajoče transkripcije, pri čemer pri izboru gradiva upoštevamo naslednja merila:
 - a. razmerje med spontanostjo in vnaprejšnjo pripravljenostjo;
 - b. video posnetek omogoča natančno opazovanje neverbalnega obnašanja govorca in po možnosti tudi sogovorca, če ta nastopa; to pomeni, da morata biti v 90 % hkrati vidna vizualni (geste, situacija) in akustični (prozodija, verbalno) del;
 - c. vrsta medija oziroma prenosnika;
 - d. vrsta stika med nastopajočim(i) in publiko, pri čemer mora biti viden odziv sogovorca, da se vidi, kam sodi gesta;
 - e. število nastopajočih;
 - f. uravnoteženost med klasično in e-komunikacijo;
 - g. znanje, potrebno za razumevanje humorja;
 - h. raznovrstnost tematike, ki naj bo čim manj vezana na obdobje nastanka gradiva, priporočena je njena izvzetost iz aktualnega konteksta, da se tako aktualnost teme ne bi zmanjšala;
 - i. primerljiva zastopanost spolov.
3. Transkripcije nastajajo po izoblikovanih standardih.
4. V transkripcijah so po predhodno oblikovani označevalni shemi označene humorne pasaže in instrumenti za doseganje humorja.

5. Merilo za vrednotenje humornosti posameznih pasaž v transkripcijah se presoja na podlagi vprašalnika, in kjer je to mogoče, delno samodejno s prevzemanjem že obstoječih oznak za smeh in ploskanje v transkripcijah.
6. Označevanje poteka v ustreznem orodju za označevanje, v katerem se lahko dodajajo oznake na jezikovni kot nejezikovni ravni.

4.1 Vprašalnik za vrednotenje humornosti posameznih pasaž v konverzaciji

4.1.1 Gradivo

Za učno izvedbo procesa samodejnega in ročnega označevanja humornih izrekov in pasaž v konverzaciji v slovenskem jeziku sva uporabili gradivo slovenskega stand-upa, podcasta ali razvedrilne TV-oddaje. V vprašalnik sva dodali pet izsekov iz video posnetkov po lastni presoji in upoštevajoč kriterije iz 4. poglavja. Podrobneje so opisani v tabeli 3.

Tabela 3: Izbrani humorni izseki za vrednotenje humornosti posameznih pasaž v konverzaciji

Vrsta gradiva	Dolžina	Leto nastanka	Št. govorcev	Dostopnost	Povezava
stand-up	1 min 45 s	2020	1	dovoljenje avtorja	https://1ka.arnes.si/uploadi/editor/doc/1680781230udar-po-mosko_1.mp4
podcast	4 min 27 s	2022	2	dovoljenje avtorjev	https://1ka.arnes.si/uploadi/editor/doc/1679393556Prepona-1.mp4
stand-up	30 s	2012	1	korpus BERTA	https://1ka.arnes.si/uploadi/editor/doc/1680781683ales-novak_komik-1.mp4
razvedrilna TV-oddaja	1 min 30 s	2008	3	korpus GOS	https://1ka.arnes.si/uploadi/editor/doc/1680781804as-ti-tud-not-padu_Kobal_1.mp4
razvedrilna TV-oddaja	59 s	2007	3	korpus GOS	https://1ka.arnes.si/uploadi/editor/doc/1680817192as-ti-tud-not-padu_1.mp4

Vir: lasten

Pri prepoznavanju humornosti pasaž v konverzaciji in pri določanju, katera jezikovna in nejezikovna sredstva so humorna, sva ugotovili, da razumevanje pasaž kot humornih niti med nama ni bilo enotno in da je bilo razumevanje pasaž/izrekov

kot humornih/smešnih odvisno tudi od najinega trenutnega razpoloženja. Tudi zato sva se odločili, da v gradivu, ker je prisotno občinstvo, njegov odziv s smehom ali ploskanjem pomeni, da so naslovniki izrečeno razumeli kot humorno/smešno. Kjer pa ga ni bilo, sva nujno potrebovali neodvisno, testno občinstvo, ki bi ob pripravljnem anketnem vprašalniku podalo svoje razumevanje, doživljanje izbranih pasaž.

4.1.2 Anketa

Anketa je bila izvedena med študenti slovenistike in računalništva. Delno ali v celoti jo je skupno rešilo 171 študentov. Izvedena je bila z namenom presoje, ali je lahko tovrstna metoda učinkovita za določanje humornih pasaž in ali lahko pomaga pri presoji označevalca, katera jezikovna ali nejezikovna sredstva vplivajo na humornost izreka oz. pasaže. S tem se želiva predvsem izogniti subjektivnim presojam posameznih označevalcev, hkrati pa želiva pri identifikaciji humorja upoštevati presojo več različnih generacij.

Anketni vprašalnik je vključeval pet posnetkov, anketa pa je potekala tako, da so si študenti najprej pogledali izbrano pasažo, nato so bili pozvani, da:

- na lestvici od »sploh ni smešno« do »zelo smešno« ocenijo, kako smešen se jim je zdel predvajani odlomek;
- izberejo enega ali več odgovorov, zakaj se jim pasaža ni zdela oziroma se jim je zdela smešna:
 - a) zaradi uporabljenih besed (zapiši besede),
 - b) zaradi odnosa govorečega do sogovorca, sogovorke,
 - c) zaradi vsebine (zapiši vsebine),
 - d) zaradi odnosa do vsebine,
 - e) zaradi mimike, kretenj, gibov,
 - f) zaradi načina, kako je bilo kaj povedano (zapiši, kaj je bilo smešnega pri načinu povedanega), in
 - g) drugo.

Odgovori pod točko a) opisujejo semantična sredstva, odgovori pod točkami b), c) in d) opisujejo pragmatične strategije in znanje, odgovori pod točkama e) in f) pa kažejo na nejezikovne kode oziroma prozodična sredstva. Točko g) sva dodali, da bi

anketiranci lahko dodali še prvine, ki jih ne bi znali ali ne bi mogli umestiti v katero od prejšnjih točk.

Na podlagi analize podanih odgovorov sva ugotovili, da je vprašalnik smiselno orodje za določanje humornih pasaž ter s humorjem povezanih jezikovnih in nejezikovnih sredstev v korpusu:

- Zapisane besede, ki so jih navedli anketiranci v svojih odgovorih in zaradi katerih se jim zdi pasaža humorna, določajo pasažo kot humorno. V humorni pasaži je lahko več govorcev in več različnih šal, ki sodijo v eno krovno humorno sekvenco. Ob tem je treba preveriti, ali so označeni deli prekrivni s smehom, ploskanjem publike.
- Zapisane besede, ki so jih navedli anketiranci v svojih odgovorih in zaradi katerih se jim zdi pasaža humorna, so del semantičnih sredstev, ki sodelujejo pri tvorjenju humorja, npr. metaforični izrazi, uporaba sopomenk ali logičnih mehanizmov za doseganje humorja, npr. za napad, pretiravanje, samoškodovanje ipd.
- Odgovori, ki opredeljujejo doseganje humornega učinka zaradi mimike, kretenj, gibov ali zaradi načina, kako je bilo kaj povedano, definirajo označena nejezikovna sredstva in prozodične prvine govora.

Kljub navedenim prednostim vprašalnika sva zaznali nekatere pomanjkljivosti pri definiranju posameznih vprašanj. Izkazalo se je namreč, da anketiranci niso dobro ločevali med posameznimi ponujenimi kategorijami; prekrivnost odgovorov sva zaznali pri ponujenih možnostih »zaradi uporabljenih besed« in »zaradi vsebine«, v manjši meri tudi pri odgovorih »zaradi mimike, kretenj, gibov« in »zaradi načina, kako je bilo povedano«. Te prekrivnosti so razumljive, saj je vsebina ubesedena z leksikalnimi sredstvi, laiki pa tudi težko razlikujejo med jezikovnim kodom, nejezikovnimi kodi in prozodičnimi prvinami govora.

5 Sklep

Čeprav je humor družbeni pojav (teorija superiornosti) in človekov notranji mehanizem premagovanja težav v življenju ter sprošča napetost v telesu (teorija sproščanja) zaradi česa nepričakovanega (teorija neskladja), je njegovo doživljanje odvisno od razpoloženja konkretnega posameznika in njegove pripadnosti določeni

skupini, lahko tudi generacijski. Prav to dodatno otežuje raziskovanje humorja v konverzaciji in tudi izdelavo korpusa, ki bi imel označene humorne pasaže. Kljub temu, da je npr. metafora po Griceovi teoriji vedno izkoriščanje maksime kakovosti, vsaka metafora v konverzaciji ni humorna in tudi ne učinkuje humorno. Anketni vprašalnik pomaga pri določanju humornih pasaj in izrekov, vendar laiki običajno ne razmišljajo o tem, zakaj je kaj humorno/smešno in zakaj ne, tj. ne razmišljajo o semantičnih scenarijih in drugih virih znanja ter nejezikovnih kodih in prozodičnih prvinah govora za tvorjenje humornega besedila. To pomeni, da bo najverjetneje potrebnih še nekaj poskusov pred dokončnim oblikovanjem anketnega vprašalnika z vedno novimi posnetki in verjetno tudi z vedno novimi sodelujočimi. Pomeni pa tudi, da bodo humorne pasaže v korpusu najverjetneje imele hkrati več korpusnih oznak, ki bodo z različnega vidika (semantičnega, pragmatičnega, nejezikovnega, prozodičnega) označevale humorno/smešno.

Literatura

- Argiris ARCHAKIS, Maria GIAKOUMELOU, Dimitris PAPAZACHARIOU, Villy TSAKONA, 2010: The Prosodic Framing of Humour in Conversational Narratives: Evidence from Greek Data. *Journal of Greek Linguistics* 10, 187–212.
- Salvatore ATTARDO, 2001: Humorous Texts: A Semantic and Pragmatic Analysis. Berlin, New York: Mouton de Gruyter.
- Salvatore ATTARDO, 2006: Cognitive linguistics and humor. *Humor: International Journal of Humor Research* 19(3), 341–362.
- Salvatore ATTARDO, Lucy PICKERING, Amanda BAKER, 2011: Prosodic and multimodal markers of humor in conversation. *Pragmatics and Cognition* 19(2), 224–247.
- Salvatore ATTARDO, Victor RASKIN, 2017: Linguistics and Humor Theory. *The Routledge Handbook of Language and Humor*. Ur. Salvatore Attardo. New York, London: Routledge. 49–63.
- Saša BABIČ, 2018: Stereotipi in predsodki o ženskah in moških v šaljivih vprašanjih. Humor u svakodnevnj komunikaciji. Ur. Renata Jambrešić Kirin [et al.]. Zagreb: Institut za etnologiju i folkloristiku. 81–95.
- Robert Alain DE BEAUGRANDE, Wolfgang Ulrich DRESSLER, 1992: *Uvod v besediloslovje*. Prevod: Aleksandra Derganc, Tjaša Miklič. Ljubljana: Park.
- Henri BERGSON, 1977: *Esej o smehtu. Filozofska intuicija. Uvod v metafiziko*. Prev. Janez Gradišnik. Ljubljana: Slovenska matica.
- Silvija BOROVIK, 1997: Humor, ironija in groteska v delih Janka Messnerja. *Slavistična revija* 45/2-1. 127–134.
- Davide BUSCALDI, Paolo ROSSO, 2007: Some experiments in humour recognition using the italian wikiquote collection. *WILF 2007. LNCS (LNAI)* 4578. 464–468.
- Luis CHIRUZZO, Santiago CASTRO, Aiala Rosá, 2020. HAHA 2019 Dataset: A Corpus for Humor Analysis in Spanish. *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille: European Language Resource Association. 5106–5112.
- Alison DIGUID, 2009. Loud signatures- Comparing evaluative discourse styles – patterns in rants and riffs. *Studies in Corpus Linguistics* 36. 289–315.
- Oswald DUCROT, 1988: *Izrekanje in izrečeno*. Ljubljana: Založba ŠKUC: Znanstveni inštitut Filozofske fakultete.

- Marta DYNEL, 2009: Beyond a Joke: Types of Conversational Humour. *Language and Linguistics Compass* 3(5). 1284–1299.
- Liana ERMAKOVA, Tristan MILLER, Fabio REGATTIN, Anne-Gwen BOSSER, Claudine BORG, Elise MATHURIN, Gaëlle LE CORRE, Silvia ARAUJO, Radia HANNACHI, Julien BOCCOU, Albin DIGUE, Aurianne DEMOY, Benoît JEANIEAN, 2022: Overview of JOKER@CLEF 2022: Automatic Wordplay and Humour Translation Workshop. *Experimental IR Meets Multilinguality, Multimodality and Interaction*, Springer International Publishing. 447–469.
- Tomaz ERZAR 1997: *Obrat Gilles Deleuza: Teorija subverzije in ideja populne ontologije v Logiki smisla in Anti-Ojdlpu*. Ljubljana: ZRC SAZU, Založba ZRC.
- Kurt FEYAERTS, Geert BRÔNE and Robin DE CEUKELAIRE, 2015: The Art of Teasing A corpus study of teasing sequences in American sitcoms between 1990 and 1999. *Cognitive Linguistics and Humor Research*. Ur. Geert Brône, Kurt Feyaerts, Tony Veale. Berlin, München, Boston: De Gruyter Mouton. 215–243.
- Sigmund FREUD, 1991: Humor. *Problemi*. 2/3. 93–96. Prevod: Peter Klepec.
- Paul GRICE, 1989: Logic and Conversation. *Studies in the Way of Words*. Cambridge: Harvard University Press. 22–40.
- Ulrike K. GÜNTHER, 2003: *What's in a laugh? Humour, jokes and laughter in the conversational corpus of the BNC*. Doktorska disertacija. Univerza v Edinburgu.
- Urška JARNOVIČ, 2008: 7. mednarodna poletna šola in simpozij o humorju in smehu, Aberdeen, 9.–14. july 2007. *Jezik in slovstvo* 53/1, 155–160.
- Marko JESENSEK, 2010: Humornost v kratki pripovedi Prežihovega Voranca. *Jezik in slovstvo* 55/3–4, 111–121.
- Ankush KHANDELWAL, Sahil SWAMI, Syed S. AKHTAR, Manish SHRIVASTAVA, 2018: Humor Detection in English-Hindi Code-Mixed Social Media Content : Corpus and Baseline System. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 1203–1207.
- Tomo KOROŠEČ, 2002: Razčlemba razžalitev v tiskanih medijih (Jezikoslovna izhodišča). Razžalitev v tiskanih medijih. Avtorji T. Korošec idr. Ljubljana. Znanstvena knjižnica. 3–30.
- Helga KOTTHOFF, 1998: *Spaß Verstehen. Zur Pragmatik von konversationellem Humor*. Reihe Germanische Linguistik 196. Tübingen: Niemeyer.
- Mira KRAJNC IVIČ, 2011: Uvod v analiziranje šaljivega v pogovarjanjih. *Meddisciplinarnost v slovenistiki* Ur. Simona Krajnc. Obdobja 30. Ljubljana: Znanstvena založba Filozofske fakultete. 265–270.
- Mira KRAJNC IVIČ, 2013: Humorno v besedilih različnih neumetnostnih funkcijskih zvrsti. *Družbena funkcijskost jezika: (vidiki, merila, opredelitve)*. Ur. Andreja Žele. (Obdobja 32.) Ljubljana: Znanstvena založba Filozofske fakultete. 227–232.
- Mira KRAJNC IVIČ, 2017: Humorja ni nikoli preveč – ali pač. *Slovenistika 10: zbornik predavanj*. Ur. Marija Bajzek Lukač. Budapest: ELTE BTK, Szlav Filológiai Tanszék. 157–168.
- Olga KUNST GNAMUŠ, 1984: Govorno dejanje – družbeno dejanje. Komunikacijski model jezikovne vzgoje. Ljubljana: Pedagoški inštitut.
- Sakis KYRATZIS, 2003: *Laughing metaphorically: metaphor and humour in discourse*. 8th International Cognitive Linguistics Conference: July 20– 25, 2003, University of La Rioja, Spain. Theme session: Cognitive-Linguistic Approaches to Humour. Dostop 1. 9. 2023 na <http://www.ling.arts.kuleuven.ac.be/iclc/Papers/Kyrtzis.pdf>.
- Cristina LARKIN-GALIÑANES, 2017: An Overview of Humor Theory. *The Routledge Handbook of Language and Humor*. Ur. Salvatore Attardo. New York, London: Routledge. 4–17.
- Rada MIHALCEA, Carlo STRAPPARAVA, 2006. Learning to laugh (automatically): Computational models for humor recognition. *Computational Intelligence* 22/2, 126–142.
- Izidor MLAKAR, Darinka VERDONIK, Simona MAJHENIČ, Matej ROJC, 2019: Towards pragmatic understanding of conversational intent: A multimodal annotation approach to multiparty informal interaction–The EVA Corpus. *Statistical Language and Speech Processing: 7th International Conference*. Ur. Carlos Martín-Vide, Matthew Purver, Senja Pollak. 19–31.
- Hilary NESI, 2012. Laughter in university lectures. *Journal of English for Academic Purposes* 11/2, 79–89.

- Lucy PICKERING, Marcella CORDUAS, Jodi EISTERHOLD, Brenna SEIFRIED, Alyson EGGLESTON, Salvatore ATTARDO, 2009: Prosodic markers of saliency in humorous narratives. *Discourse Processes*, 46(6), 517–540.
- Béatrice PRIEGO-VALVERDE, Brigitte BIGI, Mary AMOYAL, 2020: "Cheese!": a corpus of face-to-face French interactions. A case study for analyzing smiling and conversational humor. *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille: European Language Resource Association. 467–475.
- Béatrice PRIEGO-VALVERDE, 2016: *Cheese!*. Dostop 1. 9. 2023 na <https://hdl.handle.net/11403/cheese>.
- Victor RASKIN, V., 1984: *Semantic Mechanisms of Humor*. Dordrecht, Boston, Lancaster: D. Reidel Publishing Company.
- Dushyant SINGH, Gopendra VIKRAM, Asif EKBAL, Pushpak BHATTACHARYYA, 2023: MHaDiG: A Multilingual Humor-aided Multiparty Dialogue Generation in multimodal conversational setting. *Knowledge-Based Systems* 278.
- Irena STRAMLJIČ BREZNIK, 2002: *Besedoslovne lastnosti Gomilšakovega sestavka Čujte, čujte možje!* Zbornik razprav. Ur. Simona Kranjc. (Mednarodni znanstveni simpozij o Jakobu Gomilšaku). Trnovska vas: Občina (Ptuj; Senčar). 99–109.
- Mojca STRITAR, 1999/2000: Šale. Analiza besedilne zgradbe in pojasnitev poante. *Jezik in slovnost* 45/4. 148–161.
- Yuen-Hsien TSENG, Wun-Syuan WU, Chia-Yueh CHANG, Hsueh-Chih CHEN, Wei-Lun HSU, 2020: Development and Validation of a Corpus for Machine Humor Comprehension. *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille: European Language Resource Association. 1346–1352.
- Valencia.Español.Coloquial, 2014: Corpus Val.Es.Co, vodja projekta: Antonio Briz. Projekt je nastal s podporo CIPROM/2021/038 (MCIN/AEI in FEDER) »Način ustvarjanja Evrope«.
- Tony VEALE, 2013: Humorous similes. *Humor: International Journal of Humor Research* 26(1). 3–22.
- Darinka VERDONIK, Andrej ŽGANK, Simona MAJHENIČ, Izidor MLAKAR, 2000: *Govorni, dialoški in multimodalni jezikovni viri : pregled stanja*. Vsebinsko poročilo. Maribor: Fakulteta za elektrotehniko, računalništvo in informatiko. Dostop 1. 9. 2023 na <https://dk.um.si/Dokument.php?id=142047&lang=slv>.
- Franc ZADRAVEC, 1991: Cankarjeva ironija. Murska Sobota: Pomurska založba; Ljubljana: Znanstveni inštitut Filozofske fakultete.

TVORBA KORPUSŮ MLUVENÉHO JAZYKA

MILOSLAV VONDRÁČEK

Slezská univerzita v Opavě, Filozoficko-přírodovědecká fakulta v Opavě,
Česká republika
miloslav.vondracek@fpf.slu.cz

Má někdejší univerzita se v minulých letech podílela na tvorbě korpusů mluvené komunikace. Spolu se studenty jsme pořídili zvukový záznam 220 soukromých dialogických situací a jejich přepis. Při té příležitosti jsme museli řešit řadu praktických problémů. Ty vedly k formulaci podstatných teoretických otázek. K základním patří relativita jednotek mluveného jazyka. Výsledkem je metodologie tvorby korpusu, od té doby neustále zdokonalovaná. Příspěvek přináší přehled základních otázek a snaží se poskytnout teoretické odpovědi i metodiku řešení.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.11](https://doi.org/10.18690/um.ff.4.2024.11)

ISBN
978-961-286-882-6

Ključne besede:
korpus mluveného jazyka,
relativita jednotek řeči,
pravidla přepisu,
zvukový záznam,
neoficiální komunikační
situace



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.11](https://doi.org/10.18690/um.ff.4.2024.11)

ISBN

978-961-286-882-6

Keywords:

spoken language corpora,
relativity of speech units,
transcription rules,
audio recording,
unofficial communication
situations

CREATION OF SPOKEN LANGUAGE CORPORA

MILOSLAV VONDRÁČEK

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
mojca.smolej@ff.uni-lj.si

In recent years, my former university has been involved in the creation of corpora of spoken communication. The students made audio recordings of 220 private conversations. With my help, the students converted these dialogues into written text. On this occasion, we had to solve some practical problems. These difficulties led to the formulation of substantial theoretical questions. The relativity of the units of spoken language is one of the fundamental ones. The result is a corpus building methodology that has been continuously improved since then. The paper provides an overview of the fundamental questions and attempts to provide theoretical answers and a solution methodology.



TVORBA GOVORJENEGA KORPUSA

MILOSLAV VONDRÁČEK

Šlezjska univerza v Opavi, Filozofska in naravoslovna fakulteta v Opavi,
Republika Češka

miloslav.vondracek@fpf.slu.cz

V preteklih letih je moja nekdanja univerza sodelovala pri ustvarjanju korpusov govornjene komunikacije. Študenti so posneli 220 zasebnih pogovorov in jih z mojo pomočjo transkribirali. Ob tej priložnosti smo morali rešiti vrsto praktičnih težav. Te težave so privedle do oblikovanja pomembnih teoretičnih vprašanj. Eno temeljnih je relativnost enot govornjenega jezika. Rezultat tega je metodologija gradnje korpusa, ki se od takrat nenehno izboljšuje. Prispevek podaja pregled temeljnih vprašanj ter poskuša podati teoretične odgovore in metodologijo reševanja.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.11](https://doi.org/10.18690/um.ff.4.2024.11)

ISBN
978-961-286-882-6

Ključne besede:
korpusi govornjenega jezika,
relativnost govornnih enot,
pravila transkripcije,
zvočno snemanje,
neuradne komunikacijske
situacije



Univerzitetna založba
Univerze v Mariboru

1 Úvod

Tvůrci korpusů mluvené komunikace řady ORAL, vznikajícího úsilím Ústavu Českého národního korpusu Filozofické fakulty Univerzity Karlovy v Praze, se zhruba v roce 2005 obrátili na další bohemistická univerzitní pracoviště České republiky s výzvou k účasti na jejich projektu.¹ Pracoviště už tou dobou mělo zkušenost s tvorbou Pražského mluveného korpusu (2001, dále PMK²), byly tu poznatky spjaté s Brněnským mluveným korpusem (2002, dále BMK³), z aktuálního sběru materiálu pro celočeský ORAL2006,⁴ to vše na pozadí důkladné obeznámenosti s problematikou v mezinárodním měřítku. Spolupráce s mimopražskými středisky měla spočívat ve shromáždění zvukových záznamů neoficiálních komunikačních situací dostatečného rozsahu a jejich přepisu podle stanovených pravidel. Povaha spolupráce s organizačním centrem se mohla případ od případu lišit v závislosti na odborném (a nakonec i lidském) profilu prostředníka. Zpětně mohu konstatovat, že pro mne – a průkazně i pro angažované studenty⁵ – se stala tato aktivita zdrojem širokého spektra poznatků ze všech lingvistických disciplín. To je důvod, proč neváhám zkušenost s odstupem a nadhledem uplynulých let znovu sdílet.⁶

2 Povaha komunikační situace

Korpusy ORAL2008 a ORAL2013,⁷ k nimž se váže má akviziční zkušenost, jsou charakterizovány jako korpusy neformální mluvené češtiny. Sdružují „materiál představující prototypický spontánní mluvený jazyk, který se používá při bezprostřední interakci mluvčích v neformálních komunikačních situacích.

¹ Výsledky se promítají do korpusů ORAL2008 a ORAL 2013. K nim více na <https://wiki.korpus.cz/doku.php/cnk:uvod>

² Informace o PMK dostupné na <https://wiki.korpus.cz/doku.php/cnk:pmk>

³ Informace o BMK dostupné na <https://wiki.korpus.cz/doku.php/cnk:bmkmk>. PMK i BMK ovšem zahrnují zčásti i moderované formální monologické promluvy.

⁴ Informace o ORAL2006 dostupné na <https://wiki.korpus.cz/doku.php/cnk:oral2006>.

⁵ Mé tehdejší působiště: Univerzita Hradec Králové, Pedagogická fakulta, roky 2005–2011. Zapojilo se postupně přes 200 studentů. Každý pořídil jedinečnou sondu do mluvené komunikace, tj. nahrávku o délce průměrně 30 minut (reálně 20–50 minut, celkem přes 100 hodin). Student sám svůj zvukový záznam přepsal (v první etapě do textového editoru Word, později do aplikace Transcriber). Přepis poté prošel několikanásobnou kontrolou věrnosti přepisu. V prvních kolech kontrolu zajišťoval příslušný akademický pracovník univerzity (na UHK já), finální kontrolu pracovník ÚČNK FF UK. U obou spolutvořených korpusů viz úsek Poděkování (<https://wiki.korpus.cz/doku.php/cnk:oral2008>, <https://wiki.korpus.cz/doku.php/cnk:oral2013>).

⁶ O aktuálních zkušenostech z probíhajícího projektu jsem hovořil na konferenci Čestina v mluveném korpusu, Praha 2007, pro více informací viz Vondráček 2008.

⁷ Oba ve srovnání s korpusem ORAL2006 kromě Čech zahrnují i oblast Moravy a Slezska (což bylo jedním z motivů k oslovení mimopražských univerzitních pracovišť).

Hlavními kritérii pro získávání nahrávek byly: fyzická přítomnost všech mluvčích na jednom místě, dialogičnost promluv, vzájemný blízký vztah mluvčích, nepřipravenost, spontánnost, neveřejná a neoficiální komunikační situace“ (ORAL2013, kráceno).

Starší korpusy popisují neformální komunikační situace dalšími znaky: „neformální promluvy tvoří dialogy dvou, případně i více mluvčích, kteří se dobře znají“ (BMK), resp. „kteří se znají“ (PMK). V případě korpusu ORAL2006 „[v]šechny nahrávky vznikaly výhradně v neformálních situacích, což znamená, že se mluvčí vzájemně znali a měli k sobě přátelský vztah.“ Při tomto zadání se projevila nejednoznačnost v interpretaci vykání (i mezi rodinnými příslušníky). Současně se objevil jiný nečekaný situační rys – účastník komunikace je sice v přátelském vztahu ke komunikačnímu partnerovi, současně však hovoří z titulu své profese (kamarádka kadeřnice při úpravě účesu, kamarád automechanik při sjednávání opravy vozu). Průnikem obou eventualit vzniklo další pomocné kritérium, v dostupných podkladech nezaznamenané: vykání samo o sobě není v rozporu s neformálností a neoficiálností komunikační situace; současně (bez ohledu na tykání) žádný z účastníků nevystupuje v dialogu z titulu své profese. Tato charakteristika se více než osvědčila při jemném rozlišení komunikačních situací v rámci stylistických analýz komunikační sféry běžné.

3 Delimitace syntaktických jednotek

Pro první přepisy záznamů (pořizovaných již na digitální záznamníky) jsme využívali textového editoru Word.⁸ Ještě pro předchozí korpus ORAL2006 platila pravidla stanovení hranic syntaktických celků víceméně odpovídající zvyklostem psaných textů.⁹ Mírný posun přináší korpus ORAL2008 ve způsobu zaznamenání obvyklých defektů mluveného projevu.¹⁰

⁸ Složitou cestu uvědomování si a dokumentace rozdílů psané a mluvené češtiny ilustruje text J. Hoffmannové a M. Míkulecké (2011, 78–92).

⁹ „Hranice vět se vyznačují jen interpunkcí, na začátku věty píšeme malé písmeno. Větné interpunkce se užívá tak, jak je to obvyklé v textech psaných, tj. nezachycuje se přerušování věty pauzami, naopak náležitou čárku ve větách a v souvětích píšeme, i když se větné předěly pauzou nerealizují. Neukončené věty označujeme třemi tečkami s dvojtečkou odsazenými mezerou, tedy ...: Příklad, kdy je mluvčí přerušen jiným mluvčím, ale ve výpovědi později pokračuje, se v dialogu značí třemi tečkami na konci přerušené výpovědi i na začátku její navazující části.“ (archiv autora, viz též dále) „Pořizování nahrávek, jejich přepisování a označování probíhalo v souladu s obecnými zásadami uplatňovanými při přípravě všech předchozích mluvených korpusů v rámci Českého národního korpusu, zejména korpusu ORAL2006.“ (<https://wiki.korpus.cz/doku.php/cnk:oral2008>)

¹⁰ „Zvláštní rysy syntaktické stránky mluvených projevů (přeřeknutí, přerušování a změny větné perspektivy, přiřazování vět a větných úseků apod.) jsou zachyceny zjednodušeně, většinou pouze pomocí čárky. Pokud se slova opakují, jsou oddělena čárkou.“ (ORAL2008)

<3> *dovopravdy funguje.*

<2> *já vim, nó, my sme jak já, já, né my, co budu říkat my, já sem za vola.*

<3> *né, to není, není, já ti říkám, čím to bylo, vy ste podle mě tady s tím šabo, šibovali a naklonilo se to, udělala se bublina a to je jedna z alternativ, která byla možná, a pře, pak to nechladilo, protože si to musí sednout.*

<2> *ale vono, já mam takovej pocit, že s tím předtím jakoby nikdo nebejbal a najednou to přestalo jít.*

<3> *druhá varianta je, že, že se, žes v podstatě se to dostalo do nějaký mezipolohy a že, eee, třeba při tom přepínači, což mohlo být, a že začal blbnout termostat. a znovu začal fungovat. z nákejch důvodů.*

<2> *hm, hm.*

<0> *to je nejspíš:*

<2> *to je, to sou, todle, Jarečku, viš, to mi vysvětlí, viš, že já se tady ztrpňuju, rok tady máme ledničku vypnutou ... (úryvek ze sondy 05H006N)*

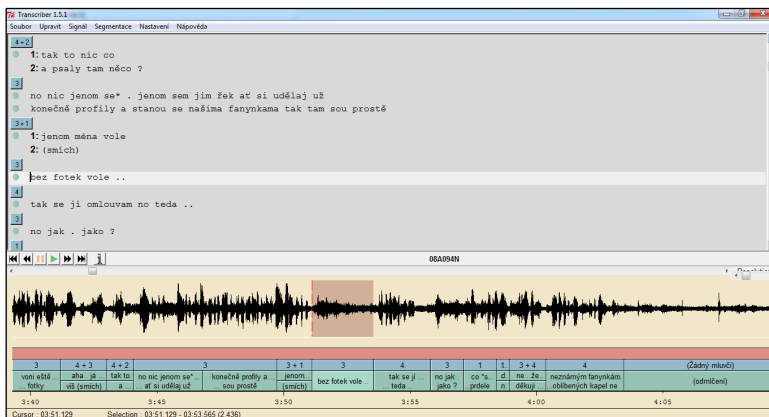
Počínaje korpusem ORAL2013 se přepis odehrává v transkripčním programu Transcriber.¹¹ To s sebou přináší – i vzhledem ke změně techniky převodu, opoře ve zrakové kontrole zvukového záznamu, odlišné segmentaci toku řeči¹² – změnu vnímání toku dialogu:¹³ systematictější a komunikačně velmi přínosně se sledují překryvy komunikantů¹⁴ (opouští se tedy snaha rozplést je do zdánlivě samostatných, nesimultánních replik), a zejména interpunkce převzatá z psaných textů je nahrazena interpunkcí pauzovou. Odlišují se tři kvality pauzy: pauza krátká, pauza delší a odmlčení; rozlišený jsou podle individuálního tempa jednotlivých mluvčích.

¹¹ Program Transcriber neslouží k automatickému převodu zvukového záznamu na psaný text. Po uložení nahrávky ale přepisovatelé umožňuje členit text do krátkých sekvencí podle kritérií formálních (max. délka, viz dále) nebo funkčních (replika jednoho mluvčího, pasáž překrývání více mluvčích ap.) Program sloužil jen pro přepis a kontrolu. Hotová, zkontrolovaná sonda převzatá pro zařazení do korpusu se dále zobrazuje v jiném nástroji (KonText, <https://www.korpus.cz/kontext/query?corpname=syn2020>, viz obr. 2) Všechny sondy vytvořené na Univerzitě Hradec Králové pod mou patronací mají na třetí pozici sedmimístné značky sondy písmeho H (př.: 09H015N = sonda vytvořená v roce 2009 v Hradci Králové jako 15. v pořadí, komunikační situace neoficiální).

¹² „[r]epliky jsou členěny na segmenty představující sémanticky, prozodicky i syntagmatically ucelenou sekvenci v průměru o 5–10 slovech (maximálně však 15).“ (ORAL2013)

¹³ K vertikální ose komunikátu viz kupř. podnětnou studii Nepravá hypotaxe v spontánních mluvených projevech (Bílková 2021)

¹⁴ K tomu např. Komrsková–Poukarová 2018, 41–56; Komrsková–Poukarová–Havlík 2019, 102–116.



Obrázek 1: Přepis zvukového záznamu v programu Transcriber

Zdroj: <https://wiki.korpus.cz/doku.php/cnk:oral2013>. Printscreensy aplikace s přepisem studentských sond jsem nepoživoval

Tradiční interpunkcí se vyznačuje pouze nápadně stoupavá intonace věty tázací, popř. vykřičníkem nápadná zvolací intonace. Důvodem rezignace na standardní značení mezivětných předělů bylo zejm. stanovisko odborníků z oblasti fonetiky a fonologie,¹⁵ že zvukové hranice takových vyšších textových jednotek lze nalézt jen stěží. Opakovaně se nám tak potvrzuje, že sémaziologická dekompozice textu probíhá primárně na základě kritérií významových.

Tabulka 1: Ukázka přepisu dialogu s pauzovou interpunkcí

Radomíra_7878	–	<i>to uplně cejtím že mně to chybí ty jo su uplně taková ..</i>
Adéla_5592	–	<i>no já taky . chcu jako něco . pořádně .</i>
Adéla_5592	–	<i>když -</i>
+ Radomíra_7878	–	<i>mně hlavně .</i>
Radomíra_7878	–	<i>já hlavně cejtím jak sem závislá na Dimim . až teď s* dycky když jakoby sem vod něj .</i>
Adéla_5592	–	@
Radomíra_7878	–	<i>teď jak sem sem jela tak sem říkala . ty jo mně se sem nechce .</i>
Adéla_5592	–	<i>(smích)</i>
Radomíra_7878	–	<i>najednou musím všechno sama vš všechno</i>
Adéla_5592	–	<i>no jasné no</i>
+ Radomíra_7878	–	<i>tak jakože ono ale</i>
Radomíra_7878	–	<i>mně to nepřide</i>

Vir: Zdroj: Korpus ORALv1, sonda 11A085N, <https://www.korpus.cz/>

¹⁵ Fonetický ústav Filozofické fakulty Univerzity Karlovy Praha, tehdy pod vedením prof. PhDr. Zdenky Palkové, CSc. (<https://fonetika.ff.cuni.cz/ustav/vyucujici/zdena-palkova/>)

4 Fenomén lingvální, paralingvální a nonlingvální

Nikoli nezajímavé je (a bylo nejen pro studenty filologických oborů) ověřování hranic žvlu lingválního a nonlingválního s přechodovou oblastí v jevech povahy paralingvální. Pravidla platná pro korpus ORAL2006 stanovila dosti jednoduše, že parazitní zvukové projevy se zaznamenávají sekvencí tří písmen: obvykle *hmm* pro zvuky spíše souhláskové, *eee* pro zvuky povahy převážně vokaliké. Smích se vyznačuje poznámkou v závorce (smích); v případě verbalizovaného, resp. i verbalizovatelného smíchu se zapisuje zvuková forma „co nejbliže slyšenému“.¹⁶

Už v průběhu sběru jazykového materiálu pro korpus ORAL2008 jsme narazili na nezbytnost odlišení komunikačně relevantních neartikulovaných nebo poloartikulovaných zvuků vyjadřujících souhlas, nesouhlas, pochybnosti, váhání apod. Zejména šlo o laryngálně-nazální zvuk (provázený případně odmítavým pohybem hlavou), graficky snad zaznamennatelný jako kombinace rázu a písmena pro konsonant *m*, tj. jako *'m-'m*. (Dlužno říci, že např. ani autoři Příruční mluvnice češtiny, dále PMC, nečinili nijak podstatného rozdílu mezi výrazy *hm*, *ehm* a *mbm*, radice je souběžně k interjekcím (první dva) a partikulím (třetí), obecně přitakacím (PMC 1995, 356 - § 593, a 365 - § 606). Garanti nově vznikajícího korpusu pochopili důležitost signalizace zásadně odlišné komunikační funkce prostředků hraniční povahy a stanovili další, strukturovanější pravidla transkripce: *hmm* pro přitakací responzní zvuk, *emem* pro nesouhlasný responzní zvuk, *mmm* pro souhláskové hezitační zvuky a *eee* pro zvuky hezitační samohláskové.¹⁷

Nejde o jev funkčně triviální. Při tvorbě hesla *ehm* pro Akademický slovník spisovné češtiny (dále ASSČ) odlišila Jana Špirudová¹⁸ sedm jeho různých komunikačních platností (v dnešní zveřejněné verzi jsou dostupné čtyři, pročež uvádím kompletní znění hesla v neoklešněné podobě):

ehm

- V1 vyjádření nejistoty, váhání, nerozhodnosti, mírných rozpaků
Ehm, stala se taková věc...; Já jsem, ehm – zkrátka, neřlobte se, že jsem vás tak otravoval.; Ehm, jak ti to říct, zkrátka kluci nevěděli, že ta bonboniéra není naše.; Ehm, nevím, vážně mě to nijak extra neláká.; „Mohl bych... *Ehm, mohl bych vás o něco požádat?*“ *vykotal.*

¹⁶ Zásady přepisu pro korpus ORAL2006. Dostupné z: https://wiki.korpus.cz/doku.php/seznamy:pravidla_2006

¹⁷ Inspirativní je zpracování výrazu *no* jako diskurzivního markeru (Bílková-Zeman 2020, 191-198)

¹⁸ Jako spoluautor ASSČ a osoba pověřená kontrolou tohoto hesla vidím informace o zakladateli (zpracovateli) hesla a autorech jeho vyšších verzí. Protože interjekce představují jednu z mých gramatických specializací, Jana Špirudová mě o revizi tohoto hesla výslovně požádala.

- V2 opatrný, váhavý souhlas, kladná odpověď na otázku
Dáte si pivo? – Ehm, díky.; Ehm, tak to můžu říct docela přesně.
- V3 zdůraznění otázky, pobídka k odpovědi
Ehm, nevádí ti, když půjdu s vámi?; Ehm a jak byste sestavil rozpočet vy?; Mohu vám, ehm, ehm, nějak pomoci?
- V4 zvuk při odkašlání, kterým chce zprav. někdo na něco upozornit (např. že se chystá něco říct)
„Ehm,“ odkašlal si ředitel.; „Ehm, ehm, ehm,“ zakašlala paní Renata a kopla bo pod stolem prudce do holeně.; „Ehm, ehm,“ odkašlala jsem si významně, jak to mám ve zvyku před klíčovým projevem na jednání s komplikovaným partnerem.
- V5 výraz zamyšlení nebo snahy posoudit návrh, event. rozvinout téma
Ehm, co kdybychom šli radši dovnitř, je tu celkem zima.; Ehm, ehm...ještě jedna věc.; Ehm – a co takhle použít polarizační filtr?
- V6 výraz nelibosti, nesouhlasu nebo mírného odmítnutí
Ehm, to teda děkuju.; Ehm, to jste myslel vážně, pane? Ehm, to jako nemám říkat?
V7 povzdechnutí, vyjádření mírné námitky
Ehm, a je to tady zase.; Ehm, tak to už tady jednou bylo.

V úzu se vedle základní podoby citoslovce *ehm* objevují varianty opakující písmeno *m*, čímž se naznačuje delší trvání zvuku a zesiluje se tak komunikační účinek: *ehmm, ehmmm*. (ASSČ, heslo *ehm*, verze z 26. 5. 2022, před umrtvením významů 5-7)

Tabulka 2: Frekvence grafických forem rezponzních a hezitačních zvuků

	Grafická forma	Absolutní frekvence ¹⁹	Frekvence na milion slov (i. p. m.)
ORALv1	<i>hmm</i>	50 095	7 874,46
	<i>emm</i>	636	99,97
	<i>ééé</i>	5	0,79
	<i>em em</i>	1	0,16
	<i>chm</i>	1	0,16
PMK	<i>hm</i>	760	927,66
	<i>ehm</i>	143	174,55
	<i>eee</i>	60	73,24
	<i>hmm</i>	35	42,72
	<i>chm</i>	27	32,96
	<i>mmm</i>	25	30,52
	<i>ééé</i>	4	4,88
	<i>emem</i>	3	3,66
	<i>em em</i>	1	1,22
	<i>ehmm</i>	1	1,22
ORTOFON	<i>hmm</i>	29 200	11 403,62

¹⁹ Absolutní frekvence je ovlivněna velikostí konkrétního korpusu. Parametr i. p. m. udává přepočtenou frekvenci na ideální jeden milion slov daného korpusu.

	Grafická forma	Absolutní frekvence ¹⁹	Frekvence na milion slov (i. p. m.)
	<i>emm</i>	579	226,12
	@ / @@	23 606	9 218,97
	<i>chm</i>	6	2,34
	<i>eee</i>	3	1,17
	<i>mmm</i>	1	0,39

Zdroj: vlastní

S ohledem na budoucí lemmatizaci byla cíleně sjednocována přirozeně oscilující forma zápisu (respektující nakonec pokyn zapisovat co nejvěrněji slyšené). Logicky se tak potlačuje možnost vyjádřit počtem grafémů či jejich skladbou délku či obecnou povahu zvuku. Do transkriptů se však přece jen individuální reflexe v jisté míře prosadila.

Současná praxe dvouúrovňového přepisu, uplatňovaná pro korpus ORTOFON, nahlíží tuto oblast komunikace zase jiným prizmatem. Pravidla²⁰ ukládají přepisovat na rovině ortografické responzní zvuky přitakávací *hmm*, responzní zvuky odmítací *emm*, citoslovce bez lexikalizované podoby jako & a hezitací zvuky znakem @ pro kratší hezitace a @@ pro delší. Není zřejmé, jestli ojedinelé odlišné zaznamenané formy vypovídají o snaze vědomě zachytit specifický jev přesahující pravidla, nebo jsou jen projevem přehlédnutí. (Pro srovnání: Korpusy psaného jazyka dokládají, že psané texty, zejm. beletristické a publicistické, pracují i s dalšími formami, v mluvených korpusech registrované nanejvýš okrajově; v SYNv11 kupř. *chm* 561x / 0,09 i.p.m., *chmm* 131x / 0,02 i.p.m., *chmmm* 4x, *chmmmmm* 1x. To může odpovídat komunikačně funkčnímu, převážně nazálnímu nelaryngálnímu výdechovému zvuku, povzdechu se sevřenými rty.)

*Víte ... **chm** ... zrovna než jste přišel ... prohlížel jsem podložní skříčka ... a přišel jsem najedno, které by vás mohlo zajímat.* (korpus SYNv11)

*. --- brácha a tatínek si dycky přál syna a syna teda měl . i když se narodil pozdějc než já . ale dycky já sem byla taková .. **mmm** .. no nechci říct úplně v koutku . ale Jirka dycky moh víc ---* (korpus ORALv1)

²⁰ Transkripcie v korpusu ORTOFON. Dostupné z: <https://wiki.korpus.cz/doku.php/cnk:ortofon:pravidla>

Tabulka 3: Ukázka přepisu dialogu s pauzovou interpunkcí a responzními zvuky

Ivana Š.	– <i>bmm</i>
Rozálie B.	– <i>víš ? .. hele fakt do mě nic nebylo hele Ilonko</i>
Ivana Š.	– <i>bmm ..</i>
Ivana Š.	– <i>bmm .. bmm .. bmm</i>
Rozálie B.	– <i>do mě nic nebylo jo</i>
Ivana Š.	– <i>bmm</i>
Rozálie B.	– <i>no a naštěstí teda když přijeli ..</i>
Rozálie B.	– <i>tak mi řekli že to byl teror že je . v devět večer přivezli na pokoje</i>
Rozálie B.	– <i>a v sedm ráno vstávali protože se zase jede ne ..</i>
Ivana Š.	– <i>jasně byli urvaný</i>

Zdroj: Korpus ORTOFONv2, sonda 18X096N, <https://www.korpus.cz/>

Tabulka 4: Ukázka přepisu s nelexikalizovanými interjekcemi (&) a hezitačními zvuky (@)

Anna V.	– <i>tady v Brně má @ Kokino se to jmenuje je to značka . to dělá nějaká slečna a ..</i>
Anna V.	– <i>má tam @ čokolortik s mořskou solí měla #s to někdy ?</i>
Miriám Š.	– <i>ehm</i>
Anna V.	– <i>jestli máš ráda čokoládku a karamel .. z tohoble by ses úplně potento protože -</i>
Miriám Š.	– <i>musíš dát potom nějaký odkaz</i>
Anna V.	– <i>&</i>
Anna V.	– <i>& . to je víc kde to mají ? . to mají na Gor* @ Gorkého . víc kde</i>

Zdroj: Korpus ORTOFONv2, sonda 14A004N, <https://www.korpus.cz/>

5 Delimitace lexikálních jednotek a jejich hláskové struktury

S relativitou textových jednotek na úrovni věty a souvětí, resp. výpovědi, jakož i s otázkou hranic jevů lingválních a nonlingválních souvisí relativita centrální jednotky lexikálního systému (a svým způsobem centrální jednotka jazyka vůbec), totiž slova. Znovu platí, že vyčlenění jednotky v proudě řeči je značně relativní, formální delimitace vůči okolním jednotkám pauzou je hypotetická. Přírozený jazyk ze své podstaty inklinuje k jednotě formy a funkce, což posiluje tendenci k formálnímu splývání původně analytických jednotek. Podobně vede změna funkcí k proměnám formy. Všechny náznaky těchto procesů v mluvené komunikaci má proto smysl registrovat.²¹

²¹ K tomu viz např. Blatná 2006, 7–19, Vondráček 2013, 79–89, Volín 2015, 44–47, Vondráček 2018, 35–110.

Tak by bylo možno sledovat ukázkovou proměnu desubstantivních forem kontaktového prostředku citoslovečného a částicového rázu *ty vole* (chápaného nejen ve školském prostředí funkčně neadekvátně jako vulgární, resp. zhrubělé²² oslovení) ve zvukové a grafické realizaci: jeho hláskovou variabilitu či erozi včetně intonačního, resp. obecně prozodického pozadí v proměnách doby a v závislosti na funkci (→ part. *tyvoe* → *tyoe* → *tye*; → interj. *tývoe* → *týe* vedle snad tabuové náhražky *týjo* → *týo*). Z dobrých důvodů ovšem tematizují samostatně dva jiné jevy podobné povahy.

Důvody pro upuštění od detailní analýzy jevu již naznačeného (deskripce slovnědruhově transpozice kontaktové fráze spočívající v pronominu + vokativu substantiva *ty vole* v kontaktovou partikuli *tye* či podivovou interjekci *týe*) jsou naznačené povahy axiologické (byť příznak vulgárnosti či zhrublosti v jazykovém materiálu nebyl v žádném případě důvodem k vyřazení sondy). Volím však dva jiné, analogické, ovšem různě složité případy funkční a slovnědruhově proměny provázené destrukcí formy: proměnu substantivního vokativu *člověče* v jednotku pronominálního typu *čěk*, v kontaktovou partikuli *čěče* a v podivovou interjekci *čěče*. Dílem jiné (tabuové, ovšem v tomto případě nábožensky tabuové) jsou pohnutky k obměnám forem (a k jejich následným deformacím) u interjekcí s funkcí povzdechu či zaklení typu *proboba*, *prokerindapána*. Podobných ilustračních typově odlišných dokladů je možno zvolit desítky; uvádím právě jen některé z těch, jež si obvykle i v přepisovacích pravidlech vyžádaly samostatné pojednání. K oběmu viz např. Laubeová 2020.

5.1 Typ *člověk* / *čěk*, *člověče* / *čoveče* / *čěče*

Starší české slovníky zaznamenávají výraz *člověče* jako vokativ substantiva, funkčně jej označují jako expresivní oslovení, zařazují jej k hovorové nebo obecné češtině. ASSČ uvádí původní vokativní formu jako samostatné heslo, slovnědruhově interpretované jako interjekci (kolokviální vyšší), k ní dále i variantní formy *čoveče*, *čěče* jako kolokviální. Právě o zaznamenání hláskové eroze slova při slovnědruhově transpozici nám jde (a zmínit je třeba i funkci částicovou, v níž výraz nevykazuje znaky větňého ekvivalentu). ASSČ také pod substantivním heslem *člověk* uvádí uplatnění ve funkci zájmena neurčitého, záporného nebo osobního s významem 'blíže neurčená osoba, někdo, kdokoliv, každý, mluvčí (já), adresát (ty, vy) apod., ve

²² SSJČ i SSČ svorně: zhrub., i nadávka, hlupák; kontaktovou frází *ty vole* však tyto zdroje ani IJP neuvádějí.

spojení se záporným slovesem žádná osoba, nikdo!. I tato slovnědruhová transpozice je v hovoru provázena destrukcí formy (*čěke*).

The screenshot shows the KonText search results for the word 'čěče'. The interface includes a search bar, filters, and a list of search results. The selected result shows the word 'čěče' in context with its frequency and other linguistic information.

Obrázek 2: Ukázka konkordancí s výrazem *čěče* ve webovém rozhraní KonText

Následující tabulka ukazuje, které z forem zachytil který z korpusů, popř. jak jsou morfologicky značkovány.

Tabulka 5: Korpusově evidované formy výrazů *člověk* / *ček*, *člověče* / *čoveče* / *čěče*

	Grafická forma	Absolutní frekvence	Frekvence na milion slov (i. p. m.)
PMK	<i>čoveče</i> / <i>čověče</i>	49 / 3	59,81
PMK	<i>člověče</i>	15	18,31
PMK	<i>čoeče</i>	1	1,22
PMK	<i>ček</i>	1	1,22
ORAL2008	<i>čoveče</i> / <i>čověče</i>	130 / 14	96,33
ORAL2008	<i>člověče</i>	16	11,86
ORAL2008	<i>čěče</i>	2	1,48
ORAL2008	<i>ček</i>	2	1,48
ORAL2013	<i>čoveče</i> / <i>čověče</i>	547 / 10	166,49
ORAL2013	<i>člověče</i>	16	4,87
ORAL2013	<i>čěče</i>	17	5,17
ORAL2013	<i>ček</i>	0	0
ORTOFONv2	<i>čoveče</i> / <i>čověče</i> (subst.)	12 / 0	4,69
ORTOFONv2	<i>člověče</i> (subst.)	233	90,99
ORTOFONv2	<i>čěče</i> (subst.)	0	0
ORTOFONv2	<i>ček</i> (subst.)	0	0

Zdroj: vlastní

... --- sou takový ještě vomámený ty ryby , že jo ? ... povídal, to sem ale m* ...: --- voni sou přibblý, **čoveče**, voni vůbec nejdou, voni zůstanou u břehu v tý trávě a vůbec nikam nejdou . no (korpus ORAL2008)

už sem nemohla . no a já říkám . zas to máš blbě . **čěče** co t* jako něco takovýho no a . tam je jedna taková šileně . mmm . šilená poděška . a vona botová úplně jako . **čěče** jo a toto . a teď se tam řebtali . brozný (korpus ORAL2013)

... no prostě, stálo jim to za to, no . jesi ... : no pokavaď, že jo, pokavaď je to fyzicky a duševně vodromná, tak **čěk** nemůže říct: stálo to za to, no . holt, já neverím, no . (korpus BMK)

... v noci v noci dyž se **čěk** probudí tak to bolí, a nevím jak si lehnout, fakt je že pak si nebudu smět dát noby křížem že jo třebaš, vsedě určitě ne, jo jak s* **ček** normálně sedí a přebodí si nobu tak (korpus ORALv1)

Stejně jako koreluje změna funkce slovní formy s omezením či ztrátou paradigmatu a s erozí fonetické struktury, dochází ke změnám u víceslovných lexikálních jednotek (historicky nejspíš takto haplogicky vznikla slova *vašnosti* ← *Vaše Jasnosti* s dodatečným Nsg *vašnosta*, *slečna* ← *slechtična*, tímž směrem se ubírala forma *pančelka* ← *paní učitelka*, v mírné podobě nakonec i dosud nekodifikované *nashledanou* ← *na sbledanou*, **sbledaná*). Také toto splývání forem a redukční procesy má smysl sledovat. I pro druhé uvedu ilustrační příklad.

5.2 Typ *probaha* / *probůh*, *prokristapána*, *propánajána*, *propánaboha*

Všechny velké slovníky češtiny 20. století registrují lexikální formu *probaha* (některé i *probůh*) jako citoslovce;²³ neuvádějí varianty psané analyticky (tudíž nemusí řešit pravopis velkých písmen vázaný na tuto funkci). Lístková excerpta²⁴ tvořící podklad Příručního slovníku jazyka českého (dále PSJČ) registrují ve třinácti dokladech i lexikální formy *prokristapána* a jeho tabuovou obměnu *prokřindapána*, vedle ní s nižším počtem dokladů i *propánajána*, *propánaboha*, vše hodnoceno jako interjekce (slovníková hesla ovšem nevznikla). Pozdější slovníky už heslo uvádějí (SSČ²⁵ *prokrista(pána)* i *pro křista pána*, SSJČ²⁶ *prokřista*, *prokřistapána*, psáno též *pro křista pána*, obdobně *propána*, *propánajána*, *propánaboha*, *propánakrále* (psáno též *pro pána boha*), rovněž jako interjekci. Psaní malého písmena lze přiřknout spíše režimním pohnutkám než snaze odlišit rouhavé brání jména Syna Božího nadarmo od vroucího obvolávání se na něj; proti tomuto směru uvažování naopak hovoří připouštěné analytické psaní. Nezdá se nepodstatné, jak je týž výraz zaznamenaný v přepisech pro mluvené korpusy sto let od vzniku PSJČ.

²³ Srov. též např. Kleňhová 2010; 2012, 238–254.

²⁴ Elektronicky jsou dostupná z webu Ústavu pro jazyk český AV ČR, v. v. i.: <https://psjc.ujc.cas.cz?>

²⁵ *Slovník spisovné češtiny pro školu a veřejnost*, 1994. Praha: Academia.

²⁶ *Slovník spisovného jazyka českého*, 1989. Praha: Academia.

Tabulka 5: Korpusově evidované formy výrazů *prokristapána / propánajána / propánaboha*

	Grafická forma	Absolutní frekvence	Frekvence na milion slov (i. p. m.)
PMK	<i>proboba / probůh</i>	7 / 0	8,54 / 0
PMK	<i>pro boba</i>	0	0
PMK	<i>prokristapána / propánajána / propánaboba</i>	3 / 0 / 0	3,66 / 0 / 0
PMK	<i>pro krista pána / pro pána jána / pro pána boba</i>	0 / 0 / 0	0 / 0 / 0
BMK	<i>proboba / probůh</i>	9 / 1	15,1 / 1,68
BMK	<i>pro boba</i>	0	0
BMK	<i>prokristapána / propánajána / propánaboba</i>	0 / 0 / 0	0 / 0 / 0
BMK	<i>pro krista pána / pro pána jána / pro pána boba</i>	0 / 0 / 0	0 / 0 / 0
ORAL2008	<i>proboba / probůh</i>	6 / 0	4,45 / 0
ORAL2008	<i>pro boba</i>	0	0
ORAL2008	<i>prokristapána / propánajána / propánaboba</i>	4 / 0 / 0	2,96 / 0 / 0
ORAL2008	<i>pro krista pána / pro pána jána / pro pána boba</i>	0 / 0 / 0	0 / 0 / 0
ORAL2013	<i>proboba / probůh</i>	53 / 0	16,13 / 0
ORAL2013	<i>pro boba</i>	7*	5,04
ORAL2013	<i>prokristapána / propánajána / propánaboba</i>	0 / 1 / 0	0 / 0,3 / 0
ORAL2013	<i>pro krista pána / pro pána jána / pro pána boba</i>	0 / 0 / 0	0 / 0 / 0
ORTOFONv2	<i>proboba / probůh</i>	23 / 2	8,98 / 0,78
ORTOFONv2	<i>pro boba</i>	1*	0,39
ORTOFONv2	<i>pro krista pána / pro pána jána / pro pána boba</i>	2 / 0 / 0	0,78 / 0 / 0
ORTOFONv2	<i>prokristapána / propánajána / propánaboba</i>	1 / 0 / 0	0,39 / 0 / 0

* (také) v adekvátní funkci předložkové substantivní vazby

Zdroj: vlastní

.. každopádně si myslím že by ženy se svými vědomostmi . @ . at už životními nebo že škol . neměly zůstat takzvaně na ocet ale **probůh** už ne žádnou další Anežku Hodinovou Vzpurnou výstřelky . @ . byly vždycky . @ . no (BMK)

mě zamkly v pokoji .. co si vůbec myslely .. že jsem .. jo a Pavel říkal no když #s jím utíkala .. tak . pr* . **prokristapána** copak měli dělat .. sestřička tě bonila po baráku a nemohla tě najít .. tak ji zamkli .. (ORTOFONv2)

. a jo , vlastně , já sem jí slíbila tu kaši , **prokristapána** . čekáš , až ti to vystydne? ne , už du . juž odcházejm . bože , bože . no , podívej . eee . (ORAL2008)

lusev + pujova a samý placení že jo

jo *právě*
 lusev + pujova *to stojí peněz potom ježkovy voči tak*
to je jistý to je pe peněz vid'*
 lusev + pujova *ž* já jim do tobo mluvit nebudu propánajána --- nákyj*
 jo to jako jistý (ORAL2013)

voduna *no to víš no tak --- nevíš jaká budeš ty (smích)*
 rehega *doufám že takováde ne*
 voduna *(se smíchem) no to nevíš tak jako*
 rehega *pro boha svatýho*
 voduna *(smích) no a tak jako (ORAL2013)*

Světlna R. *.. tak tam byla nějaká mladá holka myslím si že nějaká středoškolačka že se zaučovala ne? .. učila se*
 Růžena J. (pousmání) *jo .. ne tak to já bych asi nechtěla aby ta se zaučovala na mně probůh*
 Světlna R. (smích) *no .. a právě .. přede mnou teda byla nějaká holka .. nejdřív brali z prstu to už ti brali krev z prstu .. taký? .. nebo ne? (ORTOFONv2)*

V zájmu automatizovatelné lemmatizace a snazšího vyhledávání vychází dvouúrovňový korpus ORTOFONv2 z ortografického přepisu; v zásadě proto není divu, že emocionální výrazy zaznamenává především analyticky. Jeho (popředložkové) komponenty jsou však tagovány jako substantiva. Tím se zcela stírá rozdíl mezi interjekcí a substantivním užitím, doloženým v témže korpusu kupř. sekvencí:

...to jsou .. to je pohled prostě pro boha jako tak jo jakože jo takové jako přirovnání jakože .. (ORTOFONv2)

Tuto pasáž, věnovanou změnám forem a funkcí prostředků mluvené komunikace, jsem uvedl příkladem spojení *ty vole / tyoe / tyve / tye*, užívaného mj. jako kontaktní partikule. Při sběru materiálu pro korpusy, na nichž jsme spolupracovali, jsem považoval za potřebné všechny podobné názny funkčních změn zaznamenávat. Poslední zmíněný korpus ORTOFONv2 eviduje 2630 dokladů formy *vole* a jedinou formu *voe*. Všechny ortograficky ztvárněné výrazy jsou tagovány jako substantiva. Tímto postupem je zcela zastřena funkční diference partikule, interjekce a jména (jako takového významně frekvenčně omezeného):

Zbyněk S. *hochu ale co mě překvapil třeba Jenda NP .. tam byl a .. ty jo ale já tomu Babišovi fandím ..*
 Norbert R. (smích) *kokote jeden vole že to tak řeknu ..*

6 Závěr

Pokusil jsem se shrnout několik zásadních, dosti různorodých momentů, jež jsme seznali za hodné pozornosti při tvorbě mluvených korpusů s pomocí studentů filologických oborů. Ti, nakonec stejně jako my sami jsme se při obstarávání zvukových záznamů neformálních komunikčních situací a jejich přepisu museli zabývat značně obecnými otázkami pragmlingvistickými a sociolingvistickými (povaha komunikační situace), obecně lingvistickými (sémioticky heterogenní povaha komunikátu, počítaje v to prolínání prostředků lingválních, paralingválních a nonlingválních), lexikologickými a morfosyntaktickými (delimitace jednotek v proudu řeči a zachycení jejich relevantní formy).

Porovnáním více korpusů mluveného jazyka se snažím ukázat dynamiku sledování určitých jevů a zisky, popř. ztráty plynoucí ze změny intence. Dílčí omluvou budiž, že omezení prostoru pro individuální snahu transkriptora zaznamenat co nejdělejší obraz zvukové formy je veden objektivizačním úsilím. Právě to bylo i pohnutkou k ústupu od interpunkce psaných textů a přechodu k interpunkci pauzové. Dalším zásadním motivem bude (resp. dlouhodobě je) strojové učení a automatické zpracování mluvené formy řeči.

Literatura

- Jana BÍLKOVÁ, 2021: Nepravá hypotaxe v spontánních mluvených projevech. *Štýl – komunikácia – kultúra*. Ur. Zuzana Popovičová-Sedláčková. Bratislava: Univerzita Komenského. 417–427.
- Jana BÍLKOVÁ, Jiří ZEMAN, 2021: Diskurzní marker no a v mluvené češtině. *Lingvistika – korpus – empirie*. Praha: Ústav pro jazyk český AV ČR, v. v. i., 191–198.
- Renata BLATNÁ, 2006: *Víceslovné předložky v současné češtině*. Praha: NLN, Nakladatelství Lidové noviny: Ústav Českého národního korpusu, 9–17.
- Miroslav GREPL, 1995: Příruční mluvnice češtiny. Praha: NLN, Nakladatelství Lidové noviny, 1995.
- Jana HOFFMANNOVÁ, Marie MIKULOVÁ, 2011: Korpusy mluvené češtiny a možnosti jejich využití pro poznání rozdílných "světů" mluvenosti a psanosti. *Korpusová lingvistika Praha 2011 - 2 Vězeň a výstavba korpusů*. Ur. František Čermák. Praha: NLN, Nakladatelství Lidové noviny. 78–92.
- Eliška KLEŇHOVÁ, 2010: Interjekce v českém jazykovém systému. Kvalifikační práce magisterská, Filozofická fakulta Univerzity Karlovy.
- Eliška KLEŇHOVÁ, 2012: Postavení a užívání interjekcí v současné češtině. *Naše řeč*, roč. 95, č. 5, 238–254.
- Zuzana KOMRSKOVÁ, Petra POUKAROVÁ, 2018: Kdo, kdy a proč skáče komu do řeči aneb překryvy ve spontánní mluvené češtině. *Korpus – gramatika – axiologie*, č. 17, 41–56.
- Zuzana KOMRSKOVÁ, Petra POUKAROVÁ, Martin HAVLÍK, 2019: Překryvy replik. *Syntax mluvené češtiny*. Ur. Jana Hoffmannová, Jiří Homoláč, Kamila Mrázková. Praha: Academia. 102–116.
- Zuzana LAUBEOVÁ, 2020: *Mluvenost v dialogické elektronické komunikaci*. Kvalifikační práce dizertační,

Filozofická fakulta Univerzity Karlovy.

Jan VOLÍN, 2015: Vztah mluvené a psané formy jazyka. *Mluvnice současné češtiny. 1, Jak se píše a jak se mluví*. Ed. Václav Cvrček. Praha: Univerzita Karlova, nakladatelství Karolinum. 44–47.

Miloslav VONDRÁČEK, 2008: Diskrétnost řečových jednotek. *Čeština v mluveném korpusu*. Ur. Marie Koprřivová, Martina Waclawičová. Praha: Nakladatelství Lidové noviny / Ústav českého národního korpusu. 255–261.

Miloslav VONDRÁČEK, 2013: Vlastnosti slova a slovní druhy. *Akademická gramatika spisovné češtiny*. Ur. František Šticha. Praha: Academia, 2013. 79–89.

Miloslav VONDRÁČEK, 2018: Druhy slov. *Velká akademická gramatika spisovné češtiny*. Ur. František Šticha. Praha: Academia, 2018. 35–110.

Slovník spisovného jazyka českého, 1989. 2. vyd. Praha: Academia.

Slovník spisovné češtiny pro školu a veřejnost, 1994. 2. vyd., opr. a dopl. Praha: Academia.

Internetové zdroje:

ASSČ: Akademický slovník současné češtiny. Dostupný z: <https://slovníkcestiny.cz/uvod.php>.
Poslední přístup 15. 6. 2023.

BMK: Brněnský mluvený korpus. Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:bmk>. Poslední přístup 15. 6. 2023.

ORAL2006: Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:oral2006>. Poslední přístup 15. 6. 2023.

ORAL2008: Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:oral2008>. Poslední přístup 15. 6. 2023.

ORAL2013: Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:oral2013>. Poslední přístup 15. 6. 2023.

ORTOFONv2: Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:ortofon>. Poslední přístup 15. 6. 2023.

PMK: Pražský mluvený korpus. Dostupný z: <https://wiki.korpus.cz/doku.php/cnk:pmk>. Poslední přístup 15. 6. 2023.

PSJČ: *Přřručnř slovník jazyka českého. Kartoteka lexikálního archřvu (1911–1991)*. Dostupný z: <https://psjc.ujc.cas.cz/search.php?> Poslední přístup 15. 6. 2023.



III

Vidiki raziskovanja
govora in govorenega
diskurza

MI I NAŠI, ONI I NJIHOVI U POLITICI: OSOBNJE DEIKSE U GOVORIMA HRVATSKIH SABORSKIH ZASTUPNIKA

GORANKA BLAGUS BARTOLEC

Institut za hrvatski jezik, Zagreb, Hrvatska
gblagus@ihji.hr

Kao upućivačke jedinice deikse su sastavnica jezične strukture koje sudionicima komunikacijskoga događaja omogućuju da se postave prema okolnostima iskaza (osobe, društvo, prostor, vrijeme, diskurs). U radu se analizira upotreba zamjenica *mi i naši, oni i njihovi* kao osobnih (personalnih) deiksa u govoru hrvatskih saborskih zastupnika prema potvrdama iz korpusa ParlaMint-HR 2.0 (Croatian parliament) 2016–2020 dostupnom na korpusnoj platformi (No)Sketch Engine. Političar kao pojedinac u političkim govorima najčešće govori u prvom licu množine te se persuzivnost i intencionalnost njegove političke argumentacije u načelu temelji na stavovima društvene skupine koju predstavlja ili kojoj se ideološki priklanja. Cilj je rada opisati sintaktička i značenjska obilježja osobnih deiksa u govoru saborskih zastupnika s obzirom na referente na koje upućuju te utvrditi koliko se upotreba osobnih deiksa temelji na prototipnoj slici *nas* i *njih* kao polariziranih strana, a u kojoj su mjeri *mi/naši* i *oni/njihovi* ravnopravni, odnosno *mi* se postavlja kao subjekt koji štiti ili je na strani referenata obuhvaćenih deiksom *oni*.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.12](https://doi.org/10.18690/um.ff.4.2024.12)

ISBN
978-961-286-882-6

Ključne besede:

govornik,
hrvatski jezik,
osobne deikse,
parlamentarni govor,
politički diskurs



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.12](https://doi.org/10.18690/um.ff.4.2024.12)

ISBN
978-961-286-882-6

Keywords:
speaker,
Croatian,
person deixes,
parliamentary speech,
political discourse

WE AND OURS, THEY AND THEIRS IN POLITICS: PERSON DEIXES IN THE SPEECHES OF CROATIAN PARLIAMENTARIANS

GORANKA BLAGUS BARTOLEC

Institute for the Croatian Language, Zagreb, Croatia
gblagus@ihjj.hr

As units of reference that refer to real content and circumstances (persons, society, space, time, discourse), deixes are part of the language structure that enable the participants of a communication event to position themselves according to the circumstances of the utterance. The paper analyses the use of the pronouns *we* and *ours*, *they* and *theirs* as person deixes in the speech of Croatian parliamentarians using the ParlaMint-HR 2.0 corpus 2016–2020, which is publicly available on the NoSketch Engine tool. A politician in political speeches usually speaks in the first person plural, and his political argumentation is, in principle, based on the views of the social group he represents or to which he is ideologically inclined. The aim of the paper is to describe the syntactic and semantic features of person deixes in the speeches of parliamentarians with regard to the referents they refer to, and to determine to what extent the use of person deixes is based on the prototypical image of *us* and *them* as polarised parties, and in which *we/ours* and *they/theirs* are equal, i.e. the deixis *we* posits itself as a subject that protects or sides with the referents covered by the deixis *they*.



MI IN NAŠI, ONI IN NJIHOVI POLITIKI: OSEBNI DEIKTIKI V GOVORIH HRVAŠKIH POSLANCEV

GORANKA BLAGUS BARTOLEC

Inštitut za hrvatski jezik, Zagreb, Republika Hrvatska

gblagus@ihji.hr

Deiktiki so kot nanašajnske enote sestavina jezikovne strukture, ki udeležencem komunikacijskega dogodka omogoča umestitev glede na okoliščine izreka (osebe, družba, prostor, čas, diskurz). Prispevek analizira rabo zaimkov *mi* in *naš*, *oni* in *njihovi* kot osebnih (personalnih) deiktikov v govoru hrvaških poslancev, zbranih v korpusu ParlaMint-HR 2.0 2016–2020, dostopnem na platformi (No)Sketch Engine. Politik v političnih govorih praviloma uporablja prvo osebo množine, prepričljivost in intencionalnost njegove politične argumentacije pa načeloma temelji na stališčih družbene skupine, ki jo zastopa ali ji je ideološko naklonjen. Cilj prispevka je opisati skladske in pomenske značilnosti osebnih deiktikov v govorih poslancev glede na reference, na katere se nanašajo, in ugotoviti, v kolikšni meri raba osebnih deiktikov temelji na prototipski podobi o *nas* in *njih* kot polariziranih strank ter v kolikšni meri sta *mi/naši* in *oni/njihovi* enakovredna, tj. deiktik *mi* se postavlja kot subjekt, ki varuje ali ki se postavlja na stran referenc, vključenih v deiktik *oni*.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.12](https://doi.org/10.18690/um.ff.4.2024.12)

ISBN
978-961-286-882-6

Ključne besede:
govornik,
hrvaški jezik,
osebni deiktiki,
parlamentarni govor,
politični diskurz



Univerzitetna založba
Univerze v Mariboru

1 Uvod: deikse kao upućivačke jedinice¹

Deikse kao jezične jedinice (riječi i druge gramatičke i leksičke kategorije) kojima se upućuje na okolnosti (sudionike, prostor, vrijeme, društvo, diskurs) iskaza od 70-ih godina 20. st. (Fillmore 1975) do danas opširno su proučavane u okviru pragmalingvistike, tekstne lingvistike te kritičke analize diskursa.² U pozadini proučavanja deiktčnosti u jeziku zapravo je čovjekovo izvanjezično (kognitivno i tjelesno) iskustvo s okolinom koja ga okružuje (Diessel 2012: 2). Iz takve perspektive razumijevanje deiksa temelji se na čovjekovoj sposobnosti da se preko njih postavi u kakvu stvarnu ili fiktivnu izvanjezičnu (mjesnu, vremensku, društvenu) perspektivu koju u jezičnom iskazu izražava netko drugi (Tomasello 2006), pa je temeljna uloga deiksa povezivanje jezične strukture s izvanjezičnim kontekstom u kojemu se iskaz ostvaruje (Levinson 1983: 55). Deiktčnost se u gramatičkim opisima povezuje s kategorijom foričnosti kao svojstvu onih riječi koje se temelje na značenju upućivanja, odnosno mogućnosti zamjenjivanja drugih riječi u rečenici ili na razini cijeloga teksta, a što je temeljno obilježje zamjeničkih riječi (Silić i Pranjković 2005: 241). Upravo zbog mogućnosti da zamjenjuju ili upućuju na druge riječi deikse se u literaturi određuju kao jedinice na granici punoznačnosti s obzirom na to da se njihovo značenje uspostavlja u konkretnom iskazu (Karlić i Bago 2020, prema Kordić 2002).

Iako deikse kao jezične jedinice upućuju na različite referente, a to su osobe, prostor, vrijeme, društvo te sam tekst (ili diskurs) u kojemu se ostvaruju, u deiktčko središte postavlja se govornik prema kojemu se određuju svi drugi referenti. U središtu je dakle govornikov prostor u kojemu govori ili o kojemu govori, govornikovo vrijeme u kojemu ili o kojemu govori te govornikova komunikacijska situacija (usp. Levinson 1983: 63–64, Karlić i Bago 2020: 750). U iskazu se na govornika i na druge osobe kao adresate, a to su sugovornik i negovornik, primarno upućuje kategorijom lica, a deikse koje izražavaju kategoriju lica i koje upućuju na sudionike iskaza određene su kao osobne (personalne) deikse (Levinson 1983: 68–73, Diessel 2012: 8–10). Osobne se deikse izražavaju osobnim zamjenicama (*ja, ti, on/ona/ono*) *mi, vi, oni/one/ona*) u kojima je u hrvatskome jeziku sadržana kategorija lica (1., 2. i 3. lice)

¹ Ovaj je rad sufinancirala Hrvatska zaklada za znanost projektom MWE-Cro: *Višerječni izrazi u hrvatskome jeziku – leksikološki, računalnolingvistički i glotodidaktički pristup* [šifra projekta: IP-2022-10-7697].

² Deikse su u navedenim jezikoslovnim okvirima sustavno zastupljene i u hrvatskim istraživanjima (Kordić 1995, 1996, 2002, Podboj 2011, Katnić-Bakaršić 2018, Karlić i Bago 2020, Badurina i Hrg 2022, Blagus Bartolec 2023 itd.).

te kategorija broja (jednina i množina) i kategorija roda (muški, ženski i srednji rod u 3. l. jd. i mn.). Osobne se deikse izražavaju i odgovarajućim glagolskim oblicima koji sadržavaju kategorije lica i broja (Pranjković 2013: 12–13, Karlić i Bago 2020: 751). Koliko je važna kategorija lica, odnosno tumačenje da je govornik deiktičko središte, pokazuje se i u razumijevanju drugih deiksa, vremenskih ili prostornih, npr. prilozi *ovdje, ondje, onda, tada*, koje se konkretiziraju, odnosno obilježavaju preko kategorije lica, ponajprije govornika.³

Polazeći od navedenih postavki, cilj je rada opisati obilježja osobnih deiksa u govoru saborskih zastupnika te utvrditi u kojoj se mjeri upotreba osobnih deiksa temelji na prototipnoj slici *nas* i *njih* kao podijeljenih političkih strana, a u kojoj su *mi/naši* i *oni/njihovi* ravnopravni pa se govorni subjekt izrečen zamjenicom *mi* postavlja kao netko tko štiti ili je na strani referenata na koje se upućuje deiksom *oni*.

2 Politički diskurs i parlamentarni govor

Govori saborskih zastupnika u okviru sjednica prema *Poslovniku Hrvatskoga sabora* (dalje: *Poslovník*) ostvaruju se kao osnovno izlaganje na zadanu temu ili kao replika na čije izlaganje.⁴ Trajanje govora propisano je *Poslovníkom*. Iskazi saborskih zastupnika kao govorna forma mogu se smjestiti između monološkoga izlagačkog govora, koji političar (predsjednik, premijer, stranački predstavnik itd.) kao pojedinac izgovara na javnome mjestu (predizbornom skupu, službenom događaju) pred većim auditorijem, i političkoga dijaloga kao interaktivnoga iskaza u kojemu govornik/govornici i sugovornik/sugovornici izmjenjuju svoja mišljenja (npr. dijalog predstavnika Vlade i sindikata, dijalog visokih dužnosnika na međunarodnoj razini).

Načinom izražavanja, sadržajem te s obzirom na mjesto u kojemu se ostvaruju (Hrvatski sabor kao središnje zastupničko tijelo) parlamentarni govori pripadaju političkom diskursu kao jednom od diskursa javne komunikacije koja je dostupna

³ O odnosu kategorije lica i obilježavanju prostora Pranjković (2013: 11–12) tumači: »Kategorija lica igra vrlo važnu ulogu i u obilježavanju prostora... Vrlo se često naime prostor konkretizira preko lica, osobito preko pokaznih zamjenica, najizravnije pokaznim priložima tipa *ovdje, tu, ondje*. U tom se smislu katkada kaže da je lice »prostor koji govori«.

⁴ Prema *Poslovníku* osnovno je izlaganje »izlaganje govornika o temi o kojoj se raspravlja prema utvrđenom dnevnom redu sjednice Sabora«, a replika »je odgovor zastupnika na osnovno izlaganje«. Saborska sjednica na kojoj se provodi čitanje zakona započinje uvodnim izlaganjem predlagatelja zakona te završnim izlaganjem (govorom) predstavnika Vlade ili predlagatelja (usp. *Poslovník*).

većem broju primatelja (slušatelja/gledatelja), pri čemu govornici istupaju kao javne osobe i imaju određenu ulogu u društvu, odnosno političkoj strukturi (usp. Kovačević i Badurina 2001). Politički diskurs namijenjen je definiranju i artikuliranju općih društvenih interesa te uz administrativni, pravni i medijski diskurs pripada retoričkoj domeni javne komunikacije (Škiljan 2000: 69–75). S pragmatolingvističkoga gledišta upravo je javna uloga političara kao govornika bitna za uključivanje deiksa u iskaz, ako se u obzir uzme govornikova namjera da iskazom djeluje na javnost te zastupa interese skupine u čije ime govori, ali i da se istodobno razlikuje od drugih skupina kojima ideološki ne pripada. Stoga se politički diskurs definira kao diskurs koji ima obilježja profesionalnoga sociolekta u kojemu se, prema ustaljenoj formi i sadržaju iskaza, može prepoznati poruka koju govornik prenosi te njegova pripadnost skupini govornika koje ideološki povezuje jedan sociolekt i koji se (više ili manje) razlikuje od drugih političkih sociolekata unutar društvene zajednice (Škiljan 2000: 70). Deiktičnost izražena jezičnim jedinicama *mi/naši* i *oni/njihovi* u okviru tako postavljene perspektive svojstvena je političkomu diskursu ponajprije zbog izvanjezičnoga konteksta u kojem se politički diskurs ostvaruje, a to je demokratski (višestranački) ustrojena zajednica temeljena na polarizaciji političkih i društvenih skupina koja se preslikava i u jezični iskaz.

3 Govorni korpus ParlaMint-HR 2.0

The screenshot displays the ParlaMint-HR 2.0 web interface. The main content area shows the following statistics:

Counts		General info		Lexicon sizes	
Tokens	22,990,088	Corpus description	Document	word	236,707
Words	20,342,230	Language	Croatian	lc	211,048
Sentences	936,171	Encoding	UTF-8	norm	231,914
Paragraphs	521,449	Compiled	05/12/2021 15:55:51	lemma	85,285
Documents	16	Target	Description	lemma_lc	82,434
				pos	18
				feats	1,178
				id	815
				dep	43
				dep_head_lemma	62,471
				dep_head_pos	19
				dep_head_feats	983
				dep_head_id	783

Structures and attributes:

- name 689,064
- text 16
- note 23,687
- p 521,449
- s 936,171
- speech 124,496

Subcorpora statistics:

Subcorpus	Tokens	Words	%
test	2,795,200	~2,446,721	12.02

The interface also includes a navigation menu on the left with options like Home, Search, Word list, Corpus info, My jobs, and User guide. The bottom right corner features the CLARIN.SI logo and version information: 2.36.7 open-2.167.8 open-2.116.13.

Slika 1: ParlaMint-HR 2.0
Izvor: ParlaMint-HR 2.0 (Croatian parliament)

Građa s primjerima za ovo istraživanje preuzeta je iz korpusa ParlaMint-HR 2.0 koji sadržava transkripte rasprava na sjednicama Hrvatskoga sabora. Korpus je javno dostupan i pretraživ u alatu NoSketchEngine (Slika 1).

Osim korpusa ParlaMint-HR transkripti rasprava na sjednicama Hrvatskoga sabora dostupni su i na mrežnim stranicama Hrvatskoga sabora u okviru e-Doc baze zakonodavnih podataka (<https://edoc.sabor.hr/>). e-Doc baza opsežnija je od ParlaMint-a jer sadržava transkripte sjednica od V. do X. saziva (od 2003. do 2023.), a ParlaMint sadržava transkripte samo IX. saziva od studenoga 2016. do svibnja 2020. (Erjavec et al. 2023: 423). S obzirom na ulogu korpusa u suvremenim istraživanjima kao jezičnoga resursa s razvijenim parametrima prilagođenima za ciljanu analizu te s mogućnošću javnoga pristupa preko NoSketchEngine platforme, ParlaMint-HR odabran je kao izvor za pretraživanje primjera. Važno je istaknuti da je u kontekstu istraživanja govorenoga jezika ParlaMint-HR jedan od malobrojnih korpusa koji omogućuje uvid u govornu realizaciju suvremenoga hrvatskog jezika u javnoj domeni.⁵ ParlaMint-HR hrvatski je korpus unutar veće skupine korpusa ParlaMint koji sadržavaju transkripte rasprava na sjednicama 16 europskih parlamenata (Belgija, Bugarska, Češka, Danska, Latvija, Litva, Velika Britanija, Island, Italija, Hrvatska, Mađarska, Nizozemska, Poljska, Slovenija, Španjolska, Turska).⁶ Svi korpusi uz transkripte sadržavaju i metapodatke o govornicima, odnosno parlamentarnim zastupnicima (usp. Erjavec et al. 2023), a to su govornikova dob, spol, stranačka pripadnost, podatci o sjednici, status govornikove stranke (oporba ili vladajuća koalicija). Korpus bilježi i druge podatke koji su važni za govorni korpus – didaskalije i komentare prepisivača, praznine u transkripciji, naznaku prekida u govoru, pljesak itd.:

*Da ste dobro čitali onda biste vidjeli. [Upadica sa strane, ne razumije se.]; Kolega Maras, hvala vam lijepa. [Upadica:] Govornik se ne čuje.
... ali prije želim pozdraviti učenice i učenike gimnazije Eugena Kumičića iz Opatije sa profesoricama i profesorima koji su ovdje danas s nama. Molim vas pozdravimo ih pljeskom. /Pljesak!/.⁷*

⁵ Za hrvatski jezik izrađen je i *Hrvatski korpus govornog jezika odraslih* – HrAL dostupan u TalkBank-u. Usp. Kuvač Kraljević i Hržica 2016.

⁶ Uz inačicu 2.0 dostupna je i inačica 2.1, koja sadržava i francuski korpus i u kojoj su ispravljene neke pogreške iz prethodne inačice te su dodani novi metapodatci za dvodomne parlamente (usp. Erjavec et al. 2021).

⁷ Svi primjeri iz korpusa ParlaMint-HR 2.0 preuzeti su u izvornome obliku.

Rečenice koje izgovaraju saborski zastupnici strukturom upućuju na to da je riječ o spontanom govorenom iskazu. Mnoge potvrđene rečenice često ne slijede očekivani sintaktički ustroj svojstven pisanom jeziku, vidljiva su odstupanja od gramatičkih pravila, slijed rečenice od početka do kraja često je nepovezan, isprekidan govornikovim tijekom misli:

Izbori dakle, ovo je ovo izborna godina, nisam ovdje vidio niti jedan prijedlog da bi odgovila apsolutno izbori za godinu dana recimo iz tih sigurnosnih razloga, nego samo da li nekom odgovara da li su oni u 7. mjesecu ili u 9.-tom.

Mislim da se tu u jednom trenutku postavilo jedno pitanje to je ruševina pa bih ja vezano uz ruševine nešto rekla.

U Hrvatskoj mi zaista smo svjedoci pogotovo u krajevima koji su bili devastirani u ratu, ostatak ruševina. Evo, ja mogu samo pretpostaviti, da kolega Sokol, on će vjerojatno biti jednog dana hrvatski premjer.

Kako je fokus ovoga istraživanja usmjeren na upotrebu deiksa u parlamentarnim govorima u okviru pragmalingvistike te se ne provodi cjelovita sadržajna analiza parlamentarnog govora u kontekstu kritičke analize diskursa⁸, primjeri se u radu navode samostalno, bez podataka o govorniku, odnosno saborskom zastupniku koji ih izgovara. Ti su podatci javno dostupni u samom korpusu ParlaMint-HR 2.0 i može im se pristupiti preko NoSketchEngine platforme.

4 Osobne deikse u parlamentarnim govorima: analiza primjera

Iako se, u skladu s tumačenjem da je govornik deiktičko središte, u literaturi ističe da osobne deikse služe isticanju govornikove uloge u iskazu, ovdje polazimo od pragmalingvističkoga tumačenja da se osobnim deiksama identificiraju (kodiraju) sudionici iskaza kao komunikacijskoga događaja (Karlić i Bago 2020: 751), a naglasak je na odnosu između govornika i negovornika na koje govornik (saborski zastupnik) u svojem govoru referira. Kao osobne deikse obuhvaćene su:

- osobne zamjenice *mi* i *oni*
- posvojne zamjenice *naš/ naši* i *njegov/ njihovi*
- pojedini glagolski oblici koji izražavaju kategoriju lica (npr. *smatramo, napominjemo*)

⁸ Što bi uključivalo analizu izvanjezične, ponajprije, ideološke podloge govora pa bi s toga gledišta bio bitan i identitet govornika kao javnih osoba s političkom, tj. stranačkom pripadnošću.

- pokazna zamjenica *oni* kao množinski oblik zamjenice *onaj*. Pretraga korpusa potvrdila učestalost upotrebe te zamjenice u govoru saborskih zastupnika te se pokazalo da u takvim iskazima pokazna zamjenica *oni* ima istaknutu deiktičku ulogu. Analiza uključuje izraze *oni koji* i *jedan od onih*.

Ovdje se ne bavimo analizom govora saborskih zastupnika na razini dijaloga u kojemu govornik i sugovornik izmjenjuju svoje uloge. Iz analize je, stoga, izostavljena osobna zamjenica *vi*, koja je također učestala u parlamentarnim govorima, ali ima druga obilježja.⁹ U govorima zastupnika zamjenica *vi* upotrebljava se i u jednini i u množini te bi posebno trebalo istražiti deiktička obilježja te zamjenice ovisno o referentu – je li riječ o sugovorniku u jednini kojemu se govornik obraća zamjenicom *vi/Vi* iz poštovanja ili je riječ o skupini kojoj se govornik obraća u 2. licu množine. Također, zamjenica *vi* kad se upotrebljava u jednini, ubraja se u društvene deikse jer se njome određuje društveni status sudionika komunikacijskog događaja, što može biti temom posebnoga istraživanja.¹⁰

Analiza diskursnih obilježja osobnih deiksa u govorima saborskih zastupnika obuhvaća: 1. strukturnu razinu prema kojoj se određuje sintaktički (unutarrečenični i širi, izvanrečenični) kontekst upotrebe osobnih deiksa, 2. značenjsku (komunikacijsku i pragmatičku) razinu prema kojoj se mogu utvrditi referenti na koje se deiksama upućuje. Za razumijevanje osobnih deiksa, osim analize samoga iskaza, katkad je potrebno i šire izvanjezično (pred)znanje kako bi se odredili stvarni referenti u pozadini deiktičkoga označavanja.¹¹

⁹ Lema *mi* u korpusu je potvrđena 128 826 puta. Lema *vi* potvrđena je 116 058 puta. Lema *naš* potvrđena je 46 891 put, lema *njihov* potvrđena je 17 754 puta. Za navedene zamjenice provedena je jednostavna pretraga (Simple query). U hrvatskome su množinski oblici muškoga roda osobne i pokazne zamjenice *oni* homografi te je broj potvrda za te zamjenice dobiven dvjema pretragama. Za osobnu zamjenicu *oni* upotrijebljen je regularni izraz [lemma="oni"] [lemma="koji"] kojim je iz pretrage uz osobnu zamjenicu *oni* zdesna isključena lema odnosno zamjenice *koji* te je potvrđeno 72 919 rezultata, dok je za pokaznu zamjenicu *oni* provedena jednostavna pretraga izraza *oni koji* koja je dala 7160 rezultata. Iako takvom pretragom nisu obuhvaćeni svi rezultati za te dvije zamjenice, dobiveni su rezultati u visokom postotku objektivni jer se pokazna zamjenica *oni* u upotrebi najčešće pojavljuje uz odnosnu zamjenicu *koji* u veznome sredstvu *oni koji* u subjektivnim zavisnosloženim rečenicama (usp. Silić i Pranjković 2007: 331), dok se osobna zamjenica *oni* najčešće ne pojavljuje uz odnosnu zamjenicu *koji*. Jednostavna pretraga izraza *jedan od onih* pokazala je 218 rezultata. (izvor: ParlaMint-HR, pristupljeno 16. 6. 2023.).

¹⁰ Primjerice, govornici u saborskim raspravama često se služe zamjenicom *vi* u množini kao retoričkim sredstvom, u funkciji isticanja referenata kojima se obraća, ali bez namjere da se dijalog ostvari, odnosno izmijene dijaloške uloge, npr. *A znate li vi gospodo što znači milijun kuna koji se ne isplati ljudima koji žive od poljoprivrede?, Kad govorimo o tome na koji način upravo Zakon o otocima i njegove mjere doprinose razvoju otoka i većem broju stanovnika na hrvatskim otocima to je nešto što vi saborski zastupnici trebate znati., I zato vas dragi građani koji sad ovo gledate, pozivam, posebno vas preko 600 000 koji ne izlazite na izbore, izdajte da konačno pometemo one koji proizvode loše zakone.* (izvor: ParlaMint-HR 2.0, pristupljeno 9. 6. 2023.)

¹¹ Npr. u rečenici *Mi za njihovog mandata nažalost nismo uspjeli ništa.* potrebno je znati koja je stranka u prethodnome mandatu bila na vlasti da bi se mogao odrediti referent posvojne zamjenice *njihov* u izrazu *njihovog mandata*.

4.1 Sintaktički obrasci

Osobne deikse zastupljene u iskazima saborskih zastupnika prepoznaju se kao dva ustaljena sintaktička obrasca – korelacija i koreferencija – kojima se izražava odnos između deiksa *mi* i *oni* te odnos između osobnih deiksa i referenata.

4.1.1 Korelacija *mi/naši* – *oni/njihovi*

Sintaktička korelacija između osobnih deiksa *mi* i *oni* česta je u iskazima saborskih zastupnika. Upotreba korelacijskih rečeničnih struktura pridonosi ritmičnosti iskaza, ali ima i retoričku funkciju isticanja pozicije govornika u odnosu na negovornike koji su uključeni u govornikov iskaz, kao u primjerima 1) – 5). *Mi* i *oni* najčešće su u funkciji rečeničnoga subjekta ili objekta, ali imaju i druge rečenične funkcije:

- 1) ... *kada su dolazili do nas* i *kada smo im predstavljali što mi nudimo ili dajemo kroz zakone, oni su rekli, ti zakoni se ne trebaju usklađivati sa zakonodavstvom EU.*
- 2) *Mi o tome govorimo i raspravljamo se ovdje (...) dokazujući da smo mi bolji od onih prije nas, a oni poslije nas će dokazivati da su oni bolji od nas.*
- 3) *Štite sebe i svoje ljude, a mi štitimo građane RH.*
- 4) *Nama je bitan čovjek, nama je bitan radnik a njima je bitan kapital, njima su bitni poduzetnici.*
- 5) *Kako su oni mogli, a mi ne možemo?*

4.1.2 Koreferencija: osobna deiksa – referent

Suodnos između osobne deikse i referenta na koji se deiksom upućuje može se smatrati polaznim obrascem za uključivanje osobnih deiksa u iskaz. Saborski zastupnici osobnim se deiksama u 3. licu množine osobne i pokazne zamjenice *oni* te posvojnomo zamjenicom *njihov* redovito služe za upućivanje na negovornike (osobe ili skupinu) koji su predmetom njihova govora, pri čemu se razlikuje unutar tekstno ili endoforično upućivanje, kad je referent naveden u tekstu, te izvantekstno ili egzoforično upućivanje, kad se upućuje na referente izvan teksta. U govorima saborskih zastupnika u koreferencijalnom odnosu osobne deikse i referenta dominira osobna zamjenica i pokazna zamjenica *oni*, koja prevladava i kao endofora i kao egzofora. U literaturi se primarno egzoforama smatraju osobne zamjenice 1. i 2. lica jer upućuju na govornika i sugovornika, dakle na referente koji postoje izvan teksta, a endofore su osobna zamjenica u 3. licu i pokazne zamjenice kojima se upućuje na negovornike, odnosno referente o kojima se govori i koji su najčešće navedeni u tekstu (Kordić 1996: 58). Ovdje, međutim, polazimo od osobne zamjenice u 3. licu koja može biti i endofora i egzofora te upućivati na referente

unutar teksta i na referente koji nisu navedeni u tekstu: »Lične zamjenice trećeg lica prvenstveno su endofore, premda se koriste i egzoforički kad je potrebno ukazati na predmet ili osobu koja ne sudjeluje u razgovoru.« (Kordić 1996: 58).¹² Jednaka obilježja ima i pokazna zamjenica *oni*.

Endoforično upućivanje u govorima saborskih zastupnika ostvaruje se kao anafora, kad se osobnim deiksama upućuje na referente koji su već izrečeni u tekstu, unutar iste rečenice ili u prethodnim rečenicama (Katičić 1995, Kordić 1996). Primjeri 6) – 10) upućuju na osobnu zamjenicu u 3. licu jednine i množine kao anaforu:

- 6) ... pazite **Mađari**, koji su prošli najgori komunizam, **oni** sada zapošljavaju 300 tisuća radnika (...)
- 7) Evo, ja mogu samo pretpostaviti, da **kolega Sokol**, **on** će vjerojatno biti jednog dana hrvatski premjer.
- 8) ...s druge strane **Europska komisija**, **ona** ima neke druge planove...
- 9) I to je razlika između nas i **HDZ-a**. Nama je bitan čovjek, nama je bitan radnik, a **njima** je bitan kapital, **njima** su bitni poduzetnici.
- 10) **HNB** je postala davno anacionalna, **ona** nema nikakav epitet...

Kad osobna deiksa prethodi referentu u tekstu, ostvaruje se kao katafora jer upućuje na sadržaj koji slijedi, kako pokazuju primjeri 11), 12) i 13). Taj je obrazac potvrđen i u govorima saborskih zastupnika, pri čemu ulogu katafore najčešće ima pokazna zamjenica *onaj* ili *oni* kao sastavnica veznoga sredstva *onaj/oni koji*.

- 11) ... ali razgovaramo i **bez onog koji** je odgovoran za ovo iznješće, a to je **ministar Aladrović**...
- 12) **Ima ih onih koji pišu**, ima onih koji prepisuju. Jučer smo **ostali bez jednoga koji piše**, **ostali smo bez Predraga Matvejevića** koji je znanstvenik i književnik.
- 13) ... ljudi znaju, da su njihove najdublje nade i očekivanja **izjgrana od strane onih koji** bi o njima trebali najviše brinuti, **a to su hrvatske institucije i hrvatska Vlada**.

Egzoforično upućivanje također se ostvaruje pokaznom zamjenicom *oni*, tj. izrazom *oni koji* u funkciji veznoga sredstva u zavisnosložanim (subjektnim ili objektnim) rečenicama. U takvim strukturama govornici upućuju na izvankontekstne referente koji nisu jasno definirani, kao u primjerima 14) i 15), a izraz *oni koji* ima dvostruku ulogu – istodobno je i deiksa jer se njime upućuje na negovornika, ali je i referent jer

¹² Kao što i osobne zamjenice u 1. i 2. licu mogu biti endofore ako se u tekstu navode imena govornika i sugovornika (Kordić 1996: 58): »Premda su zamjenice prvog i drugog lica tipično egzoforične, one mogu biti endoforične u navođenju direktnog govora, u dijalozima pripovijetke (kada su imena s kojima povežujemo *ja* i *ti* iz dijaloga negdje dalje u pripovijetki) i sl.«

konkretan referent nije izrečen, nego se samo opisno ističe neka značajka koja je svojstvena negovorniku na kojega izraz *onaj/oni koji* upućuje:

- 14) *Ali u svojem radu, sam se uvjerio da upravo jedino borbom i upravo oni koji govore da se može i rade će na kraju nešto napraviti.*
- 15) *Dakle ja mislim da je dosta hrvatskom narodu onih koji pričaju, a ne rade i nisu radili nikada ništa.*

4.2 Značenjski obrasci

Upotreba osobnih deiksa na temelju potvrda iz korpusa pokazuje da je deiktičnost govora saborskih zastupnika značenjski slojevita, a može se promatrati s komunikacijsko-pragmatičkoga aspekta. Ako se u obzir uzme komunikacijska narav osobnih deiksa, one imaju ključnu ulogu u prikazu odnosa između govornika i svih drugih referenata na koje govornik upućuje, dok pragmatički aspekt otkriva izvanjezični, tj. društveni okvir u kojemu iskaz nastaje. Kao osobna deiksa ključna je pritom osobna zamjenica *mi* jer zastupnici svoj iskaz većinom iznose u prvom licu množine, čime se stvara dojam da se persuazivnost i intencionalnost njihove argumentacije temelji na stavovima društvene skupine koju predstavljaju ili kojoj se priklanjaju odnosno kritici drugih s kojima ne dijele jednake stavove, a na koje se upućuje osobnom ili pokaznom zamjenicom *oni*. Značenjski potencijal osobnih deiksa u govorima saborskih zastupnika karakteriziraju različiti obrasci kao što su miješanje gramatičkih lica, povezivanje govornika sa skupinom ili odvajanje govornika iz skupine, polariziranost ili bliskost govornika i negovornika, što u konačnici pridonosi stilogenosti i dinamičnosti iskaza.

4.2.1 Enalaga

Govornici u saborskim raspravama često svoj iskaz temelje na enalagi jer tijekom govora miješaju gramatička lica osobnih zamjenica prelazeći s 1. lica jednine na 1. lice množine. Primjena enalage u odnosu na druge deiktičke obrasce najizravnije upućuje na spontani, *ad hoc* govor koji se odvija u javnome prostoru. Govornik nastupa kao pojedinac, ali i kao govornik s javnom ulogom, kao u primjerima 16) – 21). Izmjena gramatičkih lica i prelaženje s 1. lica jednine na 1. lice množine može se tumačiti na dva načina: 1) govornik se ne želi previše izložiti, odnosno poseže za izražavanjem u množini kako bi umanjio osobnu odgovornost i kako bi se zaštitio skupinom kojoj se priklanja služeći se zamjenicom *mi*, ili se služi zamjenicom *mi* kako

bi pojačao uvjerljivost sadržaja iskaza (usp. Arambašić 2021), 2) miješajući 1. lice jednine i množine, govornik upućuje na svoju pripadnost društvenoj skupini, odnosno postavlja se u svoju javnu ulogu te iz domene privatnosti zalazi u sferu javnosti:

- 16) *A dajmo svi skupa **napravimo** nekakav red u državi. Pa to je **naše poslanstvo** ovdje koji **mi dolazimo**. Ovi koji su **me** izabrali, **mene** osobno, **vas i nas sve skupa**. **Mi to nećemo**. Najbitnije je da **kažemo** oni ne valjaju ali **ja sam dobar**. Ovi ne valjaju ali sutra ću opet **ja** biti s njim. To je pitanje politike populizma. **Mi ispadamo** ovdje populisti. Ne, u stvari česti građani koji smo došli da riješimo neke probleme.*
- 17) *Ova dva zakona **smo pripremili** i bit će u javnoj raspravi u vrlo skorom vremenu, **ovdje ćemo istaknuti** nekoliko razlika i detalja između **naših** stavova i **ja ću predstaviti** u biti ovaj **naš** prijedlog gdje **smo mi napravili** nekakve, odnosno predložili nekakve stvari kako bolje urediti ovu materiju kroz zakonske tekstove.*
- 18) *... nije to nastalo samo u razdoblju ni **kada sam ja bio ministar** nego i prije, **mi smo to doradili** recimo u segmentu gospodarenja otpadom.*
- 19) *Ono što je ovaj zakon trebao otići korak dalje, a **ja** nekako uvijek **pokušavam** i **mi** u Glasu **pokušavamo** gledati afirmativno, pokušavamo uvijek vidjeti da ipak je nekakvo poboljšanje.*
- 20) *... **ja osobno mislim** i **mi** u Klubu GLAS-a **mislimo** da je puno bolje da ostavite jednu malo širu lepezu.*
- 21) *... u vrijeme **dok sam ja išla u školu**, pa **mi smo učili** iz udžbenika **koje smo nasljeđivali** od generacijama udžbenika starijih 5, 6, 7, 8, 10 g.*

4.2.2 Odnos pojedinca i skupine

Osobne deikse u govorima saborskih zastupnika izražavaju stupanj uključenosti pojedinca u skupinu ili njegovo odvajanje iz skupine, a pri izražavanju takva odnosa izdvajaju se dva obrasca – inkluzivno *mi* i izraz *jedan/jedna od onih*.

Upotreba osobne zamjenice *mi*, kako je navedeno, ključna je u govornikovu iskazu pri upućivanju na povezanost pojedinca sa skupinom, odnosno isticanja njegove uloge u javnoj, konkretno političkoj domeni u kojoj djeluje kao saborski zastupnik. Prelaskom s osobne zamjenice *ja* na govorenje u 1. licu množine služeći se inkluzivnim *mi*, govornik u političkom diskursu izražava drukčiju namjeru u odnosu, primjerice, na upotrebu zamjenice *mi* u znanstvenome diskursu u kojemu je riječ o autorskom *mi*, često definiranom i kao skromnosno *mi*, jer autor na taj način obezličuje znanstveni tekst stavljajući naglasak na sadržaj teksta, odnosno njegovu objektivnost, a ne na sebe kao autora (Badurina i Hrg 2022: 126–127). Za razliku od znanstvenoga diskursa, u političkom diskursu, pa tako i u parlamentarnom govoru,

uloga govornika nije skrivena. Govornik ističe sebe kao kazivača, ali, služeći se pragmatičkim sredstvima, bira one jezične jedinice koje, s obzirom na sadržaj i stavove koje izriče, bolje pridonose učinku iskaza. Osobna deiksa *mi* najbolje upućuje na uključenost pojedinca u političku ili drugu javnu skupinu te mu omogućuje da se skupinom koristi za izricanje svojih osobnih stavova, kao što potvrđuju primjeri 22), 23) i 24) u kojima se govornik, rabeći deiksu *mi*, identificira sa skupinom, ali, zapravo, iznosi vlastita stajališta:

- 22) ... *jako je bitno mi donosimo strategije na nivou države, ali ih često puta ne provodimo ili te strategije stoje u ladici pa onda ne znamo ni sami što ćemo s njima.*
- 23) *Znač, bavimo se sa procedurama, a ne vidimo ono bitno, ono stvarno što nam se svaki dan dešava pred očima a to je da 10 do 15 % učenika u RH nema novaca da plati školsku užinu.*
- 24) *I ako mi vrlo skoro ne krenemo tim putem, za sve nas neće biti dobro jer neće biti dobro onako kako mi mislimo da bi trebalo biti.*

Upućivačku ulogu u parlamentarnim govorima ima i deiktički izraz *jedan/jedna od onih* koji se upotrebljava u rečeničnom obrascu *ja sam jedan/jedna od onih*,¹³ kad se govornik poistovjećuje s drugima na koje se u izrazu *jedan od onih* upućuje pokaznom zamjenicom *oni*, kako pokazuju primjeri 25) i 26). U takvim iskazima u prvom je planu, zapravo, osobna zamjenica *ja* koju govornik supostavlja pokaznoj zamjenici *oni* kako bi uputio na svoju pripadnost skupini, ali istodobno izdvaja sebe kao pojedinca u javnome prostoru:

- 25) ***Ja sam jedan od onih*** koji je jedno 50-ak puta do sada u javnim istupima javno tražio na neki način od ministarstva da to napravi.
- 26) ... ***ja sam jedna od onih*** koja se zalaže da zakoni idu kroz dva čitanja.

4.2.3 Odnos između osobnih deiksa

Osobne deikse *mi/naši* i *oni/njihovi* u govorima saborskih zastupnika postavljaju se u dva međusobno oprečna odnosa. Prevladava obrazac prema kojemu su *mi* i *oni* polarizirane strane, vidljiv u primjerima 27), 28), 29). Govornik svoj iskaz gradi na prototipnoj slici *nas* i *njih* kao oprečnih strana, pri čemu su *mi* pozitivno obilježeni, a

¹³ Na jednak bi se način mogla analizirati i sintagma s niječnim predikatom *ja nisam jedan/jedna od njih* kojom se govornik isključuje iz skupine, tj. izražava svoju nepripadnost skupini, no u korpusu ParlaMint-HR potvrđen je samo jedan primjer: ... ***nisam jedan od onih*** pokajnika kakve oni favoriziraju koji svoj kriminal osobni brišu na način da optužuju druge itd.

oni su skupina koja djeluje protiv strane kojoj se govornik priklanja u iskazu. Najčešće se taj odnos izražava suprotnim rečenicama s veznicima *a*, *ali*, *nego*:

- 27) ... ***oni mogu*** radit šta god boće, ***a mi ne smijemo***.
- 28) *svu smo tu argumentaciju* ***mi iznijeli*** ovdje tijekom rasprave, ***ali oni su gluhi*** na argumente jer ***njih*** vode drugi interesi.
- 29) ***Mi za njihovog mandata*** nažalost nismo uspjeli ništa.

Prema drugome obrascu, potvrđenom u primjerima 30) i 31), osobnim deiksama *mi/naši* i *oni/njihovi* upućuje se na bliskost između govornika i negovornika, odnosno na koheziju *nas/naših* i *njih/njihovih*. Osobnom deiksom *mi* govornik se postavlja u ime skupine koja štiti interese ili je na strani referenata označenih deiksom *oni*:

- 30) Na taj način ***bismo mogli*** i bolje braniti ***njihove*** želje, ***njihova*** prava, ***njihove*** zahtjeve.
- 31) ... ***oni su naši*** heroji i ***mi prema njima*** imamo moralni dug.

U obama obrascima, i kad je riječ o polarizaciji, potvrđenoj u primjerima 32) i 33), i kad je riječ o koheziji, potvrđenoj u primjeru 34), stilogenost iskaza postiže se ponavljanjem istih deiksa:

- 32) ***Nama je bitan*** čovjek, ***nama je bitan*** radnik a ***njima je bitan*** kapital, ***njima su bitni*** poduzetnici.
- 33) ... ***mi smo*** članica EU, ***mi smo*** članica NATO-a, ***mi imamo*** iza sebe bez obzira na sve probleme koji te dvije integracije danas prate, ***mi iza sebe imamo*** saveznike. ***Oni pokušavaju*** te saveznike naći i spremni su platiti jako veliku cijenu da bi ostvarili neke svoje interese. ***Oni imaju*** veliki problem s Kosovom, ***oni imaju*** veliki problem u BiH...
- 34) Na taj način ***bismo mogli*** i bolje braniti ***njihove*** želje, ***njihova*** prava, ***njihove*** zahtjeve...

4.3 Osobne deikse i referenti

Množinski oblici osobnih, pokaznih i posvojnih zamjenica u govorima saborskih zastupnika obuhvaćaju različite skupine stereotipnih referenata unutar izvanjezičnoga konteksta u kojemu politički diskurs nastaje. Referenti na koje se upućuje osobnim deiksama predstavljaju društvene skupine (branitelji, građani, umirovljenici, mladi, obitelj i sl.), kao u primjerima 35) i 36), državu ili narod, kao u primjerima 37) i 38), ili različite društvene subjekte (stranke, politička i javna tijela i sl.), kao u primjerima 39) i 40):

- 35) *Dakle, pred nama je zakon o osobama nestalim u Domovinskom ratu... oni [nestali branitelji] su naši heroji i mi [građani Hrvatske] prema njima imamo moralni dug. Pronađi ih i dostojanstveno pokopati.*
- 36) *... građani neće osjetiti od toga ništa. Mislim da oni [građani] trebaju, da trebaju osjetiti, da trebamo napraviti što je moguće više i što je moguće bolje u njihovom interesu.*
- 37) *... kad je došlo vrijeme da Hrvatska sanira svoju stratešku industriju i brodogradnju svojim novcem... onda zapravo mi [Hrvatska/Hrvati] moramo slušati Njemačku i njezin diktat, zapravo oni [Njemačka/Nijemci] mogu raditi šta god hoće, a mi [Hrvatska/Hrvati] ne smijemo.*
- 38) *Pa u Rumunjskoj su se do nedavno vozili i kolima po cesti i konjskim zapregama a polako će nas [Hrvatsku] sada i oni [Rumunjska] preći.*
- 39) *I to je razlika između nas i HDZ-a. Nama [oporba] je bitan čovjek, nama je bitan radnik a njima [HDZ kao vladajuća stranka] je bitan kapital, njima su bitni poduzetnici.*
- 40) *Mi [HDZ] za njihovog [SDP] mandata nažalost nismo uspjeli ništa...¹⁴*

5 Zaključak

Osobne deikse u govorima saborskih zastupnika tema su o kojoj se u jezikoslovnim okvirima može provesti plodna analiza, što je pokazalo i ovo istraživanje. Iako je deiktičnost imanentna i pisanom i govorenom tekstu, može se ustvrditi da su osobne deikse prototipno obilježje govorenoga iskaza, i dijaloškog, u kojemu dominiraju osobne zamjenice *ja* i *ti* kojima se referira na govornika i sugovornika koji izmjenjuju svoje uloge, i monološkog ili izlagačkog, kakav je parlamentarni govor, u kojemu važnu ulogu imaju osobne, pokazne i posvojne zamjenice koje upućuju na negovornike koji ne sudjeluju u iskazu, ali su predmetom govornikova iskaza.

Analiza je pokazala da se osobne deikse u parlamentarnim govorima: 1. ostvaruju kroz nekoliko sintaktičkih obrazaca, 2. na značenjskoj razini imaju višestruku komunikacijsku i pragmatičku ulogu s obzirom na društveni (politički) kontekst u kojemu govorni iskaz nastaje te 3. u samome iskazu mogu imati i stilogenu funkciju.

Opsežnija kritička analiza osobnih deiksa, koja bi bila usmjerena na opis društvenih i ideoloških parametara parlamentarnoga govora nije provedena, što svakako otvara prostor za daljnja istraživanja.

¹⁴ Iz širega konteksta referira se na prethodni mandat kad je na vlasti bila Socijaldemokratska partija Hrvatske.

Baze i korpusi:

e-Doc baza. Pristup 12. 6. 2023. na <https://edoc.sabor.hr/>.

ParlaMint-HR 2.0. Pristup 2. 5. 2023., 9. – 22. 6. 2023. na

https://www.clarin.si/noske/all.cgi/first?corpname=parlamint20_hr&reload=1&iquery=&qqueryselector=iqueryrow&lemma=&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=&fsca_text.t2ld=

Literatura

- Lidija ARAMBAŠIĆ, 2021: *Svećnost i nemoć komunikacijskog procesa: Priča o zelenom kvadratu i žutom šesterokutu*. Jastrebarsko: Naklada Slap.
- Lada BADURINA, Tihana HRG, 2022: O deiksama i deiktčnosti u akademskom diskursu. *Sarajevski filološki susreti 6: Zbornik radova 1*. Ur. Munir Drkić, Halid Bulić. Sarajevo: Bosansko filološko društvo. Sarajevo. 122–140.
- Krešimir BAGIĆ, 2012: *Rječnik stilskih figura*. Zagreb: Školska knjiga.
- Goranka BLAGUS BARTOLEC, 2023: Pragmalingvističko nazivlje. *Hrvatsko jezikoslovno nazivlje*. Ur. Milica Mihaljević, Lana Hudeček, Željko Jozić. Zagreb: Institut za hrvatski jezik i jezikoslovlje. 351–367.
- Holger DIESSEL, 2012: Deixis and demonstratives. *An international handbook of natural language meaning 3*, 2407–2431.
- Tomaž ERJAVEC et al., 2021: Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 2.1, *Slovenian language resource repository CLARIN.SI*. Dostup 12. 6. 2023 na <http://hdl.handle.net/11356/1431>.
- Erjavec et al. = Tomaž ERJAVEC, Maciej OGRODNICZUK, Petya OSENOVA, Nikola LJUBEŠIĆ, Kiril SIMOV, Andrej PANČUR, Michal RUDOLF, Matyáš KOPP, Starkađur BARKARSON, Steinþór STEINGRÍMSSON, Çağrı ÇÖLTEKIN, Jesse de DOES, Katrien DEPUYDT, Tommaso AGNOLONI, Giulia VENTURI, María CALZADA PÉREZ, Luciana D. de MACEDO, Costanza NAVARRETTA, Giancarlo LUXARDO, Matthew COOLE, Paul RAYSON, Vaidas MORKEVIČIUS, Tomas KRILAVIČIUS, Roberts DARĠIS, Orsolya RING, Ruben van HEUSDEN, Maarten MARX, Darja FIŠER, 2023: The ParlaMint corpora of parliamentary proceedings. *Lang Resources & Evaluation 57/1*, 415–448. <https://doi.org/10.1007/s10579-021-09574-0>
- Charles FILLMORE, 1975: *Santa Cruz Lectures on Deixis 1971*. Bloomington: Indiana University Linguistics Club.
- Virna KARLIĆ, Petra BAGO, 2020: Pragmatika i leksikografija: deiktici kao izazov suvremene leksikografije. *Rasprave. Časopis Instituta za hrvatski jezik i jezikoslovlje 46/2*, 749–763
- Radoslav KATIČIĆ, 1995: O rečenicama s anaforom i kataforom povezanima u diskurz. *Filologija 24–25*, 195–199.
- Marina KATNIĆ-BAKARŠIĆ, 2018: Stilistika govornika i sugovornika. *Croatica 62*, 333–342.
- Snježana KORDIĆ, 1995: *Relativna rečenica*. Zagreb: Hrvatsko filološko društvo – Matica hrvatska.
- Snježana KORDIĆ, 1996: Zamjenice u izgradnji kohezije teksta. *Radovi Zavoda za slavensku filologiju 30–31*, 55–100.
- Marina KOVAČEVIĆ, Lada BADURINA, 2001: *Raslojavanje jezične stvarnosti*. Rijeka: Izdavački centar Rijeka – Filozofski fakultet Rijeka.
- Jelena KUVAC KRALJEVIĆ, Gordana HRŽICA, 2016: Croatian Adult Spoken Language Corpus (HrAL). *Fluminensia. 28/2*, 87–102.
- Stephen C. LEVINSON, 1983: *Pragmatics*. Cambridge: Cambridge University Press.
- Martina PODBOJ, 2011: Manipulacija u političkom diskursu – kritički pristup. *Hrvatistika 5*, 123–133.
- Poslovnik Hrvatskoga sabora*. Pročišćeni tekst, 2018: *Narodne novine 81/13., 113/16., 69/17., 29/18.*

Ivo PRANJKOVIĆ, 2013: *Gramatička značenja*. Zagreb: Matica hrvatska.

Josip SILIĆ, Ivo PRANJKOVIĆ, 2007: *Gramatika hrvatskoga jezika za gimnazije i visoka učilišta*. Zagreb: Školska knjiga.

Dubravko ŠKILJAN, 2000: *Javni jezik*. Zagreb: Antibarbarus.

Michael TOMASELLO, 2006: Why don't apes point?. *The Roots of Human Sociality: Culture, Cognition, and Interaction*. Ur. Nick J Enfield, Stephen Levinson. Oxford: Berg. 506–524.

FONOLOŠKA ZMOŽNOST BOSANSKO GOVOREČIH PRISELJENK IN PRISELJENCEV

JANA LOVREC SRŠA, GJOKO NIKOLOVSKI

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija

janalovrecsrša@gmail.com, gjoko.nikolovski@um.si

Od vseh jezikovnih ravnin, ki se obravnavajo pri poučevanju nekega jezika kot neprvega, je najmanj zastopana fonetično-fonološka raven. Razlog za to je dejstvo, da je sporazumevanje možno tudi takrat, ko izgovorjava ni povsem pravilna. Uporaba komunikacijskega pristopa pri poučevanju tujih jezikov tudi vpliva na »zanemarjanje« fonetično-fonološke ravni, saj se tako učenci kot tudi učitelji zavedajo, da je sporazumevanje kljub nepravilni/neustrezni izgovorjavi možno. Prispevek obravnava fonološko zmožnost neslovensko govorečih priseljenk in priseljencev iz Bosne in Hercegovine, ki živijo v Mariboru in se za uspešno integracijo v slovensko okolje učijo slovenščino kot neprvi jezik. Za potrebe prispevka je analiziran korpus njihovih govornih besedil, v katerih so obravnavane in definirane glasoslovne težave, s katerimi se soočajo pri učenju slovenščine. Analiza temelji na posnetkih njihovih govornih besedil ter registrira sledeče težave: izgovor polglasnika, izgovor vzglasnega *v*- pred (ne)zvenečim soglasnikom, izgovor izglasnega *-v* v položaju za samoglasnikom ali *r*, izgovor predloga *v*, izgovor morfemskega *-ol-* za nekdanji zvočniški glas v položaju pred soglasnikom, izgovor izglasnega *-l* v položaju za samoglasnikom, izgovor *-l-* v položaju za samoglasnikom in pred soglasnikom, težave z naglasnim mestom, težave s kakovostjo samoglasnikov idr.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.13](https://doi.org/10.18690/um.ff.4.2024.13)

ISBN
978-961-286-882-6

Ključne besede:
glasoslovne težave,
izgovor,
slovenščina kot neprvi
jezik,
jezikovne interference,
jezikovna integracija



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.12](https://doi.org/10.18690/um.ff.4.2024.12)

ISBN
978-961-286-882-6

Keywords:
phonological problems,
pronunciation,
Slovene as a non-first
language,
language interference,
language integration

PHONOLOGICAL COMPETENCE OF BOSNIAN-SPEAKING IMMIGRANTS

JANA LOVREC SRŠA, GJOKO NIKOLOVSKI
University of Maribor, Faculty of Arts, Maribor, Slovenia
janalovrecsrša@gmail.com, gjoko.nikolovski@um.si

Of all the linguistic levels addressed in the teaching of one language as a second and foreign language, the phonetic-phonological level is the least represented. This is because communication is possible even when the pronunciation is not completely correct. The use of a communicative approach in foreign language teaching also affects the "neglect" of the phonetic-phonological level, as both students and teachers are aware that communication is possible despite incorrect/inadequate pronunciation. The paper deals with the phonological ability of non-Slovenian-speaking immigrants from Bosnia and Herzegovina who live in Maribor and learn Slovene as a second and foreign language (SSFL) to successfully integrate into the Slovenian environment. For this paper, a corpus of their spoken texts is analysed, in which the phonetic difficulties they face in learning SSFL are discussed and defined. The analysis is based on recordings of their spoken texts and registers the following problems: pronunciation of the semivowel, pronunciation of the *v*- before the (un)voiced consonant, pronunciation of the final *-v* after the vowel or *r*, pronunciation of the preposition *v*, pronunciation of the morphemic *-o/-* for the formerly voiced consonant in front of the consonant, pronunciation of the final *-l* after the vowel, pronunciation etc.



1 Uvod¹

Sporazumevalna zmožnost pri učenju neprvega jezika zajema znanje in védenje, ki učečim se omogoča, da lahko učinkoviteje uporabljajo jezik in sposobnosti, da svoje znanje uporabljajo za sporazumevanje v vsakdanjih situacijah, kar pomeni, da lahko producirajo in razumejo različna besedila v različnih govornih situacijah za različne vsakodnevne sporazumevalne potrebe (Pirih Svetina 2005: 147). Po *Skupnem evropskem jezikovnem okvirju: učenje, poučevanje, ocenjevanje* (SEJO 2011: 133) sporazumevalna zmožnost zajema: (1) jezikovne zmožnosti, (2) sociolingvistične zmožnosti in (3) pragmatične zmožnosti. Jezikovna zmožnost pa vključuje leksikalno zmožnost, slovnično zmožnost,² semantično zmožnost, fonološko zmožnost, pravopisno zmožnost in pravorečno zmožnost.³ Ravno fonološka zmožnost je po Mildner (1999: 13) med vsemi jezikovnimi ravninami, ki se obravnavajo pri poučevanju določenega jezika kot neprvega jezika, najmanj zastopana. Razlog za to je dejstvo, da je sporazumevanje možno tudi takrat, ko izgovorjeno odstopa od knjižnojezikovne norme, dodaten razlog pa uporaba komunikacijskega pristopa pri poučevanju tujih jezikov, ki zagotovo vpliva na »zanemarjanje« fonetično-fonološke ravni, saj se tako učenci kot tudi učitelji zavedajo, da je sporazumevanje kljub izgovorjavi, ki odstopa od norme, možno (Mildner 1999: 13).

V prispevku analiziramo fonološko zmožnost bosansko govorečih priseljenk in priseljencev, ki po SEJO (2011: 140) zajema poznavanje in spretnosti zaznavanja in tvorjenja, med katere sodijo: (1) fonemi jezika in njihove uresničitve v posameznih kontekstih (alofoni); (2) fonetične značilnosti, po katerih se fonemi med seboj razločujejo (zvenečnost/nezvenečnost, razločevanje glede na vrsto, mesto ovire); (3) fonetične sestave besed (struktura zlogov, zaporedje fonemov, jakostno in tonemsko naglaševanje); (4) stavčna fonetika (prozodija): stavčni poudarek in ritem ter intonacija; (5) fonetična redukcija: samoglasniška redukcija, krepke in šibke oblike, prilikovanje, izpad končnega glasu (elizija) (SEJO 2011: 140).

¹ Prispevek je nastal v okviru Raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine – vodja programa prof. dr. Marko Jesenšek), ki ga je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

² Več o slovnični zmožnosti na primeru rabe samostalnikov moškega spola pri Šumenjak in Volk (2024: 195–218).

³ Več o pravorečnih in pravopisnih vprašanjih pri poučevanju slovenščine kot tujega jezika pri Pirih Svetina (2021: 157–165).

2 Jezikovni prenosi

Na različnih ravneh znanja tujega jezika se fonološka zmožnost učečih se razlikuje in narašča z napredovanjem v znanju jezika, kot je razvidno iz tabele 1 Fonološki nadzor (SEJO 2011: 141).

Tabela 1: Fonološki nadzor

C2	Kot C1
C1	Menja intonacijo in pravilno naglašuje, da bi izrazil drobne pomenske odtenke.
B2	Usvojil je jasno, naravno izgovorjavo in intonacijo .
B1	Izgovorjava je popolnoma razumljiva, čeprav je občasno očiteno tuji naglas in se pojavljajo manjše napake v izgovorjavi .
A2	Izgovorjava je v glavnem dovolj jasna, da je razumljiva kljub opaznemu tujem naglasu , toda sogovorniki bodo morali občasno prostiti , naj izrečno ponovi.
A1	Izgovorjavo zelo omejenega nabora naučenih besed in fraz lahko z nekaj truda razumejo domači govorniki, vajeni sporazumevanja z osebami njegove jezikovne skupine.

Vir: SEJO (2011: 141)

Na nižjih ravneh (A1, A2 in B1) so na glasovni ravni karakteristična odstopanja, ki nastanejo zaradi prenosa iz prvega jezika (Filipović 1986; Požgaj Hadži, Ferbežar 2012: 139–149).⁴ V glavnem gre za »tuji naglas« in »napake v izgovorjavi« (SEJO 2011: 141), ki so preneseni iz prvega jezika. Saville-Troike, Barto (2019: 39) in Balažic Bulc (2004: 77–78) razlikujejo pozitivni jezikovni prenos (uporaba iste strukture iz prvega jezika, ki je ustrezna tudi v nepravem jeziku), kar je tipično pri učenju sorodnih jezikov, in negativni jezikovni prenos (napačna uporaba iste strukture iz prvega jezika, ki ni ustrezna v nepravem jeziku). K temu pogledu, ki sopostavlja prvi in neprvi jezik, je treba dodatno ovrednotiti in sopostaviti pogovorno obliko neprvega jezika, v katerem učeči se živijo in s katerim so v pristnem stiku, in knjižno obliko neprvega jezika, ki se ga učeči se učijo in slišijo le na tečajih. Ker predvidevamo, da nastajajo prenosi tudi iz pogovorne oblike neprvega jezika v knjižno obliko neprvega jezika, k vrstam jezikovnih prenosov dodajamo in obravnavamo vrsto prenosov iz pogovorne oblike neprvega jezika, ki jih v primeru usvajanja »naravn/e/izgovorjav/e/« (gl. opis B2 v tabeli 1), ki pa ima posamezna odstopanja od knjižne oblike, ne uvrščamo med negativne, temveč med pozitivne jezikovne prenose.

² Z jezikovnimi prenosi iz makedonščine v slovenščino na glasoslovni ravni se ukvarja Pavletić (2021: 85–101).

3 Slovenski težavnejši fonemi (polglasnik, *v* in *ʎ*)

Slovenski glasoslovni sistem obsega 29 fonemov (8 samoglasnikov⁵ in 21 soglasnikov), bosanski pa 30 (5 samoglasnikov in 25 soglasnikov). Samoglasniški sistem slovenskega jezika, ki temelji na fonološkem nasprotju med širokim in ozkim *e*-jevskim in *o*-jevskim samoglasnikom, prostim mestom naglasa in fonološkosti trajanja pri jakostno naglašeni samoglasnikih, ima osem samoglasnikov (fonemov), ki se zapisujejo s petimi črkami in izgovarjajo jasno brez reduciranja (Toporišič 2004: 48; Tivadar, Batista 2019: 19).

Tabela 2: Samoglasniki slovenskega knjižnega jezika

	Sprednji	Srednji	Zadnji
Visoki	i		u
Sredinski	e		o
	ɛ	ə	ɔ
Nizki		a	

Vir: Toporišič (2004: 48); Tivadar, Batista (2019: 22)

Samoglasniški sistem bosanskega knjižnega jezika, ki ne razlikuje širokega in ozkega izgovora, in z naglasom, ki je kombinacija treh elementov: intenzitete (naglašen/nenaglašen zlog), kvantitete (dolg/kratek zlog) in kvalitete (padajoč/rastoč ton), ima 5 samoglasnikov (fonemov) (Jahić, Halilović, Palić 2000: 88).

Tabela 3: Samoglasniki bosanskega knjižnega jezika

	Sprednji	Srednji	Zadnji
Visoki	i		u
Sredinski	e		o
Nizki		a	

Vir: Jahić, Halilović, Palić (2000: 88)

Na podlagi razlike, ki je razvidna iz tabel 2 in 3, se pri usvajanju slovenskega jezika lahko pojavijo težave pri izgovoru širokih *e* (*kmet*, *zemlja*, *univerza*) in *o* (*voda*, *gora*, *soba*) ter polglasnika (*pes*, *prvi*, *minister*). Pri analizi samoglasniškega sistema se omejujemo

⁵ Jurgec (2011: 243–268) navaja, da ima slovenščina dodatni 9. samoglasnik, in sicer srednji nizki samoglasnik [ʌ], ki ga opisuje kot kratki naglašeni *a* (*čas*, *brat*, *jant*).

in osredotočamo na izgovor polglasnika, ki tudi v mariborskem pogovornem jeziku odstopa od norme.

Slovenski jezik ima 21, bosanski jezik pa 25 soglasnikov, ki se v obeh jezikih delijo na zvočnike in nezvočnike. Slovenski zvočniški sestav šteje šest glasov: *m, n, v, j, l, r*, bosanski pa osem: *j, l, lj, m, n, nj, r, v*. Največ variant v slovenskem knjižnem jeziku imata zvočnika *v* in *l*, ki imata v mariborskem pogovornem jeziku posamezna odstopanja, kar je razvidno iz tabel 4 in 5.

Tabela 4: Položajni izgovor /l/ v knjižnem jeziku in mariborskem pogovornem jeziku

Položaj	Knjižni jezik (Slovenski pravopis 2001: 73–76; Šeruga Prek, Antončič 2014: 149; Tivadar, Batista 2019: 22)	Mariborski pogovorni jezik (Koletnik 2001: 248, 249)
/l/ pred samoglasnikom	[l]: [lipa]	
/l/ med dvema samoglasnikoma	[l]: kolo [kolo], bolezen [boleʒən]	
/l/ pred soglasnikom in za samoglasnikom	dvoglasniški u [u]: <i>molke</i> [moʊk], <i>pol kile</i> [poʊ kile], pri časovnem prislovu <i>pol pa kot</i> [l], npr. <i>pol enib</i> [pol enih]	
/l/ v izglagolskih izpeljankah - <i>l-</i> , - <i>lk-</i> , - <i>lk-</i> , - <i>lsk-</i> , - <i>lstv-</i>	[u]: <i>bralca, bralka, bralčeva, bralski, bralstvo</i> [braʊca, braʊka, braʊski, braʊstvo]	[l]: <i>bralca, bralka, bralčeva, bralski, bralstvo</i> [bralca, bralka, bralski, bralstvo]
/l/ na koncu besede	[u]: deležnik na - <i>l</i> (<i>hodil</i> [hodiʊ]), večina samostalnikov moškega spola v imenovalniku in tožilniku ednine (<i>stol</i> [stoʊ]), nekateri samostalniki ženskega spola (<i>misel</i> [misəʊ]), večina pridevnikov (<i>bel</i> [beʊ])	iz - <i>il</i> , - <i>el</i> , - <i>al</i> in polglasnika <i>z</i> /l/ je nastal ponaglasni <i>o</i> [z'ri:xtɔ], v naglašnem zlogu se izgovarja kot [u] ['da:ʊ]
morfemski - <i>al-</i>	[ou]: <i>volk</i> [voʊk]	
/l/ v zvezi - <i>ln-</i>	[l]: <i>kopalnica</i> [kopalnica]	zaradi izgube samoglasnika ob <i>n</i> nastane samoglasniški <i>n</i> [ko'pa:lɲca]
/l/ v prevzetih besedah	[l]: <i>intelektualci</i> [intelektualci]	
črkovni sklop <i>lj</i>	ločeno <i>l</i> + <i>j</i> : <i>Ljubljana</i> [Ljubljana], <i>medalja</i> [medalja], <i>prijatelj</i> [prijatelj]	palatalni /l/ se je razvil v srednji /l/ [domiš'la:va]
/l/ v črkovnem sklopu <i>lj</i> pred soglasnikom ali na koncu besede	mehčani / [l']: <i>Poljska</i> [pol'ska], <i>prijatelj</i> [prijatelj], učiteljski [učitel'ski]	samoglasniški /l/ zaradi izgube samoglasnika ob /l/ [ka:šl]

Vir: lasten

Tabela 5: Položajni izgovor *v* v knjižnem jeziku in mariborskem pogovornem jeziku

Položaj	Knjižni jezik (Slovenski pravopis 2001: 73–76; Šeruga Prek, Antončič 2014: 149; Tivadar, Batista 2019: 22)	Mariborski pogovorni jezik (Koletnik 2001: 248, 249)
<i>v</i> pred samoglasnikom	v: <i>vino</i> [vino]	
<i>v</i> med dvema samoglasnikoma	v: Slovenija [Slovenija]	
<i>v</i> pred soglasnikom in za samoglasnikom	soglasniški in dvoglasniški <i>u</i> [u]: <i>košček</i> [kouček]	zobnoustnični [v] [sve'to:vna], ustničnoustnični [u] [g'la:una]
<i>v</i> na koncu besede za samoglasnikom in pred preporom	dvoglasniški <i>u</i> [u]: <i>siv</i> [siu];	
<i>v</i> pred zvenečim nezvočnikom	ustničnoustnični zveneči <i>v</i> [w]: <i>vzajemno</i> [wzajemno]	lahko onemi [z'di:gnji] ali zobnoustnični [v] vbod [v'bot]
pred nezvenečim nezvočnikom	ustničnoustnični nezveneči <i>v</i> [ʍ]: <i>rpis</i> [ʍpis]	[f] [f'ča:six], v vzglasju lahko onemi [f'ča:six]
<i>v</i> pred zvočnikom	ustničnoustnični zveneči <i>v</i> [w]: <i>vlada, vreme, vnetje</i> [vlada, vreme, wnetje] (premislek o dopustnosti variantnega zobnoustničnega izgovora [ʋ]) (Tivadar 1999: 359)	zobnoustnični [v] [v're:me]
<i>v</i> za zvočnikom	ustničnoustnični zveneči <i>v</i> [w]: obrv [obərw]	v sklopu <i>-rv</i> je [f] [f'bərf]
<i>v</i> na začetku besede pred zvenečim nezvočnikom ali zvočnikom in na meji dveh besed, od katerih se prva ne končuje na samoglasnik	ustničnoustnični zveneči <i>v</i> [w]: <i>pred vlado</i> [predwladɔ]	zobnoustnični [v]: [v'nɛ:tje]
na začetku besede pred nezvenečim nezvočnikom in na meji dveh besed, od katerih se prva ne končuje na samoglasnik	ustničnoustnični nezveneči <i>v</i> [ʍ]: <i>ob rpisu</i> [opʍpisu]	[f]: [opf'pi:su]
v besedah s predposkim <i>u</i>	[u], [w] ali [ʍ]: udariti [udariti/wdariti/ʍdariti];	zobnoustnični [v]
predlog <i>v</i> se v knjižnem jeziku vedno izgovarja povezan z naslednjo besedo, vendar ne zobno-ustnično	[u], [w] ali [ʍ]: v banki [wbanki], je v Iraku [jeʋiraku], v hiši [ʍhiši]	zobnoustnični [v] [v 'bɔ:lnici], pred nezvenečimi soglasniki je [f] [f s'lu:žbo]

Vir: lasten

V bosanskem knjižnem jeziku ima največ variant zvočnik *l*, in sicer: (1) zobnodlesnični oz. trdi *l* (v položaju pred zadnjimi samoglasniki, pred soglasniki in na koncu besede): *laž* [laž], *lud* [lud], *loš* [loš]; *Olga* [olga], *Vltava* [vltava]; *alka* [alka];

hotel [hotel], *motel* [motel], *bordel* [bordel] in (2) dlesnični oz. polmehki *l* izgovor v položaju pred sprednjimi samoglasniki: *lice* [lice], *legalan* [legalan]. Mehki izgovor zvočnika *lj* je prisoten v vseh položajih: *ljubav* [l'ubav], *fotelja* [fotel'a] (Jahić, Halilović, Palić 2000: 89–92).

Iz pregleda je razvidno, da imata slovenska zvočnika *v* in *l* širši spekter alofonov, ki bosanskim priseljenkam in priseljencem lahko povzročajo težave pri izgovorjavi, zato se v analitičnem delu omejujemo in osredotočamo na analizo izgovora teh dveh težavnejših fonemov iz zvočniškega sistema slovenskega knjižnega jezika.

4 Metodologija raziskave

Analitični del obravnava fonološko zmožnost neslovensko govorečih priseljenk in priseljencev iz Bosne in Hercegovine, ki živijo v Mariboru in se za uspešno integracijo v slovensko okolje učijo slovenščino kot neprvi jezik. V raziskavo so bile zajete le osebe, ki dosegajo vstopno raven (A1) in so zaključile tečaj Začetna integracija priseljencev (ZIP) v obsegu 180 ur. Osebe imajo različen zaposlitveni status, različno stopnjo izobrazbe (od osnovnošolske do univerzitetne) ter različno trajanje bivanja v Sloveniji. V raziskavo je bilo zajetih več žensk kot moških, saj so le-te pogosteje udeležene tečajev iz razloga brezposelnosti. Analizirana so prosto govornjena besedila, ki so producirana na podlagi sestave ustnega dela izpita iz slovenščine na osnovni ravni, ki ga sestavljajo tri naloge, in sicer: (1) predstavitev (kandidat se predstavi in pove nekaj več o sebi na podlagi tem, ki so mu bile ponujene; od šestih tem izbere tri, ima možnost predhodne priprave); (2) igra vlog (kandidat odigra dialog na izbrano temo; ponujeni sta dve temi, izbere eno na podlagi naslova, ima možnost predhodne priprave); (3) opis slike (kandidatu sta ponujeni dve sliki, izbere eno in jo opiše, odgovori na vprašanja, povezana s temo na sliki; odgovarjanje poteka brez priprave).

Za potrebe prispevka je analiziran korpus njihovih govornjenih besedil, v katerih so obravnavane in definirane glasoslovne težave, s katerimi se soočajo pri učenju slovenščine. Analiza temelji na več kot 600 minutah posnetkov njihovih govornjenih besedil ter registrira izgovor polglasnika, izgovor vzglasnega *v*- pred (ne)zvencim soglasnikom, izgovor izglasnega *-v* v položaju za samoglasnikom ali *r*, izgovor predloga *v*, izgovor morfemskega *-ol-* za nekdanji zvočniški glas v položaju pred

soglasnikom, izgovor izglasnega *-l* v položaju za samoglasnikom, izgovor *-l-* v položaju za samoglasnikom in pred soglasnikom in težave z naglasnim mestom.

Cilji raziskave so: (1) prikaz nekaterih slovenskih težavnejših fonemov (s poudarkom na polglasniku, *v* in *l*), (2) analiza odstopanj v slovenskem jeziku bosansko govorečih priseljenk in priseljencev, živečih v Mariboru, in (3) predstavitev vzrokov za odstopanja od knjižne norme in ugotavljanje pozitivnih prenosov iz mariborskega pogovornega jezika.

4.1 Predstavitev in analiza rezultatov

V tabelah so predstavljeni primeri rabe posameznega glasu v določenem glasovnem okolju, pod katerimi nato sledi interpretacija rezultatov, ki ugotavlja, zakaj je oziroma ni prišlo do odstopanja v izgovoru posamezne besede. Številke ob posameznih primerih nakazujejo število govorcev, ki so posamezno besedo izgovorili. Kjer ob primeru številka ni zapisana, gre za enkratno pojavitev.

Tabela 6: Izgovor *l* pred samoglasnikom na začetku besede ali za soglasnikom

Izgovor po knjižni normi	'lɛ:ti (6), lo'čʉ:jemo (2), 'lɛ:ta (6), 'lɛ:t (23), 'li:stike, 'lɛ:to (5), 'li:ce
Izgovor, ki odstopa od knjižne norme	'lɛ:pa, 'lɛ:po, raz'lɔ:g (2), p'lɛ:skar

Vir: lasten

Na začetku besede se *l* pred samoglasnikom v večini primerov izgovori pravilno kot srednji *l*, medtem ko se za soglasnikom zmeraj izgovarja kot trdi *l* [l], kar je posledica jezikovnega prenosa iz bosanščine, kjer se pred *a*, *o* in *u* izgovarja trdi *l* [l]. Pri primerih napačne izgovarjave pridevnika *lepa* gre za osebo s slabšim znanjem slovenščine, ki je v Sloveniji le 5 mesecev, govori zelo počasi in velikokrat zloguje.

Tabela 7: Izgovor *l* med dvema samoglasnikoma

Izgovor po knjižni normi	ze'lɔ: (4), če'la:do, da'ri:lo po'rɔ:čila, os'ta:lo, x'va:la, po'lɛ:ti, pi'lates, ispo'sɔ:dila, če'bu:la, veli'kɔ:st, prese'li:li, 'ze:lo, izb'ra:la, skɔr'bɛ:lo, 'pa:zila, po'nu:dila, po'čʉ:tila, 'mi:sli, ku'pi:la, ko'lɔ (3), 'dɛ:lo, 'bɛ:le, o'ča:la, čoko'la:dico, ve'li:ko, nadalje'va:la, uk'va:rjala
Izgovor, ki odstopa od knjižne norme	'vɛ:lɪka, bo'la:na, 'ze:lo (2), 'že:lim, 'nu:la, texno'lɔ:gija, e'lɛ:ktro, 'dɛ:lala, 'dɛ:lo (2), že'lɛ:la, 've:lɪko, ko'si:lo, 'kɔ:lo, 'kɔ:la, ža'lɔ:stno, ža'lɔ:stni, sode'la:vci (2), de'lɔ:vni (3)

Imamo primere tako pravilne (srednji *l*) kot napačne izgovarjave. Pri nekaterih govorcih je registriran pravilen izgovor kot posledica pravilne ponovitve na podlagi izpraševalkega vprašanja (*čelado*). Trdi *l* [l] je slišen pri osebah, ki v Sloveniji bivajo krajši čas (do enega leta), ali v primerih, ki so posledica jezikovnega prenosa iz bosanščine, kjer se pred *a*, *o* in *u* izgovarja trdi *l* [l].

Tabela 8: Izgovor morfemskega -ol- za nekdanji zvočniški glas pred soglasnikom

Izgovor po knjižni normi	/
Izgovor, ki odstopa od knjižne norme	'dɔ:lga (3), 'dɔ:lge (4)

Vir: lasten

Primerov za morfemski -ol- je zelo malo, saj ga analizirani kandidati v govoru skoraj popolnoma izločijo (uporabijo ali *jabuka* ali *japka*); v primerih, ki sta podana, gre za napačen izgovor, saj se izgovarja [ɔ:l] namesto [ou]. Gre za negativen prenos iz bosanskega jezika.

Tabela 9: Izgovor izglasnega -l v položaju za samoglasnikom

Izgovor po knjižni normi	'de:u sem p'la:čala, je kon'ča:u, pri'pe:lja:u, z'na:u, 'bi:u (2), 'i:skau, spar'ki:rau, š'ta:rtau, 'ša:u (2), 'me:u
Izgovor, ki odstopa od knjižne norme	bil prob'le:m, 'za:l (6), 'ku:pil, 'bi:o, z'ri:xtu, pok'l:co, 'bi:l, ži'v'e:l, 'po:l 'polovica'

Vir: lasten

Ker je bilo manj oseb moškega spola, teh primerov ni veliko. Slišanih je bilo kar nekaj izgovorov skladnih s knjižno normo, pri čemer jih je nekaj posledica mariborskega narečja, ki so mu še posebej izpostavljeni zaposleni. Vpliv mariborskega narečja je opazen tudi pri izgovoru, ki odstopa od knjižne norme, ko govorci izgovarjajo [o] namesto [u], pri čemer imamo največ teh primerov pri osebah, ki v Sloveniji bivajo dalj časa in so v nenehnem stiku s Slovenci (služba, prijatelji). Čeprav gre v izgovoru za odstopanje od knjižne norme, pa ne gre za negativen jezikovni prenos, saj so osebe usvojile slovenski dialekt.

Tabela 10: Izgovor -l- v položaju za samoglasnikom in pred soglasnikom

Izgovor po knjižni normi	/
Izgovor, ki odstopa od knjižne norme	ga'si:lce (2), ga'si:lci, 'ta:lcev, prebi'va:lcev (12), 'fɔ:lk'lɔ:rno

Vir: lasten

Vsi govorce so namesto srednjega *l* izgovorili trdi *l* [l]. Posebej je treba izpostaviti samostalnik *prebivalci*, ki ga osebe izgovarjajo [l], kljub temu da so na tečajih velikokrat opozorjene na pravilen izgovor te besede. Tukaj gre za negativen prenos iz prvega jezika, v katerem se *l* v položaju za samoglasnikom in pred soglasnikom izgovarja trdo.

Tabela 11: Izgovor v zvezi *ln*

Izgovor po knjižni normi	'sɔ:cjalna, ig'ra:lnicu, s'ta:lno, s'ta:lnega, u'či:lnica
Izgovor, ki odstopa od knjižne norme	s'pa:lnico (2), zado'vɔ:lna, u'či:lnica

Vir: lasten

Primeri izgovora, ki je skladen s knjižno normo, so registrirani pri govorcih z daljšim bivanjem v Mariboru, višjo izobrazbo in zaposlitvijo. Primeri trdega *l* [l] so slišni pri osebah, ki v Sloveniji bivajo manj kot eno leto in so edini stik s slovenščino imeli na tečaju.

Tabela 12: Izgovor *lj* pred samoglasnikom

Izgovor po knjižni normi	o'kɔ:lje, vzgoji'te:ljica (2), nadal'je:vati, pos'ta:vljam, prip'ra:vljam, preživl'ja:vaje, Ljubl'ja:ne, pri'ja:telji, dovol'je:nja, pre'ži:vljanje (3), op're:mljeno, fkl'ju:čeno, pri'ja:teljev, pri'pe:ljaɹ, u'či:teljica, pri'ja:teljem, upo'ra:bljam, pri'ja:teljicama, Ljubl'ja:no, zas'kærbljena, zas'kærbljenosti, pri'ja:teljico, najl'ju:pšo, 'zlɔ:mljeno, op'ra:vlaju, posp'ra:vlaju, o'kɔ:lju, lju'di:, o'kɔ:lje, o'kɔ:lja, 'pe:ljat, pri'si:ljen, l'ju:di, ses'ta:vljena, živl'je:nje, 'pe:ljala, nadalje'va:la, živl'je:nske, 'ci:lje, 'ɔ:ljem, pri'ja:telja, uči'te:ljico, 'pe:ljem, pel'ja:ti, lju'di:, Ljubl'ja:na, pri'ja:telja, lju'di:, pozd'ra:ɹljeni
Izgovor, ki odstopa od knjižne norme	preživl'ja:nje, prip'ra:vljena, 'pɔ:jstle

Vir: lasten

Lj je pred samoglasnikom skoraj zmeraj izgovorjen kot v slovenskem knjižnem jeziku. Opaznih je nekaj izjem, pri čemer gre za posledico izgovora s premorom pri poudarku (*preživljanje*), zlogovanja (*pripravljena*) in mariborskega narečja (*pojstle*).

Tabela 13: Izgovor /j/ pred soglasnikom ali na koncu

Izgovor po knjižni normi	zado'vɔ:l'na
Izgovor, ki odstopa od knjižne norme	'bɔ:ljše je (3), 'bɔ:ljši je, naj'bɔ:lše, naj'bɔ:lši, 'bɔ:lʃ, naj'bɔ:l, zado'vɔ:l'na, 'bɔ:l 'bolj', naj'bɔ:l, 'da:lše

Vir: lasten

Registriran je le en izgovor mehčanega / [ʃ], in sicer pri osebi z visoko motiviranostjo zaradi potrebe po znanju jezika na višji ravni. V vseh drugih primerih se izgovarja kot srednji ali trdi / [l].

Tabela 14: Izgovor v pred samoglasnikom

Izgovor po knjižni normi	'voda, 'vɛ:dno, vi'sɔ:ka, 've:liko, ve'li:k, 'vi:dim (2), 'va:š (2), 'va:je, 've:č, ve'li:ka, 'va:s (6), 'vɛ:mo (2), 'vi:diva, 'va:si, var'ču:je (2)
Izgovor, ki odstopa od knjižne norme	/

Vir: lasten

Vsi govorniki so pravilno izgovorili *v* pred samoglasnikom kot [v], izjem ni bilo, kar ne preseneča, saj se *v* v bosanščini izgovarja kot [v].

Tabela 15: Izgovor vzglasnega v- pred (ne)zvenečim soglasnikom

Izgovor po knjižni normi	u'se: (2x), u'šɛ:č (2), us'ta:nem, u'sa:ki, u'sa:k, u'ča:six (3)
Izgovor, ki odstopa od knjižne norme	f'šɛ:č (4), f'se: mi je f'šɛ:č (2), f'sa:k (5), f'se: (2), v'šɛ:č, fkl'ju:čeno, vs'ta:nemo, f'si: (2), f'se:m, f'sa:ko, f'se: (3)

Vir: lasten

Pri izgovoru vzglasnega *v* ni slišane izgovora, ki bi bil skladen s knjižno normo. Splošno je bilo opaženo, da osebe s krajšim časom bivanja v Sloveniji (tj. manj kot eno leto) izgovarjajo [v], tisti, ki so v Sloveniji dalj časa in imajo veliko stika s Slovenci, izgovarjajo [f], [u] pa izgovarjajo tisti, ki za potrebe zaposlitve potrebujejo znanje slovenskega jezika na višji ravni ali ravni odličnosti. Primerov pravilnega izgovora ni, kljub temu da izpraševalka, s katero se pogovarjajo, uporablja pravilen izgovor, osebe za njo ponovijo [u] ali pa besedo izgovorijo, kot so se jo naučile. Pri izgovoru vzglasnega *v* kot [v] gre za posledico negativnega prenosa, saj se v bosanščini v tem glasovnem okolju *v* izgovarja kot [v], pri izgovoru kot [u] gre najverjetneje za napačno naučen izgovor oziroma za posledico napačnega izgovora s strani govornika slovenščine kot maternega jezika, saj govorniki povejo, da oni zaznavajo [u]; posledično smo te primere uvrstili med primere, ki so skladni s knjižno

normo, saj gre tukaj za minimalna odstopanja. Pri izgovoru *v* kot [f] pa gre za pozitiven prenos, saj so ti govorniki osvojili izgovorno varianto mariborskega pogovornega jezika.

Tabela 16: Izgovor *v* pred soglasnikom in za samoglasnikom

Izgovor po knjižni normi	'a:uto (5), na'ra:unost, d'ne:uno, 'ka:uč, 'A:ustriji, enos'ta:uno, s'ta:ubi, pozd'ra:uljeni (2), za'ba:uno
Izgovor, ki odstopa od knjižne norme	prip'ra:vljam, posp'ra:vljam, 'a:fta, preživl'ja:vanje, Bet'na:vski, 'u:pravnix, 'o:snovno, pre'ži:vljanje, zdrav'ni:ca, av'gu:stu, pozit'i:vno, 'ja:vno, 'da:vno, preživl'ja:nje, zd'ra:vnik, preživl'ja:nje, na'ra:vni, zdrav'ni:k, sode'la:vci, ses'ta:vljena, zd'ra:vstvenix, živl'je:nje, zd'ra:vja, živl'je:nske, s'ta:vbe

Vir: lasten

Pravilen izgovor nekaterih besed se pojavi pri osebah z višjo stopnjo izobrazbe, ki bodo za opravljanje svojega dela potrebovale izpit na višji ravni ali ravni odličnosti. Pri besedah *avto* in *pozdravljeni* gre za pravilno ponovitev za izpraševalko. Večina govorcev izgovarja [v], kar je posledica negativnega jezikovnega prenosa, saj bosanščina v tem glasovnem okolju *v* izgovarja kot [v].

Tabela 17: Izgovor predloga *v*

Izgovor po knjižni normi	/
Izgovor, ki odstopa od knjižne norme	v 'ku:xinji, v 'ro:ke, v 're:du (4), v Bet'na:vski, f ka'te:rem, v 'p'o:djetju, v bli'ži:ni, f Slo'v'e:niji (2), v 'u:'ka:'če:, v p'r'o:stem, v av'gu:stu, v 'B'o:sni, v Slo'v'e:niji (2), v Slo'v'e:nijo, f p'r'o:stem (4), v 'š'o:li, v Ljubl'ja:no, v na'ra:vi, f Slo'v'e:niju, v 'Ma:ri'b'o:ru, v ru'm'e:no, v 'i:stem (3), v sme'ti:, v f'se:m, v 'p'o:gon, v're:di, f p'r'o:sti, v Slo'v'e:niji (2), f Sep'te:mbru, v 'K'o:p:ru, v 'če:ntru, v 'i:stem, v 'pa:rk, v b'l'o:ku

Vir: lasten

Primeri izgovora predloga *v*, ki bi bil skladen s knjižno normo, ni; ali se izgovarja kot [v] ali pa kot [f]. [f] izgovarjajo osebe, ki že dalj časa bivajo v Sloveniji in imajo dnevno stik s Slovenci (prijatelji, sosedi, delovno mesto). Pri izgovoru predloga kot [v] govorimo o negativnem jezikovnem prenosu, medtem ko pri izgovoru kot [f] govorimo o pozitivnem jezikovnem prenosu, saj so govorniki usvojili izgovor besede v mariborščini. Vsi primeri izgovora predloga kot [f] so pred nezvenečimi nezvočniki, kot je to značilno za mariborski pogovorni jezik.

Tabela 18: Izgovor izglasnega -r v položaju za samoglasnikom ali r

Izgovor po knjižni normi	zdru'ži:teu, ele'me:ntou, 'e:rou, 'mē:trou, plače'va:u, 'strō:škou, 'fi:lmo:u, s'ta:rše:u, 'fa:ntou, 'tu:ristou
Izgovor, ki odstopa od knjižne norme	kolek'ti:v, mese'ce:v, zdru'ži:tev, 'mē:trov, 'e:uof (2), 'pa:rkov, pri'ja:teljev (7), Slo'vē:ncev (3), 'vi:kendov, 'ta:lcev, do'mu: (9 let v slo, m), prebi'va:lcev, zdru'ži:tev (20), nas'lō:v (3), odlo'či:tev, d'vō:jčkov (Sumeja), nas'lō:v, 'ba:rf

Vir: lasten

S knjižno normo skladen izgovor izglasnega -r je slišen samo pri tistih govornicah, ki so pravilno izgovarjali tudi r pred soglasnikom in za samoglasnikom, zato lahko sklepamo tudi na iste vzroke. Najpogosteje je registriran izgovor glasu [v], kar je posledica negativnega jezikovnega prenosa iz bosanščine. Pri govornicah, ki v Sloveniji bivajo več kot 5 let in so dnevno v stiku s Slovenci, pa slišimo tudi [f], kar je posledica mariborskega pogovornega jezika. Ker gre pri slednjem za naučeno varianto slovenskega dialekta, tako govorimo o pozitivnem jezikovnem prenosu.

Tabela 19: Izgovor ustnično-ustničnega zvonečega r [w]

Izgovor po knjižni normi	/
Izgovor, ki odstopa od knjižne norme	v're:me (6), vzgoji'te:ljica (3), v're:dno (2), vz'gō:ja, v're:čko (2), v'lō:m (2)

Vir: lasten

Ustnično-ustnični r [w] se pred zvonečim nezvočnikom in pred zvočnikom izgovarja v bosanščini kot [v]. Tukaj težko govorimo o negativnem ali pozitivnem jezikovnem prenosu, saj je isti izgovor značilen tudi za mariborski pogovorni jezik.

Tabela 20: Izgovor polglasnika

Izgovor po knjižni normi	ər'dē:či, 'vərže:jo, 'dər:vu, ər'dē:ču, ər'ja:ve (2)
Izgovor, ki odstopa od knjižne norme	v 'vərtec (4), 'se:m (vsi), 'fa:ntek, 'ni:sem (12), 'o:sem, ver'je:tno (4), 've:n (4), pu'lō:ver, pre'ci:zen, 'de:š, 'dva 'mē:sca, 'dō:ber, 'za:jterk (4), 'pe:s, se'de:m

Vir: lasten

Polglasnika, ki ga v slovenščini zapisujemo s črko *e*, analizirani govorniki ne izgovarjajo, ampak je na tem mestu izgovor [e]. Ker njihov izgovor sovpada z izgovorom v mariborskem pogovornem jeziku, težko govorimo o negativnem jezikovnem prenosu. Polglasnik pred r na začetku besede ali med soglasnikom in r so izgovorili vsi, pri čemer pa je treba poudariti, da so uporabljene besede osnovnega besedišča, ki se na tečaju večkrat izpostavijo.

Naglasno mesto: Iz primerov je opazno, da imajo bosansko govoreče priseljenke in priseljenci veliko težav z naglasnim mestom. Na tem mestu ne izpostavljamo le besed, ki so v slovenskem in bosanskem jeziku enake in se razlikujejo le v mestu naglasa, ampak govorimo tudi o slovenskih besedah, s katerimi se priseljenke in priseljenci srečajo šele s prihodom v Slovenijo in v procesu učenja slovenskega jezika. Pri slednjih je zanimivo, da se naglasno mesto iste besede od govorca do govorca razlikuje, pri čemer je treba izpostaviti, da gre za primerjavo govorcev s podobnimi karakteristikami (trajanjem bivanja, zaposlitvenim statusom, načinom učenja slovenščine, stopnjo izobrazbe).

5 Sklep

Rezultati kažejo, da je za govorke in govorce bosanščine pri učenju slovenščine kot tujega jezika problematičen izgovor glasov *v* in *l*. Kot neproblematična izgovorna položaja sta se izkazala le izgovor *lj* pred samoglasnikom in *v* pred samoglasnikom. Pri izgovoru *l* pred samoglasnikom na začetku besede ali za soglasnikom ter med dvema samoglasnikoma beležimo tako primere, ki sovpadajo s knjižno normo kot tudi primere, ki od nje odstopajo, pri čemer je sovpadanje pogosteje registrirano pri osebah z večjo motiviranostjo in daljšim časom bivanja v Sloveniji, odstopanje pa pri osebah, ki v Sloveniji bivajo manj kot eno leto in nimajo veliko stikov s Slovenci. Pri slednjih je slišen trdi *l* [l] kot posledica negativnega jezikovnega prenosa iz bosanščine, kjer se pred *a*, *o* in *u* izgovarja trdi *l* [l]. Primerov morfemskega *-ol-* ni veliko, saj ga analizirani kandidati v govoru skoraj popolnoma izločijo, v primerih, kjer je izgovorjen, pa se izgovarja trdo. Prav tako nismo zabeležili izgovora *l* v položaju za samoglasnikom in pred soglasnikom, ki bi bil skladen s knjižno normo, saj so vsi govorcev izgovorili trdi *l* [l]. Primerov izgovora izglasnega *l* v položaju za samoglasnikom ni bilo veliko, saj je večina analiziranih oseb ženskega spola. Iz primerov, ki smo jih zabeležili, je razvidno, da so osebe z daljšim časom bivanja v Sloveniji (več kot pet let) in večjo izpostavljenostjo slovensko govorečemu okolju že usvojile izgovor posameznih besed v mariborskem pogovornem jeziku in namesto [u] izgovarjajo [o], pri čemer lahko govorimo o pozitivnem jezikovnem prenosu. V zvezi *-ln-* so primeri pravilnega izgovora registrirani pri zaposlenih osebah, osebah z višjo stopnjo izobrazbe in daljšim bivanjem v Mariboru, medtem ko je registriran trdi *l* [l] pri osebah, ki v Sloveniji bivajo manj kot eno leto in so edini stik s slovenščino imele na tečaju. Izgovor *lj* pred samoglasnikom ali na koncu besede v skoraj vseh primerih odstopa od knjižne norme, in sicer kot srednji ali trdi *l*. Kot

problematični so se izkazali tudi izgovor vzglasnega *v*- pred (ne)zvenečim soglasnikom, izgovor *v* pred soglasnikom in za samoglasnikom, izgovor predloga *v*, izgovor izglasnega *-v* v položaju za samoglasnikom ali *r* ter izgovor ustnično-ustničnega zvenečega *v* [w], saj so analizirani govorniki pod vplivom prvega jezika večinoma izgovarjali [v]. Primeri pravilnega izgovora so bili registrirani pri osebah z daljšim časom bivanja v Sloveniji, z večjo motiviranostjo ali pa kot posledica ponovitve za izpraševalko, ki se je z njimi pogovarjala. Pri izgovoru predloga *v* kot [f] pred nezvonečimi nezvočniki pa je opazen tudi vpliv mariborskega pogovornega jezika, kar obravnavamo kot pozitiven jezikovni prenos. Izgovor polglasnika se iz stališča mariborskega pogovornega jezika ni izkazal za problematičnega, saj tako mariborščina kot bosanščina polglasnik, ki ga v slovenščini zapisujemo z *e*, izgovarjata kot [e], medtem ko so polglasnik na začetku besede ali med soglasnikom in *r* izgovorili. Pri kontrastivni analizi slovenščine in bosanščine na fonološki ravni ostaja še veliko odprtih vprašanj kot na primer naglasno mesto, kakovost samoglasnikov ipd. Ker je ta raven v priročnikih, učbenikih in drugih gradivih slabo zastopana, bi nadaljnje raziskave lahko vodile do objave didaktičnega pripomočka, ki bi vseboval vaje za odpravljanje izgovornih napak specifične skupine govorcev.

Literatura

- Tatjana BALAŽIČ BULC, 2004: Jezikovni prenos pri učenju sorodnih jezikov (na primeru slovenščine in srbohrvaščine). *Jezik in slovnost* 49/3/4, 77–89.
- Rudolf FILIPOVIĆ, 1986: *Teorija jezika u kontaktu*. Zagreb: JAZU.
- Dževad JAHIĆ, Senahid HALILOVIĆ, Ismail PALIĆ, 2000: *Gramatika bosanskega jezika*. Zenica: Dom štampe.
- Peter JURGEC, 2011: Slovenščina ima 9 samoglasnikov. *Slavistična revija* 59/3, 243–268.
- Mihaela KOLETNIK, 2001: Mariborski pogovorni jezik. *Časopis za zgodovino in narodopisje* 72/1/2, 245–254.
- Vesna MILDNER, 1999: Odpravljanje izgovornih napak v maternem in tujem jeziku. *Skripta 3: zbornik za učitelje slovenščine kot drugega/tujega jezika*. Ur. Marja Bešter. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovanske jezike in književnosti Filozofske fakultete. 13–21.
- Nika PAVLETIĆ, 2021: Analiza glasoslovnih napak govorcev makedonščine pri učenju slovenščine. *Jezikoslovni zapiski* 27/2, 85–101.
- Nataša PIRIH SVETINA, 2005: *Slovenščina kot tuji jezik*. Domžale: Izolit.
- Nataša PIRIH SVETINA, 2021: Od kod do kod slovenščina? Pravorečna in pravopisna vprašanja pri poučevanju slovenščine kot tujega jezika. *Jezik in slovnost* 66/2–3, 157–165.
- Vesna POŽGAJ HADŽI, Ina FERBEŽAR, 2012: Tudi to je slovenščina. *Izšrivi kontrastivnega jezikoslovja = Izšrivi kontrastivne lingvističke*. Ur. Vesna Požgaj Hadži. Ljubljana: Znanstvena založba Filozofske fakultete. 139–149.
- Muriel SAVILLE-TROIKE, Karen BARTO, 2019: *Introducing Second Language Acquisition*. Cambridge: Cambridge University Press.

- Skupni evropski jezikovni okvir: učenje, poučevanje, ocenjevanje* (SEJO), 2011. Ljubljana: Ministrstvo RS za šolstvo in šport, Urad za razvoj šolstva.
[Http://www.mizks.gov.si/fileadmin/mizks.gov.si/pageuploads/podrocje/razvoj_olstva/Jeziki/Publikacija_SEJO_komplet.pdf](http://www.mizks.gov.si/fileadmin/mizks.gov.si/pageuploads/podrocje/razvoj_olstva/Jeziki/Publikacija_SEJO_komplet.pdf).
- Slovenski pravopis*, 2001: spletna izdaja. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU.
- Cvetka ŠERUGA-PREK, Emica ANTONČIČ, 2014: *Slovenska zborna izreka: priročnik z vajami za javne govorce*. Maribor: Aristej.
- Klara ŠUMENJAK, Jana VOLK: Analiza rabe samostalnikov moškega spola pri tujih študentih Univerze na Primorskem. *Slavistična prepletanja* 5. Ur. Gjoko Nikolovski, Natalija Ulčnik. Maribor: Univerzitetna založba. 195–218.
- Hotimir TIVADAR, Urban BATISTA, 2019: *Fonetika 1*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Jože TOPORIŠIČ, 2004: *Slovenska slovnica*. Maribor: Obzorja.

VEČKODNOST V ŽANRSKI ANALIZI LITERARNEGA BRANJA

BRANISLAVA VIČAR, KATJA PLEMENTAŠ

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
branislava.vicar@um.si, katja.plemenitas@um.si

Prispevek preučuje součinkovanje različnih semiotskih kodov pri literarnem branju. Temeljna žanrska značilnost literarnega branja je izvedba oz. interpretacija literarnega dela. Opredelimo ga lahko kot vrsto žive uprizoritvene umetnosti, ki predpostavlja neposreden odnos med interpretkami_i, ki so lahko obenem tudi avtorice_ji besedila, in občinstvom. Za študijo primera je bil izbran izsek iz literarnega branja lezbične literature, poimenovanega *Veliko lezbično branje*. Študija primera pokaže na interakcijo različnih semiotskih kodov in razkrije, kako se določena dejanja uresničujejo v vzajemni kontekstualizaciji glasnega branja z neverbalnimi kodi, kot so glasba, rokovanje s predmetom, drža telesa, pogled, obrazni izrazi, kretnje idr. Analiza strukturnih enot pokaže, da literarno branje tvorijo tri zaporedna družbena dejanja, ki ustrezajo elementom žanrske strukture literarnega branja. To so: uvod oz. napoved branja, branje samo (ključna strukturna enota), koda, tj. ovrednotenje. Večkodna interakcijska analiza potrjuje, da je treba koncept načina kot registrske spremenljivke razširiti v smislu, da zaobjame tudi neverbalne kode.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.14](https://doi.org/10.18690/um.ff.4.2024.14)

ISBN
978-961-286-882-6

Ključne besede:
literarno branje,
večkodni diskurz,
žanrska analiza,
večkodna interakcijska
analiza,
kodna gostota,
zamrznjeno dejanje



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.14](https://doi.org/10.18690/um.ff.4.2024.14)

ISBN

978-961-286-882-6

Keywords:

literary reading,
multimodal discourse,
genre analysis,
multimodal interaction
analysis,
modal density,
frozen action

MULTIMODALITY IN THE GENRE ANALYSIS OF LITERARY READING

BRANISLAVA VIČAR, KATJA PLEMENITAŠ

University of Maribor, Faculty of Arts, Maribor, Slovenia
branislava.vicar@um.si, katja.plemenitas@um.si

The article examines the interaction of different semiotic modes in literary reading. The genre of literary reading is characterized by the performance of a literary work. This type of live performance art involves a direct relationship between the interpreters, who can be authors themselves, and the audience. The study of an excerpt from the literary reading of lesbian literature called *The Great Lesbian Reading* shows the interaction of different semiotic modes and the contextualization of reading aloud by nonverbal modes, such as music, handling an object, body posture, gaze, facial expressions, gestures, etc. The literary reading consists of three consecutive social actions corresponding to the elements of its genre structure: the introduction or announcement, the reading itself (key structural unit), and the coda, i.e., evaluation. A multimodal interaction analysis reveals that the concept of mode as a register variable needs to be expanded to include nonverbal modes.



1 Uvod

V prispevku¹ preučujeva součinkovanje različnih semiotskih kodov pri literarnem branju. Osrediniva se na vlogo, ki jo ima kohezija v večkodnih diskurzih, in pokaževa, da se interakcija ne uresničuje zgolj z govorom oz. glasnim branjem, ampak sloni na več semiotskih kodih, ki so koherentno povezani. Pri literarnem branju se poleg verbalnega koda enakovredno uresničujejo tudi številni drugi kodi, kot so obrazni izrazi, pogled, drža telesa, rokovanje s predmeti, pogosto tudi vizualne projekcije in zvok. S študijo primera literarnega branja pokaževa na interakcijo različnih semiotskih kodov in opiševa, kako se določena dejanja uresničujejo v vzajemni kontekstualizaciji branja z neverbalnimi kodi. Povezave med semiotskimi kodi se uresničujejo v strukturi večkodnega besedila in v tem smislu jih imenujemo tudi čezkodne kohezivne vezi.

Večkodno interakcijo opazujeva tudi z vidika žanrske strukture literarnega branja. Za osvetlitev večkodnih praks v literarnem branju sva uporabili kombinacijo večkodne interakcijske analize (Norris 2004, 2011) in registrske analize (Martin in Rose 2003). Za študijo primera sva izbrali izsek iz literarnega branja lezbične literature, poimenovanega *Veliko lezbično branje*, na katerem je nastopilo ok. 30 interpretk. V predstavljenem videoposnetku pesnica, skladateljica in lezbična aktivistka Nina Dragičević bere odlomek iz literarnega dela Djune Barnes *Damski almanah* (2009 [1928]). Z analizo si prizadevava pokazati, »kako lahko različne kombinacije semiotskih kodov součinkujejo, da tvorijo koherentne komunikacijske artefakte« (Bateman 2014). Osrediniva se na vprašanje, kako so v izbrani študiji primera literarnega branja uporabljeni različni kodi za kompleksno večkodno interakcijo med interpretko literarnega dela, interpretko spremljevalne glasbe in občinstvom. Ugotovitve raziskave osvetljujejo način, kako so semiotski kodi medsebojno povezani in kako se medsebojno dopolnjujejo v večkodni celoti.

2 Teoretski in metodološki okvir analize

Večkodna interakcijska analiza se umešča v širši okvir večkodne diskurzivne analize, ki se vse bolj široko uporablja v uporabnem jezikoslovju, sociolingvistiki, antropologiji, psihologiji in drugih uporabnih vedah. Večkodnost v večkodni analizi

¹ Prispevek je nastal v okviru raziskovalnega programa »Slovenski jezik – bazične, kontrastivne in aplikativne raziskave« (P6-0215), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

razumemo kot »besedilne kombinacije različnih semiotskih kodov in njihovo integracijo z vidika strukture, diskurzivne semantike in retorične funkcije« (Stöckl 2019: 50). Znotraj večkodne analize se je razvilo več pristopov, npr. pristop sistemsko-funkcijskega jezikoslovja (npr. O'Halloran 2004; O'Toole 2011), družbene semiotike (npr. Kress 2014; Kress in van Leeuwen 2001), konverzacijske analize (npr. Goodwin 2000; Heath 1986), vendar pa je večkodna interakcijska analiza edini interdisciplinarni pristop, ki je bil razvit posebej za analizo večkodnih interakcij. V zadnjih 15 letih je različne pristope k večkodni diskurzivni analizi integriralo tudi slovensko jezikoslovje (npr. Starc 2009, 2020, 2023; Vičar 2015, 2020), večkodna interakcijska analiza pa je bila prvič uporabljena v prispevku Vičar in Plemenitaš (2023) za analizo večkodne konstrukcije identitete priseljenca.

Avtorica večkodne interakcijske analize jezikoslovka Sigrid Norris je pri oblikovanju tega kvalitativnega pristopa (Norris 2004, 2011) izhajala iz temeljev družbeno-kulturne psihologije, sociolingvistike in uporabnega jezikoslovja, sledila pa je temeljnemu namenu, razviti pristop, ki nam bo pomagal bolje razumeti človeško interakcijo. Za raziskave, ki si prizadevajo opisati človeška dejanja onstran koda govorjenega jezika, se prav vključitev vseh kodov, ki so vključeni v družbeno dejanje, kaže kot oblika celostne in popolne analize. Večkodna interakcijska analiza torej zagotavlja metodološka orodja in teoretske koncepte za analizo družbenih dejanj v njihovi kompleksnosti, tj. omogoča integracijo verbalnih in neverbalnih kodov ter njihovo integracijo z materialnimi objekti in okoljem. Osrednjo enoto večkodne interakcijske analize predstavlja družbeno dejanje.

Norris (2004) družbeno dejanje razdeli na dejanje nižje ravni (angl. lower-level mediated action), ki označuje najmanjšo pragmatično pomensko enoto koda (npr. izrek, kretnja), in dejanje višje ravni (angl. higher-level mediated action), ki označuje povezavo različnih nizov dejanj nižje ravni (npr. sestanek, predavanje). Dejanja nižje in višje ravni v interakciji soustvarjajo druga drugo, pri čemer nobeno ni ne teoretično ne praktično predhodno drugemu. Medtem ko koncept dejanja nižje ravni omogoča, da preučujemo določena dejanja nižje ravni, ki jih en družbeni akter_ka izvaja hkrati, lahko s pomočjo koncepta dejanja višje ravni preučujemo, kako se isto dejanje višje ravni izvaja z različnimi nizi dejanj nižje ravni (Norris 2019: 236–238).

Norris (2004) vpelje tudi koncept zamrznjenega dejanja (angl. frozen mediated action) in ga opredeli kot dejanje, vključeno v objekte ali okolje; tako objekti kot okolje namreč izpričujejo predhodno izvedena dejanja družbenih akterk_jev.

Umetniški grafit na zidu mestne ulice na primer izpričuje dejanje uličnega umetnika (ustvarjanje grafita), obenem pa lahko odslikava njegov notranji svet, družbeno umeščanje ali celo družbeni aktivizem. Koncept zamrznjenega dejanja je osnovan na dejstvu, da lahko z vsakega objekta v okolju razbiramo diskurzivno zgodovino delovanja družbenih akterk_jev. Vendar pa, kot poudari Norris (2015: 44), dejanj, ki izpričujejo identiteto, ne moremo kar razbrati z objektov, ampak nam vpogled v razmerje med objektom in identiteto omogoča šele globlje razumevanje družbenih akterk_jev, s katerimi se ti objekti povezujejo.

Da bi opisali pomembnost in medsebojno povezanost posameznih kodov, ki se uresničujejo pri literarnem branju, sva v raziskavo kot analitično kategorijo vključili tudi koncept kodne gostote (angl. modal density), ki obsega dve razsežnosti: kodno intenziteto in kodno kompleksnost (Norris 2004: 79–80, gl. tudi Pirini 2014: 83–84). Kodna kompleksnost se nanaša na številne kode, ki jih pri določenem dejanju uporabljajo udeleženske_ci interakcije (pogovor v živo ima na primer višjo kodno kompleksnost kot telefonski pogovor, saj se tvori s širokim razponom kodov), kodna intenziteta pa na relativno intenziteto posameznega koda (v telefonskem pogovoru ima na primer visoko intenziteto govorniki jezika). Intenziteta oz. pomembnost posameznih kodov v interakciji je določena s situacijo, družbenimi akterkami_ji ter dejavniki okolja, zato se od interakcije do interakcije spreminja, lahko pa se spremeni tudi med posamezno interakcijo. Kompleksnost in številčnost kodov je odvisna od dejanj nižje ravni udeleženk_cev, ki tvorijo dejanja višje ravni. Vsako dejanje višje ravni se tvori v medsebojnem prepletu številnih semiotskih kodov. Število in intenziteta (pomembnost) kodov sta torej zmeraj odvisna od dejanske situacije. Glede na to, da so družbene akterke_ji navadno vključeni v več dejanj višje ravni sočasno, lahko koncept modalne gostote uporabimo tudi za prikaz različnih nivojev pozornosti/osredinjenosti posameznih dejanj. S pomočjo kontinuuma pozornosti/osredinjenosti (angl. foreground-background continuum) lahko prikažemo, koliko pozornosti posameznik_ca namenja posameznemu dejanju. Višja kot je kodna gostota posameznega dejanja, več pozornosti družbeni akter_ka namenja temu dejanju. (Norris 2004: 95–98)

3 Literarno branje kot žanr

Temeljna žanrska značilnost literarnega branja je izvedba oz. interpretacija literarnega dela. Ker interpretirano literarno delo že samo po sebi nosi oznako določenega žanra, lahko literarno branje uvrstimo med t. i. makrožanre, tj. žanre, ki

vsebujejo druge žanre kot enega izmed elementov svoje strukture (Martin in Rose 2003: 209–210). Kot celoto ga lahko opredelimo kot vrsto žive uprizoritvene umetnosti, ki se navezuje na koncepte gledališča in performansa, vendar ohranja svojo posebnost v izrazu medosebnih odnosov med avtorico_jem, interpretko_om, občinstvom in literarnim besedilom. Literarno branje je tako po svojem bistvu povezano z načinom izvedbe, ki predpostavlja neposreden odnos med interpretkami_i, ki so lahko obenem tudi avtorice_ji besedila, in občinstvom, ki je med izvedbo literarnega branja fizično prisotno. Kot ugotavlja Novak (2011), je pri tradicionalnem preučevanju literarnih besedil to dejstvo dolgo časa ostajalo spregledano. Pri žanrskem opredeljevanju literarnega branja je obenem treba upoštevati, da interpretacija literarnega besedila ni le govorjena različica zapisanega besedila, čeprav na njem temelji, temveč tvori samostojen žanr. Kadar glasno branje izvaja sam avtor_ica literarnega besedila, se za ta žanr uporablja izraz avtorsko branje (Podbevšek 2021: 373). Od bralke_ca literarnega besedila se pričakuje, »da bo ustvaril zvočno razgibano smiselno celoto, se pravi, da bo govorna izrazila (intonacijo, premore, hitrost, jakost, register, barvo idr.) uporabil ustvarjalno in v skladu z besedilom« (Podbevšek 2017: 28). Pri avtorskem branju lahko prepoznamo dvojno avtorstvo: avtorstvo zapisanega besedila in avtorstvo govorne interpretacije zapisa (Podbevšek 2021: 377).

Novak (2011) opozarja na razhajanja pri opredelitvi literarnega branja. Na eni strani je literarno branje opredeljeno kot žanr, ki vsebuje literarno besedilo, ki je bilo napisano za uprizoritev (npr. slam poezija). Literarno besedilo posledično deluje kot samostojna celota, šele ko se uresniči v uprizoritvi, tj. ko pride v stik z občinstvom v živo, podobno kot dramska dela. Na drugi strani je literarno branje opredeljeno kot žanr, ki se vzpostavi šele z uprizoritvijo literarnega besedila, tj. literarno besedilo, ki je podlaga uprizoritve, ni nujno napisano za uprizoritev. V tem smislu lahko denimo branje poezije opredelimo kot izvajanje poezije v javnosti, pri čemer se interpret_ka zaveda žive prisotnosti občinstva. Takšna opredelitev je manj reduktivna in zajema literarna branja besedil, ki primarno niso bila napisana za branje v živo in delujejo povsem samostojno tudi neodvisno od žanra literarnega branja. Pri vseh opredelitvah pa je eden bistvenih elementov, ki ta žanr loči od drugih literarnih žanrov, uresničenje besedila pred občinstvom v živo. Iz tega izhaja, da je literarno branje nemogoče žanrsko opredeliti, ne da bi pri opredelitvi upoštevali elemente večkodnosti. Žanr literarnega branja se namreč uresniči v trenutku uprizoritve. Pri literarnem branju imajo torej poleg verbalnega koda pomembno vlogo tudi drugi kodi, tj. tako kodi, ki jih izražamo s telesom, na primer drža telesa,

obrazni izrazi, intonacija, kakor kodi, ki niso izraženi s telesom, na primer zvok, vizualne projekcije idr.

V nadaljevanju besedila opredeliva žanr literarnega branja na podlagi sistemsko-funkcijskega modela žanra in registra (Martin 1992; Martin in Rose 2003). V sistemsko-funkcijskem jezikoslovju je žanr opredeljen kot stopenjski, k cilju usmerjen družbeni proces (Martin in Rose 2003: 7). Usmerjenost žanra k cilju izhaja iz točno določene kombinacije kontekstualnih spremenljivk, ki se odražajo tudi v samem besedilu. Te spremenljivke tvorijo register, ki obsega polje, ton in način (Martin 1992).

Kombinacija polja, tona in načina je uresničena tudi v sami strukturi besedila, ki se na različne načine izraža s splošno shemo *uvod, osrednji del in zaključek*. Različne kombinacije registrskih spremenljivk, tj. polja, tona in načina, odražajo namen besedila (Martin in Rose 2003). Po Martinu (1992) je polje opredeljeno kot družbena dejavnost, ki obsega dejansko dogajanje in udeležence v tem dogajanju v določeni situaciji, ter pomenska področja, ki se oblikujejo v uresničevanju družbene dejavnosti. Na ravni jezikovne metafunkcije se uresničuje s predstavnim pomenom (npr. udeležence_ci, tipi glagolskih dogodkov, objekti in okolje itd.). Polje je lahko splošno ali specializirano. Ton izraža družbeno distanco, družbeni status in vrsto interakcije. Na ravni jezikovne metafunkcije se uresničuje z medosebnim pomenom (npr. glagolski naklon, modalnost, izrazi vrednotenja, obrazni izrazi, ki izražajo čustva, sproščena ali nesproščena drža itd.). Ton je lahko formalen ali neformalen. Način vključuje vrsto prenosnika in kanala. Na ravni jezikovne metafunkcije se uresničuje z besedilnim pomenom (kohezivnost, tematska struktura itd.). Način sestoji iz dveh razsežnosti vloge jezika v govorni situaciji:

- Tip stika med govorko_cem in prejemnico_kom, ki vpliva na monološko ali dialoško naravo besedila. Stiki so lahko vidni, slušni ali haptični, eno- ali obojestranski. Od načina je odvisen tudi čas, v katerem pričakujemo odziv; ta je lahko takojšen, kot v govorjenem dialogu, lahko pa do odziva preteče precej časa, kot v primeru pisemske korespondence.
- Odvisnost besedila od konteksta situacije, ki določa vlogo jezika v govorni situaciji. Besedilo lahko ima zgolj podporno vlogo v dogajanju, lahko pa govorno situacijo konstituira. V slednjem primeru lahko sledi časovnemu sosledju polja ali je od polja neodvisno.

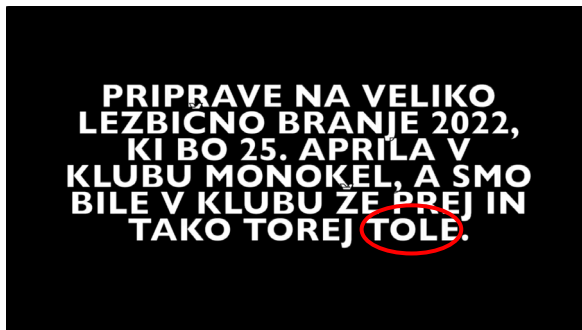
Če pojmujeemo žanr kot družbeni proces, ki se uresničuje na ravni registra v sodelovanju medosebne, predstavnne in besedilne funkcije, lahko literarno branje osvetlimo z vidika polja, načina in tona. Z vidika predstavnne funkcije je treba razlikovati med poljem literarnega branja in poljem literarnega besedila, tj. besedila, ki se bere. Samo literarno besedilo, ki ima svoje polje, ton in način, se v literarnem branju uresniči kot del polja literarnega branja, tj. z literarnim branjem se konstituira novo polje makrožanra. Literarno besedilo je večkodno povezano z drugimi sestavinami polja (npr. z interpretko_om, prostorom, zvokom itd.) in tona literarnega branja (npr. z odnosi med intepretko_om in občinstvom, odnosi med interpretkami_i). Na osnovi kontekstualnih spremenljivk polja in tona se vzpostavlja razlikovanje med literarnim besedilom kot žanrom in literarnim branjem kot makrožanrom. Polje literarnega branja vsebuje interpretke_e literarnega besedila, spremljajoče akterke_je, občinstvo, prostor, zvok, objekte in literarno besedilo. Z vidika interpretk_ov lahko razlikujemo individualna in skupinska branja. Glede na vsebino branega literarnega dela so literarna branja lahko tematsko specializirana (npr. branje LGBTIQ+ literature, branje animalistične literature, branje antropocene literature itd.). Ton literarnega branja se ustvarja s kompleksnimi medosebnimi odnosi, ki jih tvorijo interpretke_i, spremljevalne akterke_ji in občinstvo. V medosebnih odnosih ton prehaja od formalnega do neformalnega, lahko vsebuje čustva, ki se izražajo z različnimi semiotskimi kodi (neverbalnimi, kot so denimo obrazni izrazi, ali verbalnimi, kot je npr. intonacija). Polje in ton se izražata z multiplimi (verbalnimi in neverbalnimi) kodi. Z vidika načina je literarno branje monološko, s primesmi dialoškega (nagovarjanje občinstva, odzivi občinstva); medosebni stiki so slušni in vidni ter obojestranski v točno določenem času in prostoru.

4 Večkodna interakcijska analiza literarnega branja

4.1 Dejanja višje in nižje ravni v literarnem branju

Analiza strukturnih enot v študiji primera pokaže, da literarno branje tvorijo tri zaporedna družbena dejanja višje ravni, ki ustrezajo elementom žanrske strukture literarnega branja. To so: 1) uvod oz. napoved branja, 2) branje samo (ključna strukturna enota), 3) koda, tj. ovrednotenje. Vsako od dejanj višje ravni se uresničuje z nizi dejanj nižje ravni.

Uvod oz. napoved branja sestoji iz nizov dejanj nižje ravni, kot so zapisani jezik, barve, velikost črk, tipografija in premikanje besedila od spodaj navzgor.



Slika 1: Uvod oz. napoved branja

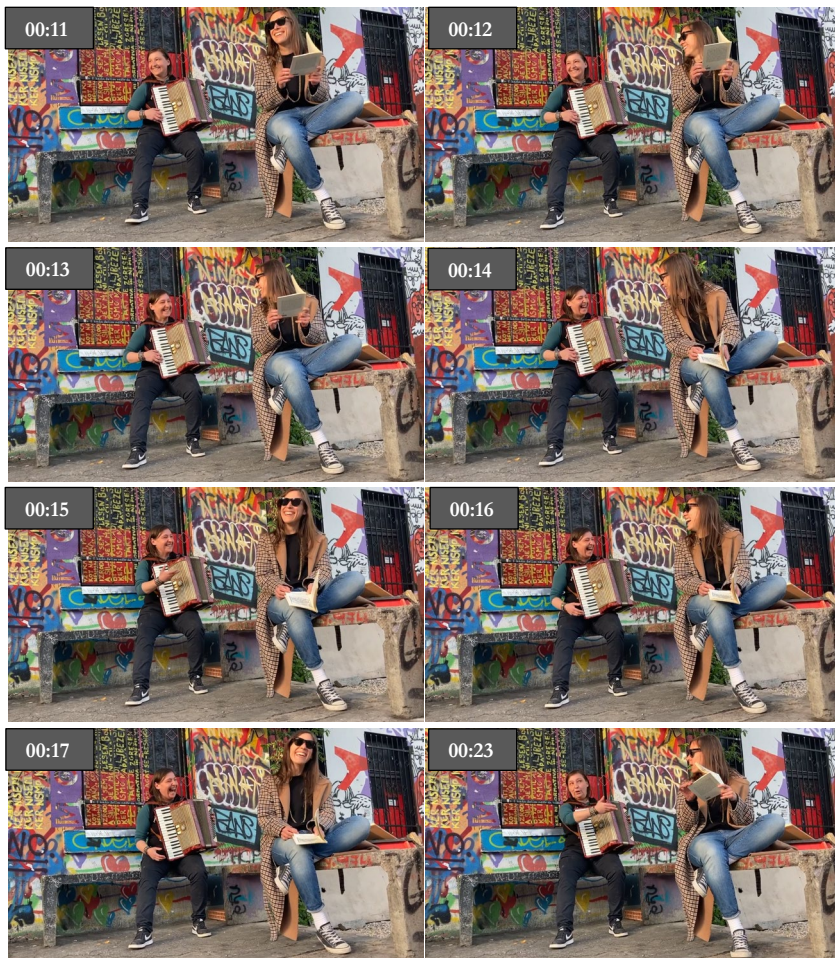
Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

V besedilu te strukturne enote je prisotna katafora *tole* («» smo bile v klubu že prej in tako torej tole«»). Imenujemo jo lahko čezkodna katafora, saj se nanaša na literarno branje kot večkodno dejanje, s tem pa se potrjuje anaforičnost kot bistveno interaktivni pojav (Lee in Stenning 1998).

V družbeno dejanje branja sta vključeni dve družbeni akterki, interpretka literarnega besedila, tj. poglavja Slutnje, znaki in znamenja iz literarnega dela Djune Barnes *Damski almanah* (2009 [1928]), in interpretka spremljevalne glasbe. Čeprav gre za na videz isto dejanje višje ravni, družbeni akterki tvorita različne nize dejanj nižje ravni in različni dejanji višje ravni. Posebnost tega literarnega branja je, da je v dejanje branja vstavljeno mikrodejanje nepravlega začetka, ki ga sproži spontani smeh in je uresničeno z nizi dejanj nižje ravni, kot so smeh, govor, kretnje, drža, obrazni izrazi, premiki glave in telesa, rokovanje s knjigo. Tudi v dejanje nepravlega začetka sta vključeni obe akterki, ki z različnimi nizi dejanj nižje ravni vzpostavljata interakcijo z občinstvom (interpretka spremljevalne glasbe z govorom, kretnjami in obraznimi izrazi, interpretka literarnega besedila s kretnjami in obraznimi izrazi). Dejanja višje ravni, ki tvorijo nepravilni začetek, niso zgolj posamezna, ampak tudi medsebojno odvisna. Akterki delujeta sočasno in se z nizi kretenj, obraznih obrazov, premikov glave in telesa odzivata druga na drugo. Verbalni in neverbalni kodi predpostavljajo solidarnost med njima. Na sliki 2² vidimo, kako interpretka literarnega besedila glavo večkrat obrne proti interpretki glasbe. Interpretka glasbe med tem desno roko večkrat usmeri proti njej, kot je razvidno s podob 5 in 8. Vsakič, ko interpretka

² Načela večkodnega transkripta so predstavljena v Norris 2019.

literarnega besedila glavo obrne proti interpretki glasbe, se srečata s pogledi, sestavni del interakcije med njima je tudi smeh.



Slika 2: Transkript gibanja glave in kretenj rok

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

Večkodna interakcija poleg solidarnosti predpostavlja lezbično identiteto interpretk. Ko slišimo glas iz občinstva »*Počinjemo ili šta?*«, interpretka glasbe odgovori: »*Počinjemo, samo da se nasmejemo*«. Interpretka literarnega besedila se s telesom obrne proti interpretki glasbe, ta pa niz izrekov zaključi z izrekom »*Pa to je lezbično, šta*«. Pri tem desno roko večkrat usmeri proti interpretki besedila in nazaj proti sebi.

Dejanje branja ima jasen začetek, ki je večkodni, tj. interpretka literarnega dela ga uresniči z rokovanjem s knjigo (tj. pomikom knjige pred oči) in govorno uresničitvijo naslova branega verznega poglavja (Slutnje, znaki in znamenja); interpretka glasbene spremljave ga uresniči z začetnim tonom na harmoniki in dvigom desne obrvi, kot je prikazano na sliki 3. Dejanje branja ima tudi jasen konec, ki je prav tako večkodni, tj. interpretka literarnega dela ga uresniči z branjem konca verznega poglavja (*»Ženske ob dekličji strani ženske«*), nasmehom, rokovanjem s knjigo (tj. spustom knjige v naročje), obratom glave in telesa proti interpretki spremljevalne glasbe ter očesnemu stiku z njo, kot je prikazano na sliki 4. Interpretka glasbene spremljave ga uresniči z zaigranim zadnjim tonom na harmoniki in dvigom obrvi. Obe akterki sta vključeni v navidezno isto družbeno dejanje in pri tem sočasno uresničujeta ključni del literarnega branja.



Slika 3: Začetek branja

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

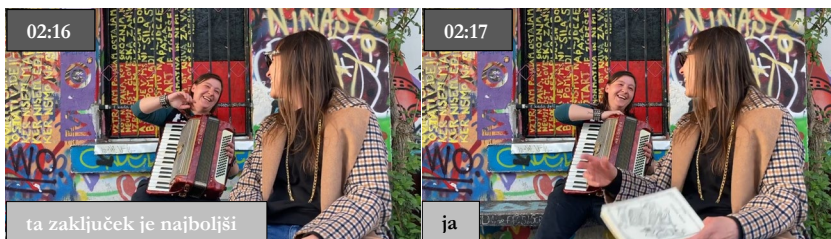


Slika 4: Konec branja

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

Koncu literarnega branja sledi ovrednotenje branja kot zadnje dejanje višje ravni. Kompleksnost večkodne interakcije odraža naslednji večkodni transkript zaključka oz. ovrednotenja branja:





Slika 5: Transkript ovrednotenja branja

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

Kot je razvidno s transkripta (slika 5), akterki ovrednotenje branja uresničita sočasno z nizi dejanj nižje ravni, ki vključujejo premike glave, ki omogočajo očesni stik (podobi 1 in 2), premike telesa, s katerimi se približata druga drugi (podobi 3 in 4), obrazne izraze, kot so nasmehi, s katerimi izražata zadovoljstvo, ter kretnje, kot sta udarec dlani v dlan (podoba 3) in stisk rok (podoba 4). Interpretka glasbe branje ovrednoti tudi verbalno, in sicer z izrekom *ta zaključek je najboljši*. Izrek je prekriven z njeno negovorno dejavnostjo, in sicer desno roko dvigne rahlo nad harmoniko in jo usmeri proti interpretki besedila (podoba 11), nato pa roko ponovno spusti na harmoniko (podoba 12). Interpretka besedila se odzove z izrekom *ja* in pri tem desno roko na hitro usmeri proti interpretki glasbe (podoba 12). V ovrednotenje branja se vključi tudi občinstvo z uresničevanjem svojega dejanja višje ravni, tj. s ploskanjem.

4.2 Kodna gostota

Interpretka literarnega dela in interpretka glasbene spremljave uresničujeta kodno gostoto različno. Razlika v izražanju kodne gostote je razvidna iz naslednjega večkodnega transkripta:

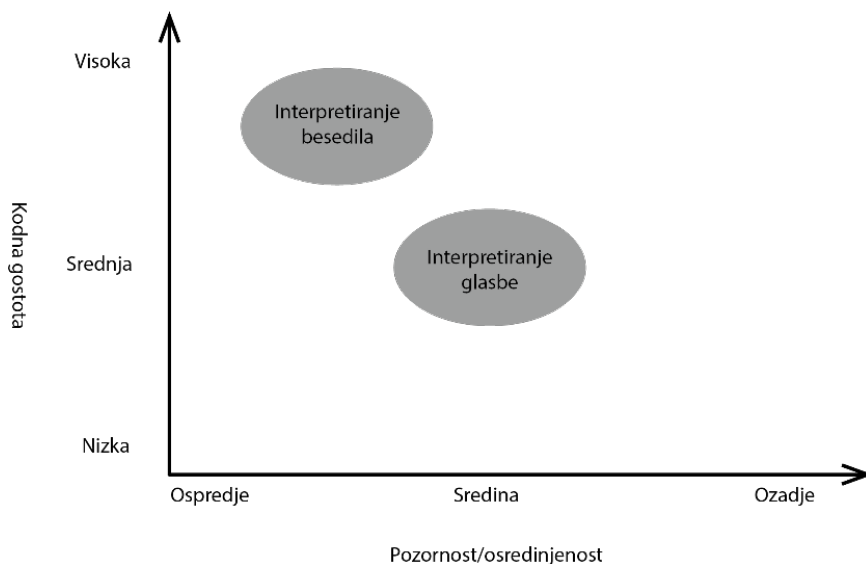




Slika 6: Transkript sočasnega izvajanja dejanj višje ravni

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

Interpretka literarnega dela uresničuje višjo kodno gostoto zlasti v ključnem delu, tj. pri samem branju, kjer uporablja določene nize dejanj nižje ravni (tako verbalne kot neverbalne) na posebej intenziven način: tonski potek, jakost glasu; obrazni izrazi, kot so stisnjene ustnice (podoba 3) ali dvig obrvi (podobe 1–3); vzravnana drža in rokovanje s predmetom, kot je trdno držanje knjige v roki, obenem pa nize dejanj nižje ravni kompleksno preplete. Interpretka glasbene spremljave uresničuje nižjo kodno gostoto kot interpretka literarnega dela, kar se kaže v manjši kodni kompleksnosti. To pomeni, da dejanje višje ravni uresničuje z manj nizi dejanj nižje ravni kot druga akterka, obenem pa med potekom branja večkrat preusmeri pozornost, in sicer vstopi v večdelno interakcijo, ki vključuje tudi interakcijo z občinstvom. To lahko opredelimo kot mikrodejanje višje ravni, ki ga izvaja sočasno z igranjem harmonike.



Slika 7: Graf pozornosti/osredinenosti, ki prikazuje umeščeno interpretiranje besedila in interpretiranja glasbe

Z analizo relativne kodne gostote lahko posamezna dejanja višje ravni, ki se izvajajo sočasno, umestimo na kontinuum pozornosti/osredinenosti. Interpretka glasbe sočasno z interpretiranjem glasbe (igranjem harmonike) komunicira z občinstvom. Čeprav igranju harmonike namenja visoko stopnjo pozornosti/osredinenosti, se kodna kompleksnost in intenziteta znižujeta, ko vzpostavlja interakcijo z občinstvom. Interpretka pri tem povezuje različne kode. Kodna kompleksnost pri komuniciranju z občinstvom se uresničuje z nizom obraznih izrazov, kot so dviganje obrvi (podoba 7) in nasmehi občinstvu (podobe 5–7), ter premiki glave proti občinstvu (podobe 5–7). S tem se zvišuje kodna gostota tega dejanja višje ravni. Intenziteta rokovanja s predmetom (tj. s harmoniko), ki je najintenzivnejši kod pri izvajanju glasbe, še zmeraj ostaja visoka, vendar pa se z drugimi dejanji nižje ravni, tj. premiki glave in ramen, zlasti pa obraznimi izrazi, dejanje komuniciranja z občinstvom premika naprej po kontinuumu pozornosti/osredinenosti.

V primerjavi z interpretko glasbe je interpretka besedila med samim branjem vključena zgolj v eno dejanje višje ravni, tj. interpretiranje literarnega besedila. Dejanje označujeta visoka kodna kompleksnost, tj. med interpretiranjem besedila povezuje različne kode, in visoka kodna intenziteta; posebej intenzivna koda sta pogled in govorni jezik (glasno branje). S tem dejanje interpretiranja besedila dosega visoko

stopnjo pozornosti/osredinjenosti. Primerjava med interpretiranjem besedila in interpretiranjem glasbe (igranjem harmonike) z vidika stopnje pozornosti/osredinjenosti je razvidna z grafa (slika 7).

4.3 Dogajališče literarnega branja

Eden od bistvenih elementov literarnega branja kot žanra je dogajališče. Literarno branje, ki je predmet analize, poteka pred kulturno-umetniškim lezbičnim klubom Monokel v AKC Metelkova mesto v Ljubljani. S stene, pred katero sedita akterki, med drugim razbiramo dejanja pisanja verzov in slikanja murala. Zamrznjena dejanja, vključena v objekte in okolje, pa ne izpričujejo samo predhodnih dejanj, ampak so tudi odraz identitete, konkretno lezbične, in omogočajo povezavo med branim besedilom in okoljem. Zamrznjena dejanja na stenah Monokla omogočajo vpogled v lezbično-feministično zgodovino ter živete ideje in politična prepričanja, ki cirkulirajo v skupnosti.



Slika 8: Dogajališče literarnega branja

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

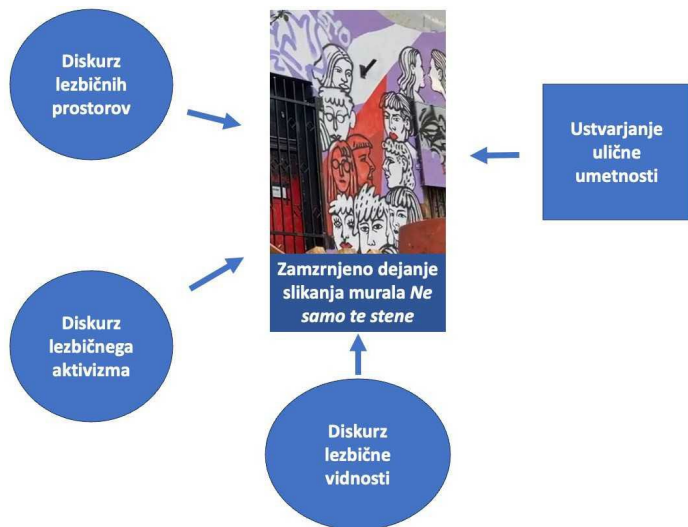
Slika 8 prikazuje interpretki in dogajališče literarnega branja, tj. vhod v lezbični klub Monokel, ki je neločljivo povezan tako z družbenim kot kulturnim in umetniškim vidikom lezbične skupnosti v Sloveniji, tega pa izpričuje tudi vizualna podoba pročelja stavbe. Kot izpostavi Konda, »ima vizualna podoba, ki se kaže 'zunanosti', poseben pomen pri percipiranju prostora« (Konda 2017: 185). Na sliki sva obkrožili tri objekte oz. elemente okolja, ki izpričujejo lezbično identiteto in jih bova natančneje predstavili, to so (od leve proti desni): na steno izpisani verzi iz pesmi

Manifest za tri minute lezbične glasbene skupine Volk za njenimi očmi,³ knjiga *Damski Almanah* Djune Barnes, ki je besedilo literarnega branja, in mural z naslovom *Ne samo te stene*, ki sta ga ustvarili umetnici Ana Lucija Šarić in Klara Kracina.

Vsi objekti oz. elementi okolja imajo lezbično-feministično sporočilo. Pesem *Manifest za tri minute* izpostavlja odtise in vpise heteronormativnega družbenega reda, evocira zatirane moduse bivanja, ki jih določa vztrajanje v času, ter poziva k solidarnosti v skupnem boju proti sistemom izključevanja in marginalizacije. Knjiga *Damski Almanah*, ki je izšla leta 1928, je satira elitističnih literarnih sodobnic Djune Barnes, ki so se v prvi polovici 20. stoletja združevale v literarnem salonu Natalie Clifford Barney. Pisateljica Bertha Harris je delo opredelila kot »dokument lezbične revolucije« svojega časa (Harris 1973: 81). Mural desno od interpretk z naslovom *Ne samo te stene* je nastal v okviru 6. festivala *Lezbična četrt*, ki je bil usmerjen v razpiranje kompleksnega prepletanja spola in seksualnosti, njunih zgodovinskih spojev in razhajanj (*Kulturni center Q*, 2020). Avtorici murala sta na stene prenesli pesem Kristine Hočevar *samo te stene* (Hočevar 2012: 89), v kateri pesnica artikulira pomen Monokla kot lezbičnega prostora v njegovih kulturnih in umetniških razsežnostih. Monokel je v pesmi vzpostavljen kot prostor, ki je vitalnega pomena za lezbično skupnost, tj. kot prostor, kjer »nihče ne more vreči / železa okrog teh gladkih vratov« (Hočevar 2012: 89), mural pa predstavlja vizualno opomenjenje tega prostora. Lezbično-feministična sporočila, ki jih razbiramo z objektov in elementov okolja, konstituira literarno branje, ki se s tem vzpostavlja kot del lezbičnega kontinuuma.

Zamrznjena dejanja, vključena v posamezne objekte oz. elemente okolja, se povezujejo z družbenimi praksami in diskurzi, ki ta dejanja omogočajo (Norris 2015: 52). V nadaljevanju se posvečava praksam in diskurzom, ki se povezujejo z dejanjem, ki je zamrznjeno v muralu *Ne samo te stene* (slika 9).

³Izpisani so naslednji verzi: Vztrajam. Pohajam. Ostajam. Pokončno. / Kot jaz. Kot napaka. Kot grožnja. Kot ni me. / Neumnost človeška zanika mi krila. / Izrodek sem. Plesen. Bolezen ... / A sila je v meni. Sila srca. / Bitje pomladi. Ljubezen mesa. / Ker nismo napaka. Že štart je zgrešen. / Ker ni normativa in ni mi vseen. / Ker nočem bit kriva za vaš "dober namen". / Nočem razlagat se, resnica je v men.



Slika 9: Povezava med zamrznjenim dejanjem ustvarjanja murala, praksami in diskurzi

Vir: <https://www.facebook.com/klubmonokel/videos/710669673718071>

Mural *Ne samo te stene* je nastal v konkretnem aktivističnem kontekstu, tj. v okviru festivala Lezbična četrt, ki je bil vzpostavljen kot platforma za lezbično ustvarjalnost na presečiščih umetnosti in aktivizma. Zamrznjeno dejanje slikanja murala je podprto z več družbenimi praksami; najtesneje se povezuje z ustvarjanjem ulične umetnosti, ki predstavlja konstitutivni element aktivističnih izrazov v javnem prostoru. Ustvarjalni izraz v skupnostnih prostorih usmerja pozornost na epistemske nepravilnosti (Bacharach 2018: 32) ter opolnomoča skupnost in posameznice_ke (*Walls of Justice*, 2021). Murale lahko, podobno kot grafite, prepoznamo tudi kot »pomembno orodje pri konceptualizaciji prostora« (Konda 2017: 170). V tem smislu se z muralom *Ne samo te stene* in družbeno prakso ustvarjanja ulične umetnosti povezuje več diskurzov; prevladujoči so diskurz lezbičnih prostorov, diskurz lezbičnega aktivizma in diskurz lezbične vidnosti (slika 9). Monokel kot lezbični prostor in mural sta povezana prek neposredne družbene angažiranosti in obenem družbeno marginaliziranega položaja. Lezbični prostori se vzpostavljajo na presečiščih zadrževanja in druženja ter kulturno-umetniške ustvarjalnosti. Prav obstoj lezbičnih prostorov omogoča avtonomen razvoj skupnosti in njene kulture – »razvijajo se lahko notranji diskurzi, tematike, polemike, raziskovanja, skratka, lastna, avtonomna, interna dinamika« (Velikonja 2011: 231). Družbena in socializacijska struktura lezbičnih prostorov predstavlja osnovo za refleksijo širših družbenih procesov, kakor se vpisujejo v življenja lezbijk, kakor tudi refleksijo njihovega

družbenega položaja in možnosti delovanja v javnem prostoru. Na tej osnovi se vzpostavlja diskurz lezbičnega aktivizma kot nepristajanje na samoumevnost marginaliziranega položaja lezbijke. Preko političnega sporočila murala se z diskurzom lezbičnega aktivizma povezuje diskurz lezbične vidnosti, ki vzpostavlja oblike samoreprezentacije, ki izstopajo iz patriarhalnega koncepta nevidnosti, in to »tako v umetniškem izrazu kot v utelesitvi sebe« (Dragičević 2015).

5 Sklep

Študija primer literarnega branja je pokazala, da večkodna interakcijska analiza omogoča večji vpogled v žanr literarnega branja kot zgolj žanrska analiza, saj veliko dejanj, ki tvorijo ta žanr, v tradicionalno žanrsko analizo ni vključenih. Pri literarnem branju se uresničujejo različni semiotski kodi, kot so govorjeni jezik, tonski potek, jakost glasu, glasba, rokovanje s predmetom, drža telesa, pogled, obrazni izrazi, kretnje; kodi se pri tem medsebojno povezujejo v kompleksno celoto. Pomemben element literarnega branja kot žanra je dogajališče. Zamrznjena dejanja, vključena v objekte in okolje, ne izpričujejo samo predhodnih dejanj, ampak omogočajo povezavo med branim literarnim besedilom in okoljem. Sistematična večkodna interakcijska analiza tudi pokaže, da imajo dejanja, ki tvorijo literarno branje, visoko kodno kompleksnost in intenziteto. Žanrska analiza sicer omogoča vpogled v žanrsko strukturo posameznega žanra, vendar ne zajame številnih neverbalnih dejanj in dogajališča; posledično se del kompleksnosti žanrskega opisa izgubi.

Večkodna interakcijska analiza potrjuje, da je treba koncept načina kot registrske spremenljivke razširiti v smislu, da zaobjame tudi neverbalne kode, da lahko različne kode in njihovo medsebojno povezovanje sistematično opišemo, s tem pa ovrednotimo tudi prispevek posameznih kodov k uresnitvi določenega žanra, kakor tudi načine, na katere se semiotski kodi medsebojno dopolnjujejo.

Literatura

- Sondra BACHARACH, 2018: Finding Your Voice in the Streets: Street Art and Epistemic Injustice. *The Monist* 101/1, 31–43.
- Djuna BARNES, 2009 [1928]: *Damski almanah*. Ljubljana: Založba Škuc. (Zbirka Vizibilija). Prev. Urška Sterle.
- Nina DRAGIČEVIĆ, 2015: Paraliza prihodnosti: pozicija in možnosti delovanja ženske in/ali lezbične glasbenice. *SIGIC*. <https://www.sigic.si/paraliza-prihodnosti-pozicija-in-moznosti-delovanja-zenske-in/ali-lezbicne-glasbenice.html> (20. 3. 2024).

- Charles GOODWIN, 2000: Action and Embodiment within Situated Human Interaction. *Journal of Pragmatics* 32/10, 1489–1522.
- Bertha HARRIS, 1973: The More Profound Nationality of their Lesbianism: Lesbian Society in Paris in the 1920's. *Amazon Expedition: A Lesbian Feminist Anthology*. Ur. Phyllis Birkby, Bertha Harris, Jill Johnston, Esther Newton, Jane O'Wyatt. New York: Times Change Press. 77–81.
- Christian HEATH, 2010: *Body Movement and Speech in Medical Interaction*. Cambridge: Cambridge University Press.
- Kristina HOČEVAR, 2012: *Na zobež aluminiij, na ustnicah kereda*. Ljubljana: Založba Škuc. (Zbirka Lambda).
- Helena KONDA, 2017: Grafiti kot vizualni simbol teritorialnih reprezentacij v Rogu. *Časopis za kritiko znanosti* 45/270, 168–190.
- Gunther KRESS, 2014: What is Mode? *Routledge Handbook of Multimodal Analysis*. Ur. Carey Jewitt. London: Routledge. 60–75.
- Gunther KRESS in Theo VAN LEEUWEN, 2001: *Multimodal Discourse: The Modes and Media of Contemporary Communication*. London, New York: Edward Arnold, Oxford University Press.
- Kulturni center Q*, 2020: 6. festival Lezbična četrt. https://www.kulturnicenterq.org/6-festival-lezbicna-cetrt/_20_3_2024.
- John LEE in Keith STENNING, 1998: Anaphora in Multimodal Discourse. *Multimodal Human-Computer Communication*. Ur. Harry Bunt, Robbert-Jan Beun, Tijn Borghuis. *Lecture Notes in Computer Science* 1374. Berlin, Heidelberg: Springer. 250–263. https://doi.org/10.1007/BFb0052322_20_2_2024.
- J. R. MARTIN, 1992: *English Text: System and Structure*. Amsterdam: Benjamins.
- J. R. MARTIN in David ROSE, 2003: *Working with Discourse. Meaning beyond the Clause*. London in New York: Continuum.
- Sigrid NORRIS, 2004: *Analyzing Multimodal Interaction: A Methodological Framework*. London: Routledge.
- Sigrid NORRIS, 2011: *Identity in Interaction: Introduction to Multimodal (Inter)action Analysis*. Göttingen: De Gruyter Mouton.
- Sigrid NORRIS, 2015: Objects, Frozen Actions, and Identity: A Multimodal (Inter)action Analysis. *Multimodal Communication* 4/1, 43–59.
- Sigrid NORRIS, 2019: *Systematically Working with Multimodal Data: Reseach Methods in Multimodal Discourse Analysis*. Hoboken, New Jersey: John Wiley & Sons.
- Julia NOVAK, 2011: *Live Poetry. An Integrated Approach to Poetry in Performance*. Amsterdam in New York: Rodopi.
- Kay O'HALLORAN (ur.), 2004: *Multimodal Discourse Analysis: Systemic Functional Perspectives*. New York: Bloomsbury.
- Michael O'TOOLE, 2011: *The Language of Displayed Art*. London, New York: Routledge.
- Jesse PIRINI, 2014: Introduction to Multimodal (Inter)action Analysis. *Interactions, Images and Texts: A Reader in Multimodality*. Ur. Sigrid Norris, Carmen Daniela Maier. Berlin, München, Boston: De Gruyter Mouton. 77–92.
- Katarina PODBEVŠEK, 2021: Avtorsko branje poezije. *Slovenska poezija. (Obdobja 40)*. Ur. Darja Pavlič. Ljubljana: Znanstvena založba Filozofske fakultete. 373–380.
- Katarina PODBEVŠEK, 2017: *Govornost literarnih besedil*. Maribor: Aristej.
- Sonja STARC, 2009: *Časopisna oglaševalska besedila, reklame: struktura in večkodnost*. Koper: Založba Annales.
- Sonja STARC, 2020: Vrednotenje v reklami in antireklami: primer alkoholnih in tobačnih izdelkov. *Slovenščina – diskurzji, zvrsti in jeziki med identiteto in funkcijo. (Obdobja 39)*. Ur. Jerica Vogel. Ljubljana: Znanstvena založba Filozofske fakultete. 67–78.
- Sonja STARC, 2023: Vpeljava vizualne slovnice v pouk slovenščine kot materinščine. *Revija za elementarno izobraževanje* 16, 137–155.
- Harmut STÖCKL, 2019: Linguistic Multimodality – Multimodal Linguistics: A State-of-the-art Sketch. *Multimodality: Disciplinary Thoughts and the Challenge of Diversity*. Ur. Janina Wildfeuer, Jana Pflaeging, John A. Bateman, Ognyan Seizov, Chiao-I Tseng. Berlin: De Gruyter. 41–68.
- Nataša VELIKONJA, 2011: *Lezbični bar*. Ljubljana: Založba Škuc. (Zbirka Vizibilija).

- Branislava VIČAR, 2015: Vloga vojaka v spomenikih prve svetovne vojne: ideološka reprezentacija vojne v tridimenzionalnih objektih. *Studia Historica Slovenica: časopis za humanistične in družboslovne študije* 15/3, 595–619.
- Branislava VIČAR, 2020: Reprezentacija živalske subjektivitete v aktivističnem filmu Green. *Pojmovanja živalskih smrti: antropocentrizem in (ne)možne subjektivitete*. Ur. Branislava Vičar. Koper: Znanstveno--raziskovalno središče, Annales ZRS. 201–228.
- Branislava VIČAR in Katja PLEMENITAŠ, 2023: "Mi vsi smo tujci." Identitetni imaginarij priseljenca: študija primera. *Naslanjanje raznolikosti v jeziku in književnosti*. Ur. Jožica Jožef Beg, Mia Hočevar, Neža Kočnik. Ljubljana: Zveza društev Slavistično društvo Slovenije. 281–294.
- Walls of Justice*, 2021: Social Justice Murals: Why Street Artists Can Create Social Change. <https://www.wallsofjustice.com/post/social-justice-murals-how-street-art-fuels-social-change> (2. 4. 2024).

GOVOR IN GOVORNA KOMUNIKACIJA V UČNIH NAČRTIH ZA OSNOVNO ŠOLO IN GIMNAZIJO TER V KATALOGIH ZNANJ

SIMONA PULKO, MELITA ZEMLJAK JONTES

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
simona.pulko@um.si, melita.zemljak@um.si

Govor je ključna človekova dejavnost, ki omogoča sporazumevanje z okolico. Za otrokov razvoj je pomembno, da že od rojstva dobi priložnosti za učenje sporazumevanja ob ustrezni podpori in zgledu. Govor ima individualno vlogo pri razvoju posameznika in družbeno vlogo pri sporazumevanju z drugimi. Eden temeljnih namenov jezikovnega pouka je razvijanje sporazumevalne zmožnosti v slovenskem (knjižnem) jeziku, tj. praktično in ustvarjalno obvladovanje vseh sporazumevalnih dejavnosti (tudi govora) in jezikovnosistemskih osnov. Z vidika prisotnosti poučevanja govora, strategij poučevanja in s poudarki na razvijanju govornih sposobnosti so analizirani učni načrti za slovenščino za osnovno šolo iz let 2011 in 2018 ter učni načrt za slovenščino za gimnazije iz leta 2008 ter katalogi znanja za slovenščino. Zanimalo naju je, kolikšen je poudarek na razvijanju pravorečne zmožnosti in na razvijanju zapisovalne zmožnosti govora, tudi v različnih načinih transkribiranja glede na stopnjo izobraževanja.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.15](https://doi.org/10.18690/um.ff.4.2024.15)

ISBN
978-961-286-882-6

Ključne besede:
slovenščina,
poučevanje govora,
učni načrt,
katalog znanja,
transkripcija govora



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.15](https://doi.org/10.18690/um.ff.4.2024.15)

ISBN
978-961-286-882-6

Keywords:
Slovene,
teaching the speech,
curricula,
catalogue of knowledge,
speech transcription

SPEECH AND SPEECH COMMUNICATION IN CURRICULA FOR ELEMENTARY SCHOOL, SECONDARY SCHOOLS AND IN CATALOGUES OF KNOWLEDGE

SIMONA PULKO, MELITA ZEMLJAK JONTES
University of Maribor, Faculty of Arts, Maribor, Slovenia
simona.pulko@um.si, melita.zemljak@um.si

Speech is a human activity that is important primarily of its communicative language function. A child's development requires appropriate support and role models from birth in order to create opportunities for learning to communicate. Speech plays an individual role in one's development and a social role in communication with the environment. One of the fundamental goals of language learning at school is the development of communication skills in the Slovene (literary) language, i.e. the practical and creative mastery of all communication activities (including speaking) and the basics of the language system. The analysis provides an overview of teaching the speech, teaching strategies and the emphasis on the development of speaking skills in the curricula for Slovene in primary schools (2011, 2018), the curricula for Slovene in secondary schools (2008) and the catalogues of knowledge for Slovene. One of the research goals was also to focus on the standard pronunciation development and the development of orthographic skills, including different types of transcription, depending on the level of education.



1 Uvod¹

Prispevek prinaša teoretična izhodišča razvoja govora in sporazumevalne zmožnosti s temeljnimi razpravami in ugotovitvami teh. Sledijo analiza govora in govorne komunikacije v učnih načrtih za slovenščino v osnovni šoli (2011 in 2018), govora in govorne komunikacije v učnem načrtu za slovenščino v gimnazijah (2008) ter govora in govorne komunikacije v katalogih znanj (*SSI* in *PTI*). Temu sledi tudi analiza transkripcije govora v istih učnih načrtih in katalogih znanj.

2 Razvoj govora in sporazumevalna zmožnost

Govorni razvoj otroka v predšolski dobi pomembno vpliva na celostni govorni razvoj otroka tudi v šolskem obdobju; kaže se v usvajanju ter razvijanju sporazumevalne zmožnosti otroka (Zemljak Jontes 2017: 45). V slovenskem šolskem prostoru je uveljavljen t. i. komunikacijski pristop k usvajanju jezika. Ta kot izhodišče za usvajanje jezika postavlja besedilo, sporazumevalna dejavnost oz. aktivnost pa predstavlja osnovno, ključno metodo dela s ciljem vzgojiti čim bolj učinkovitega uporabnika² jezika (Zemljak Jontes 2017: 45). O pomenu sporazumevalne zmožnosti v slovenskem izobraževalnem sistemu je bilo napisanih več razprav, monografij (npr. Bešter Turk (2011), Lipnik (1994), Kunst Gnamuš (1991, 1992), Krakar Vogel (2004), Zadravec Pešec (1994), Pulko (2007), Valh Lopert (2013), Zemljak Jontes (2017) idr.).

Bešter Turk (2011: 127) postavlja razvijanje sporazumevalne zmožnosti v slovenskem jeziku kot enega izmed temeljnih ciljev predmeta slovenščina na vseh stopnjah obveznega in neobveznega izobraževanja. Avtorica poudarja, da je slovenščina učni jezik in ne zgolj učni predmet, zato je treba sporazumevalno zmožnost razvijati pri vseh predmetih, ne le pri slovenščini, pri čemer se zastavlja vprašanje, ali se načrtovalci drugih predmetov, učitelji teh predmetov ter avtorji

¹ Simona Pulko je prispevek pripravila v okviru Raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine, vodja programa prof. dr. Marko Jesenšek), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Melita Zemljak Jontes je prispevek pripravila v okviru Raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine, vodja programa prof. dr. Marko Jesenšek), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna, in v okviru raziskovalnega projekta ARRS Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642), ki ga sofinancira Javna agencija za raziskovalno dejavnost Republike Slovenije (ARIS) iz državnega proračuna.

² V besedilu je kot generičen uporabljen moški spol, ki se nanaša na vse spole.

učnih gradiv zanje sploh zavedajo vloge, ki jo imajo oz. bi jo morali uresničevati pri razvijanju sporazumevalne zmožnosti. Razvijanje sporazumevalne zmožnosti v slovenskem jeziku je namreč pomembno na vseh ravneh izobraževanja in ne glede na predmetno področje (Zemljak Jontes 2017: 46). Pri predmetu slovenščina je tvorjenje govornih besedil tako moč predvideti v skoraj vseh sklopih učnih načrtov, v prispevku pa bodo obravnavani tisti, ki jih učni načrti (v nadaljevanju UN) in kataloga znanj (v nadaljevanju KZ oz. KZ (PTI)) eksplicitno izpostavljajo.

3 Govor in govorna komunikacija v učnem načrtu za slovenščino (2011)

UN za osnovno šolo (v nadaljevanju OŠ) že v poglavju *Opredeleitev predmeta*, v katerem opredeli slovenščino kot materni oz. prvi jezik za večino učencev ter kot drugi jezik oz. jezik okolja za pripadnike manjšin (ključni splošnoizobraževalni predmet v OŠ), poudari razvijanje in skrb za razvoj govora in govorne komunikacije kot enega izmed temeljnih ciljev (UN 2011: 4): »Učenci se pri njem usposablajo za učinkovito govorno in pisno sporazumevanje v slovenskem jeziku.«

Namen jezikovnega pouka je med drugim razviti sporazumevalno zmožnost v slovenskem (knjižnem) jeziku, pri čemer je poseben poudarek namenjen praktičnemu in ustvarjalnemu obvladovanju vseh štirih sporazumevalnih dejavnosti, med katere sodi tudi govorjenje (UN 2011: 4).

Že v splošnih ciljih je poudarjeno, da učenci ob pisnih dejavnostih razvijajo pripravljenost za pogovarjanje in govorno nastopanje; tako izražajo svoje znanje, misli, stališča, hotenje, čustva in izkušnje, se pogajajo ter miroljubno rešujejo različna vprašanja s pomočjo tvorjenja različnih besedil tako pri jezikovnem kot tudi pri književnem pouku. S tem pa razvijajo svojo socialno, kulturno in estetsko zmožnost.

V 1. vzgojno-izobraževalnem obdobju (v nadaljevanju VIO) učenci v sklopu *Razvijanje zmožnosti pogovarjanja* sodelujejo v igri vlog, pri čemer ponuja UN različne vrste pogovora (neuradni in uradni osebni pogovor ter neuradni in uradni telefonski pogovor, UN 2011: 12). Različne dejavnosti razčlenbe pogovora omogočajo učencem vrednotenje razumljivosti in zanimivosti pogovora, vljudnosti obeh sogovorcev in utemeljevanje lastnega mnenja. Z učiteljevo pomočjo povzemajo temeljna načela vljudnega pogovarjanja, med igro vlog izražajo svoja občutja, vrednotijo svojo zmožnost pogovarjanja in načrtujejo, kako bi jo lahko izboljšali

(UN 2011: 8). Podobne cilje najdemo v 2. in 3. VIO, pri čemer se postopoma zmanjšuje delež učiteljeve pomoči, predlagane vrste pogovora pa so iste (UN 2011: 30). Vrste pogovorov, ki so ponujene v sklopu *Vsebine* v 3. VIO, so: neuradni in uradni osebni/telefonski raziskovalni pogovor, neuradni in uradni osebni/telefonski prepričevalni pogovor, neuradni in uradni osebni/telefonski pogajalni pogovor (UN 2011: 54).

Poglavje *Razvijanje zmožnosti govornega nastopanja*³ predvideva govorno nastopanje učencev z vnaprej napovedano temo in besedilno vrsto, pri čemer je upoštevano načelo postopnosti, saj to najprej poteka vodeno, nato pa ob ponazorilih čim bolj samostojno (izhodišče predstavlja zgled podobnega, že poslušanega besedila). Pri tem učenci vrednotijo zanimivost, živost in razumljivost besedila, predlagajo popravke, izboljšave, utemeljujejo svoje mnenje in vrednotijo rabo nebesednih spremljevalcev govorenja. Z učiteljevo pomočjo povzemajo temeljna načela uspešnega govornega nastopanja, izražajo svoja občutja med govornim nastopom, vrednotijo svojo zmožnost govornega nastopanja in načrtujejo, kako bi jo lahko izboljšali (UN 2011: 9). Vrste besedil, načrtovane za govorno nastopanje v 1. VIO, so: pripoved o doživetem/videnem/slišanem, predstavitev lastnih načrtov za dani dan/konec tedna/počitnice, obnova knjige/risanke/filma, opis sebe ali druge osebe, delovnika, predmeta/igrača, živali, prostora, zgradbe, poti v šolo/katere druge poti (UN 2011: 13).

Razvijanje pravorečne zmožnosti je v UN načrtovano po naslednjem sosledju: v 1. VIO učenci zgolj posnemajo knjižno izreko, v 2. VIO spoznajo slovenske knjižne samoglasnike ter najprej vadijo in utrjujejo njihovo izreko v besedah, zahtevnost pa stopnjujejo z vadenjem in utrjevanjem knjižnega izgovora posameznih besed in povedi. V 3. VIO učenci vadijo knjižni izgovor besed, povedi in besedil, glasno izgovarjajo besede, ob tem pa še opisujejo njihovo izreko (npr. določijo mesto naglasa, naglašeni samoglasnik in njegovo trajanje, položajne variante zvočnika *v*, glasovno vrednost črke *l* in premene nezvočnikov), preseneča pa, da pri tem ne uporabljajo strokovnih izrazov in naglasnih znamenj, znanje uzaveščajo zgolj na taksonomski stopnji rabe. V vseh treh VIO učenci med govornim nastopanjem in med pogovarjanjem z učiteljem skušajo čim bolj dosledno uporabljati knjižno izreko,

³ Poglavji *Razvijanje zmožnosti govornega nastopanja* in *Razvijanje zmožnosti tvorjenja enogovornih neumetnostnih besedil* predvidevata govorno nastopanje učencev, vendar razvijanje zmožnosti tvorjenja enogovornih neumetnostnih besedil ob tvorjenju govorenih besedil vključuje tudi tvorjenje pisnih enogovornih neumetnostnih besedil.

v svoji in tuji (govorjeni) rabi prepoznavajo pravorečne napake ter jih odpravijo (UN 2011: 11, 28, 50, 51).

V sklopu *Razvijanje zmožnosti tvorjenja enogovornih neumetnostnih besedil* učenci v 2. VIO argumentirano vrednotijo svoj govorni nastop ter poznavanje besedilne vrste. Prav tako ovrednotijo svojo zmožnost govornega nastopanja /.../ enogovornih besedil in načrtujejo, kako bi jo izboljšali (UN 2011: 25, 26). Pri presojanju govornega nastopa torej ocenjujejo samo kakovost izvedbe govornega nastopa kakor tudi vsebinsko ustreznost ter poznavanje in upoštevanje značilnosti posamezne besedilne vrste. V 3. VIO učenci uresničujejo faze sporočanja ter tvorijo (tudi govorno) čim bolj ustrezna, razumljiva in jezikovno pravilna enogovorna besedila določene vrste /.../, argumentirano vrednotijo zanimivost, živost, ustreznost, razumljivost in jezikovno pravilnost svojega govornega nastopa /.../ ter odpravljajo napake ali pomanjkljivosti, vrednotijo svojo zmožnost govornega nastopanja /.../ ter načrtujejo, kako bi jo izboljšali (UN 2011: 48). Ob tem sledijo načelom praktične stilistike, kar pa v učnem načrtu ni izrecno izpostavljeno.

Razvijanje govora je v UN posebej poudarjeno tudi pri *Književnosti*, zlasti v poglavjih *Glasno interpretativno branje književnih besedil* in *Govorni nastopi*. V obeh poglavjih so za vsa tri VIO predstavljeni načini, kako naj učenci izvajajo branje oz. na kaj naj bodo pozorni pri izvedbi govornega nastopa (npr. ob predstavitvi lastne bralne izkušnje ali ob deklamaciji, recitaciji pesmi (UN 2011: 20, 21, 41, 65)). UN za 2. in 3. VIO pri *Govornih nastopih* predvideva, da učenci v vsakem razredu pripravijo en govorni nastop, pri čemer izbirajo med ponujenimi možnostmi in se odločijo za povezavo z domačim branjem, lahko pa izberejo poljubno temo, ki je ni na seznamu, a je izbor treba utemeljiti. Pri izvedbi govornega nastopa prosto in čim bolj naravno govorijo, pozorni so na jezikovno pravilnost in ustreznost (težnja h knjižni izreki), pri tem pa upoštevajo zvočne in vidne prvine govora (UN 2011: 42, 66). V 3. VIO učenci upoštevajo vsa usvojena jezikovna znanja, prepričljivost dosegajo s premišljeno pripravo govornega nastopa na ravni razvijanja teme, pri izvedbi pa upoštevajo zvočne in vidne prvine govora za izvedbo književnostrokovnih govornih nastopov (obravnavo književnega dela) in (po)ustvarjalnih govornih nastopov (UN 2011: 67).

V poglavju *Standardi znanja* so ciljem primerno določeni tudi standardi znanja, ki naj bi jih učenec dosegel. Tako ima skladno s cilji iz učnega načrta razvito zmožnost govornega nastopanja, ki jo izkazuje s samostojnim govornim nastopanjem z vnaprej

napovedano temo in besedilno vrsto ter tvori smiselna, razumljiva, sovisna in zaokrožena besedila, govori razločno, naravno in čim bolj knjižno, po govornem nastopu vrednoti svoj nastop oziroma nastope sošolcev (UN 2011: 74). Pravorečno zmožnost pokaže tako, da med govornim nastopanjem in med pogovarjanjem z učiteljem govori čim bolj knjižno, v svoji in tuji izreki prepozna neknjižne glasovne ali naglasne prvine ter jih zamenja s knjižnimi (UN 2011: 76). V 2. VIO učenec pravorečno zmožnost pokaže tako, da ob standardih, ki so predvideni že v 1. VIO, našteje slovenske knjižne samoglasnike in jih v besedah pravilno izgovori (UN 2011: 82), v 3. VIO pa sta pri pravorečni zmožnosti standardom 1. in 2. VIO dodana še dva standarda, in sicer obvladovanje knjižnega izgovora besed in povedi, opis tega ter prepoznavanje pravorečnih napak v besedilih in njihova odprava (UN 2011: 90).

4 Govor in govorna komunikacija v učnem načrtu za slovenščino (2018)

Učni načrt za slovenščino v osnovni šoli iz leta 2018 je v šolski praksi veljaven od 1. 9. 2019. Ker je glede obravnavanih vsebin tako glede vsebin kakor glede koncepta obravnave v veliki meri primerljiv z UN 2011, so v analizi izpostavljene zgolj razlike, dopolnitve in/ali morebitna nadgradnja UN 2011. V 1. VIO si učenci v sklopu *Razvijanje zmožnosti pogovarjanja* ogledajo posnetek oz. sodelujejo v igri vlog, in sicer v pogovoru določene vrste. Koncept obravnave je moč primerjati s konceptom obravnave v UN 2011, natančneje pa je le opredeljena izvedba dejavnosti pred ogledom posnetka in po njem (UN 2018: 9). Vrste besedil v 1. VIO, s pomočjo katerih se dosegajo cilji, so: osebni in telefonski (samo zvočni ali tudi vidni) pogovor med prijateljskoma, sorodnikoma ter med stranko in uradno osebo (npr. učitelj, zdravnica, prodajalka).

Sklop *Razvijanje zmožnosti govornega nastopanja* v 1. VIO predvideva govorno nastopanje učencev z vnaprej napovedano temo in besedilno vrsto. Tudi tukaj je koncept obravnave moč primerjati s konceptom v UN 2011, izpostaviti velja le nekaj dopolnitev: učitelj in učenci se pred govornim nastopom pogovarjajo o tem, kdaj in čemu govorno nastopamo, ugotavljajo, po čem se govorni nastop loči od pogovora in kako se nanj ustrezno pripravimo ter na kaj smo med samo izvedbo posebej pozorni (UN 2018: 10). Vrste besedil, ki jih UN predvideva za govorno nastopanje v 1. VIO, so primerljive z vrstami besedil v UN 2011.

Razvijanje pravorečne zmožnosti je v UN zasnovano po naslednjem sosledju: V 1. VIO so učenci med poslušanjem posnetih ali glasno branih besedil v knjižnem jeziku pozorni na knjižno izreko besed in povedi ter si jo skušajo zapomniti, nato jo posnemajo, vadijo in utrjujejo. Sledijo pojasnila razlik med knjižno in neknjižno izreko. Poseben poudarek je namenjen vadenju knjižne izreke zanje kritičnih besed, ki si jo učenci skušajo zapomniti. Učenci skušajo med govornim nastopanjem govoriti čim bolj knjižno in razločno, tudi sicer pa prepoznavajo svojo in tujo neknjižno izreko ter besede izgovorjajo knjižno (UN 2018: 10, 11). V 2. VIO razvijanje pravorečne zmožnosti sledi konceptu za 1. VIO, ob tem pa še učenci v besedah prepoznajo knjižne samoglasnike in utrjujejo njihovo izreko v besedah in povedih. Nadgradnja, ki jo prinaša UN, je, da učence ob morebitnih nejasnostih usmerja v kritično in premišljeno rabo ustreznih jezikovnih priročnikov v knjižni in elektronski obliki (UN 2018: 22). V 3. VIO UN ciljem 1. in 2. VIO dodaja še vadenje in utrjevanje knjižne izreke ob besedilih. Tako učenci glasno izgovarjajo (tudi berejo) besede in opisujejo njihovo izreko, npr. mesto naglasa in naglašeni samoglasnik, položajne variante zvočnika v, glasovno vrednost črk e, o, v in l, vendar pri tem ne uporabljajo naglasnih znamenj in strokovnih izrazov za zvočniške variante (kot v UN 2011). V UN je posebej izpostavljena primerjava slovenskih glasovnih in besednih prozodičnih prvin s tujimi; učenci svoje ugotovitve tudi predstavijo (UN 2018: 35, 36).

Sklop *Razvijanje zmožnosti enosmernega sporazumevanja* v 2. in 3. VIO pri učencih z dejavnostmi in vsebinami spodbuja razvijanje zmožnosti enosmernega sporazumevanja, in sicer s tvorjenjem besedil določene vrste, bodisi kot govorni nastop bodisi kot zapisano besedilo. Pri tem lahko najdemo veliko vzporednic s konceptom obravnave v UN 2011. Izpostaviti velja naslednje poudarke, ki jih prinaša UN 2018, poseben poudarek je namenjen strategijam pred govornim nastopom (npr. učenci in učitelj se pogovarjajo o strategijah in načelih učinkovitega govornega nastopanja, o tem, kako se pripravimo na govorni nastop /.../ (UN 2018: 21); pri dejavnostih po govornem nastopu pa ne prihaja do sprememb (UN 2018: 34).

Pri ciljih 1., 2. in 3. VIO pri *Književnosti* v poglavjih, ki se posvečajo govoru, tj. *Glasno interpretativno branje književnih besedil* in *Govorni nastopi*, v primerjavi z UN 2011 ne prihaja do sprememb (UN 2018: 18, 31, 45).

5 Govor in govorna komunikacija v učnem načrtu za slovenščino v gimnazijah (2008)

V *Opredelitvi predmeta* je zapisano, da je UN nadgradnja UN iz OŠ, zato se dijaki med drugim usposablajo za učinkovito govorno in pisno sporazumevanje (UN 2008: 5). Med temeljnimi cilji in kompetencami je med drugim zapisano, da dijaki tvorijo učinkovita, razumljiva, ustrezna in jezikovno pravilna ustna in pisna besedila (UN 2008: 6).

V sklopu *Razvijanje zmožnosti tvorjenja enogovornih besedil* dijaki pred tvorjenjem sodelujejo v vodenem pogovoru o strategijah tvorjenja enogovornih neumetnostnih besedil, načelih uspešnega tvorjenja, vrstah enogovornih neumetnostnih besedil in značilnostih tiste vrste enogovornega neumetnostnega besedila, ki ga bodo tvorili; nato se pripravijo /.../ na govorni nastop. Tudi v tem UN je poudarek namenjen strategijam pred tvorjenjem besedil. Po tvorjenju pa dijaki vrednotijo razumljivost, zanimivost, resničnost, aktualnost, živost, učinkovitost, ustreznost in jezikovno pravilnost svojega /.../ govornega nastopa in /.../ govornih nastopov sošolcev ter utemeljujejo svoje mnenje, poslušajo mnenja sošolcev o svojem govornem nastopu in izrekajo svoje (ne)strinjanje z njimi, poročajo o svoji strategiji govornega nastopanja in jo primerjajo s strategijami sošolcev ter izdelajo načrt za izboljšanje svojih zmožnosti (UN 2008: 12). Pri tem velja poudariti, da je ob strategijah po tvorjenju posebna pozornost namenjena načelom praktične stilistike, ob tem pa še kritičnemu presojanju/vrednotenju lastnih in tujih govornih nastopov z argumentiranjem (ne)sprejemanja, (ne)strinjanja podanih mnenj. S tem se sledi cilju razvijanja kritičnega govorca in kritičnega poslušalca.

Dijaki razvijajo pravorečno zmožnost ob različnih dejavnostih poslušanja knjižne izreke in z njeno uzaveščeno rabo. V 1. letniku sistematično spoznavajo, vadijo in utrjujejo knjižni izgovor posameznih besed in povedi, prepoznavajo pravorečne napake v svojih in tujih besedilih ter jih odpravljajo, med govornim nastopanjem pa skušajo govoriti čim bolj knjižno (UN 2008: 13).

V sklopu *Razvijanje zmožnosti branja in interpretiranja književnih besedil* dijaki po branju interpretirajo besedila, ob tem izražajo svoje doživljanje, razumevanje, književna besedila aktualizirajo in vrednotijo. Poleg tega sodelujejo v pogovoru, tvorijo govornjeno /.../ enogovorno besedilo, npr. govorno nastopajo o svojem branju

besedil iz vsebinskih sklopov, recitirajo ali uprizarjajo obvezna in izbirna besedila (UN 2008: 16).

Splošno sporazumevalno zmožnost razvijajo ob branju književnih besedil, ob dialoškem odzivanju nanje in ob tvorjenju enogovornih besedil o njih oziroma o književnosti, npr. zmožnost izražanja svojega mnenja, zmožnost sprejemanja drugotnih neumetnostnih besedil, zmožnost tvorjenja učinkovitih, ustreznih, razumljivih in jezikovno pravilnih drugotnih neumetnostnih besedil (UN 2008: 16).

V poglavju *Pričakovani dosežki/rezultati* iz ciljev izpeljani tudi pričakovani dosežki oz. rezultati: pravorečna zmožnost z dosledno rabo knjižne izreke med govornim nastopom; izgovarjava besed/povedi neknjižno in knjižno, primerjava obeh izgovorov in opis knjižnega izgovora; poznavanje knjižnih glasnikov in primerjava z neknjižnimi (UN 2008: 36); razvita zmožnost kritičnega sprejemanja enogovornih neumetnostnih besedil (pojasnilo kratice, bi lahko bila rabljena že prej v prispevku?) z zmožnostjo (uradnega) pogovarjanja; razvita (splošna) sporazumevalna zmožnost pa z dejavnim sprejemanjem drugotnih neumetnostnih besedil, z izreko in utemeljevanjem lastnega mnenja, s tvorjenjem učinkovitih, ustreznih, razumljivih in jezikovno pravilnih drugotnih neumetnostnih besedil (UN 2008: 38).

6 Govor in govorna komunikacija v katalogu znanja za predmet slovenščina (SSI)

Slovenščina je temeljni splošnoizobraževalni predmet tudi v srednjem strokovnem izobraževanju. Dijaki se med drugim usposablajo za učinkovito govorno sporazumevanje v slovenščini (KZ 2011: 2), razvijajo sporazumevalno zmožnost kot zmožnost kritičnega sprejemanja besedil različnih vrst ter zmožnost tvorjenja ustreznih, razumljivih, pravilnih in učinkovitih besedil različnih vrst (KZ 2011).

Med *Usmerjevalnimi/splošnimi cilji predmeta* velja pri razvijanju govora in govorne komunikacije poudariti, da dijaki razvijajo sporazumevalno zmožnost v slovenskem knjižnem jeziku, da obvladajo slovenski knjižni jezik in ga zavestno uporabljajo ter da znajo presoditi, ali je v danih sporazumevalnih okoliščinah ustrezna raba knjižnega ali neknjižnega jezika. Prav tako razvijajo zmožnost pogovarjanja, poslušanja (gledanja) in branja različnih besedil ter zmožnost ustnega in pisnega sporočanja (KZ 2011: 3).

V KZ je pri *Primerih sklopov za obravnavo* eden od dveh predlaganih sklopov za jezikovni pouk *Intervju*. Dijaki ob intervjuju presoјajo nebesedni jezik in izreko sogovorcev iz posnetega intervjuja ter ob tem spoznavajo in si uzavestijo značilnosti zvočnega prenosnika, razvijajo svojo pravorečno zmožnost, ki jo uresničujejo ob predlaganih dejavnostih, npr. presoјajo izreko sogovorcev v posnetem intervjuju, predstavijo osnovno razdelitev knjižnih glasnikov na samoglasnike in soglasnike, v besedah določijo obe nadglasovni prvini (tj. naglas in trajanje naglašenga samoglasnika), razmišljajo o svojih pravorečnih napakah in jih ponazorijo s primeri; rešujejo dodane pravorečne vaje (KZ 2011: 17). Iz ciljev so izpeljani tudi standardi znanja, ki kot minimalne standarde določajo dijakovo razvito zmožnost upoštevanja načel uspešnega pogovarjanja in knjižne izreke (KZ 2011: 18). Po predlaganem sklopu je moč uresničevati cilje za razvijanje sporazumevalne in pravorečne zmožnosti ter razvijanje govora tudi pri drugih vrstah besedil.

Za boljšo motiviranost za učenje avtorji KZ predlagajo premišljeno povezovanje obravnavanih besedil z njihovim poklicnim življenjem, s čimer vključujejo vpeljavo avtentičnih nalog v šolsko prakso.

Tudi za področje književnosti avtorji KZ predlagajo, da dijaki besedila oz. odlomke govorno interpretirajo. Priporočljivi obliki pogovora sta okrogla miza in kviz. Dijaki v pogovoru izberejo tematsko ustrezno besedilo za individualno branje oz. govorni nastop, ki mu sledi. V govornem nastopu predstavijo problematiko književnega besedila, po nastopu o njej razpravljajo z učiteljem in s sošolci, ocenjujejo pa tudi govorno nastopanje. Skupina dijakov ob koncu obravnavanega sklopa pripravi in izvede tematski recital (KZ 2011: 40). Tudi sicer pri predlogih za ocenjevanje avtorji KZ predlagajo govorni nastop kot enega od načinov ocenjevanja ob vseh predlaganih tematskih in literarnozgodovinskih sklopih.

7 Govor in govorna komunikacija v katalogu znanja za predmet slovenščina (PTI)

Slovenščina predstavlja temeljni splošnoizobraževalni predmet tudi v poklicno-tehniškem izobraževanju (PTI). Zaradi pomena, ki ga ima slovenščina kot materni oz. prvi in/ali državni jezik v osebnem, poklicnem in družbenem uresničevanju posameznika, ima predmet obsežne in razvejane naloge. Dijaki se pri predmetu

usposabljaajo za učinkovito govorno in pisno sporazumevanje v slovenskem jeziku in pridobivajo za to potrebno znanje (*KZ (PTI)*: 1).

Pri *Udejanjanju kompetenc* je zapisano, da se zaradi vsebinske in dejavnostne razvejanosti predmeta ter širokega spektra vrednot, ki jih spodbuja, razvijajo ključne zmožnosti/kompetence posameznika, med katerimi je tudi razvijanje sporazumevalne zmožnosti.

Med enajstimi predlogi sklopov, kjer je izreka posebej izpostavljena, se pojavljata dva, to sta *Pogajalni pogovor* (s sodelavcem/s poslovnimi partnerji, ob tem pa sta poudarjena izreka ter zapis besed) in *Prepričevalni pogovor* (s sodelavcem/s stranko, ob tem pa sta poudarjena izreka ter zapis povedi).

Pri učnih ciljih in dejavnostih za pouk in priporočilih ob pogajalnem pogovoru po ponovnem poslušanju/branju zapisa dela pogovora dijaki razčlenjujejo in vrednotijo izreko sogovorcev iz posnetega pogovora in jo primerjajo z zapisom ter ob tem razvijajo svojo pravorečno in pravopisno zmožnost. Dijaki presojujejo izreko sogovorcev v posnetem pogovoru. Pri tem ovrednotijo izreko vseh govorcev v javnih/zasebnih poslušanih pogovorih ter neuradnih/uradnih govornih položajih. Predstavijo osnovno razdelitev knjižnih glasnikov na samoglasnike in soglasnike; v besedah določijo obe nadglasovni prvini (tj. naglas in trajanje naglašene samoglasnika); razmišljajo o lastnih pravorečnih napakah ter jih ponazorijo s primeri; rešujejo dodane pravorečne vaje; povedi in besedila izgovarjajo tako, da zavestno uporabljajo t. i. prvine stavčne fonetike; nato vrednotijo svojo izreko in jo popravljajo; poslušajo svoje sošolce ter druge govorce, vrednotijo njihovo izreko in jo popravljajo. Pri tem velja poudariti, da je po tvorjenju posebna pozornost namenjena kritičnemu presojanju/vrednotenju lastne in tuje izreke z argumentiranjem (ne)sprejemanja, (ne)strinjanja podanih mnenj. S tem se sledi cilju razvijanja kritičnega govorca in kritičnega poslušalca.

Pri prepričevalnem pogovoru dijaki samostojno tvorijo prepričevalni pogovor in upoštevajo značilnosti uspešnega pogovarjanja ter izbirajo učinkovite strategije načrtovanja in tvorjenja, sicer pa je koncept obravnave primerljiv s konceptom ob pogajalnem pogovoru (*KZ (PTI)*: 12–13).

Govorno nastopanje se v KZ pojavlja še pri reportaži (KZ (PTI): 23), referatu (KZ (PTI): 25–27) in pri oceni ali komentarju (KZ (PTI): 35). Pri referatu dijaki pojasnjujejo razliko med govorjenjem na pamet in prosto govorno predstavitev referata z uporabo ponazoril in pripomočkov ter jo ovrednotijo. Govorna predstavitev je v KZ tudi kot predlog pri ocenjevanju; med kriteriji za vrednotenje nastopa pa je zapisan kriterij upoštevanja prozodičnih prvin in ustrezne rabe jezika.

Tudi na področju književnosti avtorji KZ pri preverjanju predlagajo samostojni govorni nastop kot poročilo o domačem branju/govorni nastop/referat/predstavitev plakata in govorne poustvarjalne interpretacije (recital pesmi/uprizoritev odlomka iz dramskega besedila/uprizoritev dramatisiranega odlomka pripovednega besedila) (KZ (PTI): 49).

8 Transkripcija govora

Razvijanje pravorečne zmožnosti je pri predmetu slovenščina ob tvorjenju govornih besedil moč predvideti v skoraj vseh sklopih učnih načrtov. V prispevku predstavljenih UN in KZ, kjer je slednje zapisano eksplicitno, se vseskozi nakazuje tudi specifične obravnave in udejanjanja pravorečja v povezavi s pravopisjem.

Pri tem je ključno fonološko oz. glasovno zavedanje, ki ima v abecednem sistemu pisave osrednjo vlogo pri učenju branja in pisanja. Ključni elementi glasovnega zavedanja so: (a) glasovno razločevanje (razločevanje dolžine besed in razločevanje glasov med seboj), (b) glasovno razčlenjevanje (členitev povedi na posamezne besede, posamezne zloge in na posamezne glasove) ter (c) poznavanje in uporaba knjižnega naglasa ter knjižne izreke. Glasovno zavedanje kot metajezikovno zmožnost glasovnega procesiranja jezika je treba razvijati na vseh strokovnih področjih, ne le na področju (slovenskega) jezika, in na vseh ravneh izobraževanja. Oblikujejo jo različne ravni, hierarhično razporejene glede na zahtevnost miselne operacije, potrebne za njihovo izvedbo, posamezne ravni pa se razvijajo kronološko in so pogojene z miselnim razvojem otrok. Sposobnosti fonološkega zavedanja se razvijajo postopoma v zaporedju od največjih glasovnih enot (stalnih besednih zvez, besed, zlogov) do najmanjših, zahtevnejših (torej od stalnih besednih zvez, besed, zlogov do glasov/fonemov). Posameznikov ontogenetski razvoj jezika že v predšolskem obdobju in še v 1. VIO zajema kot ključna elementa razvijanja glasovnega zavedanja glasovno razločevanje in glasovno razčlenjevanje, v nadaljnjem

izobraževanju pa izpostavlja predvsem razlikovanje med knjižno in neknjižno izreko (2. VIO: izpostavljeno spoznavanje razlik med knjižnimi in neknjižnimi naglasi; 3. VIO, srednješolsko izobraževanje: ločevanje in razumevanje razlik med knjižnimi in neknjižnimi naglasi ter zmožnost uporabe knjižnega naglasa oziroma knjižne izreke) (Zemljak Jontes, Bednjički Rošer (2020: 107–136); Krajnc Ivič, Zemljak Jontes (2022: 174–204)).

Ena novejših in nekako najbolj vseobsegajočih opredelitev glasovnega zavedanja (Čok in Brčaninović 2015) opredeljuje naslednje ravni fonološkega zavedanja: (1) zavedanje o dolžini besed, (2) zaznavanje in tvorjenje rim ter aliteracij, (3) členjenje povedi na besede, (4) zlogovno zavedanje (zlogovanje in združevanje zlogov v besedo), (5) zavedanje o začetkih in koncih besed (deljenje besed na začetke in konce, združevanje začetkov in koncev v besedo, tvorjenje besed s podanimi začetki in/ali konci) in (6) fonemično zavedanje z vključevanjem prepoznavanja glasov v glasovni verigi in manipuliranjem glasov (kot so: opuščanje/redukcija, dodajanje/adicija ali nadomeščanje/substitucija glasov, členjenje/segmentacija glasovne verige na glasove in spajanje glasov v besede) (Čok in Brčaninović 2015).

Pri pripravi in izvedbi ter seveda vrednotenju vseh vrst govornih besedil po UN in KZ za slovenščino je še kako pomembno, da je besedilo ustrezno zapisano in da govorniku nudi čim boljše (tudi fonetično in fonološko) pomoč za ustrezno knjižno izgovarjavo. Tako je pravzaprav smiselno način zapisa (transkripcijo) prilagoditi zmožnostim človekovega ontogenetskega razvoja. Sodobni normativni jezikoslovni priročniki (*Slovenska slovnica*, *SSKJ²*, *Slovenski pravopis*, *eSSKJ*, *ePravopis*) ponujajo prikaz nekaterih izgovornih značilnosti (npr. mesta in vrste naglasa, posebnosti v kvaliteti in/oz. kvantiteti naglašene samoglasnika, nekaterih oblikoglasnih premen, tonemskega naglasa), pri čemer so ti sicer dobrodošli, a za šolajočega se uporabnika v zapisu glede na stopnjo izobraževanja bodisi preveč zahtevni bodisi v nekaterih segmentih še premalo natančni, kar velja tudi za povprečnega splošnega uporabnika.⁴ V ta namen je smiselna uporaba prilagojenega načina zapisa po principu »Piši, kot govoriš.«, pri čemer je treba jasno in natančno zastaviti, kateri segmenti izgovarjave so (še) potrebni kot pomoč za ustrezen izgovor in kateri ne

⁴ O tem so razpravljali tudi na 1. slovenskem pravorečnem posvetu leta 2020 (<https://www.sazu.si/uploads/files/publikacije21/Rared2RAZPRAVE.pdf>) npr. H. Tivadar, L. Horjak, T. Mirtič, H. Dobrovoljc, T. Lengar Verovnik, N. Gliha Komac, M. Šekli, J. Faganel, R. Dovjak, N. Žavbi, K. Podbevšek, M. Snoj, J. Ježovnik, A. Valh Lopert, M. Zemljak Jontes.

(več). Tako lahko učenci, dijaki sami ali z učiteljevo pomočjo oblikujejo (skupne razredne ali lastne) različice t. i. poknjžženega zapisa oz. poknjžžene transkripcije. Pri tem je nujno upoštevati narečne glasovne značilnosti v primerjavi s knjižnimi ter posameznikov idiolekt, pri vsem pa je pomembna učiteljeva senzibilnost in usmerjanje otroka v prepoznavanje, zavedanje ter sprejemanje obeh glasovnih sistemov kot 'pravih', vendar v zanje ustreznih sporazumevalnih (govornih) položajih.

9 Sklep

Analiza učnih načrtov in katalogov znanj je pokazala, da so govor in razvijanje govora, govorne komunikacije ter sporazumevalne zmožnosti kot temeljni cilj(i) prisoten/-i v vseh analiziranih učnih načrtih in katalogih znanj, je pa morebiti poudarek na razvijanju (zlasti) pravorečne zmožnosti premalo eksplicitno poudarjen.

Razvijanje pravorečne zmožnosti v UN 2011 in 2018 poteka skozi premišljeno rabo, in sicer v 1. VIO med poslušanjem posnetih ali glasno branih besedil v knjižnem jeziku in usmerjanjem dejavnosti na knjižno izreko besed in povedi (učenci si jo skušajo zapomniti, nato jo posnemajo, vadijo in utrjujejo); razlike med knjižno in neknjižno izreko poskušajo pojasniti. Poseben poudarek je namenjen vadenju knjižne izreke zanje kritičnih besed, ki si jo učenci skušajo zapomniti. Med govornim nastopanjem je poudarjena čim bolj knjižna in razločna izreka. V 2. VIO razvijanje pravorečne zmožnosti sledi konceptu za 1. VIO, ob tem pa še učenci v besedah prepoznajo knjižne samoglasnike in utrjujejo njihovo izreko v besedah in povedih. Nadgradnja, ki jo prinaša UN 2018, je, da učence ob morebitnih nejasnostih usmerja v kritično in premišljeno rabo ustreznih jezikovnih priročnikov v knjižni in elektronski obliki. V 3. VIO UN ciljem 1. in 2. VIO dodaja še vadenje in utrjevanje knjižne izreke ob besedilih. Tako učenci glasno izgovarjajo (tudi berejo) besede in opisujejo njihovo izreko, vendar pri tem ne uporabljajo naglasnih znamenj in strokovnih izrazov za zvočniške variante (kot v UN 2011). V UN je posebej izpostavljena primerjava slovenskih glasovnih in besednih prozodičnih prvin s tujimi; učenci svoje ugotovitve tudi predstavijo.

V UN za gimnazije in KZ bi morebiti veljalo razmisliti o bolj dosledni kontinuiteti razvijanja pravorečne zmožnosti, saj je premalo poudarjena, zato je skrb zanjo in doslednost pri njenem razvijanju prepuščena predvsem presoji učitelja, upoštevajoč

dejstvo, da to ni zgolj ena temeljnih nalog učiteljev slovenščine, pač pa učiteljev vseh predmetnih področij.

V UN za gimnazije dijaki razvijajo pravorečno zmožnost ob različnih dejavnostih poslušanja knjižne izreke in z njeno uzaveščeno rabo, vendar pa je ta eksplicitno zapisana le v 1. letniku, kjer sistematično spoznavajo, vadijo in utrjujejo knjižni izgovor posameznih besed in povedi, prepoznavajo pravorečne napake v lastnih in tujih besedilih ter jih odpravljajo, med govornim nastopanjem pa skušajo govoriti čim bolj knjižno, v vseh drugih letnikih pa je skrb zanjo ter kontinuiteta pri njenem razvijanju in dosledni ustrežni rabi prepuščena predvsem presoji učitelja. Ista problematika razvijanja kontinuitete razvijanja pravorečne zmožnosti je prisotna tudi v obeh katalogih znanj.

Literatura

- Marja BEŠTER TURK, 2011: Sporazumevalna zmožnost – eden izmed temeljnih ciljev pouka slovenščine. *Jezik in slovstvo* 56/3, 122–127.
- Lucija ČOK, Anita BRČANINOVIC, 2015: Z igro glasov v angleškem jeziku do razvijanja fonološkega zavedanja. *Revija za elementarno izobraževanje* 8/1–2, 147–166.
- Mira KRAJNC IVIČ, Melita ZEMLJAK JONTES, 2022: Pristopi k celostni obravnavi besedil/Approaches to the holistic treatment of texts. *Pogled na šolo 21. stoletja v duhu kompetenc in pismenosti*. Ur. Tatjana Krapše, et al. Spletna izd. Ljubljana: Zavod RS za šolstvo. 174–204. http://www.zrss.si/pdf/Pogled_na_solo_21_stoletja.pdf (20. 8. 2023).
- Boža KRAKAR VOGEL, 2004: *Pogled na didaktike književnosti*. Ljubljana: DZS.
- Olga KUNST GNAMUŠ, 1991: *Sporazumevanje med željo, resnico in učinkom*. Ljubljana: Slovensko društvo raziskovalcev šolskega polja.
- Olga KUNST GNAMUŠ, 1992: *Sporazumevanje in spoznavanje jezika*. Ljubljana: Državna založba Slovenije.
- Jože LIPNIK, 1994: *Mali čvek. Zbirka besedil b knjigi Metodika govorne vzgoje*. Maribor: Založba Obzorja Maribor.
- Simona PULKO, 2007: *Sporočanje v osnovni šoli*. Maribor: Dravska tiskarna.
- Alenka VALH LOPERT, 2013: *Med knjižnim in neknjižnim na radijskih valovih v Mariboru*. Maribor: Litera.
- Renata ZADRAVEC PEŠEC, 1994: *Pragmatično jezikoslovje. Temeljni pojmi*. Ljubljana: Pedagoški inštitut, Center za diskurzivne študije.
- Melita ZEMLJAK JONTES, 2017: Čebļavčki, brbljavčki, čvekavčki in Mali čvek Jožeta Lipnika. *Med didaktiko slovenskega jezika in poezijo. Ob 80-letnici Jožeta Lipnika*. Ur. Marko Jesenšek. Maribor: Univerzitetna založba Univerze v Mariboru. 42–53.
- Melita ZEMLJAK JONTES, Barbara BEDNJIČKI ROŠER, 2020: Glasovno zavedanje: 4. gradnik. *Gradniki bralne pismenosti: teoretična izhodišča*. Ur. Dragica Haramija. 1. izd. Maribor: Univerzitetna založba Univerze: Pedagoška fakulteta; Ljubljana: Zavod Republike Slovenije za šolstvo. 107–136. <https://press.um.si/index.php/ump/catalog/view/515/631/1148-2> (20. 8. 2023).
- Tanja MIRTIČ, Marko SNOJ (ur.), 2021: *1. slovenski pravorečni posvet*. Ljubljana: Slovenska akademija znanosti in umetnosti (Razprave/Slovenska akademija znanosti in umetnosti, Razred za

- filološke in literarne vede = Dissertationes/Academia scientiarum et artium Slovenica, Classis II: Philologia et litterae; 25).
<https://www.sazu.si/uploads/files/publikacije21/Rared2RAZPRAVE.pdf> (20. 8. 2023).
- Katalog znanja slovenščina (SSI)*. [http://portal.mss.edus.si/msswww/programi2015/programi/SSI/KZ-
IK/katalog.htm](http://portal.mss.edus.si/msswww/programi2015/programi/SSI/KZ-
IK/katalog.htm) (20. 8. 2023).
- Katalog znanja slovenščina (SSI)*. [http://eportal.mss.edus.si/msswww/programi2019/programi/Ssi/KZ-
IK/katalog.htm](http://eportal.mss.edus.si/msswww/programi2019/programi/Ssi/KZ-
IK/katalog.htm) (20. 8. 2023).
- Mojca POZNANOVIČ JEZERŠEK idr., 2008: *Učni načrt. Slovenščina [Elektronski vir]: gimnazija: splošna, klasična, strokovna gimnazija*. Obvezni predmet in matura. Ljubljana: Ministrstvo za šolstvo in šport, Zavod RS za šolstvo.
[http://eportal.mss.edus.si/msswww/programi2018/programi/media/pdf/un_gimnazija/un_
slovenscina_gimn.pdf](http://eportal.mss.edus.si/msswww/programi2018/programi/media/pdf/un_gimnazija/un_
slovenscina_gimn.pdf) (20. 8. 2023).
- Mojca POZNANOVIČ JEZERŠEK idr., 2011: *Učni načrt. Slovenščina*. Ljubljana: Ministrstvo za šolstvo in šport, Zavod RS za šolstvo.
[http://www.mizs.gov.si/fileadmin/mizs.gov.si/pageuploads/podrocje/os/prenovljeni_UN/
UN_slovenscina_OS.pdf](http://www.mizs.gov.si/fileadmin/mizs.gov.si/pageuploads/podrocje/os/prenovljeni_UN/
UN_slovenscina_OS.pdf) (20. 8. 2023).
- Mojca POZNANOVIČ JEZERŠEK idr., 2018: *Učni načrt. Slovenščina*. Ljubljana: Ministrstvo za šolstvo in šport, Zavod RS za šolstvo.
[http://www.mizs.gov.si/fileadmin/mizs.gov.si/pageuploads/podrocje/os/prenovljeni_UN/
UN_slovenscina.pdf](http://www.mizs.gov.si/fileadmin/mizs.gov.si/pageuploads/podrocje/os/prenovljeni_UN/
UN_slovenscina.pdf) (20. 8. 2023).

SPLETNA APLIKACIJA SVOJILNI PRIDEVNIKI IZ PREVZETIH PRIIMKOV (SPiPP) KOT VIR ZA RAZISKAVE GOVORA

IRENA STRAMLJIČ BREZNIK

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
irena.stramljic@um.si

V prispevku predstavljamo, zakaj je tvorba svojilnih pridevnikov iz prevzetih osebnih lastnih imen eno težjih vprašanj v slovenščini kljub številnim pravopisnim napotilom v aktualnih in novonastajajočih pravopisnih priročnikih. Kot dobro alternativo predstavljamo spletni pripomoček SPiPP, ki pri tovrstnih težavah z zapisom, izgovarjavo in strnjenimi pravili na enem mestu pomaga jezikovnemu uporabniku. Hkrati aplikacijo SPiPP predstavljamo kot jezikovni vir za nadaljnje razvijanje govornih pripomočkov, saj iz njega 60 naključno izbranih francoskih priimkov preizkusimo v sintetizatorju slovenskega govora eBralcu. Namen preizkusa je spodbuditi, da bi ob trenutno dostopnih zvočnih posnetkih in pravilih o razmerjih med črkami in glasovi tujih jezikov in slovenskega jezika postopoma nadgradili eBralca tako, da bi bil specializiran za izgovarjavo tudi tujejezičnih lastnoimenskih besed in njihovih oblik, prilagojenih slovenščini.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.16](https://doi.org/10.18690/um.ff.4.2024.16)

ISBN
978-961-286-882-6

Ključne besede:
slovensko besedotvorje,
svojilni pridevniki,
tuja osebna lastna imena,
aplikacija SPiPP,
sintetizator govora eBralce



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.16](https://doi.org/10.18690/um.ff.4.2024.16)

ISBN
978-961-286-882-6

Keywords:

Slovene word formation,
possessive adjectives,
foreign personal proper
names,
SPiPP application,
eBralec speech synthesizer

WEB APPLICATION POSSESSIVE ADJECTIVES FROM ADOPTED PERSONAL PROPER NAMES (SPiPP) AS A RESOURCE FOR SPEECH RESEARCH

IRENA STRAMLJIČ BREZNIK

University of Maribor, Faculty of Arts, Maribor, Slovenia
irena.stramljic@um.si

This article discusses why the formation of possessive adjectives from adopted personal proper names is one of the most difficult problems in the Slovene language despite numerous orthographic guidelines in current and emerging orthographic manuals. As an alternative, we present the online tool SPiPP, which offers help in the formation of such adjectives by providing spelling, pronunciation and rules in one place. At the same time, the SPiPP application can also be used as a language resource for the further development of speech aids. To illustrate this usefulness, 60 randomly selected French surnames from the application were tested with the Slovene speech synthesiser eBralec. The aim of the test was to use the currently available audio recordings and the rules on the relationships between letters and sounds of foreign languages and the Slovene language to stimulate a gradual further development of eBralec, so that it would include the pronunciation of foreign proper nouns and their forms adapted to Slovene.



1 Uvod¹

Tvorba svojilnih pridevnikov iz tujih lastnih imen spada med težja pravopisna vprašanja in posega v sporazumevalne kompetence vseh uporabnikov jezika. Zahteva namreč hkratno poznavanje pravil slovenskega glasoslovja, oblikoslovja in besedotvorja. Za pravilno tvorbo svojilnega pridevnika iz tujega oziroma prevzetega priimka pa je treba poznati tudi njegov izgovor, ki se neredko zelo razlikuje od zapisa. Prevzemanje tujih lastnih imen v slovenščino je zaradi tujih in naših pravorečno-pravopisnih pravil, pregibanja in dojemanja razmerja med črko in glasom zelo zahtevno (Kocjan - Barle ³1999, 2012: 85).

2 Zakaj je tvorba svojilnih pridevnikov iz prevzetih lastnih imen težka?

2.1 Razhajanje med pisno in izgovorno podobo prevzetih lastnih imen

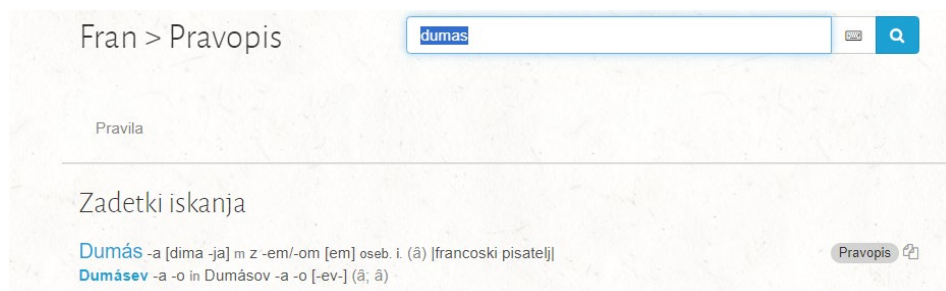
Za pravilno tvorbo svojine iz prevzetih tujih osebnih lastnih imen je treba, kot že omenjeno, poznati pravila slovenskega glasoslovja, oblikoslovja, besedotvorja in njihov izgovor. Dodatna težava pa so v slovenščini tudi dovoljene pravopisne dvojnice, ki so pri določeni skupini imen lahko samo pisne, ne pa tudi izgovorne. Ponazoritev: svojino lahko izrazimo z rodilnikom (roman *Dumasa*, izgovor [dimájaj]), še pogosteje pa s svojilnim pridevnikom, ki ima pisno dvojnico (*Dumasov* oz. *Dumasev*), izgovor pa le [dimájəu]. Izgovorno je tako normativna le oblika [dimájəu], ki jo nakazuje pisna dvojnica *Dumasev*. Prav tako dovoljena oblika *Dumasov* sledi le pisni podobi, zato je dovoljena le v zapisanih besedilih, ne pa kot izgovorna različica. Tako imenovano branje po črki oziroma upoštevanje pisne podobe v izgovoru po načelu »govôri kot piše«, torej izgovor [*dumasov] oz. [*dumasev], pa razodeva nekultiviranega in nerazgledanega govorca slovenskega jezika (Stramljič Breznik 2020: 96–117).

2.2 Osebna lastna imena v slovarju in pravilih Slovenskega pravopisa 2001

Slovenski pravopis 2001 (SP 2001) v slovarskem delu obsega 3248 iztočnic z oznako osebno ime, kar vključuje tako domača kot tuja imena in/ali priimke.

¹Prispevek je nastal v okviru raziskovalnega programa št. P6-0156 (Slovensko jezikoslovje, književnost in poučevanje slovenščine (2020–2025), vodja prof. dr. Marko Jesenšek), financer ARIS.

Iztočnica ima označeno roditeljsko končnico, v oglatem oklepaju izgovor, oznako spola, zapis težje končnice v orodniku z izgovorom, pravopisno kategorijo, ki ji sledi pomenska razlaga. Kot podiztočnica sledi svojilni pridevnik ali njegova enakovredna dvojica za vse tri spole z izgovorom in tonemom.



Slika 1: Iztočnica za francoski priimek *Dumas*

Vir: <https://fran.si>

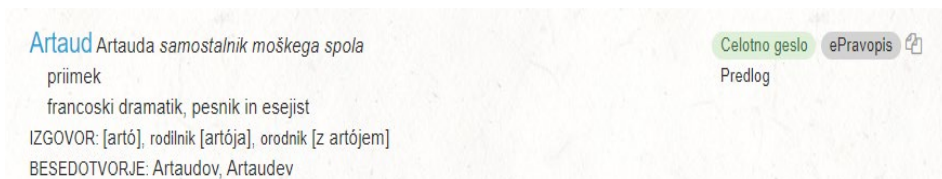
Pravopisna pravila SP 2001 prevzeta osebna lastna imena obravnavajo na različnih mestih v več poglavjih: Raba velikih in malih črk, Prevzete besede in besedne zveze, Glasoslovje (neobstojni samoglasniki), Oblikoslovje (samostalniške sklanjatve, večbesedna poimenovanja), Besedotvorje (podstave in premene v podstavi tvorjenk), Težji primeri iz besedotvorja (obrazila *-ov/-ev/-in*) (Stramljič Breznik 2020: 96–117).²

Dodatni podatki za večšega uporabnika pravopisa pa so izčrpne (61 strani, 66 paragrafov (1071–1134)) preglednice za okrog 50 tujih pisav, razdeljene v latinične in nelatinične. Znotraj vsake pisave so najprej zapisane črke tujega jezika, ki jim v drugem stolpcu sledijo črke ali črkja za slovenski fonemski izgovor in zapis, kadar je beseda pisno podomačena, v tretjem stolpcu so zgledi z dodatnimi komentarji.

² V prispevku so tudi izpisani vsi ustrezni paragrafi iz SP 2001.

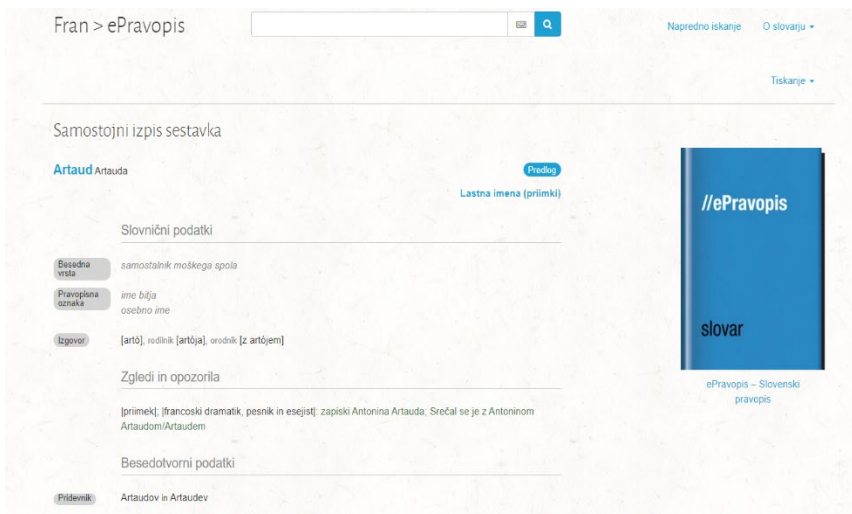
2.3 Osebna lastna imena v slovarju ePravopisa³

Geslovnik lastnih imen trenutno obsega 520 domačih in tujih priimkov. Pri tujih priimkih je ePravopis v primerjavi s SP 2001 bolj informativen in uporabniku ponuja dve možnosti: osnovne ali izčrpne podatke za geslo. Pri prvih so zajeti: zapis iztočnice, roditeljska oblika, besedna vrsta, pravopisna kategorija, pomenska razlaga, izgovor in besedotvorni podatki za tvorbo svojilnega pridevnika, kot kaže spodnji primer.



Slika 2: Osnovni prikaz iztočnice za francoski priimek *Artaud*

Vir: <https://fran.si/>



Slika 3: Celotno geslo iztočnice za francoski priimek *Artaud*

Vir: <https://fran.si/>

³ ePravopis je novonastajajoči pravopisni slovar, ki je gradivska razširitev problemsko zastavljenih pravopisnih pravil.

Klik na Celotno geslo za uporabnika sistematično in padajoče navaja posamezne slovnične podatke (besedna vrsta, pravopisna oznaka, izgovor). Sledijo zgledi in opozorila. Besedotvorni podatki predstavijo svojilni pridevnik ali njegovi morebitni dvojníci. Četrti sklop podatkov so oblikoslovni podatki, tj. sklonske oblike od imenovalnika do orodnika z možnimi dvojnicami za vsa tri števila.

V Pravopisu 8.0, ki po poglavjih vključuje Pravila novega slovenskega pravopisa za javno razpravo, najdemo v 4. poglavju z naslovom *Prevzete besede in besedne zveze* tudi pravila za pravopisno kategorijo *Lastna imena*. Znotraj nje uporabnik dobi zbrana temeljna napotila glede prevzemanja imen bitij: podomačevanje in nepodomačevanje, prečrkovanja lastnih imen ter splošna načela za prevzemanje besed (prilagoditve pisave, glasov in naglasa). Pri prikazu glasovnih in naglasnih prilagoditev so ob vsakem navedenem ponazoritvenem tujem priimku v oglatem oklepaju navedeni tudi izgovori.

Posebej izčrpno glede tega je 8. poglavje *Prevzemanje iz posameznih jezikov z latinično pisavo* (albansščina, estonščina, finščina, francoščina, italijansščina, madžarščina, poljščina, slovaščina, španščina, turščina).

Za vsakega od naštetih jezikov so navedeni podatki o njegovih črkah abecede, ločevalnih znamenjih, izgovoru s podatki o naglasnem mestu, razmerju med črkami in glasovi z vsemi morebitnimi posebnostmi. Dodana je Preglednica zapis – izgovor v slovenščini, ki vključuje zapis tuje črke, položaj, izgovor v slovenščini in zglede konkretnih primerov z izgovori. Sledi še poglavje o podomačevanju, posebej tudi za osebna imena znotraj skupine lastnih.

3 Pisno-glasovni pripomoček za tvorbo svojilnih pridevnikov iz prevzetih osebnih lastnih imen (SPiPP)

3.1 Med normo in jezikovno realnostjo

(Pravo)pisnih napotil za ustrezno tvorbo svojilnih pridevnikov iz tujih priimkov je torej veliko, a to zahteva ne samo zainteresiranega uporabnika, ki se je pripravljeno poglobiti v jezikovni problem, na katerega je naletel, ampak tudi usposobljenega uporabnika, ki v osnovi pozna ustrezna slovnična pravila in hkrati tudi vsa mesta v obstoječih jezikovnih virih, v katerih lahko poišče odgovore na svoja vprašanja.

Glede pričakovanj o ustreznem poznavanju tovrstnih jezikovnih pravil, pridobljenih na stopnji srednješolskega izobraževanja, je bil opravljen pregled aktualnih učbeniških gradiv za 2. in 4. letnik gimnazij in srednjih strokovnih šol (Ana Rotovnik Omerzu 2023), ki je pokazal, da dijaki v skladu z učnim načrtom pri jezikovnem pouku sicer razvijajo oblikoslovno in tvorbeno zmožnost v okviru poglavja o besednih vrstah in besedotvorju, a ne na primerih tujih priimkov.

Tudi rezultati ankete (Stramljič Breznik 2021: 37–56, Nadelsberger 2018)⁴ so pokazali vso zagatnost tvorbe svojilnih pridevnikov iz tujih lastnih imen, pri katerih pravila maternega jezika prenašamo na tujejezična imena, za katera je značilen razkorak med pisno in izgovorno podobo. Če uporabnik ne pozna izgovora, je velika verjetnost, da bo tvoril neustreznega svojilni pridevnik oz. lahko sledi pravilom glede na zapis, kar je lahko zavajajoče glede na pravila izgovora.

3.2 SPiPP – aplikacija, ki z zapisom, izgovorom in pravili uporabniku pomaga pri tvorbi svojilnih pridevnikov iz prevzetih priimkov

Sodobni načini komuniciranja in posredovanja vsebin tudi od jezikoslovcev zahtevajo, da so njihovi jezikovni odgovori hitri, ciljno usmerjeni in praktično na dlani⁵ za vse tiste jezikovne uporabnike, ki jim je mar jezikovna kultura.

Kot most med dvema bregovima izgovorne in pisne podobe prevzetih priimkov in njihovega ustreznega vključevanja v besedotvorni sistem slovenščine je bila izdelana pisno-glasovna spletna aplikacija SPiPP (Stramljič Breznik 2021), s katero lahko jezikoslovci v dobi vsesplošne digitalizacije pomagamo jezikovnim uporabnikom.

Nastali spletni pripomoček vsebuje 1353 priimkov, katerih nosilci so tuje osebe moškega spola iz Velikega splošnega leksikona DZS (2006). Uporabniku so na voljo informacije v pisni in zvočni obliki. Zapisani so: osnovna (imenovalniška) oblika priimka in njen fonetični zapis, poreklo in glavni podatki o nosilcu, obliki za rodilnik in izražanje svojine ter šifra pravila, po katerem je svojilna oblika tvorjena. Osnovne,

⁴ Anketo je izpolnilo 60 anketirancev, od tega 30 študentov slovenščine in 30 oseb, ki niso bili študentje. V obeh skupinah so bile v opazno večjem številu zastopane ženske. Med študenti slovenščine je bil vključen izenačen delež tistih z univerzitetno izobrazbo in tistih, ki so še študirali in imeli zaključeno srednjo šolo (43,3 %). Nekaj študentov (13,4 %) je zaključilo bolonjski magistririj. Po izobrazbeni strukturi so med neštudenti prevladovali tisti, ki so pridobili izobrazbo na višji ali visoki šoli (36,7 %), z univerzitetno izobrazbo je bilo 20 %, z doktoratom (3,3 %), z bolonjskim magistririjem 16,7 %, s končano srednjo 13,3 % in osnovno šolo 10 %.

⁵ Primer dobre prakse je Jezikovna svetovalnica: <https://svetovalnica.zrc-sazu.si/>.

rodilniške in svojilne oblike priimkov so opremljene še z zvočnimi posnetki. Brez dvojnic je zajetih 4059 zapisanih oblik in prav toliko zvočnih posnetkov, z upoštevanimi dvojnicami pa še nekaj sto več.

The screenshot shows the SPiPP web application interface. At the top, the title reads "SPiPP – svojilni pridevniki iz prevzetih priimkov" and the subtitle is "Spletna aplikacija za pomoč uporabnikom pri izražanju svojine za prevzete moške priimke". Below the title is a navigation menu with options: O PROJEKTU, ISKANJE, VIRI, PRAVILA ZA TVORBO SVOJINE, OSEBNA LASTNA IMENA, LITERATURA IN VIRI, ŠTUDIJE, POMOČ, PREDSTAVITVE, and SKRBIŠČIŠTVO. A search bar contains the text "Batteux". Below the search bar is a table with the following data:

Priimek	Poreklo	Izgovor	Rodilnik	Izražanje svojine	Pravilo
Batteux	francoski filozof umetnosti (1713–1780)	[baté] 🔊	Batteuxa 🔊	Batteuxev/Batteuxov 🔊	2c

At the bottom of the interface, there are logos for the Javni inštitut za slovenski jezik (Javni inštitut za slovenski jezik, inštitut za jezikovno in govorno kulturo Republike Slovenije), the Ministry of Education, Science and Sports of the Republic of Slovenia, and the European Union (Evropska unija, Evropski socialni sklad, NALOŽBA V VAŠO PRIHODNOST).

Slika 4: Primer pisno-zvočnih podatkov za francoski priimek *Batteux*

Vir: <https://spipp.ff.um.si/>

Jezikovno orodje je namenjeno širši javnosti, predvsem jezikovnim uporabnikom, ki se želijo pisno in govorno kultivirano izražati. S širitvijo tovrstnih jezikovnih kompetenc materinščine smo naredili korak, ki prispeva k dvigu ravni funkcionalne pismenosti vseh tipov govorcev, da bi lahko postali samozavestn(ejš)i uporabniki slovenščine, kakor predvideva Resolucija o nacionalnem programu za jezikovno politiko 2019–2025.⁶

Gre za vsestransko uporaben spletni pripomoček, v katerem sta smiselno povezana jezikoslovno in programersko znanje. Namenjen je jezikovnim uporabnikom (lektorjem, prevajalcem, radijskim in televizijskim moderatorjem, študentom, profesorjem, strokovnjakom različnih področij in vsem, ki se pri pisanju na svojih strokovnih in/ali znanstvenih področjih srečujejo s tujimi priimki in njihovimi oblikami. Njegova uporaba je primerna za vse jezikovne uporabnike, ki se želijo pisno in govorno ustrezno izražati in slediti pravilom slovenskega knjižnega jezika.

⁶ Vir: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=RESO123>.

Velja omeniti, da je bila skupna uspešna javna predstavitev rezultatov projekta SPiPP in Govornega pomočnika RTV Slovenija v februarju 2020 v prostorih Ministrstva za kulturo.

Tudi Govorni pomočnik RTV Slovenija je prosto dostopna zbirka poslovenjenih in poenoteni izgovorjav tujih imen ter drugih besed in besednih zvez z označenim mestom naglasa in naglasnim znamenjem. V prvi vrsti je namenjena poklicnim govorcem javnega medija, v pomoč pa je lahko tudi vsem drugim, ki javno nastopajo ali jih zanimajo pravila zborne izreke.⁷

4 Empirični del – primerjava izgovorov francoskih priimkov in svojilnih pridevnikov v SPiPP-u in eBralcu

4.1 Namen primerjave in njene omejitve

Namen raziskave je primerjati izgovorjavo 60 naključno izbranih francoskih priimkov in njihovih svojilnih pridevnikov iz specializiranega spletnega pripomočka SPiPP ter njihov izgovor preizkusiti v aplikaciji za sintezo slovenskega govora eBralec z govorcem Renatom. Namenoma smo se odločili za izbor francoskih priimkov, saj je znano in tudi v številnih raziskavah potrjeno (npr. Bregant 2021), da povzročajo največ težav tako pri izgovorjavi kot pri vključevanju v slovenski oblikoslovni in besedotvorni sistem.

Pričakovano je, da bo eBralec kot orodje, ki je specializirano za naravno in tekočo izgovorjavo slovenskega jezika – in to predstavlja tudi omejitev tega preizkusa –, pri tem napravil več napak. Vendar z eksperimentom ni namen kazati na slabosti eBralca, marveč spodbuditi nadaljnji razmislek, da bi ob trenutno dostopnih zvočnih posnetkih in pravilih o razmerjih med črkami in glasovi tujega in slovenskega jezika postopoma razvili tudi eBralca, ki bi bil specializiran za izgovorjavo še drugih tujejezičnih lastnoimenskih besed v slovenščini.

⁷ Spletna pripomočka Govorni pomočnik (<https://govornipomocnik.rtvlo.si/>) in SPiPP (<https://spipp.ff.um.si/>) vsak na svoj način potrjujeta stebno vlogo javnega medija pri negovanju slovenskega zbornega jezika in govorne kulture. To vlogo navajajo vsi pravni, strateški in etični dokumenti RTV Slovenija, od programskih standardov do poklicnih meril. Neusklajenost v uporabi in napake v izgovoru posameznih besed begajo občinstvo, ki od javnega medija pričakuje, da bo za zgled drugim novinarskim hišam, a tudi v oporo šolam, različnim javnim ustanovam in javnim govorcem. Vir: <https://www.rtvlo.si/varuh/aktualno/govorni-pomocnik-rtv-slovenija/513661>.

4.2 O eBralcu

Sintetizator govora eBralec (Žganec Gros, Vesnicer, Rozman, Holozan, Šef 2016: 180–185) je zelo uporaben jezikovnotehnoški pripomoček za slovenski govorniki jezik, ki sta ga razvili podjetji Alpineon in Amebis, v povezavi z Inštitutom Jožef Stefan ter Fakulteto za elektrotehniko Univerze v Ljubljani. eBralec pretvori slovensko besedilo v govor in je lahko v pomoč tako slepim in slabovidnim ter osebam z motnjo branja kot tudi tistim, ki nimajo težav z vidom, saj omogoča poslušanje besedila kjerkoli in kadarkoli. S tem poenostavi in poceni poslovni proces, olajša komunikacijo, slepim in slabovidnim ter starejšim pa omogoča dostop do besedilnih vsebin.⁸ Podatke o fonemskem prepisu besed črpa iz obsežne jezikovne baze ASEs, v primeru neznanih besed pa uporabi pravila za samodejni grafemsko-fonetični prepis (Žganec Gros, Mihelič, Pavešič 1999: 241–247).

4.3 Izsledki eksperimenta

Tabela 1: Primerjava izgovorov v SPiPP-u in eBralcu⁹

Francoski priimek	Izgovor SPiPP	Izgovor eBralec	Svojljni pridevnik SPiPP	Izgovor SPiPP	Izgovor eBralec
Bastien-Lepage	[bastjén ləpáz-]	*[béstin lepáge]	Bastien-Lepageev in (delo) Bastiena-Lepagea	[bastjén ləpázɛɥ] [bastjéna ləpáza]	*[béstin lepágeɥ] *[bajstína lepága]
Beauchamps	[bošán]	*[beaúchamps]	Beauchampsov	[bošánoɥ]	*[beaúchámsoɥ]
Balthus	[baltís]	*[bálthus]	Balthusov	[baltísou]	*[balthúsoɥ]
Barbey d'Aurevilly	[barbé dorvijí]	*[bárbidápostrof aarevillipsílon]	(delo) Barbeyja d'Aurevillyja	[barbéjadorvijíja]	*[barbíjadápostrofaaurevillipsílonja]
Barbusse	[barbís]	✓	Barbussov	[barbísou]	✓
Barthes	[bárt]	*[bárth]	Barthesov	[bártou]	✓
Basin	[bazén]	*[bazín]	Basinov	[bazénoɥ]	*[bazínou]
Bastiat	[bastjá]	✓	Bastiatev in Bastiatov	[bastjájeɥ]	✓
Batteux	[baté]	✓	Batteuxev in Batteuxov	[batéjeɥ]	✓
Bauchant	[bošán]	✓	Bauchantov	[bošánoɥ]	✓
Baudelaire	[bodlêr]	*[bodlêl]	Baudelairjev in	[bodlêrjeɥ] [bodlêrou]	✓

⁸ Vir: <https://ebralce.si/>

⁹ Pri preverjanju in zapisovanju izgovorov v eBralcu so sodelovale študentke 3. letnika smeri Slovenski jezik in književnost Lara Domej, Rebeka Krasnič in Nina Vinšek v okviru oddelčnega vključevanja študentk in študentov v znanstvenoraziskovalno in projektno delo, mentorica red. prof. dr. Irena Stramlič Breznik.

Francoski priimek	Izgovor SPiPP	Izgovor eBralec	Svojilni pridevnik SPiPP	Izgovor SPiPP	Izgovor eBralec
			Baudelairov		
Bayle	[bél]	✓	Baylov	[bélo]	✓
Bazaine	[bazèn]	*[bazèn]	Bazainov	[bazèno]	*[bazèno]
Beauharnais	[boarné]	*[beúharnáris]	Beauharnaisev in Beauharnaisov	[boarnéje]	*[beúharnáise]
Beaumanoir	[bomanoár]	*[béaumanóir]	Beaumanoirjev	[bomanoárje]	*[béaumanóirje]
Beaumarchais	[bomaršé]	✓	Beaumarchaisev in Beaumarchaisov	[bomaršéje]	✓
Bécaud	[bekó]	*[bécaud]	Bécaudev in Bécaudov	[bekóje]	*[bécaude]
Beineix	[beniks]	✓	Beineixov	[benikso]	✓
Béjart	[bezár]	*[béjart]	Béjartev in Béjartov (delo)	[bezárje]	*[béjarte]
Bernard iz Clairvauxa	[bernár is klervója]	✓	Bernarda iz Clairvauxa	[bernárda is klervója]	✓
Bernart de Ventadorn	[bernár də vantadórn]	*[bernár de ventadórn]	(delo) Bernarta de Ventadorna	[bernára de vantadórna]	*[bernárta de ventadórna]
Bidault	[bidó]	✓	Bidaultev in Bidaultov	[bidóje]	✓
Biot	[bjó]	*[bijot]	Biotev in Biotov	[bjóje]	*[bijoto]
Bissière	[bisjèr]	*[bisjé]	Bissièrejev in Bissièreov	[bisjèrje] [bisjèro]	[bisjèrje] in *[bisjéjo]
Blanchard	[blanšár]	✓	Blanchardev in Blanchardov	[blanšárje]	*[blanšáro]
Boffrand	[bofrán]	✓	Boffrandov	[bofráno]	✓
Boieldieu	[boaldjé]	*[bojeldjé]	Boieldieujev	[boaldjéje]	*[bojeldjéje]
Boileau- Despréaux	[boaló depreó]	✓	Boileau- Despréauxev in Boileau- Despréauxov in (delo) Boileauja- Despréauxa	[boaló depreóje] [boalója depreója]	✓ ✓
Boilly	[boaji]	*[bójli]	Boillyjev	[boajije]	*[boilje]
Boisbaudran	[boabodrán]	*[boisbaúdran]	Boisbaudranov	[boabodráno]	*[boisbaudráno]
Boissieu	[boasjé]	✓	Boissieujev	[boasjéje]	✓
Bontemps	[bontán]	✓	Bontempsov	[bontáno]	✓
Boucher	[bušé]	✓	Boucherev in Bouchеров	[bušéje]	✓
Bougainville	[bugenvíl]	✓	Bougainvillov	[bugenvílo]	✓
Boulangier	[bulonžé]	*[bulanžé]	Boulangerev in Boulangierov	[bulonžéje]	*[bulanžéje]
Bouquet	[buké]	✓	Bouquetev in Bouquetov	[bukéje]	✓
Bourdaloue	[burdalú]	*[bordaloúe]	Bourdalouejev	[burdalúje]	*[bordaluéje]
Bourdieu	[burdjé]	✓	Bourdieujev	[burdjéje]	✓

Francoski priimek	Izgovor SPIPP	Izgovor eBralec	Svojljni pridevnik SPiPP	Izgovor SPiPP	Izgovor eBralec
Bourget	[buržé]	*[boúrget]	Bourgetev in Bourgetov	[buržéjeu]	*[bourgétou]
Boutroux	[butrú]	✓	Boutrouxev in Boutrouxov	[butrújeu]	✓
Cabet	[kabé]	*[kabèt]	Cabetev in Cabetov	[kabejeu]	✓
Cahusac	[kaj(i)zák]	*[cahusác]	Cahusacov	[kaj(i)zákou]	*[cahusácou]
Caillaux	[kajó]	✓	Caillauxev in Caillauxov	[kajójeu]	✓
Daguerre	[dagèr]	*[dagèl]	Daguerrjev in Daguerrov	[dagèrjeu] [dagèrou]	*[dágvuvèrverjeu] ✓
Fabius	[fabijis]	*[fabís]	Fabiusov	[fabijisou]	*[fabísou]
Gabin	[gabèn]	✓	Gabinov	[gabènou]	✓
Jacob	[žakób-]	✓	Jacobov	[žakóbou]	✓
Jacque	[žák]	*[jacqué]	Jacquov	[žákou]	*[jacquéou]
Jammes	[žám]	*[jámés]	Jammesov	[žámou]	*[jamésou]
Lachelier	[lašeljé]	*[lachelier]	Lachelierev in Lachelierov	[lašeljéjeu] [lašeljéro]	*[lachelierou]
Mably	[mablí]	*[mábli]	Mablyjev	[mablíjeu]	*[máblíjeu]
Pagnol	[panjól]	✓	Pagnolov	[panjólou]	✓
Queneau	[kenó]	✓	Queneaujev	[kenójeu]	✓
Rabelais	[rablé]	✓	Rabelaisev in Rabelaisov	[rabléjeu]	✓
Rouget de Lisle	[ružé də líl]	*[roužet de líse]	(delo) Rougeta de Lísle	[ružéja də líla]	*[roužéta de lísla]
Saint Laurent	*[sènt lorán]	*[sèntlorán]	Saint Laurentov	[senloránou]	*[sent loránou]
Saint-Pierre	[sempjèr] ¹⁰	*[sènt pjèr]	Saint-Pierrjev Saint-Pierrov	[sempjèrjeu] [sempjéro]	*[sènt piorjeu] *[sent pjèrou]
Taine	[tèn]	*[tèn]	Tainov	[tènou]	*[tènou]
Vailland	[vaján]	✓	Vaillandov	[vajánou]	✓
Yersin	[jersèn]	*[jersén]	Yersinov	[jersènou]	*[jersènou]

Vir: lasten

Izmed šestdeset naključno izbranih francoskih priimkov iz pripomočka SPiPP je eBralec enako kot v aplikaciji prebral 25 osnovnih oblik priimkov in njihovih svojilnih pridevnikov, kar znaša slabih dvainštirideset odstotkov (41,7 %). Pri 6 priimkih (10 %) je enako prebral ali osnovno (1 priimek) ali svojilno obliko oz. eno od različic (5 priimkov), medtem ko je bil pri preostalih priimkih (48,3 %) neuspešen.

¹⁰ Glede na smernice prevzemanja iz francoščine, predstavljene v Pravopisu 8.0, bi bil v primeru francoskega lastnega imena, zapisanega z vezajem, ustrezen izgovor [sèn-pjèr] in ne [sempjèr], kar je morda v SPiPP-u tipkarska napaka. SP 2010 v primerih francoskih lastnih imen s sestavino "Saint" v zapisu izgovora ne predvideva naglase, npr. Saint-Germain [sen-žermén], čemur smo sledili v SPIPP-u. Medtem ko so podobni primeri v ePravopisu, kot se konkretno vidi za predhodni primer Saint Laurent, vselej dvonaglasni in zapisani dvobesedno [sèn lorán].

Zanimalo nas je, za kakšne vrste priimkov gre v primerih, ko sta bila tako osnovna oblika kot njegov svojilni pridevnik ustrezno izgovorjena. Predpostavljali smo, da gre bodisi za bolj znane in zato v rabi pogostejše priimke bodisi za priimke, za katere najdemo v jezikovnih priročnikih zapise in izgovarjavo. Najprej smo poiskali njihovo frekvenco v Gigafida 2.0¹¹ in nato morebitno vključenost v katerega od e-virov portala Fran.si.

Tabela 2: Pogostnost z eBralcem ustrezno prebranih priimkov in njihove svojine v korpusu Gigafida 2.0 ter zapis izgovora v jezikovnih virih portala Fran.si

Francoski priimek	Pogostnost v Gigafidi 2.0	Zapis izgovora na Fran.si
Barbusse	9	Pravopis
Bastiat	5	/
Batteux	5	/
Bauchant	1	/
Bayle*	25	/
Beaumarchais*	141	Pravopis
Beineix	23	/
Bernard iz Clairvauxa	44	/
Bidault	9	/
Boffrand	/	/
Boileau-Despréaux	/	/
Boissieu	12	/
Bontemps	17	/
Boucher	339	/
Bougainville	43	Pravopis
Bouquet	146	/
Bourdieu	310	/
Boutroux	7	/
Caillaux	6	/
Gabin	100	Pravopis
Jacob	3052	/
Pagnol	51	/
Queneau	211	Pravopis
Rabelais	253	Pravopis
Vailland	4	/

Vir: lasten

Iz preglednice je razvidno, da je večina priimkov – z izjemo le dveh – zastopana v korpusu Gigafida 2.0, in to z najmanj eno (*Bauchant*) ali tudi več tisoč oblikami (*Jacob*).¹² Na portalu Fran.si je iz skupine najdenih le šest priimkov.

¹¹ V korpusu smo iskali po vseh oblikah, kar je vključevalo vse sklonske oblike priimka (*Rabelais, -a, ...*), vse pridevniške oblike in njihove sklonske oblike (*rabelaisovski ...*). Upoštevan je tudi zapis samo z velikimi črkami (*RABELAIS*) ali malo začetnico, čeprav gre za napako (*rabelais(ol ejvo)*) ali prislovi (*rabelaisovsko*).

¹² Upoštevatni velja, da gre lahko tako za priimek kot tudi ime, zato morda tolikšna pogostnost.

Katere so na preizkušanih francoskih priimkih zaznane težave eBralca? Strnjeno skušamo predstaviti nekaj opažanj:¹³

a) črkovni izgovor samoglasniških sklopov, npr.:

- eau namesto o (gl. Beauchamps);

b) črkovni izgovor soglasniških sklopov, npr.:

- ch namesto š (Beauchamps);

c) napačni izgovor zvez samoglasnik + zvočnik, samoglasnik + soglasnik, npr.:

- el namesto er (Daguerre);
- ey kot i namesto e (Barbey d'Aurevilly);

č) črkovni izgovor tudi nemih soglasnikov, npr.:

- th namesto t (Barthes), rt namesto r (Béjart), izgovor končnega nemega t (Cabet);

d) neupoštevne pravil o prevzemanju soglasnikov, npr.:

- j namesto ž (Béjart);
- g namesto ž (Bourget);
- c namesto k (Cahusac);

e) v celoti napačen izgovor po črkah, celo z vključenim izgovorom znamenja (apostrofi) ali črke y, npr.: (Barbey d'Aurevilly; Beauharnai; Bécaud; Boilly; Boisbaudran; Cahusac; Jacque; Jammes; Lachelier; Rouget de Lisle);

f) napačno naglasno mesto, npr. (Mably).

¹³ Več in podrobneje v Pravopisu 8.0.

5 Sklep

Zagatnost tvorbe svojilnih pridevnikov iz tujih lastnih imen izhaja iz dejstva, da pravila maternega jezika prenašamo na tujejezična imena, pri katerih pa nastaja razkorak med pisno in izgovorno podobo. Če uporabnik ne pozna izgovora, je velika verjetnost, da bo tvoril neustrezni svojilni pridevnik oz. lahko sledi pravilom glede na zapis. To pa je lahko zavajajoče glede na pravila izgovora. Med dvema bregovoma izgovorne in pisne podobe je most izdelava zvočno-pisne aplikacije SPiPP, s katero lahko jezikoslovci v dobi vsesplošne digitalizacije pomagamo jezikovnim uporabnikom.

Ker je nabor v aplikacijo vključenih priimkov omejen, bi bilo smiselno razmišljati, da bi s pomočjo strojnega učenja (Križaj, Dobrišek, Mihelič, Žganec Gros 2022: 248–251) iz tovrstnih dostopnih govornih virov s fonetičnimi prepisi in sistematično zbranih pravopisnih pravil prevzemanja v slovenščino eBralca naučili tudi izgovorov tujih priimkov v slovenščini. eBralec bi moral pred tem razviti sposobnost samodejne prepoznave izvornega jezika lastnega imena, ki vpliva na pravilen izgovor lastnega imena (denimo soglasniški sklop *cb* se v francoščini, italijanščini, nemščini in angleščini lahko izgovori različno). Preizkus šestdeset naključno izbranih francoskih priimkov iz SPiPP-a je pokazal, da za dobrih štirideset odstotkov izmed njih izmed njih ustrezno izgovori tako osnovno kot svojilno obliko. Rezultati so za začetek dovolj obetavni in spodbudni za nadaljnji razvoj eBralca tudi v tej smeri.

Literatura

- Ivan BELE idr., 2006: *Veliki splošni leksikon: priročna izdaja v dvajsetih knjigab*. Ur. Aleš Pogačnik. Ljubljana: DZS.
- Petra BERGANT, 2021: *Svojilni pridevniki na -(o) e) v iz tujih priimkov literatov svetovne književnosti: magistrsko delo*. Maribor. Mentorica Irena Stramlijič Breznik.
- Gigafida* 2.0. Dostop 26. 7. 2023 na <https://viri.cjvt.si/gigafida>.
- Govorni pomočnik RTV Slovenija*. Dostop 26. 7. 2023 na <https://govornipomočnik.rtvlo.si/>.
- Jezikovna svetovalnica*. Dostop 26. 7. 2023 na <https://svetovalnica.zrc-sazu.si/>.
- eBralec – sintetizator govora za slovenska besedila*. Dostop 26. 7. 2023 na <https://ebralce.si/branje/>.
- ePravopis: Slovar slovenskega pravopisa 2014–*. Dostop 26. 7. 2023 na www.fran.si.
- Fran.si*. Dostop 26. 7. 2023 na www.fran.si.
- Marta KOCJAN - BARLE, 1999: *Abeceda pravopisa, Preglednice, vaje, rešitve*. Ljubljana: Državna založba Slovenije.
- Marta KOCJAN - BARLE, 2012: Končaj v tujih moških lastnih imenih iz evropskih jezikov, zapisanih v latiničnih pisavah. *Pravopisna stikanja: razprave o pravopisnih vprašanjih*. Ur. Nataša Jakop in Helena Dobrovoljc. Ljubljana: Založba ZRC. 85–100.

- Janez KRIŽAJ, Simon DOBRIŠEK, Aleš MIHELIC, Jerneja ŽGANEC GROS, 2022: Uporaba postopkov strojnega učenja pri samodejni slovenski grafemsko-fonemski pretvorbi. *Jezične tehnologije in digitalna humanistika*. Ur. Darja Fišer, Tomaž Erjavec. Ljubljana: Inštitut za novejšo zgodovino. 248–251. (https://nl.ijs.si/jtdh22/pdf/JTDH2022_Proceedings.pdf).
- Lora NADELSBERGER, 2018: *Tvorba svojilnih pridevnikov iz osebnih lastnih imen: magistrsko delo*. Maribor. Mentorica Irena Stramljič Breznik.
- Pravopis 8.0: Pravila novega slovenskega pravopisa za javno razpravo*. Dostop 26. 7. 2023 na www.fran.si.
- Resolucija o nacionalnem programu za jezikovno politiko 2019–2025*. Dostop 26. 7. 2023 na <http://www.pisrs.si/Pis.web/pregledPredpisa?id=RESO123>.
- Ana ROTOVNIK OMERZU, 2023: *(Ne)podaljševanje večložnih tujih priimkov moškega spola z osnovo na -r pri dijakih in dijakinjah v 3. letniku splošnih gimnazij: magistrsko delo*. Maribor. Mentorica Irena Stramljič Breznik.
- Slovenski pravopis 2001*. Dostop 26. 7. 2023 na www.fran.si.
- Irena STRAMLJIČ BREZNIK (vodja projekta), 2019–2020: *SPiPP – Svojilni pridevniki iz prevzetih priimkov: spletna aplikacija za pomoč uporabnikom pri izražanju svojine za prevzete moške priimke*. Dostop 26. 7. 2023 na <https://spipp.ff.um.si/>.
- Irena STRAMLJIČ BREZNIK, 2020: Svojilni pridevniki iz prevzetih priimkov v spletni aplikaciji SPiPP. *Slovensko jezikoslovje, književnost in poučevanje slovenščine; Slovar slovenskega knjižnega jezika 16. stoletja; Veliki madžarsko-slovenski spletni slovar*. Ur. Marko Jesenšek. Maribor: Univerzitetna založba Univerze (Mednarodna knjižna zbirka Zora, 135). 96–117.
- Irena STRAMLJIČ BREZNIK, 2021: Spletna aplikacija SPiPP 1 in SPiPP 2 – od teorije k praksi – od pravil k jezikovnemu izobraževanju z zvokom in zapisom. *Slovenščina na dlani 4*. Ur. Natalija Ulčnik, Špela Antloga Maribor: Univerza v Mariboru, Univerzitetna založba. 37–56.
- Jerneja ŽGANEC GROS, Boštjan VESNICER, Simon ROZMAN, Peter HOLOZAN, Tomaž ŠEF, 2016: Sintetizator govora za slovenščino eBralec. *Zbornik konference Jezične tehnologije in digitalna humanistika*. Ur. Tomaž Erjavec, Darja Fišer. Ljubljana: Znanstvena založba Filozofske fakultete. 180–185. (<http://nl.ijs.si/isjt16/JTDH-2016-Proceedings.pdf>).
- Jerneja ŽGANEC GROS, France MIHELIC, Nikola PAVEŠIČ, 1999: Rules for automatic grapheme-to-allophone transcription in Slovene. *Text, speech and dialogue: second international workshop, TSD'99*. Ur. Václav Matoušek, et al. Berlin [etc.]: Springer. 241–247.
- Jerneja ŽGANEC GROS, Varja CVETKO-OREŠNIK (urednik), 2000: *Samodejno tvorjenje govora iz besedil: postopek za izdelavo sintetizatorja slovenskega govora*. Ljubljana: Založba ZRC (Zbirka Linguistica et philologica 3).

АНАЛИЗ СПОНТАННОЙ УСТНОЙ РЕЧИ КАК СПОСОБ ИССЛЕДОВАНИЯ СТРАТИФИКАЦИОННОЙ ВАРИАТИВНОСТИ ЯЗЫКОВЫХ КОДОВ НА ПОЛЬСКО-БЕЛОРУССКОМ ПОГРАНИЧЬЕ

КАТАЖИНА КОНЧЕВСКА

Институт польского языка Польской академии наук, Краков, Польша
katarzyna.konczewska@ijp.pan.pl

Предметом статьи является анализ спонтанной устной речи как способ исследования стратификационной вариативности языковых кодов в ареале со сложной социолингвистической ситуацией. Данные для исследования были собраны автором в ходе одиночных полевых экспедиций, проведенных в 2015–2021 годах в малонисследованном полиэтническом, поликультурном, полилингвальном микроареале по обеим сторонам польско-белорусской границы. Его особенностью является сохранившееся до наших дней деление на шляхетские (мелкодворянские) околицы и крестьянские деревни, возникшее в XVI в. Результаты исследования отмечают выраженную корреляцию между современной социолингвистической ситуацией и историческими процессами освоения данного микроареала, а также особенностями ее сословной составляющей. На основании анализа собранного в ходе полевых экспедиций корпуса устной речи обращено внимание на локальность как фактор социально-когнитивной значимости, выделены выступающие в исследуемом микроареале языковые коды, сосредоточено внимание на языковых проявлениях социальной стратификации и перцептивных особенностях восприятия своего языка носителями отдельных кодов. Предлагаемый подход является продуктивным методом изучения языковых контактов в пограничных ареалах со сложной исторической и современной социолингвистической ситуацией.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.17](https://doi.org/10.18690/um.ff.4.2024.17)

ISBN
978-961-286-882-6

Ключевые слова:
спонтанная речь,
устная речь,
языковые контакты,
стратификационная
вариативность,
пограничье



Univerzitetna založba
Univerze v Mariboru

DOI
[https://doi.org/
10.18690/um.ff.4.2024.17](https://doi.org/10.18690/um.ff.4.2024.17)

ISBN
978-961-286-882-6

Keywords:

spontaneous speech,
spoken language,
language contacts,
stratification variability,
borderland

ANALYSIS OF SPONTANEOUS SPOKEN LANGUAGE AS A METHOD FOR INVESTIGATING THE STRATIFICATION VARIABILITY OF LANGUAGE CODES IN THE POLISH-BELARUSIAN BORDERLAND

KATARZYNA KONCZEWSKA

Institute of Polish Language Polish Academy of Sciences, Krakow, Poland,
katarzyna.konczevska@ijp.pan.pl

The paper focuses on analysing spontaneous speech as a method for investigating the stratification variability of language codes in an area with a complex sociolinguistic situation. The data for the study were collected by the author during field expeditions carried in the years 2015–2021 in a little-studied multi-ethnic, multicultural, and polylingual micro-region on both sides of the Polish-Belarusian border. Its peculiarity is the division between gentry (petty nobility) settlements and peasant villages, which has been preserved to this day and was established in the 16th century. The research results point to a pronounced correlation between the current sociolinguistic situation and the historical processes of land settlement in the micro-region in question, as well as the peculiarities of its class structure. Based on the corpus of spoken speech collected during field research, this paper addresses locality as a factor of social and cognitive significance, highlighting the language codes prevalent in this micro-region, and focusing on language manifestations of social stratification, and native speakers' perceptions of their language of individual codes. The approach I propose is a productive method for investigating linguistic contacts in border areas characterised by a complex historical and contemporary sociolinguistic situation.



University of Maribor Press

ANALIZA SPONTANEGA GOVORA KOT METODE RAZISKOVANJA STRATIFIKACIJSKE VARIABILNOSTI MED JEZIKOVNIMI KODI NA POLJSKO-BELORUSKEM OBMEJNEM OBMOČJU

KATARZYNA KONCZEWSKA

Inštitut za poljski jezik, Poljska akademija znanosti, Krakov, Poljska
katarzyna.konczevska@iip.pan.pl

Članek se osredotoča na analizo spontanega govora kot metode za raziskovanje stratifikacijske variabilnosti med jezikovnimi kodi na območju z zapletenim sociolingvističnim ozadjem. Podatke za raziskavo je avtorica zbrala med enournimi terenskimi raziskovalnimi odpravami med 2015 in 2021 na malo raziskanem večetničnem, večkulturnem in večjezičnem območju na obeh straneh poljsko-beloruske meje. Posebnost tega območja je do danes ohranjena delitev na plemiška (malomeščanska) naselja in kmečke vasi, ki je bila vzpostavljena v 16. stoletju. Rezultati raziskave kažejo na izrazito povezanost med sedanjim sociolingvističnim položajem in zgodovinskimi procesi poselitve obravnavanega območja ter posebnostmi njegove razredne strukture. Na podlagi analize korpusa, zbranega med terenskimi raziskovalnimi odpravami, članek obravnava lokalnost kot dejavnik družbenega in kognitivnega pomena, izpostavlja jezikovne kode, ki prevladujejo na tem območju, in se osredotoča na jezikovne manifestacije družbene slojevitosti ter značilnosti dojemanja svojega jezika s strani govorcev posameznih kodov. Predlagani pristop je produktivna metoda raziskovanja jezikovnih stikov na obmejnih območjih s kompleksnim zgodovinskim in sodobnim sociolingvističnim položajem.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.17](https://doi.org/10.18690/um.ff.4.2024.17)

ISBN
978-961-286-882-6

Ključne besede:
spontani govor,
govorjeni jezik,
jezikovni stiki,
stratifikacijska variabilnost,
obmejna območja



Univerzitetna založba
Univerze v Mariboru

1 Введение

Спонтанная речь является реальным показателем уровня языковой компетенции говорящего, поскольку реализуется в переменных коммуникативных условиях. Как преобладающая в реальной речи, она является своеобразным лингвистическим феноменом и уникальным исследовательским материалом. Спонтанную речевую деятельность можно понимать как разновидность неподготовленной устной речи, которая отличается «высоким уровнем идиоматичности, семантическим синкретизмом, синтаксической эллиптичностью, предикативностью, широким использованием готовых клишированных речевых форм в системе специфических моделей интегрирования и просодического оформления смысловых единиц в единое речевое целое» (Яковлева 2016: 4). Именно поэтому анализ спонтанной речи является, на наш взгляд, наиболее продуктивным инструментом исследования стратификационной вариативности языковых кодов, выступающих в периферийных контактных ареалах, характеризующихся сложной социолингвистической ситуацией. В свою очередь, с речевой осведомленностью говорящего коррелируют лингвистические переменные (Labov 1966: 4–22), являющиеся ключевой концепцией в социолингвистических исследованиях (Labov 1972).

Изучаемый нами микроареал польско-белорусского пограничья представляет собой зону не только периферийных балтийско-славянских контактов, но и активных языковых инфильтрационных процессов внутри различных славянских языковых подгрупп (польский, белорусский, русский). Процесс субстратного заимствования нередко проходил здесь в условиях дву-, а иногда и трехязычия, поэтому субстратно-адстратные отношения требуют чрезвычайно осторожного подхода (Rembiszewska, Siatkowski 2018). При изучении диалектологического ландшафта ареала с так сложной ситуацией мы предлагаем учитывать, прежде всего, социолингвистические переменные, особенно стратификационную вариативность, поскольку она является одним из универсальных свойств языка, особенно ярко проявляющихся в процессе его функционирования, рассматривается как один из факторов развития языка и проявляется на всех его уровнях.

В процессе проведения исследований было обращено внимание на локальность как фактор социально-когнитивной значимости (Jensen 2016), определены выступающие в исследуемом микроареале языковые коды, выделены основные индикаторы и маркеры вариативности, сосредоточено внимание на языковых проявлениях социальной стратификации и перцептивных особенностях (Preston 1989) восприятия своего языка носителями отдельных кодов.

2 Методология и состояние исследований

Выбор темы исследования был мотивирован недостаточной изученностью данного участка белорусско-польской границы (Konczewska 2021a). Проблематика влияния границ на процессы языковой конвергенции и дивергенции является весьма актуальной (Dialect change 2005, On the Border of Language and Dialect 2018). Граница ингерует в социальные контакты и, следовательно, в развитие языка по обеим ее сторонам: внутренние устные разновидности подвергаются взаимным влияниям и развиваются в сторону конвергенции, а устные разновидности, оставшиеся по другой стороне, отдаляются и значительно видоизменяются. Предметом наших исследований являются языковые контакты в микроареале по обеим сторонам белорусско-польской границы, установленной в 1948 г. в результате ятинских договоренностей и разделившей регион, становящийся единое целое в течение более 500 лет. Исследования микроареалов видятся нам очень важными в изучении языковых контактов на пограничье и социолингвистической ситуации в целом, поскольку позволяют выделить локальные особенности и определить ареал языковых явлений, а также обратить внимание на пограничный регион как историческую единицу, подчеркнуть важность видения истории с точки зрения периферии, а не, как обычно, центра. Более того, материалы полевых интервью являются важным источником истории повседневности *Alltagsgeschichte / History of Everyday Life* (Lütke 1995: 3–40) и языковой исторической памяти (Chlebda 2011: 83–98, 2019: 147–164). Войцех Хлебда верно заметил, что «пограничье, определенное государственной границей, не было единственным, которое формировало ментальность и духовность» (Chlebda 2015: 43). Осознание важности поиска этих внутренних «границ пограничья» повлияло на то, что, поставив цель изучения в полном

объеме говором в исследуемом микроареале и определения его диалектного ландшафта, мы не могли проигнорировать следующих важных моментов:

Каким образом установление границы повлияло на процессы интеграции и интерференции местных говором, их оценку носителями языка и уровень языкового сознания носителей?

Какое влияние на язык местных жителей имели общественные и цивилизационные перемены?

Какие корреляции имеют место между двумя частями единого ранее региона?

Для реализации поставленных целей нами были исследованы все населенные пункты по обеим сторонам белорусско-польской границы, входящие в состав бывшего исторического Уснарского католического и православного приходов: белорусско- и польскоязычные православные и католические деревни, а также польскоязычные шляхетские околицы. В ходе исследования были проведены интервью в 26 населенных пунктах, с автохтонами-представителями различных вероисповеданий и социальных групп, в количестве 51 человек: 30 по польской стороне, 21 по белорусской стороне; 30 женщин, 21 мужчина; 27 жителей шляхетских околиц, 24 жителей деревень (12 из католических, 12 из православных); возраст интервьюируемых от 38 до 99 лет, возрастные информаторы имели очень хорошую память. Исследование проводилось в рамках единоличного авторского проекта, в период с 2017 по 2019 г. – в сотрудничестве с Этнографическим музеем имени Северина Удели в Кракове. Полевые экспедиции были прерваны в 2020 г. в связи с политическим кризисом в Беларуси и ограничениями, введенными во время пандемии COVID-19, а также в связи с последующим эмигрантским кризисом на польско-белорусской границе в 2021 г., в результате которого польское правительство ввело чрезвычайное положение в пограничной зоне и запрет на пребывание в ней.

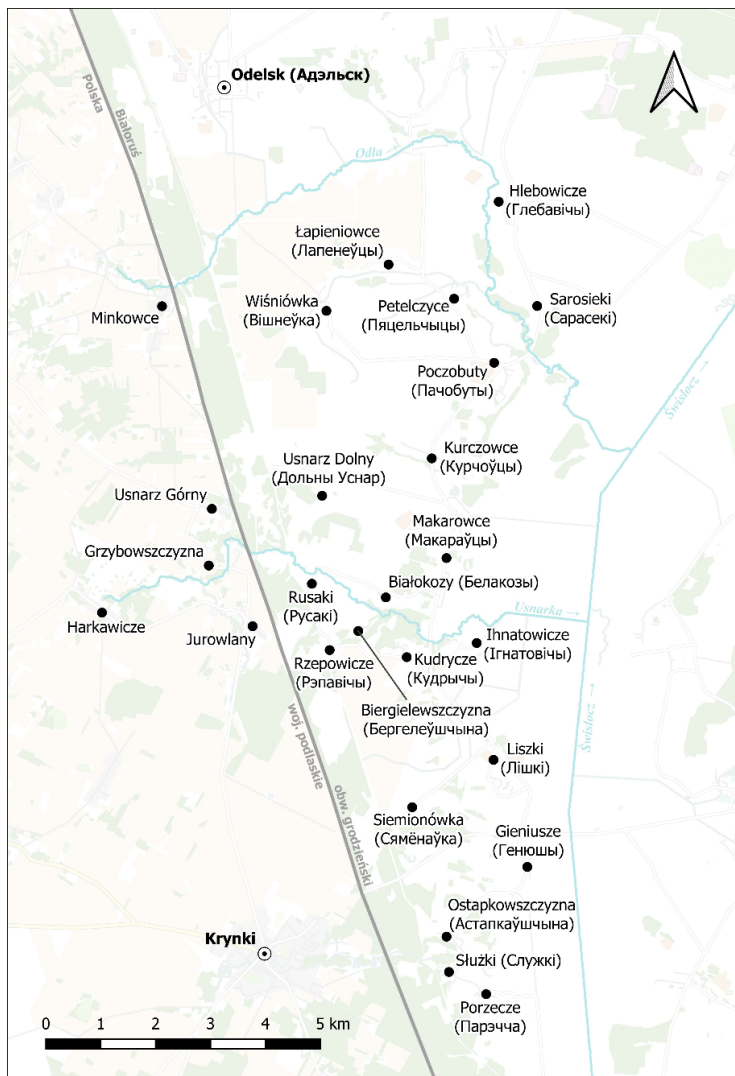
Для реализации поставленных целей нами был опробован авторский метод проведения расширенных полевых социолингвистических экспедиций, основанный на использовании в качестве исследовательского инструментария музейной коллекции – как материальных экспонатов, так и архивных документов (Konczewska 2021b: 165–174). Практика показала, что

использование данного метода как качественного (ср. Johnstone 2000) значительно облегчает первичный контакт с информаторами и позволяет не только восстановить историческую социолингвистическую ситуацию исследуемого микроареала, но и определить его современный диалектный ландшафт. При анализе языковых вариантов для выявления процессов конвергенции и дивергенции использовался метод сопоставительного анализа, а для определения специфики реализации языковых вариантов – описательный метод.

Необходимо отметить, что исследуемый микроареал не был ранее предметом диалектных и социолингвистических исследований. По польской стороне диалектологические исследования Белостокского воеводства начались только в 60-х гг. XX в. и преимущественно не охватывали территорий, граничащих с современным Гродненским и Берестовицким районами по белорусской стороне (Glinka, Smulkowa 1965: 137–161; Glinka, Smulkowa 1981: 373–383); кроме того, проводимые исследования усложняла неоднородная социолингвистическая карта пограничья (Zdaniukiewicz 1992: 171–176). Экспедиции Отдела белорусской филологии Института славяноведения Польской академии наук, осуществленные в 111 пунктах в ареале от Августовского канала по Буг с целью подготовки Атласа восточнославянских говоров Белосточчины, также не охватывали населенных пунктов нашего исследовательского микроареала. В конце 90-х гг. XX в. изучение белорусских говоров польско-белорусского пограничья осуществлял американский исследователь Курт Вулхайзер (Woolhiser 2005: 236–262). В 2017 г. с целью изучения белорусских говоров Подляшья была осуществлена совместная диалектологическая экспедиция Отдела диалектологии Института белорусского языка Белорусской академии наук и Института славистики Польской академии наук, однако в центре ее внимания были территории вне нашего исследовательского микроареала. По белорусской стороне не проводилось диалектологических экспедиций в приграничной зоне; небольшой корпус диалектной лексики приграничных местностей составил Апанас Цыхун, который в качестве школьного инспектора посещал населенные пункты района (Цыхун 1993).

3 Характеристика микроареала исследования

Данные для исследования были собраны нами в ходе полевых экспедиций, проведенных в 2015–2021 годах в малоисследованном полиэтническом, поликультурном, полилингвальном микроареале по обеим сторонам польско-белорусской границы.



Карта 1: микроареал полевых исследований

До окончательного установления польско-советской границы, утвержденной на данном участке только в 1948 г., территория микроареала становила единое целое. Ее уникальной особенностью является сохранившееся до наших дней деление на шляхетские (мелкодворянские) околицы и крестьянские деревни, возникшее в XVI в. Культурная самоидентичность микроареала сформировалась под влиянием трех монотеистических религий: христианства, ашкеназийского иудаизма, ислама. Уже к XVII в. данный микроареал сформировался как многонациональный: литвины (жители Великого княжества Литовского), русины (жители Руси Литовской), поляки (прибывшие с центральной Польши в XV и начале XVI в.), евреи (проживающие с XIV в.), татары (с XIV в. – пленники Витовта и с XVII в. – осадники на землях, дарованных Яном III Собеским за военную службу); многоконфессиональный: православие, католицизм, ислам, иудаизм и многословный: бояре (закрепленные «Уставом на волоки» 1557 г.), крестьяне, шляхта. В настоящее время в границах исследуемого микроареала проживают представители православной и католической конфессии, преимущественно поляки и белорусы по национальности, являющиеся потомками крестьян и малоземельной шляхты. Неоднородная социолингвистическая картина региона является следствием исторических и политических процессов. Данный ареал заселяли ятвяжские племена, которые были вытеснены в результате балтийской и славянской экспансии. В период с XII по XIV века здесь существовали русские княжества. С XIV в. территория входила в состав Тракайского уезда Великого княжества Литовского, впоследствии – I Речи Посполитой, а после ее раздела, с XVIII в. – в состав Российской империи (Слонимская (1796–1801) и Гродненская губернии). С 1921 г., на основании Рижского договора, эти земли вернулись в состав Речи Посполитой. С 17 сентября 1939 г., в результате нападения Советского Союза на Польшу, они вошли в состав новообразованной Белостокской области; с 22 июня 1941 г., в результате нападения нацистской Германии на Советский Союз, и до 1944 г. находились в составе округа *Bezirk Bialystok*, части Восточной Пруссии. 20 сентября 1944 года, по итогам ялтинских договоренностей, территория исследуемого микроареала вошла в состав Гродненской области БССР в составе СССР, однако граница на этом участке была окончательно установлена только в марте 1948 г.

Важное значение для изучения диалектного ландшафта исследуемого ареала имеют выводы Яна Якубовского, автора карты Гродненского уезда в XVI в. (Jakubowski 1935: 99–114). На основании географической номенклатуры ареала исследователь утверждал об однородном, белорусском характере сельского населения; отмечал наличие в данном ареале топонимов литовской этимологии; обращал внимание на поселения мелкой шляхты с русскими и литовскими родовыми фамилиями; определял этнический состав центральной части уезда как смешанный литовско-белорусский, с численным преобладанием белорусского элемента; отмечал, что уже в XVI в. произошла полная ассимиляция обеих этнических групп, а разговорным языком стал белорусский; утверждал, что в северной, «запушчанской» части уезда, присоединённой в 1566 г., проживало литовское население, а в расположенных при границе прусских осадах – частично польское, мазурское; обращал внимание на то, что «польский элемент немногочисленный, но влияние польской культуры значительно, [...] польский язык уже тогда был разговорным для высшего класса» (там же). О сложности и неоднородности процесса расселения и разнообразии этническо-конфессионального компонента данного участка польско-белорусского пограничья свидетельствуют его топонимы: Гуды, Гудевичи, Гудишки; Дайнова, Жидомля, Латыши, Литвины, Литвинки; Мазуры, Мазурки, Москали, Русаки, Русачки, Русиново, Русиновцы, Русота, Татарка, Татаровщина, Татарье, Цыгановка, Ятвезь (Ятвези).

4 Результаты исследования

Одним из компонентов кластерного изучения микроареала было исследование речевого поведения как коммуникативного взаимодействия, что позволило трактовать спонтанные высказывания собеседников (например, монологи о предложенных для распознавания музейных экспонатах) как сложные коммуникативные семиотические образования, включающие языковые и экстралингвистические факторы, подчиненные целому ряду принципов и условий функционирования.

Вступительный анализ спонтанной речи, проведенный на основании корпуса зарегистрированных интервью, позволил выделить в изучаемом микроареале следующие языковые коды.

Белорусские православные деревни: активно: «па-просту», пассивно: «польский».

Белорусские католические деревни: активно: «ро prostu», менее активно «ро polsku».

Польские православные деревни: активно «ро posku», «па-просту», менее активно «па-беларуску».

Польские католические деревни: активно «ро polsku», «ро prostu».

Белорусские и польские шляхетские околицы: активно «ро polsku», менее активно «ро prostu».

На основании проведенного сопоставительного анализа языковых кодов были выделены следующие основные социолингвистические маркеры:

возраст носителей языка (старшее, среднее, младшее поколение),
социальная принадлежность (шляхта, крестьяне),
«высокие» (религия) и «низкие» (быт, семья) темы,
сфера и среда использования языка: внутрисемейные отношения, соседи, сакральная сфера.

Важным индикатором в исследуемом микроареале является социолингвистическая переменная, указывающая на социальное положение носителя языка. Данный индикатор являлся постоянным при варьировании компонентов социально-коммуникативной ситуации.

Поскольку ситуативность и субъективная окрашенность являются обязательными характеристиками спонтанной речи, важной задачей во время полевых экспедиций было создать доброжелательную атмосферу во время интервью, вызвать у собеседника спонтанные воспоминания, чему способствовал представленный выше авторский метод, базируемый на музейной коллекции. Такие спонтанные высказывания являлись своеобразным отражением какого-то определенного «кусочка» действительности, связанного с предъявленным музейным объектом.

Во всех интервью ярко прослеживался индикатор социального положения. Несмотря на десятилетия советской власти, по белорусской стороне границы изучаемого микроареала сохранилось четкое разделение на «шляхту» и «хлопов / хамов / мужиков».

Jaka to była różnica! Nawet księża widzieli różnice między wiejskich ludzi a szlachta.

Различия между шляхетской околицей (так назывались местности, в которых проживали представители различных шляхетских родов) и крестьянской деревней проявлялись в различных областях: в архитектурном оформлении поселений (в околице преобладает свободная застройка, а в деревне – расположение домов вдоль улицы); в строительстве жилого дома (шляхетский дом и крестьянский отличаются как внешне, так и внутренним убранством, расположением и предназначением комнат); в способе ведения хозяйства (например, возле шляхетского дома обязательно разбит фруктовый сад, чего нет на крестьянском подворье); в быту (например, в домах шляхты чай подают обязательно в чашках с блюдцами; оборудована отдельная, добротна обставленная комната, обязательно с фикусом, в которой принимают гостей). До Второй мировой войны крестьяне и шляхта отличались также внешним видом:

I ubierali się my inaczej. Kapelusze taki fajny z kwiatami. A już wiejska ona nie miała prawa nalożyć kapelusza, bo każdy szlachcic mógł patkę rzucić jej. Mogła tylko chusta.

Данное разделение нашло свое отражение также в языке, чему способствовало тот факт, что вплоть до образования колхозов, которое здесь имело место только в конце 50-х – начале 60-х гг. XX в., жители шляхетских околиц и деревень, даже католических, не общались между собой (хотя и те, и другие идентифицировали себя как поляков):

*zawsze była różnica w mowie, wiejski członek i szlachta,
jeszcze za Niemcami ta różnica była.*

Польский «шляхетский» язык для жителей околиц, оставшихся по белорусской стороне, был важным, нобилитирующим маркером их сословной принадлежности:

*my wszyscy po polsku pięknie rozmawiali;
język mój od dzieciństwa trochę delikatniejszy;
w domu i tu w okolicy wszyscy mówili tylko po polsku;
po polsku zawsze rozmawiali między sobą;
tu wszędzie była szlachta i rozmawiali po polsku;
bardzo ładnie rozmawiali u nas po polsku;
my nie wiedzieliśmy, jak to rozmawiać po prostu;
tu w okolicach wszędzie po polsku rozmawiali.*

Этот языковой код был средством общения в кругу семьи и соседских контактов жителей шляхетских околиц, и только в связи с насильственным инкорпорированием в колхозы они стали постепенно приобщаться к коду «па-просту» (говор с белорусским субстратом), на котором общались с жителями деревень:

*kiedy do kolchozów nas zaganiali, to nie było już innego wyjścia;
tam z nimi że wsi to już trzeba było po prostu rozmawiać;
jak gdzieś w polu z muzykami tam, to po prostu.*

Однако языковой код «па-просту» трактовался жителями шляхетских околиц как низкий, «некультурный»:

*brzydki taki, nieładny;
niekulturalny, ot taki sobie;
tam nie umieli po polsku dobrze rozmawiać, ot tak, namieszają wszystko;
na wsi to oni tak jakoś grubo rozmawiali, tam inaczej rozmawiali, nie było tak, że jednakowo.*

Для языкового кода шляхты характерны своеобразные формулы этикета:

czy mama pójdzie, czy tatuś pozwoli, czy pani puści swoja córka na zabawę;

форма обращения к родителям предполагает использование 3 лица. На сохранение архаичности языкового кода жителей шляхетских околиц по белорусской стороне повлияло установление приграничной зоны, ограничивающее в значительной степени миграцию извне. Это способствовало также сохранению локальной лексики (приведем примеры из некоторых лексико-семантических групп в сопоставлении с литературными вариантами польского языка):

земледелане: *aktar* 'hektar';

tkачество: *bierdo* 'plocha', *czolnek* 'czólenko tkackie', *kapa* 'samodziałowa, tkana narzuta na łózko', *kowrotek* 'kolowrotek', *nitrzenica* 'układ nici z oczkami, przez które przeciąga się nitki osnowy', *pakul* 'wełna czesana, nieprzędzona', *praćki*, *sprzątki* 'drewniany pręcik, na którym obraca się cewka w czólenku tkackim';

хозяјство: *bijak* 'narzędzie do zbijania', *kurcia* 'kura', *plyta* 'piec do przygotowania potraw', *prześniaczek* 'praśny chleb', *przystępy*, *pójść w przystępy* 'przyjść na gospodarę żony', *szmutki* 'barachło';

жиные строения: *klebania* 'plebania', *mieszkanie* 'dom';
семья: *dziatki* 'dzieci';

общество: *kacaporowcy* 'sowieci', *najezdny* 'przyjezdny', *pijanica* 'pijak', *praniec* 'ktoś cwany'.

Отметим, что для языкового кода жителей польских шляхетских околиц характерно большое количество архаизмов, большинство которых восходит к варианту старопольского языка XVI в., например:

Прилагательные: *drzewniany* 'drewniany', *pamiętny* 'ten, kto ma dobrą pamięć', *rochmany* 'łaskawy, spokojny', *rządowa* 'wódka gorzelnii państwowej'.

Наречия: *fajno* 'bardzo dobrze', *lecko* 'lecko', *na stojącźka* 'na stojąco', *rychtyk* 'akurat, właśnie', *wprzód* 'najpierw'.

Глаголы: *dobijać się* 'domagać się', *dychać* 'oddychać', *gapować* 'myśleć', *hodować* 'wychowywać', *mastygować* 'majsterkować', *matać* 'nawijać', *nakrąjać* 'pokroić', *odmierać* 'wymierać', *patrzeć* 'doglądać', *popłukać* 'umyć', *porzynać* 'posiekać', *przebiekać* 'przecinać', *przyjeżdżać* 'przyjeżdżać', *rozmówić się* 'długo rozmawiać', *ruszać* 'ruszać', *towarzyszyć* 'przyjaźnić się', *upiekać* 'piec', *wzusić* 'złożyć' (buty), *zaznajomić* 'zapoznać', *znajomić się* 'zapoznawać się', *znaleźć się* 'urodzić się'.

Паремии: *pani kochanieńka* (обращение); *bywaj zdromieńka* (тост); *kiedy łaska* 'пожалуйста'; *żywa woda* (о энергичном человеке); *jak kacźka po kaluźkach* 'неуклюже'; *noży zostawić* 'разуться'.

Вступительный анализ зарегистрированной спонтанной речи позволяет сделать выводы, что в изучаемом микроареале преобладает продуктивный билингвизм с диглоссией. Выбор языка в ситуации билингвизма и форма реализации вариативности в ситуации диглоссии зависят от сферы (хозяйственная деятельность, религия) и среды (шляхта / крестьяне, католическая деревня / православная) функционирования языка. Можно

выделить следующие субъективные идиомы информаторов, определяющие акты речи: «ро polsku», «ро prostu», «па-просту», «па-беларуску».

В изучаемом микроареале по обеим сторонам границы православные молятся в церкви на русском и церковнославянском языках, а католики в костеле – на польском. Польский язык является языком внутрисемейных и соседских контактов для жителей шляхетских околиц, в том числе по белорусской стороне, где этот язык, несмотря на советские репрессии, передают детям и внукам.

Языком, который информаторы определяют как код «па-просту», пользуются в ежедневном общении жители католических и православных деревень как по белорусской, так и по польской стороне границы, а также жители шляхетских околиц в контактах с жителями деревень; причем информаторы различают коды «па-просту» и «па-беларуску».

Языковой код «па-просту» является в их субъективном понимании языком, которым они пользуются в ежедневных внутрисемейных и соседских контактах; он может иметь как белорусский, так и польский языковой субстрат (ср. Bieder 2002: 223–230), а вариативность зависит от изложенных выше социолингвистических маркеров.

По белорусской стороне границы код «па-просту» жителей православных деревень содержит высокий процент русизмов, что, на наш взгляд, можно объяснить высоким уровнем русификации Беларуси. В коде «па-просту» жителей католических деревень преобладает польская интерференция на грамматическом и лексическом уровнях.

По польской стороне границы код «па-просту» жителей православных деревень представляет собой хорошо сохранившиеся белорусские гродненские говоры, относящиеся к северно-западной группе диалектов. Для кода «па-просту» жителей католических деревень характерен субстрат белорусских гродненских говоров с высоким уровнем польской лексической интерференции.

В диалектологическом ландшафте изучаемого микроареала преобладает неравноценная интерференция с преобладанием полонизмов; особое место занимает архаичность шляхетского варианта польского языка по белорусской стороне границы.

Социолингвистическую ситуацию в исследуемом микроареале можно определить как экзогlossную, несбалансированную, четырёхкомпонентную: Польша / Беларусь, шляхетская околица / крестьянская деревня, шляхта / крестьяне, православные / католики.

5 Выводы

Проводимые нами исследования и вступительная обработка их результатов позволяют на утверждение, что при исследованиях в трансграничном периферийном ареале наиболее продуктивным является сочетание традиционной лингвистической методологии с наработками этнографов, антропологов и этноисториков (ср. Wiemer 2003: 212–229), а также поиск новых методов полевых экспедиций – например, представленный нами авторский метод на основе музейной коллекций. Такие исследования можно отнести к области лингвистической антропологии и лингвопалеонтологии.

Изогlossы в пограничных ареалах следует рассматривать, на наш взгляд, как границы интенсивности контактов, связанные с фактическими географическими и политическими границами и границами культурных влияний. Диалектный ландшафт пограничья косвенно отражает историческую изменчивость политических и культурных границ.

Представленные методы исследования пограничного микроареала могут служить для изучения процессов инфильтрации в балто-славянском пограничном регионе (ср. Wiemer, Erker 2011: 184–216), сопоставительного описания диалектов в приграничных микроареалах и определения влияния политических факторов на процессы языковой конвергенции и дивергенции, а также помогают получить новые данные о языковых контактах в малоизученных периферийных диалектных ареалах и описать пограничный региолект.

Rezultaty našix issledovanij otmečajuť vyražennuju korreľacuju mežu sovremennoj sociolingvističeskoj situaciej i istoričeskimi procesami osvoenija danogo mikroareala, a takže osobennostiami ee soslovnoj sostavljajušej. Predlažаемый нами podhod javljaetsja produktivnym metodom izučenija jazykovyx kontaktov (sr. Trudgill 1986) v pograničnyx arealax so složnoj istoričeskoj i sovremennoj sociolingvističeskoj situaciej.

Литература

- Hermann BIEDER, 2002: *Polszczyzna północnokresowa jako zjawisko kontaktowe polsko-białorusko-litewskie. Uwagi metodologiczne. Acta Polono-Ruthenica* 7, 223–230.
- Wojciech CHLEBDA, 2011: Szkiecy do językowego obrazu pamięci. *Pamięć jako wartość. Etnolingwistyka* 23, 83–98.
- Wojciech Chlebda, 2015: Pogranicza i pograniczność w polskich perspektywach oglądu. *Pograniczność i pogranicza w perspektywie nauk społecznych i humanistycznych*. Ur. Wojciech Chlebda, Ivana Dobrotová. Opole: Wydawnictwo Uniwersytetu Opolskiego, 43–77.
- Wojciech CHLEBDA, 2019: O wyzwaniach i zadaniach pamięcioznawstwa lingwistycznego. *LingVaria* 14/28, 147–164.
- Peter AUER, Frans HINSKENS, Paul KERSWILL, 2005: *Dialect change. Dialect change: Convergence and divergence in European languages*. Ur. Peter Auer, Frans Hinskens, Paul Kerswill. Cambridge: Cambridge University Press.
- Stanisław GLINKA, Elżbieta SMUŁKOWA, 1965: Dorobek językoznawstwa dotyczący Białostoczczyzny w latach 1944–1964. *Rocznik Białostocki* 5, 137–161.
- Stanisław GLINKA, Elżbieta SMUŁKOWA, 1981: Stan i perspektywy badań dialektologicznych regionu Białostockiego. *Rocznik Białostocki* 14, 373–383.
- Jan JAKUBOWSKI, 1935: Powiat grodzieński w XVI w. (mapa z tekstem). *Prace Komisji Atlasu Historycznego Polski* III. Kraków: Polska Akademia Umiejętności, 99–114.
- Marie M. JENSEN, 2016: Linking Place and Mind: Localness As a Factor in Socio-Cognitive Saliency. *Frontiers in Psychology* 7. Dostop <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2016.01143/full>, 25. 3. 2024.
- Barbara JOHNSTONE, 2000: *Qualitative Methods in Sociolinguistics*. Oxford: Oxford University Press.
- Katarzyna KONCZEWSKA, 2021a: *Polacy i język polski na Grodzieńszczyźnie*. Prace Instytutu Języka Polskiego PAN 160. Kraków: Instytut Języka Polskiego Polskiej Akademii Nauk.
- Katarzyna KONCZEWSKA, 2021b: Dialektologiczne badania terenowe na podstawie kolekcji muzealnej jako metoda budowania nowych kontekstów kulturowych. *Gwary Dzisiaj* 14, 165–174.
- William LABOV, 1966: The linguistic variable as a structural unit. *Washington Linguistics Review* 3, 4–22.
- William LABOV, 1972: *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Alf LÜDTKE, 1995: *The History of Everyday Life: Reconstructing Historical Experiences and Ways of Life*. Princeton: Princeton University Press.
- Marjatta PALANDER, Helka RIIONHEIMO, Vesa KOIVISTO, 2018: On the Border of Language and Dialect. *Studia Fennica Linguistica* 21. Ur. Marjatta Palander, Helka Riionheimo, Vesa Koivisto. Helsinki: Finnish Literature Society.
- Dennis R. PRESTON, 1989: *Perceptual dialectology*. Dordrecht: Foris.
- Dorota REMBISZEWSKA, Jerzy SIATKOWSKI, 2018: *Pogranicze polsko-wschodniostonińskie. Studia wyrazowe*. Warszawa: Wydział Polonistyki Uniwersytetu Warszawskiego.

- Peter TRUDGILL, 1986: *Dialects in Contact*. Oxford: Blackwell Publishers.
- Björn WIEMER, 2003: Zur Verbindung dialektologischer, soziolinguistischer und typologischer Methoden in der Sprachkontaktforschung (am Beispiel slavischer und litauischer Varietäten in Nordostpolen, Litauen und Weißrußland). *Zeitschrift für Slavistik* 48/2, 212–229.
- Björn WIEMER, Oksana ERKER, 2011: Manifestations of areal convergence in rural Belarusian spoken in the Baltic-Slavic contact zone. *Journal of Language Contact* 4/2, 184–216.
- Curt WOOLHISER, 2005: Political Borders and Dialect Divergence/Convergence in Europe. *Dialect Change. Convergence and Divergence in European Languages*. Ur. Peter Auer, Frans Hinskens, Paul Kerswill. Cambridge: Cambridge University Press, 236–262.
- Alojzy ZDANIUKIEWICZ, 1992: O powstaniu i rozwoju języka polskiego na Kresach Wschodnich – polemicznie. *Z polskich studiów slawistycznych* VIII, 171–176.
- Апанас ЦЫХУН, 1993: *Скарбы народнай мовы*. Гродна: Гродзенскі дзяржаўны ўніверсітэт імя Янкі Купалы.
- Эмма ЯКОВЛЕВА, 2016: *Речевые hesitation: формальный и функциональный аспекты*. Москва: Российская академия наук.

KONTEKSTI SNEMANJA GOVORJENEGA DISKURZA V SOCIOLINGVISTIKI

MAJA BITENC

Univerza v Ljubljani, Filozofska fakulteta, Ljubljana, Slovenija
Maja.Bitenc@ff.uni-lj.si

Prispevek se osredotoča na kontekste snemanja govornega diskurza, predvsem s sociolingvističnega vidika. Prinaša pregled metodoloških pristopov v različnih tujih in slovenskih raziskavah, ki preučujejo govorno variantnost tako pri posameznem govorniku kot v govorni skupnosti. Predstavljeni so pionirski sociolingvistični intervjuji Williama Labova v ZDA z izvajanjem različnih govornih stilov ter načini pridobivanja posnetkov in analiziranja govornega jezika v sodobnejših raziskovalnih projektih v jezikovnih skupnostih, v katerih so sociolingvistični profili bolj sorodni slovenskemu, npr. v Avstriji, Nemčiji, Belgiji in na Madžarskem. Ti vključujejo npr. branje besedila oz. seznama besed, prevod iz narečja v standard oz. iz standarda v narečje, neformalni pogovor s prijateljem iz istega kraja oz. z drugega narečnega področja, formalni sociolingvistični intervju z raziskovalcem, posnetke prijateljskih pogovorov, posnetke, pridobljene za druge namene, ter opazovanje z udeležbo. Ovrednotene so tudi dosedanje študije variantnosti govorne slovenščine, ki so govor zajele v različnih bolj ali manj avtentičnih vsakodnevnih situacijah.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.18](https://doi.org/10.18690/um.ff.4.2024.18)

ISBN
978-961-286-882-6

Ključne besede:
snemanje govora,
konteksti jezikovne rabe,
jezikovna variantnost,
sociolingvistični intervju,
opazovalčev paradoks



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.18](https://doi.org/10.18690/um.ff.4.2024.18)

ISBN

978-961-286-882-6

Keywords:

speech recording,
contexts of language use,
linguistic variation,
sociolinguistic interview,
observer's paradox

CONTEXTS OF RECORDING SPEECH IN SOCIOLINGUISTICS

MAJA BITENC

University of Ljubljana, Faculty of Arts, Ljubljana, Slovenia
Maja.Bitenc@ff.uni-lj.si

The paper focuses on the contexts in which spoken language is collected and recorded, primarily from a sociolinguistic perspective. It provides an overview of methodological approaches in various foreign language and Slovene studies that analyse speech variation both at the level of the individual speaker and within speech communities. It presents William Labov's pioneering sociolinguistic interviews in the United States, which include different speech styles as well as methods for eliciting and analysing spoken discourse in more recent research projects in language communities with sociolinguistic profiles closer to Slovene, such as Austria, Germany, Belgium and Hungary. These include activities such as reading texts and word lists, translation from dialect to standard and vice versa, informal conversations with friends from the same and different dialect areas, formal sociolinguistic interviews with researchers, recordings of friendly conversations, recordings for other purposes and participant observation. The existing studies on language variation in Slovene, which record speech in various more or less authentic everyday situations, are also analysed.



1 Uvod¹

Sociolingvistika od samih začetkov oz. od pionirskih raziskav Williama Labova v ZDA in Petra Trudgilla v Veliki Britaniji v 60. in 70. letih 20. stoletja poudarja vpliv situacijskega konteksta na posameznikov govor: govorniki izbirajo različne variante in varietete iz svojega govornega repertoarja glede na številne pragmatične dejavnike, kot so sogovorniki, morebitni drugi udeleženci, kraj in tema pogovora, formalnost situacije, želen učinek na naslovnik in pozornost, posvečena govoru. V različnih situacijah govorniki uporabljajo različne deleže jezikovnih variant (različic) posameznih variabel (spremenljivk). Tako je pri vsaki analizi jezikovnega sistema in njegove variantnosti ključen podatek o tem, v kateri situaciji oz. s kom govorniki določene variante uporabljajo in v kolikšni meri (prim. Fanta-Jende 2020). Stilistične oz. slogovne razlike imajo preprost družbeni korelat: formalnost se povečuje v neposrednem razmerju s količino družbenih razlik med udeleženci jezikovne interakcije. Posameznikovo sposobnost za presojo formalnosti komunikacijskega stika s stališča percepcije v veliki meri določajo jezikovni dejavniki, hkrati pa formalnost komunikacije s stališča produkcije vpliva na fonološke in druge vidike varietete in naglasa (Chambers 2003: 4–5).

Pričakovano je torej, da informanti v neformalnih okoliščinah, tudi znotraj testne situacije, uporabljajo več narečnih variant. Labov (1972) meni, da takšne slogovne spremembe sproži predvsem količina pozornosti, ki jo posamezniki namenjajo svojemu govoru – pri manj pozornem govoru navadno pogosteje uporabljajo značilnosti manj formalnega stila (ang. *attention to speech hypothesis*); Bell (1984) pa slogovne premike pojasnjuje z opažanjem, da se govorniki v glavnem prilagajajo jezikovni rabi svojega naslovnika oz. občinstva (ang. *audience design approach*). Zelo verjetno je, da pri različnih raziskovalnih metodah (npr. formalni intervju oz. neformalni pogovor med prijatelji) na variantnost vplivata oba dejavnika, tako pozornost na govor kot navzoče občinstvo (npr. neznani raziskovalec oz. tesni prijatelj) (prim. Vergeiner idr. 2022).

¹ Prispevek je nastal v okviru projekta Sociolingvistična variantnost govorne slovenščine na primeru mobilnih govorcev: Viri, metode in analiza (Z6-9371), ki ga financira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije (ARIS) iz državnega proračuna.

V variantnostnih raziskavah je vir za pridobivanje govornega materiala pogosto t. i. sociolingvistični intervju, ki ga utemeljitelj Labov (1984: 32–33) opredeli kot serijo hierarhično strukturiranih sklopov vprašanj z več cilji, med katerimi sta najpomembnejša ta, da posnamemo uro ali dve posnetka in pridobimo širok nabor demografskih podatkov za vsakega govorca. Sali A. Tagliamonte (2006: 37–49) meni, da je poimenovanje neustrezno, saj pogosto ne gre za tipični intervju – pri tem poudarja pomen primerno zastavljenih in relevantnih vprašanj, ki govorca spodbudijo, da govori čim bolj prosto.

Problematična pri intervjujih je prisotnost raziskovalca, ki praviloma vpliva na jezik informanta, cilj pa je proučevati, kako ljudje govorijo, kadar niso opazovani. Ta t. i. opazovalčev paradoks (ang. *observer's paradox*) zbuja pomisleke o avtentičnosti govora, ki je med intervjuji posnet kot primer govorčevega naravnega jezikovnega obnašanja (Labov 1972: 209, 61). Pri terenskem raziskovalnem delu k preseganju opazovalčevega paradoksa, torej k manj nadzorovanemu in bolj avtentičnemu govoru, lahko prispevajo različni dejavniki: prekinitve in premori, zaradi katerih informant predvideva, da trenutno ni intervjuvan, zastavljanje vprašanj in spodbujanje tem, ki zbujejo močna čustva, intervjuvanje v parih ali skupinah, domačnost spraševalca z informanti, spraševalčeva raba varietete, ki je informantu blizu, ter podobnost spraševalca in informanta glede družbenih meril, kot so starost, družbeni status, etničnost ipd. (prim. Labov 1972: 209–210; Wilson 2010: 73–75). Treba se je zavedati, da so odgovori v intervjuju vedno rezultat interakcije med spraševalcem in informantom. Govor informantov je tako priporočljivo opazovati tudi v njihovem naravnem družbenem okolju, npr. družini ali skupini vrstnikov, pa tudi sicer posvečati pozornost rabi jezika v situacijah brez izrecnega proučevanja (Labov 1972: 43).

Pri morebitnem (samo)snemanju avtentičnega govora v vsakodnevnih okoliščinah je relevantna tudi teorija domen jezikovne rabe, ki jo je na podlagi osnovnega Greenfieldovega modela za analizo diglosije razvil Joshua Fishman (prim. Fishman 1986). Za govorce so domene jezikovne rabe ali jezikovne izbire² relevantne in odločilne pri izbiri jezika oz. jezikovne varietete. Definirane so kot abstraktni konstrukti, ki jih določajo pripadajoči kraji, odnosi in teme. Vrsta in število domen sta različna glede na govorno skupnost in kulturo, primeri zanje so družina, sooseska, delovno mesto, cerkev, javna uprava (Werlen 2004: 335). Na jezikovno rabo v

² Albina Nečak Lük (1989: 79) jih imenuje domene/področja jezikovnega obnašanja oz. govorne domene.

posamezni domeni tako bistveno vplivajo posamezni odnosi in vloge znotraj skupnosti ter kraj pogovora, ki ima lahko velik vpliv na temo pogovora in odnos med sogovorci (prim. Fishman 1986: 443–444).

V osrednjem delu prispevka je v 2. poglavju pregled kontekstov snemanja v tujih in v 3. poglavju v domačih raziskavah.³

2 Pionirski sociolingvistični intervju Williama Labova

William Labov je v svoji pionirski študiji na primeru reprezentativnega vzorca newyorške govorne skupnosti proučeval variantnosti petih fonoloških spremenljivk. Na podlagi njegovih raziskav se je v sociolingvистики uveljavila klasifikacija petih različnih kontekstov in s tem stilov,⁴ generiranih v okviru sociolingvističnega intervjuja – gre za jezikovno variantnost pri posameznem govorcu (ang. *intra-speaker variation*), ki je povezana s stopnjo samonadzora oz. pozornosti, ki jo nameni govoru (Labov 1972: 79–94; prim. tudi Chambers 2003: 6–7):

A Vsakdanji kontekst: običajni (ang. *casual*) stil – Govor v neformalnih situacijah je deležen najmanj samonadzora in naj bi predstavljal vsakdanji običajni oz. vernakularni jezik (ang. *vernacular*), ki ga je Labov prepoznal kot najbolj regularnega in systemskega. Ta običajni oz. spontani stil se v okviru intervjuja pojavi, ko se informant pogovarja z drugo osebo, npr. družinskim članom, ko predstavlja otroške pesmi ali izštevanka in ko prosto pripoveduje o poljubni temi, posebej o izkušnji smrtne nevarnosti – ob podoživljanju resnosti in napetosti situacije se govorec zelo vživi v pripovedovanje in ob tem zmanjša zavestni nadzor nad govorom.

B Intervju: pazljivi (ang. *careful*) stil – Prisotna je določena mera samonadzora, saj intervjuvanec odgovarja na vprašanja v intervjuju, obenem pa veliko pozornosti posveča vsebini odgovorov.

³ Več o opisanih raziskavah in njihovih izsledkih tudi v Bitenc 2016 – o sociolingvističnem intervjuju 44–47, 49–50, o opisanih prvih treh tujih raziskavah 83–92, o raziskavah govorne slovenščine 123–131.

⁴ Pomen stila je prepoznal že John L. Fischer v svoji študiji leta 1958: ugotovil je, da izbira končnice [in] za [ɪŋ] v deležnikih kot *walkin'*, *talkin'* ipd. v govoru šolskih otrok v Bostonu variira od skoraj izključne rabe -ing v formalnih situacijah do prevlade -in v neformalnih intervjujih (prim. Chambers 2003: 5).

C Branje besedila – Intervjuvanec se bolj posveča izgovarjavi, še vedno pa se osredotoča tudi na vsebino. Labov je informantom v branje predložil dve zgodbi; v prvi so bile prisotne glavne fonološke variable, v drugi pa minimalni pari (prim. točko D^o). Obe sta bili napisani kot pripoved najstnika v pogovornem stilu, da bi se bralec kar najbolj vživel in bral čim bolj naravno.

D Branje seznamov besed: formalni (ang. *formal*) stil – Branje besed v izolaciji predstavlja še dodaten korak v smeri bolj formalnega konteksta. Za proučevanje variabil (r), (eh) in (oh) je Labov intervjuvancem v branje dal imena dnevo in mesecev, seznam besed s podobnim ali enakim segmentom in seznam minimalnih parov iz besedila, ki so ga že prej prebrali, potem pa je informante prosil za presojo, ali so besede izgovorili enako, kot jih izgovarjajo sicer.

D^o Branje minimalnih parov – Besede so razvrščene v pare glede na fonološke podobnosti, zato je vsa govorčeva pozornost usmerjena na izgovor elementa, ki se v besedah posameznega para razlikuje.

3 Sodobnejše tuje sociolingvistične raziskave

Predstavljene so izbrane sodobnejše raziskave v jezikovnih skupnostih, v katerih so sociolingvistični profili bolj sorodni slovenskemu in ki nudijo vpogled v različne načine pridobivanja posnetkov za proučevanje govorne variantnosti pri posameznih govornikih. Brit Mæhlum (1986) in James Wilson (2010) sta pri raziskavi govora Novežanov s podeželja v Oslu oz. Moravce v Pragi vzpostavila dva konverzacijska konteksta, v katerem sta dve različni osebi v različnih varietetah z informanti vodili vsaka en pogovor, pri raziskavi govora prebivalcev nemškega podeželskega kraja Erftstadt-Erp je bil govor zajet v treh različnih kontekstih z različno stopnjo formalnosti (Besch 1981, 1983), pri nizozemskih informantkah iz treh narečnih področij Flandrije v petih govornih situacijah (Ghyselen 2016; Ghyselen in De Vogelaer 2022), pri Avstrijcih različnih narečnih skupin v šestih (Lenz idr. 2019; Fanta-Jende 2020), pri prebivalcih Budimpešte pa so izvedli sociolingvistične intervjuje in uporabili nekatere druge raziskovalne pristope (Kontra 1995, Váradi 1998).

3.1 Dva konverzacijska konteksta pri Norvežanih v Oslu in Moravcih v Pragi

Na podoben način sta raziskavo osnovala Brit Mæhlum (1986) in James Wilson (2010), ki sta proučevala govorno variantnost Novežanov s podeželja v Oslu oz. Moravcev v Pragi. Oba sta vzpostavila dva konverzacijska konteksta, v katerem sta dve različni osebi v različnih varietetah z informanti vodili vsaka en pogovor. Pri Mæhlum (1986) je v prvem intervjuju njena prijateljica iz Osla uporabljala standardno norveščino, v drugem pa avtorica sama podobno narečje kot informanti. Pogovarjali so se o izkušnjah s preselitvijo v mesto, o jeziku in podobno. Proučevala je trinajst izbranih variabel pri dveh moških in dveh ženskah. Pri Wilsonu (2010) je prvi dvajsetminutni neformalni klepet z moravskimi študenti v študentskih domovih v Pragi vodila govorka v splošni češčini, drugega pa raziskovalec sam v standardni češčini z nekaj nestandardnimi oblikami, ki so v rabi tako na Češkem kot na Moravskem. Slednji je bil razdeljen na dva dela: prvi je vseboval pogovor o življenjskem stilu in različnih vidikih informantovega družbenega življenja, na podlagi katerega so izračunali indeks mrežne integracije v gostiteljski skupnosti (*ang. network integration*); drugi se je posvečal stališčem glede ljudi in kulture na Češkem in Moravskem, različnim varietetam češčine in samooceni govora po selitvi v Prago. Analiza se je osredotočala na šest variabel, reprezentativni vzorec 39 študentov je bil stratificiran glede na štiri neodvisne družbene variable: izvorno regijo, spol, čas bivanja v Pragi in družbeno mrežo.

3.2 Trije konteksti pri prebivalcih nemškega podeželskega kraja Erftstadt-Erp

V okviru projekta Erp ali Jezikovno obnašanje v podeželskih skupnostih (prim. Besch 1981, 1983) so raziskovalci dokumentirali in analizirali govor nemškega podeželskega kraja Erftstadt-Erp, iz katerega veliko prebivalcev dnevno migrira v Köln in druga bližnja mesta. Prvi kontekst za snemanje govora je bil prosti vsakdanji pogovor najmanj dveh znancev ali prijateljev; drugi intervju, voden v visoki nemščini; tretji pa test govorne ocene in govorne zmožnosti. Pri prvem so informante spodbujali k čim bolj običajnemu pogovoru, v katerega se je raziskovalec poskušal čim manj vpletati, da bi dobili res čim bolj spontani in vsakdanji način govora med domačini. Pri drugem so se morali govorniki osredotočati na teme, ki jih je uvajal raziskovalec, pri čemer naj bi sprememba sogovorca, teme in oblike pogovora povzročila tudi spremembo jezikovne varietete v smer standardnega

jezika. Tretji kontekst je vseboval uvodno predavanje, predvajanje testnih posnetkov in razvrščanje različnih komunikacijskih situacij, vprašanja o jeziku, jezikovnih stališčih, (samo)oceni jezikovne rabe, vrednosti in funkciji različnih varietet, posebej narečja, v različnih okoliščinah, tudi navajanje dnevov tedna in številke od 1 do 10. Kvantitativno analizo 10 jezikovnih variabel so opravili na govorni produkciji 20 govorcev, izbranih iz večjega reprezentativnega vzorca, ki je vključeval delavce, uslužbenke, uradnike, ljudi v samostojnih poklicih in študente.

3.3 Pet govornih situacij z različno stopnjo formalnosti pri nizozemskih informantkah iz treh narečnih področij Flandrije

V okviru doktorske raziskave Anne-Sophie Ghyselen (2016) so nizozemske informantke posnely v petih govornih situacijah: (1) test narečja – 5 min, (2) test standardnega jezika – 5 min, (3) pogovor s prijateljem iz istega mesta – 1h, (4) pogovor s prijateljem iz drugega narečnega področja – 1h, (5) sociolingvistični intervju z neznanim sogovorcem iz drugega narečnega področja – 30–45 min. V testih narečnega in standardnega jezika so informanti slišali primere povedi v standardni nizozemščini oz. lokalnem narečju, ki so jih morali prevesti v narečje starejših ljudi njihovega kraja oz. standardno nizozemščino, kot jo lahko slišijo v informativnih oddajah. Med sociolingvističnim intervjujem so bili pridobljeni podatki o jezikovnem ozadju govork in njihovi percepciji lastne jezikovne rabe in jezika v Flandriji na splošno. Korpus vključuje posnetke 30 visoko izobraženih flamskih žensk iz treh krajev na treh različnih narečnih področjih Flandrije in dveh starostnih skupin – polovica je bila v času raziskave starih med 25 in 35 let in polovica med 50 in 65 let. Preučevali so distribucijo 29 fonoloških in morfosintaktičnih variabel, in sicer s korespondenčno in grozdno analizo. Kvantitativno analizo so dopolnili s kvalitativno analizo podatkov iz intervjujev, v katerih so obravnavali izbrane teme o lastnem govoru, jezikovnih varietetah in stališčih do njih (prim. tudi Ghyselen in De Vogelaer 2022).

3.4 Šest govornih situacij pri Avstrijcih različnih narečnih skupin

V okviru projekta Nemščina v Avstriji: Variantnost – Kontakt – Percepcija⁵ so raziskovalci uporabili različne metode izvabljanja govora, da bi zajeli tako »standardizirane« kot »proste« govorne oblike. Poslušili so se uveljavljenih pristopov,

⁵ <https://www.dioe.at/en/>, dostop 23. 6. 2023.

kot je snemanje pogovorov in branja, ter bolj inovativnih, kot so računalniško podprti testi jezikovne produkcije (ang. *language production tests*). Metodologija je podobna kot v drugih variantnostnih projektih za nemščino (npr. nemškem projektu o regionalnih jezikih REDE⁶), kar omogoča primerljivost med študijami; obširnejša metodološka razprava je v Schmidt in Herrgen 2011. Govor informantov so posneli v šestih različnih situacijah z različno stopnjo formalnosti: gre za dve bralni nalogi (besedilo *Severni veter in sonce*⁷ ter seznam besed v naključnem vrstnem redu), dva prevoda (t. i. Wenkerjevih stavkov,⁸ ki so jih predvajali na posnetku, in sicer iz narečja v standard in iz standarda v narečje) in dva pogovora (formalni intervju z akademikom in neformalni pogovor med prijatelji brez prisotnosti raziskovalca). Med drugim se raziskovalci posvečajo variantnosti in spremembam narečnih varietet v realnem in navideznem času, govornim repertoarjem in variantnostnim spektrom med narečnimi in standardnimi varietetami ter vplivni moči mest, posebej Dunaja in Gradca. Proučujejo vse jezikovne ravnine, od izgovorjave (fonetika in fonologija) do slovnice (skladnja in morfologija) in besedišča, za analizo uporabljajo različne statistične metode. Prvotno so raziskovalci pridobili posnetke 160 narečnih govorcev iz dveh starostnih skupin – mlajši izobraženci v starosti od 18 do 35 let in govorce, starejše od 65 let, ki ustrezajo kategoriji »NORF« (ang. *non-mobile, old, rural, female* – nemobilne, starejše, podeželske, ženske) na 40 lokacijah v Avstriji, da bi analizirali raznolikost in dinamiko posameznih narečij (Lenz idr. 2019; Fanta-Jende 2020).

3.5 Sociolingvistični intervjuji in drugi raziskovalni pristopi pri prebivalcih Budimpešte

Budimpeški sociolingvistični intervju, BUSZI/BSI⁹ je obsežen projekt, v katerem so zbirali podatke o madžarskem govornem jeziku v Budimpešti s pomočjo posnetih intervjujev, izvedenih na reprezentativnem vzorcu prebivalcev Budimpešte. Projekt vključuje različne raziskovalne pristope in instrumente: sociolingvistične intervjuje, avdio in video posnetke skupinskih srečanj (npr. kartanja), posnetke, ki so bili

⁶ <https://www.regionalsprache.de/en/>, dostop 23. 6. 2023.

⁷ Gre za starodavno pripovedko, ki se v jezikoslovju pogosto uporablja za ilustracijo in izvajanje različnih glasovnih realizacij. Obstaja v prevodih za številne jezike in je bila prvotna uporabljena za ilustracijo mednarodne fonetične abecede (International Phonetic Alphabet, IPA) leta 1949 (Fanta-Jende 2020: 220).

⁸ Gre za metodo pridobivanja gradiva za narečne raziskave, ki jo je uvedel Georg Wenker v 19. stoletju. Šolam je poslal vprašalnike s standardnimi povedmi, ki so jih lokalni informanti prevedli v narečje in jih zapisali s splošno abecedo; pri tem so vključevali določene fonološke in slovnicične značilnosti (<https://regionalsprache.de/en/contents-wenker-questionnaires.aspx>).

⁹ <http://www.nytud.hu/buszi/bsi.htm>, dostop 23. 6. 2023.

prvotno pridobljeni za druge namene (npr. posnetke podjetja za mestni taksi, ki vsebujejo najbolj sproščen govorni stil), analizo jezikovnega sistema govornih skupnosti z opazovanjem z udeležbo in eksperimentalno analizo jezikovne rabe z nekaj hitrimi in anonimnimi anketami. Glavnino posnetega korpusa predstavljajo sociolingvistični intervjuji, ki vključujejo dopolnjevanje povedi, branje odlomkov (počasi in hitro), branje seznamov besed in minimalnih parov, ocenjevanje (ne)pravilnosti določenih elementov, presojanje o enakosti ali različnosti pomena danih parov, opisovanje dogajanja, kot bi poročal radijski poročevalec, ter najmanj pol ure trajajoči pogovor v skladu z določenim pogovornim modulom (ang. *conversation module*), ki zaobjema seznam tem in vprašanj – med njimi so npr. osebni podatki, otroštvo, družina, vera, prijateljstvo, šola, zaposlitev, splav, šale, narodne manjšine, jezik. Raziskovalci so analizirali jezikovno variantnost govorcev iz različnih družbeno-ekonomskih skupin z natančnim sociološkim profilom in preučevali različne jezikovne stile oz. variacije, ki so odvisne od stopnje nadzora nad govorom. Leta 1987 je bilo opravljenih 50 pilotnih intervjujev s po 10 učiteljev, starejših od 50 let, študentov, delavcev, prodajalcev in poklicnih pripravnikov v starosti 15–16 let. V letih 1988–1989 so intervjujali 200 informantov iz budimpeškega podvzorca nacionalnega vzorca (Kontra 1995; Váradi 1998).

4 Raziskave (variantnosti) govorne slovenščine

V slovenskem jezikoslovju se je večina raziskav govorne slovenščine osredotočala na eno od govornih varietet v govornem repertoarju izbranih govorcev, redki pa so primeri analize govorne variantnosti glede na okoliščine. Slovenska dialektologija se večinoma posveča raziskavam čim bolj tradicionalnih narečij. Kot piše Karmen Kenda-Jež (2002: 161), je izbor narečnih govorcev v slovenskih narečnih raziskavah v zadnjih desetletjih blizu idealnemu modelu klasične oz. tradicionalne evropske (in ameriške) dialektologije, veda je še vedno pretežno usmerjena v sinhroni opis jezikovnih sestavov za namen diahronih jezikovnih raziskav. To je v nasprotju z novejšo usmeritvijo tako v zahodnoevropskih kot vzhodno- oz. srednjeevropskih slovanskih državah, kjer se ta veda tesno prepleta s sociolingvistiko, uporablja sociolingvistične postopke izbire govorcev ter raziskuje govorno variantnost glede na znotraj- in zunajjezikovne dejavnike, razvoj vseh nestandardnih varietet, njihovo vlogo in rabo v sodobnih medzvrstnih jezikovnih razmerjih (več o razmerju med sociolingvistiko in dialektologijo tudi v Bitenc 2016: 118–123).

V nadaljevanju so predstavljene raziskave s konteksti snemanja govora, ki se zdijo zanimive s sociolingvističnega vidika. Pri Slovenskem govornem korpusu GOS so različni govorniki posneti v štirih tipih diskurza (Verdonik in Zwitter Vitez 2011), pri dijakinjah iz Mengša in Beltincev v Ljubljani pri šolskih nastopih in v intervjujih (Škofic-Guzej 1994), pri Ljubljančanih v neformalnem javnem in zasebnem sporočanju (Makarova 2004), pri gimnazijcih iz petih narečnih skupin pri šolskih nastopih (Zemljak Jontes in Pulko 2011), pri geografsko mobilnih osebah z Idrijskega v različnih govornih domenah v lokalnem in ljubljanskem okolju ter v sociolingvističnih intervjujih (Bitenc 2016a, 2016b) in pri mladih Korošcih v sociolingvističnih intervjujih (Kotnik 2012).

4.1 Štirje tipi diskurza pri različnih govornikih v Slovenskem govornem korpusu GOS

Osrednjeslovenska varieteta Ljubljane in okolice je bila predmet fonoloških raziskav na nereprezentativnih vzorcih govorcev z namenom standardizacije in načrtovanja govornega standarda od druge polovice 20. stoletja naprej. Tudi gradivo za Slovenski govorni korpus GOS večinoma vključuje govor posameznega govornika samo v eni, sicer avtentični situaciji. Glede na tip diskurza oz. govora gre pri 35 % od skupno 1.032.775 zaobjetih besed za javni informativno-izobraževalni diskurz (televizija, radio in osebni stik), 22 % javni razvedrilni (televizija in radio), 15 % nejavni nezasebni (osebni stik in telefon) in 28 % nejavni zasebni (osebni stik in telefon). Pri demografski opredelitvi so regijsko pripadnost pri pridobivanju podatkov o govornikih uskladili z registrskimi območji, pri čemer pa je treba upoštevati, da pod določena registrska območja lahko spada več narečnih skupin (Verdonik in Zwitter Vitez 2011: 47–50, 53).

4.2 Šolski nastopi dijakinj iz Mengša in Beltincev v Ljubljani in intervjuji z njimi

Variantnosti pri posameznih geografsko mobilnih govornikih v različnih okoljih in okoliščinah se je posvečala Jožica Škofic, ki je v magistrski raziskavi (Škofic-Guzej 1994) preučevala oblikovanje pogovornega jezika pri govorkah, ki so iz narečnega govornega okolja, iz Mengša in Beltincev, prišle v srednjo šolo v Ljubljano. Njihov govor je posnela pri šolskem ustnem spraševanju in predstavitvi referatov, osredotočala se je na spremembe na glasoslovni, oblikoslovni in skladenjski ravni. Pri analizi se zdi metodološko in teoretsko najbolj problematično iskanje »napak« in

»napačne rabe« glede »na knjižni zborni jezik« (Škofic-Guzej 1994: 575–576). Avtorica je z informantkami opravila tudi intervjuje o odnosu do slovenščine in njenih varietet ter preklapljanju v različnih govornih položajih.

4.3 Neformalno javno in zasebno sporočanje pri Ljubljančanih

Irina Makarova (2004) je analizirala jezikovno variantnost sodobne ljubljansčine na podlagi več posnetkov v neformalnem javnem in zasebnem sporočanju – potopisnem predavanju in pogovoru oz. intervjuju; osredotočala se je predvsem na realizacijo ljubljanskih lokalizmov. Za potrditev in ponazoritev ugotovitev avtorica podaja rezultate ankete med sto študenti 1. in 3. letnika slavistike FF UL o jezikovnem prilagajanju sogovorniku. Opis govora samo v enem položaju ne omogoča resnične študije narave in obsega variantnosti oz. primerjave govora posameznika glede na okoliščine. Pri anketi je treba upoštevati tudi vprašljivo zanesljivost samoocen in specifičnost študentov jezikoslovja kot anketirane skupine.

4.4 Šolski nastopi gimnazijcev iz petih narečnih skupin

Melita Zemljak Jontes in Simona Pulko (2011) sta preučevali, ali se mladostniki različnih narečnih govornih področij izbranega vzorca v šoli med seboj sporazumevajo z rabo lastnega narečja ali pri tem prestopijo h knjižni ali drugi normi. Posnetih 10 različnih besedil, od katerih jih je bilo pol pripravljenih vnaprej in nato branih, pol pa pripravljenih konceptualno in izvedenih bolj ali manj spontano. Opis govornih dogodkov, ki so jih v sklopu obravnave socialnih zvrsti jezika pri urah slovenščine snemali s kamero in minidiskom, zbuja pomisleke glede spontanosti in avtentičnosti proučevanega govora, skopi podatki o govorcih (gimnazijci vseh štirih letnikov iz petih narečnih skupin) pa ne povejo prav dosti o stopnji in načinu njihovega (ne)prilagajanja v obravnavanih okoliščinah.

4.5 Različne govorne domene v lokalnem in ljubljanskem okolju pri govorcih z Idrijskega

Prva variantnostna študija v slovenščini je doktorska raziskava Maje Bitenc (2016) o govorni variantnosti pri govorcih z Idrijskega, ki se šolajo ali delajo v Ljubljani. Z namenom proučevanja vsakdanjega spontanega govora in zmanjšanja vpliva opazovalčevega paradoksa je pet informantov, tri ženske in dva moška, trije odrasli in dva dijaka, samosnemalo svoj govor s pomočjo digitalnega diktafona, ki so ga

nosili v torbici okoli vratu, in sicer en dan od jutra do večera. Raziskovalni korpus sestavlja izbor izsekov iz celotnega nabora posnetkov glede na različna merila: predvsem sogovorce in temo pogovora, pa tudi čas, kraj in formalnost situacije.

Ker so vsakodnevna dinamika in življenjske vloge pri posameznih informantih zelo različne, so različne tudi okoliščine nastanka posnetkov in opredeljene domene. Za namen analize so bili izbrani karseda reprezentativni izseki za različne okoliščine, v skupni dolžini 2 ur, 31 minut in 30 sekund od skupno skoraj 47 ur posnetkov. Pri odraslih informantih sem večino posnetkov uvrstila v štiri domene: *sodelavci, vsakdanja tema in sodelavci, strokovna tema* – ti posnetki so bili posneti v dopoldanskem času v službi v Ljubljani ter *lokalni bližnji in otroci* s posnetki v popoldanskem času v lokalnem okolju. Nekaj pogovorov z drugimi sogovorniki, kot so npr. natakarji, prodajalci, smo glede na njihov izvor in kraj pogovora uvrstili v domeni *lokalni neznanci* in *neznanci od drugod*. Pri dijakih imamo zaradi različnih okoliščin bolj različne domene; pri obeh *lokalni bližnji* in *učitelji javno* (pri dijaku tudi *učitelji zasebno* in *molitev* s kratkim posnetkom skupne molitve v cerkvi), pri dijakinji še domeni *sostanovalka* in *vrstniki od drugod*, pri dijaku pa za vrstnike od drugod *šolski vrstniki* in *vrstniki v glasbeni šoli*.

Kot dopolnitev k variantnosti študiji so bili z informanti opravljeni sociolingvistični intervjuji, v katerih smo se posvečali zaznavi in refleksiji lastnega govornega obnašanja informantov, njihovim jezikovnim stališčem, povezavi med jezikom in identiteto ter izkušnjam z jezikovno rabo.

Velika količina in raznovrstnost posnetkov, pridobljenih s tehniko celodnevne samosnemanja, omogoča analizo avtentičnega govora v vsakodnevni situacijah z različnimi sogovorniki. Obdelava tako obsežnega korpusa je po drugi strani zamudna, precej je tudi neuporabnih delov z veliko hrupa. Obenem glede na različne življenjske dinamike informantov večkrat težko dobimo primerljivo količino gradiva in primerljive kontekste. Pozornost je pri tovrstnem zajemu podatkov treba nameniti tudi etičnim pomislekom. Posnetki pri izbranih idrijskih informantih so ponujali obilo možnosti za natančnejše študije vsakdanjega stila oz. neformalnega spontanega govora, manjkalo pa je gradiva za proučevanje bolj formalnih govornih stilov.¹⁰

¹⁰ V podoktorski raziskavi, v kateri analiziram govor migrantov iz Ribnice in Maribora v Ljubljani, je pri nekaterih informantih, npr. profesorjih in novinarjih, več tovrstnega gradiva.

4.6 Sociolingvistični intervjuji z mladimi Korošci

Katja Kotnik (2012) je v raziskavi variantnosti govorjene slovenščine s šestimi mobilnimi mladimi Korošci, tremi moškimi in tremi ženskami, izvedla sociolingvistične intervjuje; govor je analizirala na vseh jezikovnih ravneh. Treba je poudariti, da gre za analizo variantnosti znotraj intervjuja, ki jo prepleta z opisovanjem variantnosti v življenju na podlagi samoocen informantk in informantov. Dobrodošlo bi bilo preveriti, kako informanti v resnici govorijo v različnih situacijah, z različnimi sogovorci, in za ta namen definirati variable oz. spremenljivke.

5 Zaključek

V prispevku so predstavljeni različni načini in konteksti snemanja oz. opazovanja govorjenega jezika za raziskovalne namene v sociolingvistiki, od pionirskih sociolingvističnih intervjujev Williama Labova v 60. letih 20. stoletja do pristopov v sodobnejših raziskovalnih projektih. Glede na namen in obseg posameznih študij se ti seveda razlikujejo. V opisanih tujih raziskavah gre za snemanje v različnih situacijskih kontekstih z različno stopnjo formalnosti, posebej ustvarjenih za namen raziskave. Pri Labovu isti raziskovalec znotraj intervjuja generira pet različnih kontekstov, ki naj bi pri govorcu pripomogli k različni stopnji pozornosti na govor in izvajali različne stile, od čim bolj prostega (po)govora z vsakdanjim stilom do branja minimalnih parov s formalnim. V drugih opisanih študijah spremembo jezikovne rabe, večinoma s poudarkom na kontinuumu od narečne do standardne varietete, vključujoč jezikovno zmožnost tako v najbolj akrolektalnem kot bazilektalnem govornem stilu informantov, skušajo spodbuditi z različnimi pristopi, s spremembo sogovorca in teme pogovora oziroma komunikacijske oblike. Raziskovalci ustvarijo od dva do šest kontekstov, npr. dva pogovora z dvema različnima sogovorcema, ki uporabljata različni varieteti, ali različne situacije od branja besed ali besedil preko intervjuja do prostega pogovora z lokalnimi prijatelji. Nekatere raziskave (npr. budimpeškega govora, Kontra 1995; Váradi 1998) vključujejo opazovanje jezikovne rabe v avtentičnih situacijah. Na kontinuumu med narečjem in standardnim jezikom se načeloma z naraščajočo formalnostjo in javnostjo situacije ter večjo pozornostjo, posvečeno govoru, viša delež standardnih variant, ki je vsakič odvisen tudi od udeležencev v komunikaciji.

Večina dosedanjih študij v variantnostni sociolingvistiki, tudi opisane v tem prispevku, se osredotoča pretežno na proučevanje fonoloških variabel. Za analizo drugih ravnin, npr. sintaktične, so potrebne prilagoditve in razširitve teoretičnih in metodoloških pristopov. Med novejšimi je npr. test jezikovne produkcije, ki omogoča pridobitev zadostne količine podatkov tudi za tovrstne analize (prim. Lenz idr. 2019).

V opisanih slovenskih raziskavah so raziskovalke govorce opazovale oz. snemale v avtentičnem okolju, v različnih vsakodnevnih situacijah, med drugim v šoli, na potopisnem predavanju ali med prijatelji, oz. so govorci svoj govor posneli sami.

Pri analizi je vedno treba upoštevati način pridobivanja podatkov in pri skonstruiranih kontekstih tudi verjeten vpliv eksperimentalne situacije. Tudi pri samosnemanju avtentične vsakodnevne komunikacije (kot pri Bitenc 2016), kjer je učinek opazovalčevega paradoksa zmanjšan na najmanjšo možno mero – informanti po svojih zagotovilih na snemanje hitro pozabili in mu niso posvečali posebne pozornosti – njegovega učinka ne moremo povsem izključiti, o čemer pričajo tudi posamezni komentarji informantov.

Proučevanje variantnosti pri manjšem številu informantov v različnih okoliščinah in z različnimi sogovorci skupaj z metajezikovnimi podatki iz sociolingvističnih intervjujev omogoča poglobljene in celostne študije primerov, upoštevajoč kompleksnost socio- in psiholingvistične situacije pri vsakem posamezniku. Za proučevanje večjih vzorcev pa je treba zajeti različne čim bolj primerljive kontekste, ki omogočajo raziskovanje karseda avtentičnega govora in jezikovnih repertoarjev posameznih govorcev kot proučevanje vzorcev variantnosti med različnimi govorci oz. v govorni skupnosti.

Literatura

- Allan BELL, 1984: Language Style as Audience Design. *Language in Society* 13/2, 145–204.
- Werner BESCH (ur.), 1981: *Sprachverhalten in ländlichen Gemeinden: Ansätze zur Theorie und Methode: Forschungsbericht Erp-Projekt: Band 1*. Berlin: E. Schmidt.
- Werner BESCH (ur.), 1983: *Sprachverhalten in ländlichen Gemeinden: Dialekt und Standardsprache in Sprecherurteil: Forschungsbericht Erp-Projekt: Band 2*. Berlin: E. Schmidt.
- Maja BITENC, 2016: *Z jezikom na poti med Idrijskim in Ljubljano*. Ljubljana: Znanstvena založba Filozofske fakultete.
- Jack K. CHAMBERS, 2003 [1995]: *Sociolinguistic Theory: Linguistic Variation and Its Social Significance*. Malden, Oxford: Blackwell.

- Johanna FANTA-JENDE, 2020: Varieties in Contact: Horizontal and Vertical Dimensions of Phonological Variation in Austria. *Variationist Linguistics meets Contact Linguistics*. Ur. Alexandra N. Lenz, Mateusz Masek. Göttingen: Vienna University Press. (Wiener Arbeiten zur Linguistik, 6). 203–240.
- Joshua A. FISHMAN, 1986 [1972]: Domains and the Relationship between Micro- and Macrosociolinguistics. *Directions in Sociolinguistics: The Ethnography of Communication*. Ur. John J. Gumperz, Dell Hymes. Oxford, New York: Basil Blackwell. 435–453.
- Anne-Sophie GHYSELEN, 2016: *Verticale structuur en dynamiek van het gesproken Nederlands in Vlaanderen: Een empirische studie in Ieper, Gent en Antwerpen*. Gent: Universiteit Gent, Faculteit Letteren, Wijsbegeerte Vakgroep Taalkunde.
- Anne-Sophie GHYSELEN, Gunther DE VOGELAER, 2018: Seeking Systematicity in Variation: Theoretical and Methodological Considerations on the “Variety” Concept. *Frontiers in Psychology* 9/385, 1–19.
- Jožica GUZEJ, 1989: Vpliv migracij na jezik in govor posameznika. *Jezik in slovstvo* 35/3, 52–57.
- Karmen KENDA-JEŽ, 2002: Model idealnega govorca v slovenskih dialektoloških raziskavah. *Med dialektologijo in zgodovino slovenskega jezika*. Ur. Marko Jesenšek, Bernard Rajh, Zinka Zorko. Maribor: Slavistično društvo Maribor. (Zora 18). 150–165.
- Miklós KONTRA, 1995: On Current Research into Spoken Hungarian. *International journal of the sociology of language* 111/1, 9–20.
- Katja KOTNIK, 2022: *VARIANTNOST GOVORJENE SLOVENŠČINE PRI GEOGRAFSKO MOBILNIM MLADIM GOVORCIL, GOVORKAB IZBRANEGA KOROŠKEGA NAREČJA: Magistrsko delo*. Maribor: Filozofska fakulteta Univerze v Mariboru.
- William LABOV, 1972: *Sociolinguistic Patterns*. Philadelphia, Pennsylvania: University of Pennsylvania Press.
- William LABOV, 1984: Field methods of the project on linguistic change and variation. *Language in Use: Readings in Sociolinguistics*. Ur. John Baugh, Joel Sherzer. Englewood Cliffs, New York: Prentice-Hall. 28–54.
- Alexandra N. LENZ, Ludwig Maximilian BREUER, Matthias FINGERHUTH, Anja WITTIBSCHLAGER, Melanie E.-H. SELTMANN, 2019: Exploring syntactic variation by means of “Language Production Experiments”: Methods from and analyses on German in Austria. *Journal of linguistic geography* 7/2, 63–81.
- Irina MAKAROVA, 2004: Mehanizem jezikovnega prilagajanja in variantnost sodobne ljubljansčine. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*. Ur. Erika Kržišnik. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete Univerze v Ljubljani. (Obdobja, 22). 287–296.
- Brit MÆHLUM, 1986: *Språklige variasjonsmonstre hos innflyttere i Oslo*. Oslo: Novus.
- Albina NEČAK LÜK, 1989: O temeljnem izrazju pri proučevanju jezikov v stiku. *Simpozij Slovenski jezik v znanosti 2: Zbornik prispevkov*. Ur. Ada Vidovič-Muha, Nace Šumi. Ljubljana: Znanstveni inštitut Filozofske fakultete. 121–128.
- Natalie SCHILLING-ESTES, 2008 [2002]: Investigating Stylistic Variation. *The Handbook of Language Variation and Change*. Ur. J. K. Chambers, Peter Trudgill, Natalie Schilling-Estes. Malden, Oxford: Blackwell Publishing. 375–401.
- Jürgen Erich SCHMIDT, Joachim HERRGEN, 2011: *Sprachdynamik: Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: Erich Schmidt Verlag.
- Jožica ŠKOFIC-GUZEJ, 1994: O oblikovanju slovenskega pogovarjalnega jezika. *Slavistična revija* 42/4, 571–578.
- Sali A. TAGLIAMONTE, 2006: *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tamás VÁRADI, 1998: Manual of The Budapest Sociolinguistic Interview Data. Budimpešta: Linguistics Institute, Hungarian Academy of Sciences. (Working Papers in Hungarian Sociolinguistics, 4).
- Darinka VERDONIK, Ana ZWITTER VITEZ, 2011: *Slovenski govorni korpus GOS*. Ljubljana: Trojina, zavod za uporabno slovenistiko.
- Philip C. VERGEINER, Dominik WALLNER, Lars BÜLOW, 2022: Language change in real-time: 40 years of lectal coherence in the Central Bavarian dialect-standard constellation of Austria.

- The Coherence of Linguistic Communities: Orderly Heterogeneity and Social Meaning*. Ur. Karen V. Beaman, Gregory R. Guy. New York: Routledge.
- Iwar WERLEN, 2004: Domäne / Domain. *Sociolinguistics / Soziolinguistik: Volume 1 / 1. Teilband*. Ur. Ulrich Ammon, Norbert Dittmar, Klaus Mattheier, Peter Trudgill. Berlin, New York: Walter de Gruyter. 335–340.
- James WILSON, 2010: *Moravians in Prague: A Sociolinguistic Study of Dialect Contact in the Czech Republic*. Frankfurt am Main: Peter Lang.
- Melita ZEMLJAK JONTES, Simona PULKO, 2011: Govorica mladostnikov v šoli – narečna ali nenarečna? *Globinska moč besede: red. prof. dr. Martini Orožen ob 80-letnici*. Ur. Marko Jesenšek. Maribor: Filozofska fakulteta, Mednarodna založba Oddelka za slovanske jezike in književnosti. 407–420.



IV

Med govorjenim
in pisnim diskurzom

JEZIK INFLUENCERA U KONTEKSTU NOVIH NOVIH MEDIJA

BORKO BARABAN, SNJEŽANA BARIĆ-ŠELMIĆ

Sveučilište Josipa Juraja Strossmayera u Osijeku, Akademija za umjetnost i kulturu,
Osijek, Hrvatska
bbaraban@aukos.hr, sbaric@aukos.hr

Velik tehnološki napredak digitalnog i umreženoga društva omogućilo je stvaranje novih novih medija koji su iznjedrili novu publiku – *prosumere*. Novi novi mediji omogućuju njihovu korisniku istovremeno stvaranje i konzumiranje sadržaja. Upravo taj trenutak obilježit će uspon *influencera* – utjecajnih osoba. Utjecajne osobe, odnosno kreatori javnoga mišljenja, nisu novost, no za razliku od utjecajnih osoba u prošlosti, kraljevskih obitelji, plemstva, političke elite, sportaša i drugih, današnje utjecajne osobe svoju popularnost i utjecaj stječu prvenstveno zahvaljujući razvoju tehnologije, a stavove publike oblikuju stvaranjem objava, *vblogova*, *tweetova* i drugih kanala mrežnih društvenih medija. U ovom se radu jezik influencera analizira u dvije kategorije: leksičku i sintaktičku. Metodološki okvir rada usmjeren je kvalitativnoj metodi analize sadržaja koja se može kvantificirati. To je istraživačka tehnika kojom se klasificira i opisuje komunikacijski sadržaj prema unaprijed određenim kategorijama, odnosno komunikacija će se analizirati na sustavan i kvantitativan način. Nastavno istraživanju koje je provela Styria i agencija Nielsen, prema metodologiji NielsenMedia i alatu InfluenceScope (2021), određen je uzorak istraživanja.

DOI

[https://doi.org/
10.18690/um.ff.4.2024.19](https://doi.org/10.18690/um.ff.4.2024.19)

ISBN

978-961-286-882-6

Ključne besede:
jezik influencera,
mrežna komunikacija,
kvalitativna analiza
sadržaja,
novi, novi mediji,
sintaksa



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.19](https://doi.org/10.18690/um.ff.4.2024.19)

ISBN

978-961-286-882-6

Keywords:

influencers' language,
network communication,
qualitative content analysis,
new new media,
syntax

THE LANGUAGE OF INFLUENCERS IN THE CONTEXT OF THE NEW NEW MEDIA

BORKO BARABAN, SNJEŽANA BARIĆ-ŠELMIĆ

University of Josip Juraj Strossmayer in Osijek, Academy of Arts and Culture, Osijek,
Croatia

bbaraban@aukos.hr, sbaric@aukos.hr

The enormous technological progress of the digital and networked society has enabled the creation of new new media that create a new audience – prosumers. New new media enable users to produce and consume content at the same time. It is precisely this dynamic that characterises the rise of influencers. Influencers or creators of public opinion are not a novelty, but unlike influencers in the past – royalty, nobility, political elite, athletes and others – today's influential persons gain their popularity and influence largely thanks to the development of technology, and they shape audience attitudes through the creation of announcements, vblogs, tweets and other online social media channels. In this paper, the language of influencers' is analysed in two categories: lexical and syntactic. The methodological framework is centred on a qualitative method of content analysis that can be quantified. This is a research technique used to classify and describe communication content according to predetermined categories, i.e. the communication is analysed in a systematic and quantitative way. Following the study carried out by Styria and the Nielsen agency, a research sample was defined using the NielsenMedia methodology and the InfluenceScope tool (2021).



JEZIK INFLUENCERA U KONTEKSTU NOVIH NOVIH MEDIJA

BORKO BARABAN, SNJEŽANA BARIĆ-ŠELMIĆ

Univerza Josipa Juraja Strossmayerja v Osijeku, Akademija za umetnost in kulturo,
Osijek, Republika Hrvatska
bbaraban@aukos.hr, sbaric@aukos.hr

Velik tehnološki napredek digitalne in omrežne družbe je omogočil nastanek novih novih medijev, ki ustvarjajo novo občinstvo – *prosumerje*. Novi novi mediji uporabnikom omogočajo, da hkrati ustvarjajo in porabljajo vsebine. Prav ta dinamika je značilna za vzpon vplivnežev. Vplivneži ali ustvarjalci javnega mnenja niso novost, vendar za razliko od vplivnežev v preteklosti – kraljeve družine, plemstva, politične elite, športnikov in drugih – današnje vplivne osebe svojo priljubljenost in vpliv pridobivajo predvsem zaradi razvoja tehnologije, stališča občinstva pa oblikujejo z ustvarjanjem objav, vblogov, tвитov in drugih spletnih besedil na družbenih omrežjih. V tem prispevku je jezik vplivnežev analiziran z dveh vidikov: leksikalnega in skladnjskega. Metodološki okvir je osredotočen na kvalitativno metodo analize vsebine, ki jo je mogoče kvantificirati. To je raziskovalna tehnika, ki se uporablja za razvrščanje in opisovanje vsebine komunikacije glede na vnaprej določene kategorije, tj. komunikacija se analizira na sistematičen in kvantitativen način. Po študiji, ki sta jo izvedli Styria in agencija Nielsen, je bil raziskovalni vzorec opredeljen z uporabo metodologije NielsenMedia in orodja InfluenceScope (2021).

DOI
[https://doi.org/
10.18690/um.ff.4.2024.19](https://doi.org/10.18690/um.ff.4.2024.19)

ISBN
978-961-286-882-6

Ključne besede:
jezik vplivnežev,
spletna komunikacija,
kvalitativna analiza
vsebine,
novi mediji,
skladnja



Univerzitetna založba
Univerze v Mariboru

1 Uvod – o razvoju medija, medijske, masovne i digitalne kulture

Osnovna je svrha medija prenijeti informacije široj javnosti (Jurčić 2017: 127). Razvojem informacijske i komunikacijske tehnologije mediji su proširili svoje djelovanje i to je donijelo značajne promjene u gotovo svim aspektima ljudskoga djelovanja. Mediji su se od tradicionalnih i konvergiranih tradicionalnih razvili u nove, a potom i u tzv. nove nove medije (Wikipedija, YouTube, Facebook, Twitter i druge društvene mreže). Tako su prvobitni masovni mediji (novine, radio, televizija te kasnije novinski portali, radio i televizija na zahtjev) – koje određuje komunikacija „jedan prema mnogima“ – prešli ponajprije na komunikaciju „jedan prema jedan“ (e-pošta, blogovi, videoigre, *web*), a zatim, posredovanjem novih novih medija, na komunikaciju „mnogi prema mnogima“ (Zgrabljic Rotar 2017: 63). Usporedno s razvojem medija razvijala se i mijenjala i medijska kultura. Nju čine ljudi i ponašanje te ona predstavlja „aktivno i etično djelovanje u procesu medijske komunikacije“ pri čemu se razmjenjuju informacije, umjetnički sadržaji i dr. (Zgrabljic Rotar 2017: 64). Uz medije usko je povezana i masovna kultura i u novije vrijeme digitalna kultura. Ona je pak stvorila novu, interaktivnu publiku koja ima drukčiji odnos prema medijima i koja „se sastoji od mnoštva pojedinaca ili grupa sa specifičnim interesima, ponašanjima i navikama korištenja medija“ (Zgrabljic Rotar 2017: 64). Zbog toga su medijski sadržaji danas demokratizirani i individualizirani i time više podložni zakonima tržišta (Jurčić 2017: 135). Drugim riječima, mediji nude ono što traži tržište, a ono danas u velikoj mjeri traži zabavu i ugodu.

Paul Levinson u svojoj knjizi *New new media* (2012: 2-5) ističe ključne razlike između tradicionalnih, novih i novih novih medija. Naime, tradicionalni mediji, kao i konvergirani i novi mediji (poput novina *The New York Times* i internetskog portala www.nytimes.com) još uvijek imaju uređivački postav u kojem nekolicina ljudi poput urednika i producenata, odnosno ono što prema medijskoj teoriji nazivamo *gatekeeperi*, donose sve odluke u vezi sadržaja koji će se objaviti. Koje informacije, novosti i zabava će se objaviti na stranicama, radijskim postajama, odnosno televizijskim ekranima, u rukama je uredničkog postava. Publika tradicionalnih, konvergiranih i novih medija može s lakoćom primati i konzumirati informacije, ali ih ne može objavljivati. Levinson (2012: 6) daje ključne odrednice novih novih medija koje stubokom mijenjaju poziciju publike i konzumacije medijskoga sadržaja. Temeljna odlika novih novih medija, uz to što su besplatni, jest da omogućuju kreiranje, čitanje i izdavanje sadržaja bilo komu tko je prisutan na mreži. Upravo su ti kreatori, čitatelji, izdavatelji jedna te ista osoba, to su osobe koje proizvode sadržaj

s jedne strane, a s druge strane ga konzumiraju. To su konzumenti i producenti, tzv. *prosumeri* kojih ima na stotine milijuna u svijetu. To je upravo srž novih novih medija i po tome se razlikuju od novih medija poput Amazona, iTunesa, Spotifyja, www.nytimes.com jer daju producentima ogromnu moć. Osnajivanje svih, kao producenata i diseminatora informacija, nastavlja mijenjati način na koji živimo, radimo i zabavljamo se, a u digitalnom vremenu novi novi mediji daju svakomu mnogo izravniji glas u odlučivanju o društvu (Levinson 2012: 6-8.). Novi novi mediji besplatni su za konzumenta, a vrlo često i za proizvođača sadržaja, npr. Facebook, Instagram, TikTok, Snapchat, YouTube. Uspješnost novih novih medija leži u njihovoj međusobnoj simbiozi, čini se kako oni rade jedni za druge. Sadržaj objavljen na blogu s umetnutim YouTube videozapisom može se podijeliti na nekoj od društvenih mreža. No ne treba se zavaravati da se oni ne natječu za našu pozornost. Oni jesu kompetitivni kao što su i sinergijski povezani s konvergiranim i novim medijima.

2 Jezične promjene u novim medijima

Jozić, Baraban i Filipić (2021: 76) navode da je uporaba novih medija prouzročila promjene i u samoj strukturi komunikacije. Osim što je, kao i dosad, ona temeljena na govorenoj i pisanoj riječi, u novije vrijeme gradivnim komunikacijskim elementima postaju slikovni i videomaterijal kojima se prenosi određeni sadržaj. Time se komunikacija usmjerava vizualnom sadržaju koji se lakše i brže uočava i procesuiru čime se uvelike postiže prijenos veće količine informacija u što kraćem vremenu što i jest tendencija moderne civilizacije. Upravo je fotografija postala jednom od najvećih promjena strukture komunikacije jer velikim dijelom nadomješta tekst. Štoviše, u nekim slučajevima ona može čak zauzeti primat u strukturi teksta pa se posljedično doživljava dijelom teksta (Jozić, Baraban i Filipić 2021: 89). Danas se također često govori o interkulturalizmu i multikulturalizmu te o interkulturalnoj komunikaciji, a zatim i o interkulturalnoj kompetenciji koja nužno dovodi do prožimanja različitih jezika i kultura do te mjere da se granice između njih brišu. Tako „jezik postaje kulturom, a kultura jezikom“ (Granić 2009: 22). Novi mediji uvelike doprinose izgradnji interkulturalne komunikacije (Carev 2021: 1), a time i pospješuju nastajanje novih načina komunikacije i promjena u samom jeziku (Mihaljević Djigunović, Cergol i Li 2006). Jedna je od tih promjena tzv. anglicizacija koja je sada već prisutna na svim jezičnim razinama. Strane riječi izjednačavaju se „s posuđenicama, iako nisu prošle nijednu fazu u procesu prilagodbe hrvatskome jeziku“ pa su već potpuno uvriježeni izrazi poput „hendlati, kliknuti, sejvati,

downloadati, izbedirati, spikati, iskulirati, brejknuti, mobingirati, bulingirati...“ (Sapunar Knežević i Togonal 2012). Za sve veći utjecaj engleskog jezika zaslužni su upravo novi mediji (prije svega društvene mreže) na kojima je glavni jezik sporazumijevanja upravo engleski. Osim vrlo česte uporabe anglizama, na leksičkoj razini primjetna je uporaba ostalih posuđenica te neologizama (Filipan-Žignić 2012: 32). U jeziku novih medija također su zastupljeni različiti varijeteti: dijalektizmi, žargonizmi i drugi oblici jezika mladih i jezika različitih struka (Filipan-Žignić 2012: 79-80). Crystal (2006 prema Filipan-Žignić i Turk Sakač 2016) takav jezik naziva *netspeakom* za koji također navodi da ima obilježja i govora i pisma, a Storrer (2001, prema Filipan-Žignić i Turk Sakač, 2016) naziva ga otipkanim razgovorom.

Jedno od prvih istraživanja utjecaja komunikacijskih tehnologija na hrvatski jezik provedeno je od 2000. do 2013. na više od 8 000 tekstualnih poruka (Žic Fuchs, Broz, Tuđman Vuković 2013). Već je ono pokazalo da poruke teže ekonomičnosti, odnosno upotrebljavanju kraćih formi, što je posebno vidljivo u češćoj uporabi (u hrvatskom govoru već gotovo arhaičnog) aorista pa se vrlo često javljaju primjeri poput *Vidjeh te pred faksom.* umjesto *Vidjela sam te pred faksom.* Razlog češće pojave aorista u tekstualnim porukama autori istraživanja vide u njegovoj kratkoj formi pri čemu se istodobno štede i vrijeme i prostor. Specifičnost u oblikovanju SMS poruka vide i u tome što su one najvećim dijelom upućene dobro poznatim osobama (rođacima, prijateljima) pa su skraćivanje poruka i upotreba elipsa uvjetovani i time koliko se sugovornici dobro poznaju. Neka kasnija istraživanja ističu značajnu kreativnost u jezičnom oblikovanju stila, posebice kad je riječ o komunikaciji preko društvene mreže Facebook čime se „nadoknađuje nedostatak pisanoga stila u nekonvencionalnoj komunikaciji u odnosu na govorni“, prije svega stvaranjem novih načina „intoniranja, afektiranja, naglašavanja te izražavanja drugih neverbalnih poruka pisanim načinom“ (Šaravanja i Rajle 2015). Najnoviji radovi (Brzak, 2022) ističu da je jezik društvenih mreža vrlo sličan jeziku blogera.

Jezik se oduvijek mijenjao pod različitim društvenim, političkim i drugim utjecajima, no novi komunikacijski mediji iznjedrili su pojavu specifičnih jezičnih struktura i oblika koje su prihvatili ne samo mladi nego i odrasli, neovisno o spolu, obrazovanju, struci i zanimanju. Tako ostvarena komunikacija prouzročila je u vrlo kratkom vremenu velike promjene „na svim jezičnim razinama: na pravopisnoj, grafostilističkoj, morfološkoj, sintaktičkoj, semantičkoj i leksičkoj razini“ (Filipan-Žignić i Turk Sakač 2016). Pojava prvobitnih SMS poruka, a kasnije i komunikacija putem društvenih mreža poput Facebooka, Twittera, Instagrama i Tik-Toka sve do

YouTubea pokazala je da korisnici u jeziku teže ekonomičnosti i što bržem prijenosu informacija što se najviše očituje u ignoriranju pravopisa, uporabi brojnih kratica, emotikona i anglizama, višestrukih interpunkcijskih znakova ili izostavljanja interpunkcije i drugih grafostilističkih sredstava (Žic Fuchs, Broz, Tuđman Vuković 2013; Filipan-Žignić i Turk Sakač 2016). Za takav je oblik komunikacije specifičan razgovorni stil, odnosno stil u kojem se isprepleću govor, pismo i slika (Turk Sakač i Filipan-Žignić 2019). Nerijetko su prisutni i akronimi, elipse, ortografski pokušaji onomatopeje, vulgarizmi te neleksički izričaji „kojima se uspješno /.../ prenosi dobar dio prozodijaskoga i emocionalnoga sadržaja“, poput naizmjeničnog pisanja malih i velikih pisanih slova, udvostručavanja slova i slogova (Šaravanja i Rajle, 2015), pisanja samo malim ili samo velikim slovima, izostavljanja dijakritičkih znakova i dr. (Šaravanja i Rajle, 2015; Turk Sakač i Filipan-Žignić, 2019). Jozić, Baraban i Filipić (2021) ističu da postojanje različitih oblika objava unutar različitih mreža dokazuje i postojanje objave kao samostalne kratke tekstne vrste. Pritom navode elipsu kao jednu od najučinkovitijih rješenja kojim se odgovara na potrebe same forme objava koje prije svega zahtijevaju kratkoću (posebice na Twitteru, koji je čak ograničio broj znakova u objavi). Autori navode da ona omogućuje skraćivanje, odnosno „sa što manje tekstualnoga sadržaja opisuje što više značenja“. Također ističu da je uvođenje novih elemenata teksta (pr. znakovi # i @) odlika isključivo objave što ju ujedno i obilježava.

3 Standardni jezik i novi novi mediji

Zbog težnje za bržim prijenosom informacija, gramatička i pravopisna točnost u komunikaciji novih novih medija stavljena je u drugi plan. Jezik na društvenim mrežama stoga odstupa od standardnojezične norme na svim razinama (Granić 2009: 24). Štoviše, na društvenim mrežama prevladava, a čak je i poželjna, neformalna komunikacija i razgovorni stil pa u takvom okružju inzistiranje na jezičnoj točnosti nije ni prirodno. Neka starija istraživanja pokazala su da ne postoji bojazan o tome da bi učenici mogli „pokvariti svoju pismenost“ pod utjecajem novih medija (Filipan-Žignić i Turk Sakač 2016), ali neke studije pokazuju i negativne rezultate što pokazuje da je za konkretniji odgovor na pitanje o utjecaju novih medija na pismenost mladih potrebno provesti dodatna sustavna istraživanja (Turk Sakač i Filipan-Žignić 2019). No zasigurno se može postaviti pitanje postaju li mladi indiferentni prema jezičnim pravilnostima u situaciji kad je očito da formalni oblici učenja i jezično uređeni tekstovi sve slabije dopiru do njih, a time na njih imaju i sve manji utjecaj.

Zbog sve veće popularnosti i velikog broja sadržaja različite tematike, ali i mogućnosti istodobnog i neograničenog primanja i izmjena informacija te mogućnosti komuniciranja bilo s kim i bilo gdje, što novi mediji nude (Carev 2021: 24–28), tradicionalni su mediji mladoj publici danas nezanimljivi te se ona sve više usmjerava na društvene mreže i, u novije vrijeme, jutjubere i tiktokere, takozvane influencere (Biloš, Budimir i Jaška 2021). Oni komuniciraju preko blogova, vlogova, *tweetova* i drugih kanala te podržavaju stavove svoje publike, poistovjećuju se s njom, ali i uvelike utječu na njezine stavove (Biloš, Budimir i Jaška, 2021: 59). Mogu se baviti potpuno različitim temama i objavljivati sadržaje iz različitih područja: svakodnevnog života, mode, hrane, kućnih ljubimaca, ljepote, majčinstva, glazbe, sporta... (Brzak, 2022: 11). Još su Katz i Lazarfeld (1955., prema Biloš, Budimir i Jaška, 2021: 59) zaključili da mišljenje lidera (*opinion-leader*) kao posrednika komunikacije ima veći utjecaj od poruke prenesene masovnim medijima. Pritom je *opinion-leader* „osoba koja je autoritet u nekom području i u koju pojedinac ili grupa ima povjerenja...“ i ima moć utjecati na odluke drugih (Katz i Lazarfeld, 1955, prema Biloš, Budimir i Jaška, 2021: 59). Danas su tu ulogu velikim dijelom preuzeli influenceri.

Istraživanje koje su proveli Biloš, Budimir i Jaška od 25. rujna do 1. listopada 2019. na 107 ispitanika, kojemu je cilj bio istražiti značaj influencera iz perspektive internetskih korisnika i društvenih mreža u Republici Hrvatskoj, pokazalo je da ispitanici „nisu osobito skloni vjerovati“, odnosno „da je povjerenje u preporuke influencera na ispodprosječnoj razini“, a više od 65 % ispitanika ocijenilo ih je u većoj mjeri neiskrenima (Biloš, Budimir i Jaška, 2021: 63). Ipak, činjenica je da njihovo mišljenje i djelovanje danas utječe na mišljenje velikog broja ljudi, posebice mladih (Brzak, 2022: 14). Istraživanja jezika novih i novih novih medija uglavnom se zasnivaju na pisanim izvorima. Rijetko analiziraju jezik influencera, i to uglavnom stavljajući naglasak na njihovu uporabu anglizama.

3.1 Jezična analiza pisanih i transkribiranih objava influencera

U ovom radu izdvajaju se odabrana jezična odstupanja dvaju influencera na temelju kojih se, zbog ograničenosti količine teksta, ne mogu izvoditi opći zaključci, ali navedeni primjeri mogu poslužiti kao predistraživanje za neku opsežniju analizu. U pitanju su pisane i transkribirane (pisane objave koje prati zvučni zapis, tj. autor ih i izgovara) objave Elle Dvornik (38 objava) i Marka Vuletića (14 objava). Odabir tih

influencera temeljen je na istraživanju koje je provela Styria i agencija Nielsen, prema metodologiji NielsenMedia i alatu InfluenceScope (2021) (Media Mareting, 1.8.2023.). Do rezultata su došli dubinskom analizom i velikim istraživanjem koje je Styria napravila u suradnji s agencijom Nielsen. Najutjecajniji influenceri izabrani su u devet različitih kategorija, i to na trima društvenim mrežama – Instagram, YouTube i TikTok.

Alat InfluenceScope temelji se na jedinstvenoj metodologiji i algoritmu koji je razvio Nielsen Media. Scorecard se sastoji od četiriju glavnih varijabli, *reach* – odnosi se na ukupan doseg influencera i kvalitetu pratitelja, *relevance* – odnosno relevantnost, odnosi se na stupanj podudaranja između baze pratitelja pojedinog influencera i ciljane demografije klijenta, *resonance* – uzima u obzir *engagement rate* po objavi i pridruženi sentiment te dimenzija *return* – prosječna procijenjena vrijednost (u novčanom ekvivalentu) po objavi u odnosu na prosječnu procijenjenu vrijednost po sponzoriranoj objavi.

Rezultati su pokazali da su na društvenoj mreži Instagram najutjecajniji influenceri Marko Vuletić i Ella Dvornik. Analizirane su objave na Instagramu i to u razdoblju od siječnja do svibnja 2023. godine. Iz prikupljenih materijala isključene su sponzorirane objave, odnosno oglasi.

Jezična je analiza potvrdila sljedeća odstupanja od hrvatskoga jezičnoga standarda na svim jezičnim razinama:

- izostanak interpunkcijskih i dijakritičkih znakova (nebilježenje točke, upitnika i uskličnika na kraju izjavnih, upitnih i uskličnih rečenica, bilježenje sibilanata umjesto palatala, izostanak zareza u jednostavnim i zavisnosloženim rečenicama...)
- malo početno slovo na početku rečenice i u tekstu objave / neopravdano pisanje velikim početnim slovom općih imenica (*Ponedjeljak, Veljača, split...*)
- neprovođenje smjenjivanja *ije* u *je* u komparativima pridjeva (*svijetao > svijetliji*)
- pisanje prijedloga *sa* uz imensku riječ – zamjenicu, bez odvajanja (*samnom*)
- pisanje negacije sastavljeno s glagolskim pridjevom radnim (*neponovila*)

- dijalektni oblici glagolskoga pridjeva radnoga u muškom rodu (*zaželili, volija, ima, zavalija, posta*), upitnih zamjenica (*šta, ko*), glagola (*nemrem, objavljujen, peren, san, veselin, ulazin, doručkujen*), priloga (*uvik, oduvik*), imenica (*dite, dućan*)
- netočan položaj zanaglasnice (*Niche parfemi su...; ... dva jela su me...*)
- nebilježenje zareza iza zavisne surečenice (*Kad to kažem onda mislim...; Da bi osvojile ovaj poklon trebate...*) kao ni iza imenske riječi u vokativu (*Stari moj sretan ti rođendan...*)
- nepotrebno pisanje zareza u uzročnim rečenicama (*Nije se oglasio, jer nije prihvatilo...*)
- krnji infinitivi (*ić, kupovat, vratit, ispružit...*)
- uporaba nehrvatskih leksema – uglavnom neprilagođenih anglizama (*pauza, outfiti, lineup, slide, staple, caption, postala* (u značenju objavila), *very, proud, turn, page, unisex...*)
- upotreba krnje vezničke skupine umjesto složenoga veznika dopunskih rečenica (*... s obzirom da sam im ambasador...* umjesto *... s obzirom na to da sam im ambasador...*)
- pisanje upitne rečenice s pomoću čestične/vezničke skupine *da li* umjesto čestice *li* kojoj prethodi predikat ili prediktna kopula ako je riječ o imenskom predikatu (umjesto *da li* točno je upotrijebiti ustrojstvo predikat / predikatna kopula + *li*): *da li bi vi probali ovo?* umjesto *Biste li (vi) probali ovo?*
- nerazlikovanje aoristnih oblika glagola *biti* u jednini i množini (1. lice jednine: *ne bi naručila; bi ih pojela* – uporaba oblika *bi* umjesto točnoga oblika *bih*; 2. lice množine: *kako bi osvojili* – uporaba oblika *bi* umjesto točnoga oblika *biste*)
- neprovođenje vokalizacije (*vesel*).

Na temelju se preuzetih objava poznatih influencera može zaključiti da u brznoj komunikaciji, primarno onoj na društvenim mrežama, postoje jezični trendovi koji su postali obilježni: skraćenice, emotikoni, sleng (neformalni izrazi, dijalektni oblici), anglizmi i dr., no za razliku od uporabe hrvatskoga jezika u medijima i novim medijima, ovdje je riječ o novim novim medijima kojima su uglavnom predstavnici utjecajne osobe (tzv. influenceri) pa se postavlja pitanje imaju li oni uistinu dovoljan utjecaj i na jezičnu (ne)kulturu mladih koji su im uglavnom ciljna publika. Nepobitna je činjenica da jezična kultura influencera odražava posebne jezične obrasce, norme i vrijednosti koje su prisutne u njihovu stilu komunikacije. Jezik influencera često je

neformalan i opušten kako bi se postigla bliskost s publikom. Neformalnost omogućava stvaranje osjećaja pripadnosti i prijateljstva s pratiteljima. Influenceri često pokazuju kreativnost u jeziku što može uključivati izmišljanje novih riječi, kreiranje originalnih rečeničnih konstrukcija ili upotrebu igre riječima kako bi privukli pozornost i stvorili jedinstven stil komunikacije. Često svojim objavama (upravo onim izgovorenim) potiču interakciju s publikom kroz jezik kojim se koriste. Postavljaju pitanja, traže povratne informacije, potiču pratitelje na dijeljenje svojih mišljenja i sudjelovanje u različitim aktivnostima. To stvara osjećaj uključenosti i zajedništva. Budući da se komunikacija influencera odvija na digitalnim platformama s ograničenim prostorom ili vremenom, jezik influencera često je prilagođen brznoj komunikaciji i svjetskim trendovima: influenceri su često svjesni trendova i promjena u jeziku koji se rabi na društvenim mrežama i digitalnim platformama. Oni prate te trendove i prilagođavaju svoj jezik kako bi ostali relevantni i angažirali svoju publiku. No koliko ta angažiranost u cilju približavanja ciljnoj publici, uz očito zanemarivanje jezične kulture, može utjecati na jezičnu kulturu mladih? Dakako da jezična kultura influencera može varirati ovisno o specifičnom influenceru, ciljnoj publici i tematici kojom se bavi. Svaki influencer ima svoj stil i način komunikacije koji odražava njegovu osobnost i privlači njegovu publiku.

Valja istaknuti da, za razliku od slavnih osoba koje su obilježene temporalnošću rijetkosti (u smislu medijske neredovite pojavnosti), mrežne su utjecajne osobe, odnosno influenceri obilježeni temporalnošću izobilnosti. Naime, utjecajne osobe, nastavno fizičkoj udaljenosti od svojih obožavatelja, svoj odnos s obožavateljima održavaju putem masovnih medijima u kojima se pojavljuju rijetko, odnosno neredovito i ne svakodnevno. Mrežne slavne osobe pak obilježene su ne samo temporalnošću izobilnosti nego i stalnoga osvježavanja i aktualiziranja svojih digitalnih novomedijskih sadržaja. To doprinosi njihovoj neposrednosti i autentičnosti koja se rađa u njihovoj trenutačnosti što im pak omogućuje stalna povezanost i uvezanost na mreži. Mrežne slavne osobe producenti su, ali i konzumenti *online* sadržaja te proizvode sebe i kao brend i kao robu. U svakome slučaju, njihov interes nije zaštititi svoju intimu i udaljiti se od svoje publike i pratitelja, nego upravo suprotno, njihova je autentičnost u tome što svu svoju svakodnevnicu i intimu prikazuju trenutačno onako kako se ona odvija u realnome vremenu. Njihova privilegija nije distancija, nego upravo bliskost i neposrednost sa svojom publikom i pratiteljima tijekom cijeloga dana, svakoga dana. Takva bliskost i neposrednost stvara predispozicije razvijanja zamišljenog intimnog odnosa, tzv. parasocijalnog odnosa (Rojek 2001: 389). Slavne osobe, pa tako i mrežne slavne

osobe, odnosno ‘persone’, središte su parasocijalnih odnosa. One mogu zahtijevati i postići intimnost s mnoštvom nepoznatih ljudi. Publike poznaju persone kao što poznaju svoje prijatelje. Persona, kroz svoju ulogu, nudi takvoj vezi kontinuitet, redovnost u smislu pojavljivanja te integraciju u rutinu svakodnevice. Većina je parasocijalnih odnosa normalna i zdrava. Te pseudointerakcije uče članove publike grupnim interakcijama i društvenosti. Najpodložniji takvim odnosima jesu oni najusamljeniji i izolirani (Horton i Wohl 1956: 216–219). No, u svakom slučaju influenceri imaju itekako snažan utjecaj (kako im i sam naziv govori) na svoje pratitelje.

Videoposnetek je odličan način za podkrepitve stališč. Ko kliknete »Spletni videoposnetek«, lahko prilpite vdelano kodo za videoposnetek, ki ga želite dodati. Vnesete lahko tudi ključno besedo in v spletu poiščete videoposnetek, ki najbolj ustreza vašemu dokumentu.

4 Umjesto zaključka – pitanja za daljnja istraživanja

S obzirom na dokazan utjecaj influencera na mlade ljude, u nekom bi se opsežnijem istraživanju moglo odgovoriti na sljedeća pitanja: 1. Utječu li objave prepune odnaka od hrvatskoga jezičnoga standarda na pismenost mladih ljudi i mogu li u velikoj mjeri utjecati na usvojena jezična znanja? 2. Ako su mladi ljudi izloženi takvim objavama bez dovoljno kritičkog razmišljanja, može li to utjecati na njihovu pismenost, ali i na opću kulturu? Osim toga, važno je promisliti i o odgovoru na pitanje: Mogu li društvene mreže utjecati na uspostavljanje novih jezičnih normi koje bi bile obilježne upravo Instagramu, TikToku i sličnim platformama ili su te norme već uspostavljene pa je naučena jezična kultura postala nepoželjnim načinom izražavanja kada je riječ o navedenim platformama?

Sukladno tipologiji influencera koji se dijele na mega-, makro- i mikroinfluencere (Biloš et al. (2021), valjalo bi daljnja istraživanja usmjeriti vrstama objava koje se pojavljuju, a nisu bile predmetom ovoga istraživanja. To bi svakako uključilo različite sponzorirane kampanje, influencerski marketing te uporabu jezika u takvim slučajevima. Zanimljivo bi bilo istražiti različite forme komunikacije te uporabe jezika u objavama koje nisu dio tržišnih komunikacija u odnosu na plaćene ili sponzorirane objave. Također, valjalo bi analizom sadržaja istražiti koji su oblici komunikacije najčešći, npr. *story* objave, *reels* ili *post* objave te analizirati upotrebu videosadržaja u odnosu na objavu sa slikom ili ilustracijom uz tekst.

Literatura

- Antun BILOŠ, Bruno BUDIMIR, Sanja JAKŠA, 2021: Pozicija i značaj *influencera* u Hrvatskoj. *CroDiM* 1/4, 57–67.
- Nikola, BRZAK, 2022: Engleski izrazi i anglicizmi u jeziku mađarskih influencera. Neobjavljeni diplomski rad. Zagreb: Sveučilište u Zagrebu, Filozofski fakultet.
- Matea CAREV, 2021: Novi mediji i interkulturalna komunikacija. Neobjavljeni diplomski rad. Split: Sveučilište u Splitu, Filozofski fakultet.
- David CRYSTAL, 2006: *Language and the Internet*. Cambridge: University Press.
- Blaženka FILIPAN-ŽIGNIĆ, 2012: *O jeziku novih medija: Kvare li novi mediji suvremeni jezik?* Split: Redak.
- Blaženka FILIPAN-ŽIGNIĆ, Marija, TURK SAKAČ, 2016: Utjecaj novih medija na jezik mladih u pisanim radovima. *Slavistična revija* 4/64, 436–474.
- Jelena MIHALJEVIĆ DJIGUNOVIĆ, Kristina CEGOL, Qingmin LI, 2006: Utjecaj medija na nenamjerno usvajanje engleskog vokabulara. *Jezik i mediji – jedan jezik: više svjetova* Ur. Jagoda Granić. Zagreb, Split: Hrvatsko društvo za primijenjenu lingvistiku. 445–452.
- Jagoda GRANIĆ, 2009: Jezik kulturom – kultura jezikom. *Jezici i kulture u kontaktu: zbornik radova*. Ur. Igor Lakić in Nataša Kostić. Podgorica: Institut za strane jezike. 18–26.
- Donald HORTON, Richard WOHL, 1956: Mass Communication and Para-social Interaction: Observations on Intimacy at a Distance. *Psychiatry* 19, 215–229.
- Željko JOZIĆ, Borko BARABAN, Lobel FILIPIĆ, 2021: Linguistic Changes and Communication Structure under the Influence of Social Networks. *Movements – European realities, New Developing Trends*. Ur. Dubravka Oraić Tolić, Ivana Bestvina Bukvić, Ivana Đurđević Babić i dr. Osijek. 74–95.
- Daniela JURČIĆ, 2017: Teorijske postavke o medijima – definicije, funkcije i utjecaj. *Mostariensia* 1/21, 127–136.
- Elihu KATZ, Paul F. LAZARSELD, 1955: *Personal Influence. The Part Played by People in the Flow of Mass Communication*. New York: Free Press.
- Paul LEVINSON, 2012: *New New Media*. London: Pearson.
- Chris ROJEK, 2001: *Celebrity*. London: Reaktion Books.
- Andrea SAPUNAR KNEŽEVIĆ, Marijana TOGONAL, 2012: Hrvatski jezični standard u jeziku javne komunikacije. Uloga medija u oblikovanju hrvatske jezične svijesti i kulture. *Medianali* 12/6, 17–34.
- Angelika STORRER, 2001: Getippte Gespräche oder dialogische Texte? Zur kommunikationstheoretischen Einordnung der Chat-Kommunikation. *Sprache im Alltag: Beiträge zu neuen Perspektiven in der Linguistik*. 439–66.
- Lidija ŠARAVANJA, Lorna RAJLE, 2015: Jezične funkcije i stil fejsbučkoga govora osnovnoškolaca. *Dijete i jezik danas : dijete i mediji = Children and language today : Children and media : proceedings* Ur. Dubravka Smajić in Valentina Majdenić. Osijek: Fakultet za odgojne i obrazovne znanosti Sveučilišta Josipa Jurja Strossmayera u Osijeku. 251–272.
- Marija TURK SAKAČ, Blaženka FILIPAN-ŽIGNIĆ, 2019: Pisanje osnovnoškolaca u novim medijima – izazov za školu i učenike. *Školski vjesnik* 2/68, 439–453.
- Nada ZGRABLJIĆ ROTAR, 2017: Novi mediji digitalnog doba. *Informacijska tehnologija i mediji*. Ur. Ljubica Josić. Zagreb: Sveučilište u Zagrebu. Dostopno na <https://tikz.unizd.hr/Portals/46/PROPI%20I%20DOKUMENTI/Zbornik%20INFORMACIJSKA%20TEHNOLOGIJA%20I%20MEDIJI%202016%20final%20web.pdf>
- Milena ŽIČ FUCHS, Vlatko BROZ, Nina TUĐMAN VUKOVIĆ, 2013: Communication Technologies and Their Influence on Language: the Notion of *Convention* Revisited. *Jezikoslovlje* 1/14, 65–8.

Internetski izvor

<https://www.media-marketing.com/vijesti/oni-su-top-hrvatski-influenceri/> (1. 8. 2023)

KOMENTARJI NOVIC REGIONALOBALA.SI MED GOVORJENIM IN PISNIM DISKURZOM

MAŠA ROLIH

Univerza na Primorskem, Znanstveno-raziskovalno središče Koper, Slovenija
masa.rolih@zrs-kp.si

Komentarji spletnih novic so z razmahom spletnih komunikacijskih okolij, med katere spadajo družabna omrežja, postali način izmenjave mnenj med različnimi uporabniki. Spletni časopisi delijo svoje novice na svojih Facebook profilih in s tem omogočajo komentiranje. Komentarji so večinoma zapisani, jezik teh besedil pa odstopa od jezikovne norme, saj komunikacija pogosto poteka v neknjižnem jeziku. Vloga spletnih komunikacijskih okolij je povezovalna oz. praktičnosporazumevalna, komunikacija med uporabniki pa vsebuje elemente spontanega govora. Komentarji novic, ki jih uporabniki družabnih omrežij, kot je Facebook, zapišejo pod novico, poleg zapisa po izgovoru, leksike različnih socialnih zvrsti, rabe emotikonov in gif-ov, vsebujejo tudi sredstva besedilne semantike (deikte, diskurzne označevalce), ki so pogostejša v govorjenem diskurzu, in posebno skladnjo, značilnejšo za govorjeni diskurz. V zapisanih komentarjih spletnih novic lahko tako opazujemo značilnosti govorjenega jezika, v prispevku pa bodo prikazani predvsem primeri besedilne semantike in skladnje govorjenega jezika.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.20](https://doi.org/10.18690/um.ff.4.2024.20)

ISBN
978-961-286-882-6

Ključne besede:
komentarji spletnih novic,
govorjeni jezik,
spletni diskurz,
besedilna semantika,
skladnja

DOI
[https://doi.org/
10.18690/um.ff.4.2024.20](https://doi.org/10.18690/um.ff.4.2024.20)

ISBN
978-961-286-882-6

Keywords:
web news comments,
spoken language,
online discourse,
semantics,
syntax

REGIONALOBALA.SI NEWS COMMENTS BETWEEN SPOKE AND WRITTEN DISCOURSE

MAŠA ROLIH

University of Primorska, Science and Research Centre, Koper, Slovenia
masa.rolih@zrs-kp.si

The enormous technological progress of the digital and networked society has enabled the creation of new new media that create a new audience – prosumers. New new media enable users to produce and consume content at the same time. It is precisely this dynamic that characterises the rise of influencers. Influencers or creators of public opinion are not a novelty, but unlike influencers in the past – royalty, nobility, political elite, athletes and others – today's influential persons gain their popularity and influence largely thanks to the development of technology, and they shape audience attitudes through the creation of announcements, vblogs, tweets and other online social media channels. In this paper, the language of influencers' is analysed in two categories: lexical and syntactic. The methodological framework is centred on a qualitative method of content analysis that can be quantified. This is a research technique used to classify and describe communication content according to predetermined categories, i.e. the communication is analysed in a systematic and quantitative way. Following the study carried out by Styria and the Nielsen agency, a research sample was defined using the NielsenMedia methodology and the InfluenceScope tool (2021).



1 Uvod – o razvoju medija, medijske, masovne i digitalne kulture

Prispevek obravnava prvine govornega jezika v komentarjih novic spletnega časopisa Regionalobala.si, objavljenih na družabnem omrežju Facebook. Zanima nas, v kolikšni meri so komentarji zapisani na način spontanega govora in katere značilnosti govornega diskurza, predvsem semantike in skladnje, kažejo. Sredstva govornega diskurza bodo analizirana s pomočjo objavljenih komentarjev novic Regionalobala.si (objavljenih na Facebook profilu časopisa), zapisanih v mesecu maju 2023.

Značilnosti zapisov komentarjev spletnih novic lahko razložimo primerjalno s spontanim govorom oz. spontanim govornim diskurzom, v zvezi z zapisi v preučevanih okolij uporabljamo predvsem pojme spletni diskurz, spontani diskurz, spletna komunikacija in spontana komunikacija. Spletni diskurz se uporablja kot realizacija spletne komunikacije, spontani diskurz pa je realizacija spontane komunikacije. Za zapisana sporočila v okviru spletnih komunikacijskih okolij ne moremo uporabiti izraza spontani govor oz. spontani govorni diskurz, saj gre pravzaprav za zapise, primerjalno s spontanim govorom pa jih razlagamo zato, ker vsebujejo ključne značilnosti spontanega govornega diskurza in delujejo kot njegova reprodukcija.

Za spletna komunikacijska okolja je značilna uporabniška, socialno- in funkcijskozvrstna heterogenost, v njih pa uporabniki zavzemajo različne identitetne položaje, vloge in nagovarjajo druge uporabnike, ki so z njimi v enakovrednem in/ali neenakovrednem družbenem razmerju, tako komentarji spletnih novic pogosto vsebujejo tudi značilnosti neknižnih jezikovnih zvrsti (pokrajinski pogovorni jezik, sleng, narečja), značilnejši za govorni diskurz, saj svoj način komuniciranja približujejo neposredni (govorni) komunikaciji. Za splet je značilna tudi prenosniška heterogenost, saj splet omogoča posredovanje tako pisne kot tudi govorne komunikacije. Spletni komentarji tako pogosto vsebujejo veliko referenčnih izrazov (npr. prvoosebni in drugoosebni zaimkov), eliptičnih stavkov (tj. nezapolnjenih stavčnih vzorcev, prekinitiv, premorov, prekrivajočega govora ipd.) deiktov, besedilnih aktualizatorjev in diskurznihi označevalcev.

2 Značilnosti spletnega diskurza

Za splet je značilna heterogenost tako v smislu socialne in funkcijske zvrstnosti kot tudi prenosniške, po drugi strani pa gre predvsem v spletnih komunikacijskih okoljih tudi za mešanje javnega in zasebnega diskurza, vse navedeno pa se kaže tudi v značilnostih jezika (Rolih 2014: 79–81). Opazimo lahko, da elementi neposredne komunikacije prehajajo v jezik na spletu, predvsem družabna omrežja in ostala komunikacijska okolja (npr. forume), tako pa uporabniki dobivajo občutek pristnosti komuniciranja, kljub temu, da to poteka preko spleta (Rolih 2014: 79–81). Splet zaradi možnosti multimedijske narave in sočasnosti komuniciranja ustvarja navidezno zmanjšanje distance uporabnikov, ki se lahko povezujejo in komunicirajo neodvisno od geografskih omejitev, zato imajo občutek neposredne udeležnosti v komunikacijski proces.

Susan C. Herring (v: Schiffrin, Tannen, Hamilton 2006: 622–623) pravi, da je za spletni diskurz značilna raba komunikacijskih sredstev, ki ponazarjajo neposredno komunikacijo oz. elemente nebesedne komunikacije, najbolj znani pa so t. i. emotikoni, ki opisujejo obrazno mimiko (npr. 😊 😞) in tako prikažejo ostalim svoje čustveno stanje, kot npr. veselje, žalost itd. Ponazarjanje ekspresij uporabniki izvajajo tudi tekstovno, s tem da natipkajo izraze čustvovanja, kot npr. <grin>, <yawn> itd., ki ponazarjajo nebesedno in neposredno komunikacijo (Herring v: Schiffrin, Tannen, Hamilton 2006: 623).

Crystal (2001: 6) govori o jezikovni raznolikosti na spletu (»language variety«), saj zanj velja sistem raznolikih lingvističnih značilnosti izražanja, ki jih določajo situacijski faktorji, kot so npr. pisanje in govor, dialekti in sociolekti, žanri itd. Za različna spletna okolja veljajo drugačna pravila in načini izražanja. Množica diskurzov na spletu uporabnikom tako določa različne načine vstopanja v komunikacijsko sfero in različne načine rabe jezika, ki je prilagojen konkretni situaciji.

2.1 Značilnosti spletnih komunikacijskih okolij (družabnih omrežij)

Spletna družabna omrežja so ustvarila nove razsežnosti diskurza in komuniciranja, hkrati pa tudi nove načine druženja, oblikovanja ter vzdrževanja odnosov in načine manifestacije identitete. Moč spleta je tako ravno v preseganju oddaljenosti in možnosti prisotnosti v različnih sferah ter komuniciranja z različnimi akterji.

Posameznik na spletu lahko tako hkrati npr. prebira dnevne novice, se pogovarja s prijateljem z drugega konca sveta, nakupuje, pošilja službeno elektronsko pošto itd.

V diskurzu komentarjev dnevnih novic, deljenih v okviru družabnih omrežij, akterji komunicirajo preko zapisanega koda, pri katerem se govorno dejanje zgodi z zamikom, med katerim se dogaja pisanje oz. pisno sporočanje. Ilokucija je v tem primeru enaka kot pri govornem neposrednem sporočanju, različen je le čas, v katerem se ustvarja sporočilo oz. izmenjava informacij. V tem času imata udeleženca pogovora čas za razmislek o tem, kaj in na kakšen način bosta predala sporočilo, da bo čim bolj ustrezalo ilokucijskemu namenu, pri neposredni komunikaciji pa tega časa ni, nasprotno pa so prisotni razni elementi nebesedne komunikacije, kot npr. govornica telesa, barva glasu, register itd., ki lahko veliko povedo o akterju samem, njegovem sporočilu in namenu komunikacije. Dejstvo je, da splet čedalje bolj postaja medij, ki ni namenjen le hitremu predajanju informacij, ampak tudi druženju in izoblikovala so se spletna družabna omrežja, spletne kavarne in forumi, klepetalnice itd. Cilj oz. ilokucijska vloga komunikacije v teh okoljih je približanje komunikacije neposredni govorni komunikaciji. Udeleženci pogovora hočejo imeti občutek, da se med sabo družijo, čeprav so skupaj le virtualno, spletna komunikacijska okolja pa tako združujejo večinoma uporabnike, ki med sabo komunicirajo na način neposredne komunikacije in objavljajo zapise, ki dajejo vtis simultane odziva. Starostna, regionalna, socialna itd. raznolikost udeležencev v družabnih omrežjih se kaže v rabi različnih jezikovnih izbir, ki so odvisne od konteksta in ilokucijske vloge. Tako so načela vljudnosti, slovnice in prehoda med govornim in pisnim jezikom pogojena s številnimi kontekstualnimi in diskurzivnimi elementi.

Crystal (2001: 16) pravi, da funkcija socialnih omrežij, klepetalnic itd. služi potrebi po večji predvidljivosti, zaupnosti, domačnosti, vse to pa vpliva na jezik. Crystal (2001) za tipično govornico spletnih komunikacijskih okolij uporablja izraz »Netspeak« (tudi »Weblish«, »Internet Language«, »Cyberspeak«, »Computer Mediated Communication« (CMC) itd.), ki označuje interaktivne in dialoške elemente v spletnem jeziku in jezikovne izbire, ki nastanejo zaradi zavedanja prisotnosti občinstva. Gre zlasti za razne termine spletne tehnologije, ki so postali del spletnega slenga (t. i. žargonski slengizmi), npr. 'spam', 'cancel' (izpeljanka 'skenslati') itd., pogovorno/narečno/sociolektalno izrazje, pisanje po izgovoru itd.

S. Kranjc (2003) poudarja, da je za splet (še posebej forume, klepetalnice, družabna omrežja) značilno prodiranje govorne forme v zapisano besedo. Struktura sporočil v diskurzu komentarjev spletnih novic je večinoma precej podobna govorjenim besedilom, še vedno pa gre pri tem za zapisano besedo. Kranjc (2003: 436) tako navaja elemente, v katerih se razlikujejo od besedil, nastalih preko govornega kanala. Kot primer vzemimo drugačen vzorec menjavanja vlog (npr. pri spletnem komuniciranju ni simultane odziva naslovnika).

Besedila v okviru spletnih komunikacijskih okolij imajo tako značilnosti pisnega, govorjenega in multimedijskega diskurza, način njihove realizacije pa je odvisen od posameznega uporabnika (kakšen je namen njegovih sporočil, kako izraža svojo identiteto in vlogo, kolikšna je stopnja njegovega jezikovnega znanja in komunikacijske kompetence itd.) in naslovnika, ki je javnost (ki jo lahko sestavljajo osebe, ki jih sporočevalec pozna, npr. prijatelji na družabnem omrežju Facebook, katerih objave javno komentira, ali pa nepoznane osebe, npr. politiki itd.

2.2 Zapisi v spletnih komunikacijskih okoljih med govorjenim in pisnim diskurzom

Komentarji dnevnih novic, ki nastanejo kot simultani odziv na novico, ki tvori skupno referenco uporabnikov, lahko vsebujejo veliko jezikovnih elementov, ki so značilnejši za govorjeni diskurz, npr. neknjižni izrazi, skladenjske posebnosti, več izrazov neposredne reference, zato združujejo značilnosti obojega, tako zapisanega kot tudi spontanega govorjenega diskurza.

Za spontani govorjeni diskurz je značilno, da uporabnik sproti v komunikacijskem procesu tvori besedila, razen v primerih, ko gre za vnaprej pripravljena besedila, ko gre za branje ali recitiranje/govorjenje na pamet, ki je pravzaprav realizacija oz. govorno podajanje predhodno zapisanega besedila. V procesu pisanja je daljši načrtovalni čas besedila, pri govoru pa krajši, zato so v spontanem govoru pogosta mašila, ponavljanja, premori, elipse, nedokončani stavki, nepravilne stavčne strukture itd. Kljub navidezni spontanosti pa izbiro jezikovnih sredstev pri govoru določa stopnja govorčeve kognicije, saj jezikovne izbire prilagaja stopnji svojega znanja, zanimanja in konteksta oz. upošteva maksimo relevantnosti (Grice 1975: 45–46). Pri tem teorijo relevantnosti, ki temelji na dveh temeljnih pravilih, in sicer: 1. človeška kognicija stremi k maksimalni relevantnosti, pri čemer je pozornost avtomatsko usmerjena k informaciji, ki je dovolj relevantna v določenem kontekstu;

2. vsaka izjava predpostavlja svojo optimalno relevantnost, pri čemer sporočevalec pri naslavljanju naslovnikov sporoča, da je njegova izjava relevantna glede na kontekst, njegove sposobnosti, izbire in posledično vredna naslovnikove interpretacije (Sperber, Wilson 2004: 615–620).

Halliday (1989: 79) pravi, da je govornjeni jezik veliko kompleksnejši kot pisni, neformalna spontana konverzacija pa je slovnično najkompleksnejši diskurz, saj se pojavljajo posebnosti, značilne za govornjeni jezik, kot so npr. vrinjene stavčne strukture, stavčne elipse, samopopravki, premori ali prekinitve, razni miselni preskoki itd. Po drugi strani pa Halliday (1989: 79) omenja leksično gostoto, ki naj bi bila v govornjenem jeziku manjša (npr. manj zloženih besednih zvez, več stavkov/pastavkov/polstavkov), v pisnem pa večja. V pisnem diskurzu, ko sporočevalec in naslovnik nista hkrati prisotna v procesu nastajanja besedila, mora uporabnik uporabiti več besed, besednih zvez, utemeljitev itd., da bo besedilo za naslovnika jasno in razumljivo. V zapisanih besedilih sporočevalec tudi ne ponavlja besed, zato za isto stvar uporabi drugo besedo ali besedno zvezo, pri čemer mora za uspešno komunikacijo upoštevati naslovnikovo poznavanje reference, da bo razumel sopomenskost, medtem ko so v govornjenem diskurzu ponavljanja pogostejša, referenco dopolnjujejo tudi nebesedni elementi, kot je govorica telesa, mimika ipd.

Spontani govor oz. spontani govornjeni diskurz razumemo kot govor, ki je vnaprej nepripravljen in nastaja sproti v komunikacijskem procesu. Toporišič v Slovenski slovnici (2004: 32–35) omenja dve vrsti govornjenega jezika, in sicer neposredno ubesedovalno (t. i. prosto) govornjenje in reproduktivno govornjenje. Neposredno ubesedovalno oz. prosto govornjenje je najtežje in vsebuje pojave, kot so pretrganost, preskakovanje iz ene skladenjske oblike v drugo, delna nestrnjenost in slabša povezanost posameznih delov besedila, njihova medsebojna slabša usklajenost, ponovitve oz. negladkost sporočanja.

J. Miller in R. Weinert (2009: 22–23) navajata značilnosti spontanega govora, ki izvirajo iz sočasnosti tvorjenja sporočila ter kratkoročnega spomina sporočevalca in naslovnika (npr. spontani priklic neke informacije in takojšnja ubeseditiv, kar je v zapisanih besedilih redkeje. Spontani priklic informacije s takojšnjo ubeseditivjo večkrat opazimo v zapisanih komentarjih novic (npr. *Aja, točno, še to moram povedati*), gradijo pa ga glasovni elementi (ki jih v zapisanih komentarjih spletnih novic ponazarjajo npr. velike tiskane črke, ki označujejo kričanje) in neglasovni elementi,

ki jih v zapisanih komentarjih npr. ponazarjajo emotikoni. Miller in Weinert (2009: 22–23) omenjata tudi bolj fragmentarno naravo spontanega govornega jezika, kar se kaže v manj kompleksni skladnji, elipsah, izpuščanju glasov ipd.

V pisnem diskurzu mora sporočevalec eksplicitno podajati informacije in pojasniti okoliščine, ki so naslovniku zaradi oddaljenosti nedostopne, pri govornem sporočanju pa si lahko pomaga z različnimi že prej omenjenimi referenčnimi izrazi, govornico telesa itd. V komentarjih spletnih novic je referenca novica, ki predstavlja skupno védenje uporabnikov, nejezikovno sporočanje pa nadomeščajo drugi nebesedni elementi, kot so emotikoni, kopičenje ločil idr., referenčni izrazi in skladenjske značilnosti pa spominjajo na spontani govorni diskurz, kot bo prikazano v nadaljevanju.

2.2.1 Sredstva besedilne semantike govornega diskurza v spletni komunikaciji

J. Zemljarič Miklavčič (2008) navaja značilnosti sodobnega spontanega govornega jezika, ki jih lahko uporabimo tudi za prikaz značilnosti komentarjev spletnih novic, pri čemer se osredotočamo na sredstva besedilne semantike. Avtorica mednje uvršča deikte, besedilne aktualizatorje in diskurzne označevalce (Zemljarič Miklavčič 2008: 95–97). Deikti so tisti izrazi, ki svoj polni pomen dobijo šele v kontekstu oz. konkretnem govornem položaju in so v tem smislu neke vrste spremenljivke, odvisne od konteksta (Levinson 1983: 139, Levinson 2003: 54, Zemljarič Miklavčič 2008: 95). Z deikti govorec tako usmeri poslušalčevo pozornost na stvarnost, ko gre za pogovor v živo, si lahko pomaga tudi z govornico telesa. Deikti so kot sredstva besedilne koherence prisotni tudi v zapisanih besedilih, se pa v zapisanih komentarjih spletnih novic pojavljajo predvsem tisti, ki so značilnejši za govorni diskurz in referirajo na hkratno udeleženo sporočevalca in naslovnika v govorni situaciji (npr. pri komentiranju novic so prisotni tisti deikti, ki referirajo na sporočevalcev in naslovnikov skupni čas in okoliščine tvorjenja ter sprejemanja sporočila, primer: *V ponedeljek odprejo novo trgovino /.../*, pri čemer se »v ponedeljek« nanaša na dan, ko je bilo sporočilo tvorjeno, zato predvideva sporočevalčevo in naslovnikovo skupno védenje, deiktičnost pa kaže na sočasni diskurz. D. Biber idr. (1999: 267) navajajo, da je ravno za spontani govor zelo značilna deiktičnost v referenci, saj se konverzacija odvija v situaciji, ki si jo govorec (sporočevalec) in poslušalec (naslovník) delita (Biber idr. 1999: 267), torej je več deiktičnega referiranja v situacijah, ko imata sogovornika več skupnega védenja.

Drugo pomembno sredstvo besedilne semantike so besedilni aktualizatorji. V zvezi z besedilnimi aktualizatorji Smolej (2006: 195) pravi, da pretvarjajo (spreminjajo) slovarsko vrednost leksema, pred katerim stojijo, v konkretno referenco (pomen) oz. opravljajo vlogo usmerjevalnih kazalcev na konkretnost/nekonkretnost oz. splošnost predstavljalne reference. Besedilni aktualizatorji so torej besede, ki določijo, ali je leksem konkreten ali splošen, nastopajo pa običajno le v govoreni komunikaciji, saj predvidevajo skupno znanje obeh govorcev. To so besede ali besedne zveze, kot npr. en tak, tisti/tist/uni/un/oni, tale/ta/toti, taprvi, taglavni itd.

Med sredstva semantike govorenega jezika ter koherence in kohezije uvrščamo tudi diskurzne označevalce, ki so eno- ali večbesedni izrazi, ki opravljajo različne funkcije: organizirajo diskurz (označujejo začetek ali konec, povezovanje besedila, menjavo vlog itd.), so sredstvo za preverjanje in izkazovanje pozornosti poslušanja, kažejo na odnos govorca do predmetnosti, njegovo čustveno angažiranost in povezanost diskurza s kontekstom (Zemljarič Miklavčič 2008: 96). Taki izrazi so npr. ne (a ne), ja, aja, mhm, tako da, mislim, (a) veš itd., ki pa jih lahko prepoznamo kot diskurzne označevalce šele v kontekstu.

Pisanski Peterlin (2007: 8–10) diskurzne označevalce pojmuje kot metabesedilna sredstva, ki k vsebini ne dodajajo novih informacij, ampak vstopajo v dialog s prejemnikom besedila in mu vsebinski del pomagajo ustrezno sprejemati. Metabesedilna sredstva so torej tisti elementi, ki se navezujejo na besedilo samo in ne na zunanji svet. Metabesedilo je v tem primeru sopomenka diskurza, ki združuje značilnosti udeležencev komunikacijskega procesa, odnose med njimi, njihovo skupno vedenje in okoliščine.

D. Verdonik (2008: 25–29) navaja naslednje vrste diskurznih označevalcev, in sicer ideacijske (vezniki, prislovi; npr. pol pa, tko da), interakcijske (npr. ne, a ne, a, veš, no, okej), označevalce procesov tvorjenja (npr. bom reku, mislim) in interpretacijske (npr. in tko naprej, pa to, kaj pa to).

2.2.2 Skladenjske značilnosti govorenega jezika v spletni komunikaciji

J. Toporišič (2004: 32–35) navaja ključne skladenjske posebnosti (predvsem frazeološke) neposredno ubesedovalnega govorenja, in sicer 1) popolne ponovitve (npr. »pridi, pridi«), 2) ponovitve nekako soodnosnega pomena (»strudim se, delam«), 3) zveze ponovljenih izrazov s členkoma pa in ne (»prideš, ne prideš, vseeno mi je«),

4) osebek in povedek sta isto (»služba je služba«), 5) zveza dovršnika in nedovršnika (»gledam, ne vidim«), 6) drugi izraz je stopnjevan (»temno, pretemno«), 7) zveza samostalnika in pridevnika iz iste podstave (»avša avšasta«), 8) ponovitve z vezniki (»hodili smo in hodili«), 9) ogovorilni glagolski izrazi (»lej, ne zanima me«), 10) členki (»le kje imaš pamet«), 11) poudarni zaimki, členki in podobno (»prav treba mi je bilo tega«).

Načelo ekonomičnosti sporočanja se v spontanem govornem diskurzu oz. konverzaciji kaže v krajšanju stavkov, tvorjenju eliptičnih struktur in uporabi zaoblik (Zuljan Kumar 2007: 29). Hitrost komunikacije je odvisna od hitrosti kodiranja in dekodiranja sporočil, ki je premo sorazmerna z obsegom skupnega védenja udeležencev interakcije (Zuljan Kumar 2007: 29). V govornem diskurzu govorci sporočajo več kot samo jezikovno sporočilo (npr. če govorci kriči, pomeni, da je razburjen) in uporabljajo kontekstualne namige, ki so sredstva signalizacije kontekstualnih informacij (prozodija, oklevanja, premori, prekrivajoči govor, razlike v hitrosti in glasnosti, zamenjava jezika ali jezikovne zvrsti) (Zuljan Kumar 2007: 29). Eliptične strukture so v okviru spletne konverzacije seveda manj prisotne, ker so si udeleženci relativno oddaljeni oz. si ne delijo skupnega prostora in je njihovo skupno védenje manj obsežno (razen v primerih, ko npr. zapisu, ki je v eliptični obliki, dodajo fotografijo, video ali povezavo).

3 Značilnosti govornega jezika v komentarjih dnevnih novic

Spletni komentarji ob objavah dnevnih novic delujejo na način konverzacije, in sicer neformalne ter do neke mere spontane, kot bo podrobneje razloženo v nadaljevanju. To pomeni, da uporabniki teh okolij ne upoštevajo nujno vseh pravil pisnega diskurza. Načeloma razložijo stvari toliko eksplicitno, da bodo ciljni uporabniki to razumeli. Nekateri zapisi vsebujejo podatke, ki jih morajo zapisana besedila vsebovati, da jih bralci, ki niso prisotni v istem prostoru oz. situaciji, razumejo, drugi pa so implicitni in v tem smislu spominjajo na govorna besedila, pri katerih si udeleženci komunikacije delijo isto situacijo, ki jo v primeru komentiranja novice tvori ravno vsebina novice, skupen čas, ko uporabniki do nje dostopajo,¹ skupen prostor, ki je družabno omrežje ipd.

¹ Ni nujno, da uporabniki novice berejo hkrati, je pa objava hkrati dostopna vsem, velika verjetnost je, da bo več uporabnikov novice prebralo hkrati.

3.1 Komentarji spletnih novic kot besedilna vrsta

Komentarji spletnih novic delujejo kot odzivna besedila, ki povezujejo med sabo različne udeležence komunikacijskega procesa, zato se med njimi vzpostavi interakcija, ki spominja na forum, pri katerem poteka izmenjava mnenj, stališč ali védenj o določeni temi. V. Mikolič (2007: 352) komentar uvršča med sporazumevalna/interaktivna besedila, med katera spadajo tudi prikazovalna, pozivna, zavezovalna in povezovalno-izrazna besedila. Med povezovalno-izrazna besedila uvršča opravičilo in komentar, ki je pravzaprav mnenjsko besedilo, zaradi spodbujanja novih naslovnikov k odzivu oz. nadaljnjim komentarjem pa sodi med pozivna argumentativna besedila (Krajnc Ivič 2018: 70).

S. Levinson (v McElroy 2013: 756) spletne komentarje definira kot najpogostejšo obliko zapisanega diskurza v novih medijih. Komentiranje novic na spletu je zelo neposredno, uporabnikom omogoča, da z ostalimi uporabniki delijo svoja mnenja, perspektive, stališča in znanje o vsebini, ki jo ustvarjajo novinarji.

Tako spletni časopisi z deljenjem novic na družabnih omrežjih, večinoma Facebooku, omogočajo dvosmerno komuniciranje, pri čemer je skupna referenca uporabnikov novica oz. novinarski prispevek, ki ga komentirajo. Na uporabo jezika v spletnih komentarjih pa vplivajo tudi dejavniki samega medija, pri čemer slog spletnih komentarjev odraža tudi tehnološke danosti, možnost uporabe emotikonov, hitrost tipkanja, hkrati pa tudi značilnosti sporočevalca, kot so izobrazba, starost, inovativnost itd. Trdimo lahko, da so spletni komentarji posebna vrsta besedila, ki združuje nekatere skupne lastnosti s klasičnimi komentarji (npr. mnenjskost, argumentativnost ipd.), hkrati pa se od njih razlikuje zaradi interaktivnosti, možnosti komentiranja različnih uporabnikov in s tem različnih osebnih stilov, obvladovanja jezikovne norme, perspektive, hitrosti tipkanja, odnosa do izhodiščnega besedila itd.

3.2 Značilnosti spontanega govornega diskurza v pisni podobi komentarjev novic spletnega portala Regionalobala.si

V namen prispevka smo pregledali komentarje uporabnikov, ki so se zvrstili pod dnevnimi novicami portala Regionalobala.si (maj, 2023), objavljenimi na Facebook strani portala, in opazovali, kako se kažejo značilnosti govornega jezika. Pregledali smo komentarje vseh dnevnih novic v navedenem obdobju, v nadaljevanju je zapisanih le nekaj primerov iz preučevanega gradiva, ki nakazujejo značilnosti

govorjenega diskurza, so pa te prvine vidne v večini komentarjev, ki smo jih opazovali. Primerov nismo spreminjali, zato so zapisani na način, kot so se pojavili (pogosto je neupoštevanje slovničnih in pravopisnih pravil). V sprotnih opombah so navedeni naslovi novic, ki so objavljene na Facebook strani spletnega časopisa Regionalobala.si, do njih pa dostopamo s pomočjo datuma objave, ki je naveden ob posameznem naslovu.

Med sredstvi besedilne semantike so izpostavljeni deikti in diskurzni označevalci, ki so se pojavljali v največji meri, veliko pa je bilo tudi primerov, ki kažejo skladske značilnosti govorenega jezika. V spodnjih primerih so vse navedene značilnosti zapisane ležeče. Z nadaljnjim pregledom gradiva bi zagotovo opazili tudi več primerov besedilnih aktualizatorjev, jih pa nismo upoštevali pri analizi, ker jih v zapisih nismo opazili v tolikšni meri in v taki obliki, da bi bili relevantni za prikaz značilnosti govorenega diskurza.

Sredstva besedilne semantike:

1. Deikti:

a) Vidla. 🙄 grozno²

b) Daj še podatke za Hrvaško po srbski agresiji če si že pri številkah³
Spet 1 članarino več cepci, 🙄 smo tk nepomembni da nas ne bo nihče resno jemal 🙄

c) Ljudje dragi kajse dogaja natej nasi planeti zaustavimo to vse hudobije⁴
Tudi pri nas bi mogli preventivno v šolskih ustanovah uvesti varovanje
Prav ima, tako bi bilo treba narediti, smrtno kazen, ker morilci si ne zaslužijo nič drugega
Se ne strinjam.

² Prometna nesreča zaprla primorsko avtocesto: Nastajajo zastoji, ne vozite po reševalnem pasu. <https://www.facebook.com/regional.si>, 9. 5. 2023

³ Bo Slovenija vstopila v Varnostni svet? Tekmujemo z Belorusijo, ki od leta 2007 kot kandidatka ni imela nasprotnika. <https://www.facebook.com/regional.si>, 9. 5. 2023.

⁴ Novo streljanje, zdaj na jugu Srbije: Mladenič z zračno puško buje ranil vrstnika. <https://www.facebook.com/regional.si>, 5. 5. 2023.

Deikti *vidla, daj, smo, nas, nihče, ljudje dragi, nasi⁵, prav ima, se ne strinjam* referirajo na hkratno udeležnost akterjev v govorni situaciji, skupen pa jima je kontekst (tj. novica). Pri samo pisnem diskurzu bi si sporočevalec in naslovník izmenjala več ključnih informacij o okoliščinah sporočanja. Sporočevalec, ki je npr. zapisal *Prav ima*, bi v pisnem diskurzu verjetno podal informacije o tem, kdo ima prav, v diskurzu spletnih novic pa novica tvori skupen kontekst oz. skupno védenje sporočevalca in naslovníka, zato ta informacija ni potrebna. Deikti, kot so npr. *vidla, daj, se ne strinjam*, tudi ne podajajo ključnih informacij o sporočevalčevi identiteti (razen npr. spola, npr. *vidla*), ki je sicer večinoma razvidna iz uporabnikovega Facebook profila (razen v primeru anonimnega komentiranja), kljub temu pa pri komentiranju objav ni ključnega pomena, saj je spletna identiteta fluidna in v komentarjih spletnih novic sporočevalci zavzemajo drugačne govorne položaje (v tem kontekstu so enakovredni udeleženci komunikacijskega procesa, ki se odvija pod objavljeno novico) kot pri npr. pisanju uradnih e-poštnih sporočil, v katerih morajo svoje podatke eksplicitno navesti. Uporabniki se večinoma med sabo niti ne nujno poznajo, kljub temu pa se pogovarjajo na način, kot bi se pogovarjali v živo oz. na način neposredne udeležnosti v pogovoru (npr. *vidla* ne izrazi, kaj je sporočevalka dejansko videla, saj vsi udeleženci komunikacijskega procesa zaradi objave novice vedo, da deikt referira na njeno vsebino). S tikanjem (npr. *daj, si*) sporočevalec na nek način krši načelo vljudnosti (Leech 1983), po drugi strani pa spletno okolje tvori poseben diskurz, v katerem udeleženci komunikacije (čeprav se med sabo večinoma ne poznajo oz. so celo anonimni) tvorijo nekakšno skupnost (kar izražajo tudi deikti *Ljudje naši, nas ipd.*), osebna distanca med nepoznanimi akterji pa se kljub fizični oddaljenosti manjša⁶.

2. Diskurzni označevalci:

a) *Bom povedal to ..brez odobritve varnostnega sveta je Nato 1999 in os zla bombardiral Srbijo / ...*⁷

b) *Sprašujem za prijatelja, ma kej je uon pržgov luči nad Dekani*  ⁸

c) *Čisto tako, mimogrede. Če kdo kaj ve...?*⁹

⁵ Mišljeno je »našem«, ampak je zapis pravopisno in slovnično neustrezen.

⁶ Verjetno je ravno fizična oddaljenost razlog za tak način komunikacije, saj bi bili pri pogovoru »v živo« udeleženci komunikacije, ki se med seboj ne poznajo, vljudnejši, med seboj bi se vikali itd.

⁷ *Bo Slovenija vstopila v Varnostni svet? Tekmujemo z Belorusijo, ki od leta 2007 kot kandidatka ni imela nasprotnika.* <https://www.facebook.com/regional.si>, 9. 5. 2023.

⁸ *Jansa danes v obalnih mestih, tudi v Ankaranu.* <https://www.facebook.com/regional.si>, 8. 5. 2023.

⁹ *Vesela novica za lastnike štirinožcev: Letošnje poletje bo na Obali še več psom prijaznih plaž.*

Ni šanse da grem čez dan v največji vročini mučit žival

č) *Važno da je domače*¹⁰

Vse bio domače pridelano v Piranski občini *sam vprašam*

Pa kaj

Ja in??

d) *tko da skoda se obremenjevat*¹¹

točno to!!! Danes je poudarek samo na pravicah, dolžnosti pa ne.

...dej res..brez besed. to je uno....

Med zgoraj navedenimi diskurznimi označevalci, ki v besedilih komentarjev izražajo učinek neposredne govornje konverzacije, se kažejo naslednji: ideacijski (npr. *tko da*), interakcijski (*bom povedal to, sam vprašam, sprašujem za prijatelja, če kdo kaj ve*) in interpretacijski (*full dobro, čisto tako, mimogrede, ni šanse, važno da, pa kaj, ja in, točno to, dej res, brez besed, to je uno*). Iz navedenih primerov opazimo, da izstopajo interpretacijski, ki označujejo odnos sporočevalca do predmetnosti, kar je hkrati ena ključnih značilnosti komentarja, ki je pravzaprav mnenjsko besedilo. Opazna je tudi ironija, o kateri lahko sklepamo na podlagi celotnega konteksta, tj. novice (npr. diskurzni označevalec *važno da* (je domače) se navezuje na komentar ob pismu bralca o prodajanju sadja brez dovoljenja v Portorožu¹²). Diskurzna označevalca *Sprašujem za prijatelja* in *sam vprašam* glede na zgradbo uvrščamo med označevalce procesa tvorjenja (tj. glagol, ki izraža proces tvorjenja v prvi osebi edine v sedanjiku), glede na kontekst pa razumemo oba izraza kot pogovorna frazema (pogosto opažena v spletnih komentarjih) za izražanje ironije, v tem smislu pa bi ju tako lahko uvrstili tudi med interpretacijske diskurzne označevalce.

Skladenske značilnosti govornega jezika:

Primeri:

a) *lepo vas prosim, kaj je z vami ljudje?! Zatajili starši, zatajila šola, sošolci katastrofa, država in sistem katastrofa, deček s travmami, ki jih skriva v sebi, deležen posmebovanja in ustrabovanja vrstnikov. Ta otrok si je sodil že sam in potrebuje pomoč. Otresel se misli na storjeni poboj nikoli ne bo, s starostjo bo trpljenje le*

<https://www.facebook.com/regional.si>, 8. 5. 2023.

¹⁰ Bralec ogorčen: "V Portorožu ob cesti prodajajo sadje brez dovoljenja", pa je temu res tako? <https://www.facebook.com/regional.si>, 8. 5. 2023.

¹¹ Ravnateljica OŠ Koper Ingrid Poropat: "Če želimo zdrave in zadovoljne otroke, jim postavimo meje." <https://www.facebook.com/regional.si>, 8. 5. 2023.

¹² Bralec ogorčen: "V Portorožu ob cesti prodajajo sadje brez dovoljenja", pa je temu res tako? <https://www.facebook.com/regional.si>, 8. 5. 2023.

*večje. Mladina, večino nje, brez empatije do drugačnih, ne bi smela prosto hoditi po tem svetu. In kdo jih vzgaja?! STARŠI BREZ EMPATIJE DO ČLOVEKA...in potem pride tudi do žrtev, ki so povsem nedolžne. Naj počivajo v miru.*¹³

V zapisu opazimo več sredstev, ki nakazujejo skladnjo govornega jezika. Sporočevalec izraža svoje čustveno stanje z retoričnima vprašanjema (*lepo vas prosim, kaj je z vami ljudje?! In kdo jih vzgaja?!*), ki jima sledi označevalec vzklične intonacije, izražen z *?!*, podobno učinkujejo tudi velike črke (*STARŠI BREZ EMPATIJE DO ČLOVEKA*), ki v spletnem diskurzu označujejo kričanje, premor, označen s tremi (ali več) pikami,¹⁴ pa lahko signalizira tudi sporočevalčev premislek pred nadaljevanjem pisanja komentarja oz. organiziranje misli. Ponovitve oz. paralelizmi (*Zatajili starši, zatajila šola, sošolci katastrofa, država in sistem katastrofa*) stopnjujejo sporočevalčev čustven odziv na novico. Pojavi se tudi napačen začetek (*in potem pride ...*), kar je značilnejše za govorni diskurz. Opazimo tudi pravopisne napake (npr. *trpljenje*) in neustrezen besedni red (*Otresel se misli na storjeni poboj nikoli ne bo*), kar signalizira bodisi sporočevalčevo pomanjkljivo jezikovno kompetenco bodisi značilnost diskurza spletnega komentiranja, ki je med drugim tudi hitrost odzivanja na novico s tipkanjem in izmenjave komentarjev v najkrajšem možnem času.

*b) Dobro jutro....12 al 10...al 15 nacrtoval je...ucil se je in vedel točno kam.streljat..in imel.spisek..to je psibicno bolan otrok in nima kaj delat na prostosti..in ni cudno da ce kao ne morejo mu nic ga bojo drugi zrihtali..zbudite se...res...za 9 zivljenj in kao nikomu nista...dej res..brez besed.*¹⁵

Primer vsebuje več premorov, označenih s tropičjem, ki označujejo počasnejši ritem, lahko pa kot v prejšnjem primeru signalizirajo piščev premislek. Opazimo tudi skladijsko neustrezne stavčne strukture in napačne začetke (npr. *in ni čudno da če kao, in kao nikomu ništa*), z glagolsko obliko za drugo osebo množine v velelnem naklonu, ki je hkrati izraz deiktčnosti (*zbudite se*) pa referira na ostale udeležence v komunikacijskem procesu oz. komentatorje. Pojavi se tudi raba kratkega nedoločnika, ki je značilnost govornega diskurza (*nima kaj delat*). Poudarjanje z izpostavljenimi členki (npr. *...res,dej res..*) je tudi značilno predvsem za govorni diskurz, v spletnem pa je izraz odnosa sporočevalca do predmetnosti in težnje po učinku spontane govorne konverzacije.

¹³ *Sporočilo na vratih doma 13-letnika po streljanju na šoli: Grožnja, ob kateri zaledeni kri v žilah.* <https://www.facebook.com/regional.si>, 8. 5. 2023.

¹⁴ Pri tem ni upoštevanja pravopisnih pravil, ki narekujejo nestičnost tropičja (razen v primeru glasovne elipse), prav tako tropičje večkrat signalizira več pik.

¹⁵ *Sporočilo na vratih doma 13-letnika po streljanju na šoli: Grožnja, ob kateri zaledeni kri v žilah.* <https://www.facebook.com/regional.si>, 8. 5. 2023.

c) *Kurjerva*¹⁶ je *zastonj*. *Arriva je poceni*. *Taxi tudi ni astronomsko drag*. *Peš je še tudi možnost*.¹⁷
hahahaba ja ja res res

V zgornjem primeru sporočevalec niza enostavčne strukture, ki dajejo vtis odsekanega govorjenja, s čimer se zapis ponovno približuje spontanem govoru. V odzivu na komentar opazimo medmet (*hahahaba*) in ponovitvi (*ja ja res res*), oboje pa se v spletnem diskurzu pojavlja zaradi čustvenega odnosa udeležencev komunikacijskega procesa do upovedenega, hkrati pa sta to značilnosti spontanega govora.

3.3 Razprava

Na podlagi analize primerov lahko opazimo, da se v spletni komunikaciji dejansko pojavljajo značilnosti spontanega govorjenega jezika, zato se zdi analiza komentarjev spletnih novic ena izmed možnosti preučevanja značilnosti pisno-govorjenega diskurza na leksikalni, semantični in skladenjski ravni. V obravnavanem gradivu tako najdemo predvsem deikte in diskurzne označevalce, na skladenjski ravni pa so pogostna predvsem ponavljanja in nekatere prozodične prvine. Deikti in diskurzni označevalci predvidevajo poznavanje konteksta (in diskurza) udeležencev komunikacijskega procesa, zato so praviloma omejeni na govorjeno komunikacijo. V komentarjih spletnih novic se pojavijo zaradi vsebine novice same, slikovnega in video gradiva ter povezav, ki tvorijo skupni kontekst in s tem skupno védenje, zato sta navedeni skupini referenčnih izrazov pogosti v diskurzu spletnih komentarjev.

Ponavljjanja (na skladenjski ravni) se pojavljajo predvsem z namenom poudarjanja pomembnih vsebinskih enot in pridobivanja časa pri tvorjenju sporočila. Pogoste so prekinitve in premori, ki so načeloma označeni s tropičjem ter signalizirajo pridobivanje časa, razmišljanje, implicitno napoved presenetljivega podatka itd., lahko pa so posledica spletne komunikacije, ki poteka zelo hitro (uporabniki avtomatsko zapišejo tropičje, ko urejajo informacije oz. razmišljajo, kaj želijo napisati). Po drugi strani ima zapis tropičja slogovno vrednost, in sicer ustreza označevanju spontanega govorjenja v pisni komunikaciji, tako pa občutku neposredne udeležnosti, v nekaterih primerih pa gre lahko zgolj za modno muho

¹⁶ Tudi zapis po izgovoru (*kurjerva*) in pogovorni izrazi (npr. *kaa*) so pogostejši v govorjenem diskurzu, vendar ti jezikovni elementi niso predmet te raziskave.

¹⁷ V Kopru od danes dražje parkiranje: V Žusterni, ob Severni obvoznici in na Trgu Brolo so se cene podvojile. <https://www.facebook.com/regional.si>, 8. 5. 2023.

ali poseben slog sporočevalca (nekateri npr. uporabljajo tropičje po vsaki besedi, drugi pišejo besede s kombinacijo malih in velikih tiskanih črk ipd.).

Sporočevalci, ki komentirajo spletne novice, pogosto uporabljajo velike tiskane črke, kopičijo ločila, kar označuje npr. kričanje, čustveno vznemirjenost ipd. Ko beremo zapise, imamo velikokrat občutek, da »slišimo« sporočevalca. Ne moremo seveda govoriti o pravem zvočnem učinku (npr. ni dejanskega kričanja, saj besedila le pišemo oz. beremo), ampak o signaliziranju slednjega z npr. velikimi tiskanimi črkami ali pomnožitvijo ločil.

Skladenjsko neustrezni zapisi (npr. napačen besedni red) se pojavljajo zaradi več dejavnikov, bodisi zaradi neustrezne jezikovne kompetence bodisi zaradi neobremenjenosti s skladenjskimi pravili in težnje po hitrem sporočanju. Uporabniki načeloma podajajo informacije zelo hitro in pri tem ne razmišljajo, če je določen zapis skladenjsko ustrezen, saj gre večinoma za neformalne okoliščine.

4 Sklep

Besedila spletnih komunikacijskih okolij odražajo živahno in dinamično podobo sodobne komunikacije, ki združuje raznolike jezikovne in diskurzivne značilnosti, zato se preučevanje teh besedil izkaže za nadvse pomembno v današnjem digitalnem svetu, ki ga zaznamujejo hitrost pretoka informacij, široka dostopnost do digitalnih vsebin in raznolikost komunikacijskih platform. Zavedanje, da ta besedila vsebujejo značilnosti govorenega jezika, dodatno poudarja potrebo po temeljiti analizi, saj so ti jezikovni elementi ključni pri razumevanju vsebine, konteksta in družbenih interakcij na spletu. V komentarjih spletnih novic se pisni diskurz prepleta z elementi govorenega jezika, kar omogoča bolj živo, interaktivno in pristno izmenjavo mnenj ter informacij med uporabniki, ki daje udeležencem komunikacijskega procesa občutek neposredne udeležnosti v konverzaciji in kljub fizični oddaljenosti zaradi narave medija razbija distanco med njimi.

Literatura

- Douglas BIBER, Stih JOHANSSON, Geoffrey LEECH, Susan CONRAD, Edvard FINEGAN, 1999. *The Longman Grammar of Spoken and Written English*. London in New York: Addison Wesley Longman.
- David CRYSTAL, 2001. *Language and Internet*. Cambridge: Cambridge University Press.

- H. Paul GRICE, 1975: Logic and Conversation. *Syntax and Semantics*. Ur. Peter Cole, Jerry L. Morgan. New York: Academic Press. 45–46.
- Michael A. K. HALLIDAY, 1989: *Spoken and written language*. Oxford: Oxford University Press..
- Mira KRAJNC IVIČ, 2018. Besedilne vrste v slovenskem besediloslovju. *Jezik in slovnstvo*, 63/2–3, 75–86.
- Simona KRANJC, 2003. Jezikovna zvrstnost v sodobnih medijih. *Aktualizacija jezikovnozvrstne teorije na Slovenskem. Mednarodni znanstveni simpozij Obdobja 22 – Metode in zvrsti*. Ur. Erika Kržišnik. Ljubljana: Center za slovenščino kot drugi in tuji jezik, Filozofska fakulteta. 395–405.
- Geoffrey LEECH, 1983. *Principles of Pragmatics*. London, New York: Longman.
- Stephen C. LEVINSON, 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Stephen C. LEVINSON, 2003. *Space in Language and Cognition. Explorations in Cognitive Diversity*. Nijmegen: Max Planck Institute for Psycholinguistics
- Kathleen McELROY, 2013. Where old (gatekeepers) meets (new media). *Journalism Practice*, 60/7, 755–771. Dostop 8. 4. 2024 na <https://doi.org/10.1080/17512786.2013.774117>.
- Vesna MIKOLIČ, 2007. Modifikacija podstave in argumentacijska struktura besedilnih vrst. *Slavistična revija*, 55/1–2. 341–355
- Jim MILLER, Regina WEINERT, 2009: *Spontaneous spoken language: Syntax and discourse*. Oxford: Oxford University Press.
- Agnes PISANSKI PETERLIN, 2007. Raziskave metabesedilnosti v uporabnem jezikoslovju: pregled področja in predstavitev raziskovalnega dela za slovenščino. *Jezik in slovnstvo*, 52/3–4, 7–19.
- Maša ROLIH, 2014. Različni vplivi na jezik mladih«. *Jezikoslovni zapiski*, 20/1, 79–92.
- Maša ROLIH, 2017. *Sleng in pogovorni jezik v spletni komunikaciji*. Doktorska disertacija. Koper: Univerza na Primorskem, Fakulteta za humanistične študije.
- Deborah SCHIFFRIN, Deborah TANNEN, Heidi E. HAMILTON (ur.). 2006. *The Handbook of Discourse Analysis*. Malden: Blackwell Publishing.
- Mojca SMOLEJ, 2006. *Vpliv besedilne vrste na uresničitev skladenjskih struktur. Primer narativnih besedil v vsakdanjem spontanem govoru*. Doktorska disertacija. Ljubljana: Filozofska fakulteta.
- Dan SPERBER, Deirdre WILSON, 2004. Relevance theory. *Handbook of pragmatics*. Ur. Laurence R. Horn, Gregory Ward. Oxford: Blackwell. 607–632.
- Jože TOPORIŠIČ, 2004. *Slovenska slovnica*. Maribor: Obzorja.
- Darinka VERDONIK, 2008. Označevanje vrste diskurzivnih označevalcev. *Zbornik 6. konference Jezikovne tehnologije*. Ur. Tomaž Erjavec, Jerneja Žganec Gros. Ljubljana: Inštitut Jožef Stefan. 25–28.
- Jana ZEMLJARIČ MIKLAUVČIČ, 2008. Iskanje odgovorov na vprašanja govornega jezika. *Jezik in slovnstvo*, 53/1, 90–106.
- Danila ZULJAN KUMAR, 2007. Narečni diskurz: diskurzivna analiza briških pogovorov. Ljubljana: ZRC SAZU.

Spletni viri

- Novo streljanje, zdaj na jugu Srbije: Mladenič z zračno puško buje ranil vrstnika.*
<https://www.facebook.com/regional.si>, 5. 5. 2023.
- Janša danes v obalnih mestih, tudi v Ankaranu.* <https://www.facebook.com/regional.si>, 8. 5. 2023.
- Vesela novica za lastnike štirinožcev: Letošnje poletje bo na Obali še več psom prijaznih plaž.*
<https://www.facebook.com/regional.si>, 8. 5. 2023.
- Bralec ogorčen: "V Portorožu ob cesti prodajajo sadje brez dovoljenja", pa je temu res tako?*
<https://www.facebook.com/regional.si>, 8. 5. 2023.
- Ravnateljica OŠ Koper Ingrid Poropat: "Če želimo zdrave in zadovoljne otroke, jim postavimo meje."*
<https://www.facebook.com/regional.si>, 8. 5. 2023
- Sporočilo na vratih doma 13-letnika po streljanju na šoli: Groznjia, ob kateri zaledeni keri v žilab.*
<https://www.facebook.com/regional.si>, 8. 5. 2023.
- V Kopru od danes dražje parkiranje: V Žusterni, ob Severni obvoznici in na Trgu Brolo so se cene podvojile.*
<https://www.facebook.com/regional.si>, 8. 5. 2023.

Prometna nesreča zaprla primorsko avtocesto: Nastajajo zastoji, ne vozite po reševalnem pasu.
<https://www.facebook.com/regional.si>, 9. 5. 2023.

Bo Slovenija vstopila v Varnostni svet? Tekmujemo z Belorusijo, ki od leta 2007 kot kandidatka ni imela nasprotnika. <https://www.facebook.com/regional.si>, 9. 5. 2023.



V

Govor v gledališki
umetnosti

UPORABA MIKROFENOMENOLOŠKEGA INTERVJUJA PRI RAZISKOVANJU IGRALČEVEGA GOVORA

MARTIN VRTAČNIK

Univerza v Ljubljani, Akademija za gledališče, radio, film in televizijo,
Ljubljana, Slovenija,
martin.vrtacnik@agrft.uni-lj.si

Kognitivna znanost in področje gledališke igre sta povezana že desetletja. Kljub temu gledališki ustvarjalci za razliko od nevroznanstvenikov v raziskavah govora ne uporabljajo tehnologije (npr. funkcionalne magnetne resonance, pozitronske emisijske in računalniške tomografije), temveč predvsem opis svojega izkustva. Ker subjektivne izkušnje potekajo tudi pod pragom zavesti, je njihovo opisovanje zapleteno. Človeškega uma po mnenju nevrobiologa Francisca J. Varele ne moremo raziskovati le z nevroznanstveno tehnologijo, zato je oblikoval nevrofenomenološki program, utemeljiteljica mikrofenomenologije Claire Petitmengin, ki nadaljuje njegovo delo, pa je oblikovala mikrofenomenološki intervju, ki nam omogoča, da intervjuvanca pripravimo do tega, da se zave svoje subjektivne izkušnje in jo opiše. Čeprav je bil mikrofenomenološki intervju zasnovan za raziskovanje epilepsije, intuicije in meditacije, predpostavljamo, da bi s to metodo lahko dodatno raziskali tudi najmanj raziskano področje odrskega govora – ustvarjanje igralčevega glasovnega sloja.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.21](https://doi.org/10.18690/um.ff.4.2024.21)

ISBN
978-961-286-882-6

Ključne besede:
mikrofenomenološki
intervju,
fenomenologija,
empirizem,
odrski govor,
gledališki lektor

DOI

[https://doi.org/
10.18690/um.ff.4.2024.21](https://doi.org/10.18690/um.ff.4.2024.21)

ISBN

978-961-286-882-6

Keywords:

microphenomenological
interview,
phenomenology,
empiricism,
stage speech,
language consultant

THE USE OF MICROPHENOMENOLOGICAL INTERVIEW IN THE RESEARCH OF ACTOR'S SPEECH

MARTIN VRTAČNIK

University of Ljubljana, Academy of Theatre, Radio, Film and Television, Ljubljana,
Slovenia
martin.vrtacnik@agrft.uni-lj.si

Cognitive science and the field of theatre play have been connected for decades. In contrast to neuroscientists, however, theatre creators do not use technology (e.g. functional magnetic resonance, positron emission tomography and computed tomography) in their speech research, but primarily use descriptions of their own experience. As subjective experiences occur below the threshold of consciousness, it is complicated to describe them. The neurobiologist Francisco J. Varela pointed out that the human mind cannot be studied using neuroscientific technology alone, which is why he developed a neurophenomenological programme. Claire Petitmengin, the founder of microphenomenology, developed a microphenomenological interview that enables us to make the interviewee aware of their own subjective experience and to describe it. Although the microphenomenological interview was developed for the study of epilepsy, intuition and meditation, we believe that this method could also be used to further explore the least researched area of stage speech – the production of the actors' vocal layer.



1 Uvod

Odrski govor¹ in njegova estetika se ves čas spreminjata (Podbevšek 2010), zato ju je treba nenehno in neprekinjeno raziskovati. Zaradi pomanjkanja specifičnih znanstvenih metod je to področje tako v teatroloških kot v jezikoslovnih razpravah zapostavljeno.² Rešitev za probleme, ki nastanejo zaradi specializiranosti v posamezni stroki, ponuja interdisciplinarno raziskovanje, ki zagotavlja ohranjanje tradicionalno vzpostavljene povezave med znanostjo in umetnostjo (Podbevšek, Žavbi Milojević 2013).

Sodobni gledališki lektor (v nadaljevanju lektor) pri oblikovanju odrskogovornega koncepta³ in tudi sicer pri svojem delu uporablja najrazličnejše (kodifikacijske) priročnike, na podlagi katerih dramskemu igralcu (v nadaljevanju igralec) svetuje v zvezi z jezikom in govorom, obenem pa posamezne jezikovne izbire pri uresničevanju uprizoritvenega koncepta⁴ utemeljuje tudi ob upoštevanju načel gledališke estetike⁵ in jezikovnega čuta. Ta sposobnost zaznavanja oz. odnos do jezika je blizu konceptu utelešene teatrologije,⁶ ki zadeva gledališki odnos v celoti, tj. način igralčevega mišljenja in odločanja ter gledalčev način zaznavanja in sodelovanja pri uprizoritvi (Sofia 2016: 49–50). Pri oblikovanju optimalnega odrskega govora je ključno tvorno sodelovanje lektorja in igralca, pri vzpostavljanju tega odnosa pa je treba upoštevati, da je »govor človekova intimna zadeva« (Trefalt 2009). Da bi ugotovili, na kakšen način in na kateri točki ustvarjalnega procesa se ta odnos vzpostavi, katere postopke in načine uporabljata lektor in igralec za doseg cilja ipd., bi bilo treba v prihodnje podrobneje študijsko opazovati in opisati ta vzajemno-zaupni odnos med omenjenima poklicnima profiloma.⁷ Izsledki teh raziskav bi pripomogli k boljši pripravljenosti na izvajanje v nadaljevanju predstavljenega mikrofenomenološkega intervjuja.

¹ Odrski govor je »govor, ki temelji na razločni izreki, ustrezni slišnosti, glasovni izraznosti, usklajen z besedilnimi in odrskimi okoliščinami, gledališko estetiko« (*Gledališki terminološki slovar*; v nadaljevanju *GTS*).

² Izjeme so prispevki nekaterih lektorjev v gledaliških listih (Vrtačnik 2012: 109–110), vendar to niso akademski, strogo strokovni oz. znanstveni članki.

³ Odrskogovorni koncept je izraz, ki smo ga naredili po analogiji s terminom uprizoritveni koncept (gl. op. 4), označuje pa zamisel gledališkega lektorja o oblikovanju odrskega govora.

⁴ Uprizoritveni koncept je »režiserjeva, dramaturgova zamisel uprizoritve, ki se navadno pripravi pred začetkom vaje« (*GTS*).

⁵ Gledališka estetika je »celota umetniških postopkov, načel, značilnih za dramatika, režiserja, igralca, določeno delo, dobo, gledališko zvrsto« (*GTS*).

⁶ Izraz je predlagal teatrolog Marco De Marinis, ki je na podlagi teorij utelešene kognicije Francisa J. Varele predlagal vključitev raziskovalčevega telesa oz. njegove subjektivnosti v teatrologijo (Sofia 2016).

⁷ Npr. s pristopom praksa kot raziskava (angl. *practice as research*), pri katerem je praksa ključna metoda raziskovanja in obenem pomemben dokaz raziskave (Nelson 2013: 8–9).

Skratka, govor kot tak je del človekove intimne, odrski govor pa je posledično tudi del intimne posamezne dramske osebe, zato lahko relevantne napotke o procesu preoblikovanja besedila⁸ iz pisno-vidnega prenosnika v govorno-slušnega podajo igralci sami.

2 Metode raziskovanja odrskega govora

Po letu 2000 se je število raziskav umetniškega govora povečalo, kar je tudi posledica organiziranja strokovnih srečanj o govoru na AGRFT. Če se osredinimo na pregled raziskav odrskega govora, ugotovimo, da gre na splošno za:

- proučevanje lektorske poetike posameznega lektorja z analizo lektorske knjige in teoretsko argumentacijo, izsledki pa so predstavljeni v gledaliških listih in drugih strokovnih publikacijah (Podbevšek 2010: 213–215);
- proučevanje odrskega govora uprizoritve s slušnozaznavno metodo,⁹ ki je nadgrajena z instrumentalno analizo s programom za fonetične analize Praat (Žavbi Milojević 2013).

Raziskave odrskega govora so si zaradi različnih namenov raziskovalcev in raziskovalnih ciljev med seboj metodološko neenotne, jih pa večina prinaša izsledke o estetiki odrskega govora kot gledališkega znaka.¹⁰ Te raziskave so dokument časa in evalvacija dela različnih lektorjev ter osnova za nadaljnje raziskovanje odrskega govora, zlasti o njegovem glasovnem sloju, saj je teh prispevkov v strokovni literaturi najmanj. Pri tem bo treba v prihodnje upoštevati vsaj dve dejstvi:

- da je »govorna podoba posamezne vloge (in celotne uprizoritve) [...] rezultat tako individualnega kot interdisciplinarnega skupinskega ustvarjanja« (Podbevšek 2008: 53), zato naj bodo metodološko tako interdisciplinarne¹¹ kot tudi empirične;
- da je treba – sploh po t. i. kognitivnem obratu v teatrologiji (McConachie, Hart 2006) – v središče raziskovanja po načelih kognitivne znanosti

⁸ Izraz besedilo razumemo v širšem pomenu (z besedami izražene misli – SSKJ).

⁹ Dejstvo je, da je »človeški organ za sluh (uho) z možganskim centrom [...] zelo precizen aparat za razpoznavanje zvoka« (Tivadar 2004).

¹⁰ Gledališki znak je »vsako odrsko in igralsko izrazilo, ki gledalcu kaj sporoča« (GT3).

¹¹ Npr. rekonstrukcijski intervju kot metoda za opredelitev estetike igre (govora in telesnega izraza) na Odru 57 (Poi 2020).

postaviti človeka, torej lektorja in igralca, ki bosta svoj delovni odnos v uprizoritvenem procesu še poglobila.

Raziskavam umetniškega oz. odrskega govora je skupen tudi teoretski diskurz, v katerem je v rabi znanstveni in metaforični jezik (Podbevšek 2010: 215). Na razširjenost uporabe le-tega je v gledališki semiotiki vplivala tudi terminološka metaforizacija, zato nekateri teatrologi omenjeno teorijo zavračajo (Balme 2010: 81), drugi menijo, da ni popolna (States 1985: 8). Teatrolog Bert O. States zato semiotični pogled, ki je uporabljen za opis semantičnih vidikov uprizoritve, združuje s fenomenološkim, ki se usmerja na neposredno doživljajske vidike uprizoritve (Balme 2010: 85–87; States 1985). Prav zato je treba pri empiričnih raziskavah odrskega govora poudariti tudi večji pomen t. i. metode intuicije; ta tradicionalno velja za neznanstveno metodo, ker temelji na samoumevnosti in očitnih stališčih (Ivanko 2007: 2), prav tako obstajajo težnje po izločitvi introspekcije iz empiričnih postopkov (Perenič 2014: 47), čeprav je ravno tovrstno raziskovanje za področje dramske igre (v nadaljevanju igre) ključno.

3 Interdisciplinarnost – neuroestetika – empirija

Interdisciplinarnost, na področju gledališkega lektorstva pričakovana in uveljavljena, je spodbudila nastanek novih ved, tudi neuroestetike.¹² Ugotovitve te vede lahko na področju odrskega govora idejno-teoretično utemeljijo splošna prizadevanja gledaliških ustvarjalcev »za izpopolnjevanje govornih spretnosti in razvijanje govorne ustvarjalnosti« (Podbevšek 2016: 47). Čeprav neuroestetika uporablja tako nevroznanstvene tehnologije kot tudi empirične metode raziskovanja možganov (Berlot Pompe 2016: 7), pa je v posameznem uprizoritvenem procesu v dramskem gledališču z vidika dostopnosti in praktičnosti primernejša in uporabnejša empirična metodologija (Perenič 2014), z vidika katere bi ob že uveljavljenih jezikoslovno-teatroloških praksah in teorijah lahko proučevali tudi odrski govor (Podbevšek 2010: 199).

¹² O metodoloških težavah nekaterih neuroestetških raziskav – češ da ni povsem jasno, ali merijo estetsko izkušnjo ali prepiranje o estetski vrednosti – in o rešitvi teh težav več v Kordeš (2016). Gl. tudi op. 25.

3.1 Popis igralčeve subjektivne izkušnje o odrskem govoru

Razmišljanja o igralčevem govoru najbolj celostno predstavlja igralec in pedagog Kristijan Muck (Podbevšek 2006; Podbevšek 2010; Podbevšek 2017). Gre za nadaljevanje misli o odrskem govoru, ki sta jih pred njim ubesedovala Oton Župančič in Mirko Mahnič (Vrtačnik, Tivadar 2017), pa tudi Jože Tiran, ki je poudaril, da je »izvor govora notranje dogajanje v človeku in da je smisel in cilj govora izražanje našega notranjega psihičnega dogajanja« (Tiran 1965: 52). Podobno kot Muck je tudi Tiran svetoval, kako v ustvarjalnem procesu iz posamezne izjave besedila izluščiti notranjo vsebino: govorec naj si ustvari »čim bolj natančno in točno začrtano sliko, ki jo bo potem med izvajanjem opazoval in podoživljak« (Tiran 1965: 62); da bi si ustvaril to sliko, je pomembno »doživljanje in pravilno čustvovanje« (Tiran 1965: 64). Tiran pri opisu tega postopka uporablja pojme, kot so podzavest, delovanje živčevja in možganov ter vzbujanje čustev (Tiran 1965: 65–66). Podobno besedje v prispevku *Blodnjak in tehtnica* uporablja tudi Muck (2000), zato smo na prispevek pogledali z vidika kognitivne (nevro)znanosti. Muck pri pisanju izhaja tako iz dela s študenti kot tudi samoopazovanja, njegovo osrednje zanimanje pa je odnos med igralcem in besedilom.

Muck ob primerih – ob pozdravu »dober večer« in verzih »da b' uka žeja me iz tvoj'ga sveta / speljala ne bila, golj'fiva kača« – opisuje postopek notranje posvojitve izjave v besedilu, ki igralcu omogoča izbrati ustrezno intonacijo, pomen in smisel izjave. Muckovo izhodišče je nezavedno oz. spomin, ki hrani podatke, ki so povezani s čustvi, ti podatki pa v določeni situaciji/odnosu prihajajo v zavest. Izjavi deli dihotomično: na pomensko raven in raven izrekanja. Pri pozdravu »dober večer« se Muck osredotoča zlasti na raven izrekanja in ugotavlja, da je iskrenost izrekanja mogoče ovrednotiti na podlagi igralčeve ustrezne/pričakovane intonacije, neenakomerna porazdelitev ritma in tona med besedama v izbranem izrazu pa lahko vpliva na spremembo pomena. Nadalje ugotavlja, da se posamezni »pojmi-besede« (Muck 2000: 5) v zavesti pojavljajo kot »zvokovne podobe« (Muck 2000: 5), in sklene, da je »spontanost pojavljanja govora [...] odvisna od sproščenosti, zbranosti, sposobnosti imaginacije oziroma uvida v bistvo izjave« (Muck 2000: 6). Pri analizi Prešernovih verzov pa se osredotoča predvsem na pomensko raven (tudi z vidika individualne interpretacije, povezane s konciznostjo ali čustvenostjo) in poskuša vzpostaviti pojavnost v biti govora, kar občuti kot združitev telesa in duha.

Muck zaključí, da je predstavljeni postopek mogoče modificirati, pomembno je le, da igralec po opravljeni pomenski analizi izjave v praksi preizkuša njeno iskrenost in spontanost ter je ob tem pozoren na izkustvo, in če je pri tem sproščen, zbran, pozoren na celoto in posameznosti ter dihanje, bo brez razmisleka o tem, kako izjavo izgovoriti, zadel »pravo intonacijo, razviden pomen in smer celovitega smisla« (Muck 2000: 8).

Ključni igralski pojmi, ki jih uporablja Muck, so spomin, domišljija in podoba. Spomin in domišljija sta osnovni igralčevi orodji (Blair 2008: 21). Po ugotovitvah kognitivne znanosti je spomin kot nevrokemični dogodek zaradi svoje spremenljivosti in dinamičnosti selektivna rekonstrukcija preteklosti, zato se pri igri uporablja kot orodje domišljije, ki igralcu omogoča ustvarjalnost (Blair 2008: 70–76). Do domišljije, osnovnega dela zavesti, igralci dostopajo s pomočjo čutil, torej ima tako psihološko kot tudi fizično razsežnost (Blair 2008: 62). Zavest pa se po ugotovitvah nevroznanosti začne pojavljati ob združitvi čutnih podob in podob o sebi; igralci se praviloma osredotočajo na podobe, ki so povezane z epizodičnim spominom, z njimi manipulirajo tako, da jim omogočajo želeno vedènje in občutenje (Blair 2008: 78–79). Muck se pri samoopazovanju torej uspešno izogiba nekaterim glavnim težavam te metode, to so ohranjanje stalne pozornosti, izogibanje pristranskosti in vedènje, kaj iskati (Petitmengin 2006: 235–236).

To je kratek prikaz, kako sta kognitivna znanost in igra (oz. poročanje o usvajanju besedila z metodo samoopazovanja) povezani. Pri samoopazovanju ali samoizpraševanju (gre za prvoosebno raziskovanje v ožjem pomenu) je raziskovalec obenem tudi raziskovani. Gre za znani pristop, ki še ni del znanstvenega korpusa znanja, čeprav naj bi z globinskimi samoopazovalnimi kvalitativnimi raziskavami že prišli do splošnih ugotovitev, ne zgolj specifičnih individualnih (Kordeš 2013: 37–38). Sicer so tudi nekateri drugi igralci poročali o oblikovanju svojega odrskega govora¹³ v obliki popisa subjektivne izkušnje. Ker subjektivne izkušnje potekajo tudi pod pragom zavesti, je njihovo opisovanje zapleteno, zato si znanstveniki prizadevajo razviti stroge metode za njihovo natančno proučevanje (Petitmengin 2006).

¹³ Npr. Majda Grbac in Tone Kuntner (v Podbevšek, Gubenšek 2000) ter Branko Jordan in Darja Reichman (v Podbevšek, Gubenšek 2006).

3.2 Gledališko ustvarjanje z vidika kognitivne znanosti

V nadaljevanju predstavljamo praktična primera iz tujine, ki razmislek o ustvarjanju vloge¹⁴ in uprizoritvenem procesu v dramskem gledališču podajata ob upoštevanju spoznanj kognitivne znanosti. Na podlagi opisanega želimo ugotoviti, ali so izsledki teh razmislekov lahko vodilo pri nadaljnjem raziskovanju in opisu odrskega govora.

3.2.1 Prvi primer: metaforično opisovanje ustvarjanja vloge

John Lutterbie (2006) je ob spoznanjih nevroznanosti in kognitivne filozofije analiziral igralsko tehniko dveh igralk iz različnih okolij. Deborah Mayo iz ZDA se je na področju dramskega gledališča šolala v Angliji, Margarita Espada-Santos iz Portorika pa se je na področju fizičnega gledališča¹⁵ izpopolnjevala na Kubi. Igralki sta svoje delo opisali z uporabo podobnih binarnih metafor, ki so v gledališču – kljub ugotovitvam Bertolta Brechta, Antonina Artauda in Jerzyja Grotowskega, ki se oddaljujejo od Descartesovega dualizma in telo pojmujejo »organsko, kot medsebojno povezano mrežo potencialov« (Lutterbie 2006: 150) – večinoma še vedno splošno sprejete, tj. ločevanje kognitivnega mišljenja od nekognitivnega raziskovanja. Prva metafora je metafora razuma in čustev (odložiti racionalno mišljenje za doživljanje podob in čustvenih odzivov), druga je pojmovanje človeškega telesa kot posode (izprazniti posodo in jo napolniti z novo vsebino), povezuje pa ju metafora, ki pri igralskem procesu opredeljuje način bivanja, tj. »biti v trenutku«, ki se kaže v izogibanju diskurzivnemu mišljenju ter usmerjanju pozornosti fizičnim in čustvenim impulzom.

Nevroznanstveniki ugotavljajo, da so čustva in občutki del intelektualnih procesov in so zanje pomembna, celo nujna (Lutterbie 2006: 159). Če to drži, pravi Lutterbie, je naslednje vprašanje, kako razumeti metafore igralcev, ki jih uporabljajo pri opisu svojega dela. Lutterbie se osredotoči na dve področji raziskav delovanja možganov: na lateralizacijo in navzkrižno aktivacijo. Izsledki kažejo, da do navzkrižne aktivacije pride med nevroni in področji možganov, ki opravljajo različne funkcije – torej bi tudi metafore umetnikov lahko temeljile na tej navzkrižni aktivaciji nevronov. Prouči pa tudi področje metafor: zapletene človeške izkušnje lahko podrobneje

¹⁴ Vloga je v tem kontekstu »igralska predstavitev dramskega lika« (GTS).

¹⁵ Fizično gledališče je »gledališče, ki poudarja telo, njegovo moč, lepoto, spretnost, ranljivost, navadno brez besed« (GTS).

predstavimo le z uporabo metafor. Na metaforah temelji tudi razlikovanje med čustvi in razumom (npr. čustva so mehka/odprta, razum je trd/omejen).

Lutterbie sklepa, da igralce, ki pri opisovanju svojega procesa ustvarjanja uporabljajo tovrstne metafore, bolje razumemo, če upoštevamo, da gre pri njihovem pojmovanju za vsaj dva pristopa. Prvi je »intelektualni« pristop, v ospredju katerega je ostra, trda usmerjenost pozornosti na kognitivne funkcije, ki aktivirajo tiste dele možganov, ki si prizadevajo za specifičnost. Drugi pa je »čustveni« pristop, ki omeji zaznavanje in pozornost usmerja k čustvom. Gre za bolj splošno, dolgotrajnejšo obliko raziskovanja, pri katerem na zastavljena vprašanja odgovorimo z upoštevanjem več raznolikih dražljajev. Lutterbie sklene, da moramo v igralski umetnosti sprejeti vrednost asociativnih in metaforičnih procesov ter intelekta, ob tem pa igralce spodbujati k raziskovanju njihovega ustvarjalnega potenciala (Lutterbie 2006: 164–165).

Lutterbie z izbranimi igralkama opravi intervju, a svoje metodologije – čeprav piše z vidika fenomenologije – natančneje ne predstavi. To kaže na to, da je igralcem poglobljeno razmišljanje imanentno oz. povezano z njihovim poklicem ter da so se ne glede na tip intervjuja sposobni poglobiti v bistv(en)o. Tako Mayo kot Espada-Santos pravita, da je bivati v trenutku nevtralnno stanje ali stanje, ki ni obremenjeno (Lutterbie 2006: 155).

3.2.2 Drugi primer: kognitivna nevroznanost v uprizoritvenem procesu

Rhonda Blair (2006) opiše ustvarjalni proces – režiranje besedila *Boy Gets Girl* Rebecce Gilman –, v katerem je uporabila spoznanja nevroznanosti. Navaja, da je cilj igralcev v dramskem gledališču ustvariti prepričljive like. Ker so bili, kot navaja Blair, elementi uprizoritvenega procesa v preteklosti osnovani na spornih in nenatančnih temeljih ter tudi mistificirani, izhaja iz ugotovitve, da je v središču igre pojem kompleksne zavesti – igralec utelešeno, koherentno artikulirano bitje ustvari na podlagi »besedilnih okoliščin¹⁶ [...] in z upravljanjem vedênja, domišljije, pozornosti, čustev in spomina« (Blair 2006: 168) – zato v članku predstavi temeljne pojme nevroznanosti, ki so za igro ključni.

¹⁶ Blair (2006: 168) zapiše »given circumstances« oz. »the facts of the script«, kot so »what, where, when, what« in »specific language«. V slovenščini se za navedeno uporablja termin besedilne okoliščine, in sicer v obeh pomenih: »1. okoliščine dogajanja dramskega besedila, npr. kraj, čas, družbeni položaj oseb, čustvena razmerja med njimi; 2. slovnične, pravorečne, besedilnofonetske zakonitosti, ki oblikujejo zgradbo replike in njenih delov« (GT3).

Blair je od igralcev zahtevala, da se pri analizi in učenju besedila posvetijo besedam in jih obravnavajo, kot da gre za Shakespearovo besedilo; predpostavlja, da če igralcu ni jasen jezik, posledično ni jasna njegova podoba, niti prizor. S tem izrazito tehničnim pristopom je dosegla, da so si igralci na podlagi besedila ustvarili določen tok (p)osebni podob, ki so jih vodile iz prizora v prizor. Glavna igralka Ellen Locy je poročala, da je uprizoritveni proces zanjo postajal »globlji, polnejši in bogatejši«: začutila je potrebo, da natančno vidi vse, o čemer govori, in v zavesti ustvarjala specifične vidne, slšne, doživljajske in druge podobe; tako je lažje vstopala v interakcijo s soigralci. Za njeno delo je bilo ključno, da sta imela vsak referent in vsaka replika osebni pomen zanjo kot igralko oz. lik, obenem pa je opazila, da jo je ta pristop odvrnil od običajnega in vsakdanjega branja¹⁷ besedila. Tako so se igralci od »pripovedovanja« preusmerili k občutju v smislu biti prisoten pri zgodbi značaja in v sedanjem svetu (Blair 2006: 168–169).

Blair (2006: 179) na odlomku omenjenega besedila prikaže, da je za uprizoritev pomembnejše, če se igralci osredotočijo na občutek o tem, kar počnejo, namesto da čutijo, kaj počnejo. Torej, kar včasih pri igranju opredeljujemo kot čustveno ali psihološko vprašanje, je na globlji ravni bolje razumeti kot vprašanje fizičnega ali telesnega stanja (Blair 2006: 182).

4 Na poti do mikrofenomenološkega intervjuja

Predstavljeni primeri kažejo, da subjektivne, intuitivne zamisli igralec lahko izrazi z avtorskim, kontemplativnim, esejističnim zapisom ali v obliki poročila, pogovora, intervjuja s soudeležencem ustvarjalnega procesa. Vprašanje je, s katero metodo čim bolj objektivno predstaviti te zamisli in ugotoviti, na kakšen način igralec pri svojem delu vključuje podzavest, čustva, misli in domišljijo. Kljub prepričanju, da »v človekovi notranjosti med procesom govorjenja obstaja del dogajanja, ki se izmika znanstvenim metodam« (Podbevšek 2017: 31), si prizadevamo najti metodo za znanstveno objektivizacijo rezultatov na področju raziskovanja odrskega govora.

Ustrezno empirično metodo smo iskali v priročnikih o metodologijah (Flere 2000; Ivanko 2007; Kordeš, Smrdu 2015) in v učbeniku o empirični literarni znanosti (Perenič 2014), ki med strategijami izpraševanja in opazovanja ne predstavi metode, ki bi bila za proučevanje subjektivne izkušnje v teatrologiji najustreznejša. Ker je

¹⁷ Branje je »razbiranje pomenov v dramskem besedilu, oblikovanje zamisli o uprizoritvi« (GTS).

govor rezultat kognitivnih procesov, smo izhajali iz ugotovitev (nevro)biologije. O bioloških pojavih kognicije je med prvimi razmišljal čilski biolog Humberto R. Maturana, ki je začel sodelovati s čilskim nevrobiologom Franciscom J. Varelo; skupaj sta obravnavala tudi človeško zavest, jezikovna področja smisla, komunikacijo in širše področje družbenih pojavov (Perenič 2014: 14–15). Ker človeškega uma ne moremo raziskovati le z nevroznanstveno tehnologijo, je Varela (2013) združil sodobno kognitivno znanost in disciplinarni pristop k človeškemu izkustvu v nevrofenomenološki program. Njegovo delo nadaljuje Claire Petitmengin z znanstveno disciplino, imenovano mikrofenomenologija.¹⁸ Na svoji spletni strani zapiše, da se mikrofenomenologija ukvarja z natančnim in zanesljivim raziskovanjem težje opisljivih življenjskih izkušenj na različnih raziskovalnih področjih, tudi umetniškem, kar kaže, da je precej prilagodljiva metoda pri načrtovanju eksperimentov (Heinmann et al. 2023). Mikrofenomenološke raziskave na umetniškem področju proučujejo predvsem dva procesa: 1) nastanek ustvarjalne ideje in 2) izkušnjo spremljanja umetniškega dela. Čeprav so jo številni raziskovalci povezali z različnimi umetniškimi področji,¹⁹ za zdaj še ni povezana s teatrologijo ali področjem proučevanja govora.

Zastavlja se vprašanje, kako so raziskave na področju te discipline sploh koncipirane. Fenomenološko raziskovanje opazuje živo, neposredno izkustvo, tj. prvoosebno doživljanje, ki je primarno – misli, pomeni, opisi, vizije, slike, občutki, čustva ipd. so del zavestnega doživljanja, torej živega izkustva. Fenomenološke študije so med najpogostejšimi kvalitativnimi pristopi, med načini zbiranja kvalitativnih podatkov pa so poleg zbiranja virov in opazovanja za raziskovanje na področju igre najustreznejši intervjuji. Poleg strukturiranega in manj strukturiranega intervjuja ločimo še odprti oz. narativni, nedirektivni in fenomenološki intervju (Kordeš, Smrdu 2015: 40–44). Glede na naš kontekst raziskovanja smo izbrali fenomenološki intervju, katerega glavna usmeritev je fenomenološka redukcija Edmunda Husserla (Varela 2013: 93); Husserl je poudaril, da je izkustvo primarno in vse, kar imamo, zato je predlagal fenomenološko redukcijo, to pomeni, da je pomemben opis izkustva, ne pa njegova razlaga (Kordeš 2008: 12–13). Intervjuvanec se mora osredotočiti na več področij doživljanja (telo, čustva, čutila, misli, energijo ipd.), kar

¹⁸ Spletna stran Micro-phenomenology (dostopno na <https://www.microphenomenology.com/home>) združuje mednarodne raziskovalce, da bi ta disciplina postala širše prepoznavna in da bi nastali novi raziskovalni projekti.

¹⁹ Na podstrani strani Micro-phenomenology (<https://www.microphenomenology.com/artistic-projects>) so predstavljene raziskave, povezane z vizualno umetnostjo, plesom, oblikovanjem spletne strani, likovno pedagogiko ipd.

pomeni, da »se nanaša na implicitno kakovost prvoosebne zavesti, tj. implicitnega zavedanja, da se vse izkušnje artikulirajo v prvoosebni perspektivi kot »moja izkušnja« (Kordeš, Smrdu 2015: 42).

Osnovni vrsti fenomenološkega intervjuja sta eksplikativni intervju po Claire Petitmengin²⁰ in deskriptivno vzorčenje izkustva po Russllu T. Hurlburtu (Kordeš, Smrdu 2015); glede na predstavitev obeh vrst smo izbrali eksplikativni intervju (fr. *L'entretien d'explicitation*) psihologa Pierra Vermerscha, Petitmengin pa je njegovo tehniko intervjuja prilagodila raziskavam na področju kognitivne znanosti in jo nadgradila tako, da omogoča zbiranje podrobnih opisov subjektivne izkušnje, ponuja pa tudi tehnike za njihovo sistematično primerjavo. Petitmengin (2006: 229–232) poudarja, da se z metodo eksplikativnega intervjuja oseba zave svoje subjektivne izkušnje in jo opiše čim bolj precizno. Prvoosebni opis kognitivnega procesa je natančnejši kot posredni opis, vendar pa to zahteva notranje napore, ker se velik del naše subjektivne izkušnje odvija pod pragom zavesti.

Za termin *L'entretien d'explicitation* se v literaturi pojavljata dva izraza: eksplikativni (Kordeš 2013: 35) in elicitacijski intervju (Kordeš 2016: 16), od leta 2017 pa tudi mikrofenomenološki intervju (Bitbol, Petitmengin 2017). Metodo mikrofenomenološkega intervjuja²¹ bi lahko uporabili pri raziskovanju odrskega govora, da bi opisali psihično dogajanje v igralcu oz. pogledali v (ne)zavedni igralčev proces. To smo za potrebe gledališkega lektorstva verbalizirali na naslednji način: raziskati igralčevo doživljanje ob oblikovanju odrskega govora lika, ki ga oblikuje in predstavlja.

Čeprav lahko metodo mikrofenomenološkega intervjuja uporabimo z neizkušenimi intervjuvanci, je za kakovostno zbrane podatke nujno, da so intervjuvarji usposobljeni in celo izkušeni (Heinmann et al. 2023). Za izvajanje mikrofenomenološkega intervjuja in objavljanje izsledkov raziskave je treba imeti certifikat Mednarodnega združenja za mikrofenomenologijo – ker usmerjanje pozornosti na zavest in njeno opisovanje zahteva notranji napor in zato posebno

²⁰ Petitmengin je proučila nekatere neuroestetske raziskave, ki so uporabljale algoritmične oz. vsebinsko zaprte intervjuje, in ugotovila, da na oblikovanje prepričanj udeležencev raziskav bistveno vpliva situacija, v katero so v raziskavi postavljeni. Uvedla je novo tehniko empiričnega raziskovanja doživljanja, tj. eksplikativni intervju, s katerim poskušamo »kultivirati stik udeležencev z lastnim (nereflektiranim) izkustvom in jim skozi proces »evokacije« pomagati poustvariti čim bolj natančno sliko preteklega doživljanja« (Kordeš 2016: 15–17), udeleženec pri tem pa ni informator, temveč soraziskovalec.

²¹ Konkreten primer intervjuja je predstavljen v Petitmengin (2006: 261–266).

usposabljanje in strokovno znanje, tudi metaznanje intervjuvanca (Petitmengin 2006: 230, 237), in ker se je metodičnega vodenja, ki zagotavlja zanesljive rezultate, mogoče priučiti (Petitmengin 2006: 255). Gre za vprašanje ustreznega doseganja omenjene fenomenološke redukcije, o čemer so bili v stroki izraženi pomisleki, obenem pa je bilo predlagano, naj mikrofenomenološki intervju, ki je že precej dobro zasnovan, v dobro stroke uporablja čim več raziskovalcev (Strle 2013: 49–51).²²

Večino temeljnih konceptov, ki jih uporabljamo pri raziskovanju doživljanja, je definiral Husserl, zato se je za omenjeno raziskovanje uveljavilo ime fenomenološko raziskovanje; ker pa se je Husserl upiral zbiranju podatkov na podlagi opazovanja, je bil poimenovanju dodan pridevnik empiričen, pojavlja pa se tudi izraz prvoosebno raziskovanje (Kordeš 2013: 30). Izraz empirična fenomenologija naj bi se uporabljal za vse vrste empiričnega raziskovanja doživljanja, ki poteka na fenomenološki način, zajema pa vse pristope k raziskovanju izkustva, ki poskušajo omejiti predsodke, prepričanja in sodbe o izkustvu; prvoosebna metoda pa je vsaka raziskovalna metoda, s katero zbiramo prvoosebne oz. fenomenološke podatke, torej podatke o doživljanju, pri tem pa gre pogosto za intervjuje (Kordeš, Demšar 2023).

Če sledimo tej širši opredelitvi empirične fenomenologije, je najustreznejše poimenovanje eksplikativni intervju (Kordeš 2013: 35; Strle 2013: 47; Kordeš, Smrdu 2015: 42–43). Ker je ta v literaturi (Kordeš, Smrdu 2015) skopo opisan, je priporočljivo izhajati tudi iz izhodišč, ki jih predstavlja Petitmengin (2006). Značilnost kvalitativnega raziskovanja namreč je, da je kompleksno, dinamično in neurejeno ter da povezuje različne teorije, metode, strategije in tehnike (Kordeš, Smrdu 2015: 27), sprejeti pa moramo tudi dejstvo, da pri kvalitativnem raziskovanju lahko uporabimo različne metode in eksperimente²³ oz. da lahko prihaja do nezaželenega povezovanja različnih metod (Dragan 2022: 10). To povezovanje pa razumemo kot povezovanje komplementarnih metod, ki vodi v integracijo znanj (Dragan 2022: 14).

²² Prav dobra zasnovanost oz. utemeljenost mikrofenomenološkega intervjuja kot metode in poziv k njegovi čim širši uporabi sta ključna razloga, da smo v tem prispevku za termin *L'entretien d'explicitation* izbrali poimenovanje mikrofenomenološki intervju.

²³ Iris Indigène je v diplomskem delu *Enquête sur le processus de création en art contemporain* (2021) pri raziskovanju nastanka ustvarjalne ideje sodobnih vizualnih umetnikov uporabila več metod: poleg dveh različnih eksperimentov tudi t. i. dolgi intervju, pri čemer se je osredotočala na besedišče, ki so ga intervjuvanci uporabljali pri opisovanju svojih ustvarjalnih procesov, ter mikrofenomenološki intervju (povzeto po <https://www.microphenomenology.com/artistic-projects>).

S tega vidika naj mikrofenomenološki intervju z igralcem opravi npr. lektor, ki pozna (osnovne) tehnike nevrolingvističnega programiranja za odkrivanje kognitivnih procesov in strategij, ki omogočajo lažje vplivanje na te procese; med izhodišči mikrofenomenološkega intervjuja so namreč poleg različnih teorij in metod tudi tehnike nevrolingvističnega programiranja (Petitmengin 2006: 231, 244). Poleg tega naj ima intervjuvar profesionalne praktične izkušnje na področju dela z igralci in lektorstvom,²⁴ torej da je raziskovalec sestavni del sistema, ki ga opazuje (Kordeš 2008: 12), posledično pa naj ima tudi določene veščine za kvalitativno raziskovanje, ki jih zgolj s proučevanjem literature ni mogoče usvojiti (Kordeš, Smrdu 2015: 15). Osnovna motiva intervjuvarja – izhajajoč iz strateških principov kvalitativnega raziskovanja – naj bosta radovednost ter želja obvladovati in rešiti praktični problem, obenem pa je treba v raziskavi zavzeti receptivno držo, kar pomeni biti dojemljiv, pozoren poslušalec in opazovalec (Kordeš, Smrdu 2015: 16).

5 Zaključek

S prilagodljivo metodo mikrofenomenološkega intervjuja ustrezno usposobljen intervjuvar z receptivno držo lahko na podlagi natančnega opisa izkustva zbere znanstveno objektivne podatke o igralčevem prvoosebne, subjektivnem doživljanju ob oblikovanju odrskega govora lika, ki ga oblikuje in predstavlja. S to metodo, ki temelji na empirični fenomenologiji, lahko preverimo, kako igralčev mentalni proces vpliva na udejanjenje odrskega govora na različnih ravneh jezika: od glasoslovja do stavčne oz. besedilne fonetike (izbira intonacije ter posredovanje pomena in smisla). Ob tem se zavedamo dejstva, da vsak dramski igralec in gledališki lektor v uprizoritvenem procesu deluje na specifičen način, ocenjujemo pa, da bi bila metoda mikrofenomenološkega intervjuja poskus približevanja k najmanj raziskanemu področju odrskega govora – ustvarjanju glasovnega sloja, in to na znanstveno objektivni način.

Literatura

- Christopher B. BALME, 2010: *The Cambridge introduction to theatre studies*. New York: Cambridge University Press.
- Michel BITBOL, Claire PETITMENGIN, 2017: Neurophenomenology and the Micro-phenomenological Interview. Susan Schneider, Max Velmans (ur.): *The Blackwell Companion to Consciousness*. Chichester: Blackwell. 726–739.

²⁴ Kordeš in Smrdu (2015: 21) navajata naslednji moto: »Kadar razumemo bolje, bolje vprašamo.«

- Uršula BERLOT POMPE, 2016: Nevroumetnost, neuroestetika in vprašanje zavesti. Uršula Berlot Pompe, Barbara Predan, Petra Černe Oven (ur.): *Umetnost v času nevroznanosti: kritika, estetika, perspektive*. Ljubljana: Časopis za kritiko znanosti XLIV/265. 39–52.
- Rhonda BLAIR, 2006: Image and Action: Cognitive Neuroscience and Actor-Training. Bruce McConachie, F. Elizabeth Hart (ur.): *Performance and Cognition: Theatre Studies and the Cognitive Turn*. London, New York: Routledge. 167–185.
- Rhonda BLAIR, 2008: *The Actor, Image, and Action: Acting and Cognitive Neuroscience*. New York: Routledge.
- Oskar DRAGAN, 2022: Primerjava mikrofenomenološke in konstruktivistične poskusne teoretske analize. *Magistrsko delo*. Ljubljana: Universität Wien, Univerzita Komenského v Bratislave in Eötvös Loránd Tudományegyetem. Dostopno na <http://pefprints.pef.uni-lj.si/id/eprint/7115> (10. 8. 2023).
- Sergej FLERE, 2000: *Sociološka metodologija: temelji družboslovnega raziskovanja*. Maribor: Pedagoška fakulteta.
- Gledališki terminološki slovar*. Pregledana in dopolnjena izdaja. Dostopno na www.fran.si (21. 6. 2023).
- Katrin HEIMANN, Hanne BESS BOELSBJERG, Chris ALLEN, Martijn van BEEK, Christian SUHR, Annika LÜBBERT, Claire PETTMENGIN, 2023: The lived experience of remembering a 'good' interview: Micro-phenomenology applied to itself. *Phenomenology and the Cognitive Sciences* 22, 217–245. Dostopno na <https://doi.org/10.1007/s11097-022-09844-4> (20. 7. 2023).
- Štefan IVANKO, 2007: *Raziskovanje in pisanje del: metodologija in tehnologija raziskovanja ter pisanja strokovnih in znanstvenih del*. Kamnik: Cubus image.
- Urban KORDEŠ, 2008: Fenomenološko raziskovanje v psihoterapiji. *Kairos: slovenska revija za psihoterapijo*, 2/3–4, 9–21.
- Urban KORDEŠ, 2013: Negotova pot: od izkustva do empiričnega podatka. *Analiza: časopis za kritično misel* 17/1–2. 27–40.
- Urban KORDEŠ, 2016: Dve vrsti svobode. Uršula Berlot Pompe, Barbara Predan, Petra Černe Oven (ur.): *Umetnost v času nevroznanosti: kritika, estetika, perspektive*. Ljubljana: Časopis za kritiko znanosti XLIV/265. 13–22.
- Urban KORDEŠ, Ema DEMŠAR, 2023: Horizons of becoming aware: constructing a pragmatic-epistemological framework for empirical first-person research. *Phenomenology and the cognitive sciences* 22/2, 339–367.
- Urban KORDEŠ, Maja SMRDU, 2015: *Osnove kvalitativnega raziskovanja*. Koper: Založba Univerze na Primorskem.
- John LUTTERBIE, 2006: Neuroscience and creativity in the rehearsal process. Bruce McConachie, F. Elizabeth Hart (ur.): *Performance and Cognition: Theatre Studies and the Cognitive Turn*. London, New York: Routledge. 149–166.
- Bruce McCONACHIE, F. Elizabeth HART (ur.), 2006: *Performance and Cognition: Theatre Studies and the Cognitive Turn*. London, New York: Routledge.
- Kristijan MUCK, 2000: Blodnjak in tehtnica. Katja Podbevšek, Tomaž Gubensšek (ur.): *Kolokvij o umetniškem govoru*. Ljubljana: AGRFT. 3–8.
- Robin NELSON, 2013: *Practice as Research in the Arts: Principles, Protocols, Pedagogies, Resistances*. London: Palgrave Macmillan.
- Urška PERENIČ, 2014: *Empirija v literarni vedi*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- Claire PETTMENGIN, 2006: Describing one's Subjective Experience in the Second Person: An Interview Method for the Science of Consciousness. *Phenomenology and the Cognitive Sciences* 5, 229–269.
- Katarina PODBEVŠEK, 2006: *Govorna interpretacija literarnih besedil v pedagoški in umetniški praksi*. Ljubljana: Slavistično društvo Slovenije.
- Katarina PODBEVŠEK, 2008: Govor kot gledališko izrazilo (na primeru Flisarjevega Akvarija). Mateja Pezdirc Bartol (ur.): *Slovenski jezik, literatura, kultura in mediji* (44. SSJLK). Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete. 51–59.

- Katarina PODBEVŠEK, 2010: Spreminjanje odrske govorne estetike v slovenskem gledališču 20. stoletja. Barbara Sušec Michieli, Blaž Lukan, Maja Šorli (ur.): *Dinamika sprememb v slovenskem gledališču 20. stoletja*. Ljubljana: AGRFT, Maska. 195–238.
- Katarina PODBEVŠEK, 2016: Kreativno govorjenje. Blaž Lukan (ur.): *Zbornik ob 70-letnici Akademije za gledališče, radio, film in televizijo Univerze v Ljubljani*. Ljubljana: AGRFT. 46–48.
- Katarina PODBEVŠEK, 2017: *Govornost literarnih besedil*. Maribor: Aristej.
- Katja PODBEVŠEK, Tomaž GUBENŠEK (ur.), 2000: *Kolokvij o umetniškem govoru*. Ljubljana: AGRFT.
- Katarina PODBEVŠEK, Tomaž GUBENŠEK (ur.), 2006: *Kolokvij o umetniškem govoru 2*. Ljubljana: AGRFT.
- Katarina PODBEVŠEK, Nina ŽAVBI MILOJEVIĆ (ur.), 2013: *Govor med znanostjo in umetnostjo*. Maribor: Aristej.
- Eva PORI, 2020: Primer Oder 57: poskus rekonstrukcije estetike igralskega govornega in telesnega izraza. *Jezik in slovnost* 65/2. 47–61.
- Gabriele SOFIA, 2016: Introduction: Towards an Embodied Theatrolology? Clelia Falletti, Gabriele Sofia, Victor Jacono (ur.): *Theatre and Cognitive Neuroscience*. London, New York: Bloomsbury Methuen Drama.
- Bert O. STATES, 1985: *Great Reckonings in Little Rooms: On the Phenomenology of Theater*. Berkeley, Los Angeles, London: University of California.
- Toma STRLE, 2013: Uganka izkustva: vloga prvo- in drugoosebnih metod v kognitivni znanosti. *Analiza: časopis za kritično misel* 17/1–2. 41–57.
- Jože TIRAN, 1965: *Umetniško pripovedovanje*. Ljubljana: Knjižnica MGL.
- Hotimir TIVADAR, 2004: Podoba in funkcija govornega knjižnega jezika glede na neknjižne zvrsti. Erika Kržišnik (ur.): *Aktualizacija jezikovnozvrstne teorije na Slovenskem*. Ljubljana: Center za slovenščino kot drugi/tuji jezik pri Oddelku za slovenistiko Filozofske fakultete (Obdobja 22). 437–452.
- Miha TREFALT, 2009: Govor je človekova intimna zadeva. *SiGledal, portal slovenskega gledališča*. Dostopno na <https://veza.sigledal.org/prispevki/govor-je-clovekova-intimna-zadeva> (21. 6. 2023).
- Francisco J. VARELA, 2013: Nevrofenomenologija: metodološka rešitev za težki problem. *Analiza: časopis za kritično misel* 1–2, 85–108.
- Martin VRTAČNIK, 2012: Gledališki lektor – njegova funkcija in namen v sodobnosti. *Jezik in slovnost* 57/3–4, 101–114.
- Martin VRTAČNIK, Hotimir TIVADAR, 2017: Sodobni pristop k jezikovnemu svetovanju v gledališču na Slovenskem. *Slavia Centralis* 10/1, 61–75.
- Nina ŽAVBI MILOJEVIĆ, 2013: Analiza odrskega govora – primer Bergerjeve uprizoritve Hlapcev (komentirana izdaja). *Slavistična revija* 61/4, 651–664.

RAZISKOVANJE GOVORJENEGA UMETNIŠKEGA JEZIKA

NINA ŽAVBI

Univerza v Ljubljani, Akademija za gledališče, radio, film in televizijo, Ljubljana,
Slovenija,
nina.zavbi@agrft.uni-lj.si

V prispevku poskušam prikazati razvoj raziskovanja umetniškega (odrskega) govora na Slovenskem. Osredotočam se na sodobni model raziskovanja, pri katerem kombiniramo slušnozaznavno in akustično analizo. Opredeljujem prednosti takšnega načina in predstavljam strategijo raziskovanja, ki odrski govor preučuje v več korakih: na začetku v razmerju dramskega in uprizoritvenega besedila (ki od dramskega bolj ali manj odstopa zaradi dramaturškega črtanja, dopisovanja, spreminjanja jezikovne zvrstnosti, posodabljanja besedila ipd.), v drugem koraku pa kot govorno izvedbo na odru. Odrski govor se v okviru analize razume kot le eden od uprizoritvenih dejavnikov. Raziskovalne rezultate interdisciplinarno kontekstualizira in tako osmišlja preučevanje umetniškega fenomena odrskega govora, ki povezuje znanost in umetnost.

DOI
[https://doi.org/
10.18690/um.ff.4.2024.22](https://doi.org/10.18690/um.ff.4.2024.22)

ISBN
978-961-286-882-6

Ključne besede:
jezik,
govor,
umetniški govor,
akustična analiza,
Praat



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024.22](https://doi.org/10.18690/um.ff.4.2024.22)

ISBN

978-961-286-882-6

Keywords:

language,
speech,
artistic speech,
acoustic analysis,
Praat

RESEARCHING ARTISTIC SPEECH

NINA ŽAVBI

University of Ljubljana, Academy of Theatre, Radio, Film and Television, Ljubljana,
Slovenia

nina.zavbi@agrft.uni-lj.si

The paper aims to show the development of research into artistic (stage) speech in Slovenia. It focuses on the contemporary research model that combines auditory-recognition and acoustic analysis. It defines the advantages of the method and presents a research strategy in interrelated steps: in the first step, the researcher compares the dramatic script with the transcript of the stage speech (that differs from the dramatic script due to dramaturgical cuts, the addition of texts, changes in the language genre, updated language, etc.); in the next step, the researcher analyses the realisation of the stage speech. In the context of the analysis, the stage speech is understood as only one of the staging factors. Interdisciplinary research contextualises the results and gives meaning to the study of the artistic phenomena of stage speech, which combines science and art.



1 Uvodne besede ali uvid v pomembnost raziskovanja gledališkega govora

Raziskovanje govora je v primerjavi z raziskovanjem jezika precej zapostavljeno, še bolj pa je zapostavljeno raziskovanje umetniškega govora. Govor kot zvočna realizacija jezika se dogaja v določenih besedilnih in zunajbesedilnih okoliščinah. Poleg zvočne ubeseditve ga spremlja tudi vidna, npr. mimika, geste, premikanje telesa v prostoru – šele z opazovanjem obeh lahko v polnosti razumemo in analiziramo vse pomembne govorne prvine.

Govorjeni umetniški jezik se pogosto naslanja na zapisano predlogo, na umetniško besedilo, ki je nato govorno interpretirano pred publiko; pogosto ga ustvarjalno oblikujejo govorni profesionalci (npr. igralci). To drži tudi za odrski govor, ki je govorna izvedba dramskega besedila (predvsem v dramskem gledališču, v nekaterih drugih oblikah odrske umetnosti pa se lahko tudi zelo približa spontanemu, zasebnemu itd.). Gre za govor, ki skuša dajati vtis sprotne tvorjenosti, naravnosti, spontanosti, kljub temu da to ni. Osebe, ki besedilo izgovarjajo v gledališki uprizoritvi, v dobri izvedbi zvenijo tako, kot da ga ubesedujejo prvič, torej spontano, v resnici pa je govor vnaprej pripravljen po besedilni predlogi. Vsi jezikovni in govorni vidiki (zvrstnost; besedje; pravorečni elementi – naglasi, izgovor glasov; prozodična sredstva – hitrost, glasnost govora, intonacija, premori, register itd.; tudi vidna nebesedna govorna sredstva – mimika, geste ipd.) so v resnici do popolnosti premišljeni, ozaveščeni in izvedbeno fiksirani.

2 Metodologija raziskovanja v preteklosti in danes

Pri raziskovanju odrskega govora so v zgodovini uporabljali različne metode, predvsem slušnozaznavno analizo (npr. Podbevšek 2008). Pri tej je raziskovalec poslušal govorno umetniško besedilo in ga poskušal čimbolj objektivno analizirati. Za večjo zanesljivost sta pogosto iste posnetke poslušala dva raziskovalca in primerjala rezultate.

V sodobnosti se s slušnozaznavno povezuje akustična (fonetična) analiza z različnimi računalniškimi programi (npr. Praat). Model raziskovanja (Žavbi 2016, 2019) slovenskega odrskega govora se zgleduje po raziskavah odrskega govora nekaterih

hrvaških raziskovalcev (Varošaneč-Škarić 2005, Vrban Zrinski 2022) in slovenskega medijskega govora (Tivadar 2004, Huber 2017).

Podrobnejše znanstvene analize govornega umetniškega jezika so (v zgodovini) postale mogoče komaj takrat, ko je bilo tehnično omogočeno ponovno poslušanje govora – razvoj tehnologije in dostopnost različnih posnetkov. Pred tem so raziskovalci izhajali le iz ogleda in poslušanja posamezne predstave, torej se je govor popisovalo bolj celostno, tudi po spominu ter na podlagi kritičkih zapisov. Avdioposnetki so omogočili večkratno poslušanje, zato tudi že poglobljeno slušno analizo. Kasneje dostopni videoposnetki so omogočili hkratno preučevanje tudi vidnih spremljevalcev govora. Za akustično analizo, ki prinese večjo mero objektivnosti (s preverjanjem slušnih zaznav), potrebujemo primerne računalniške programe, npr. Praat, ki so prosto dostopni in dokaj enostavni za uporabo, za doseganje kriterija relevantnosti pa je po opravljenih meritvah ključna poglobljena interpretacija rezultatov oziroma postavljanje pridobljenih podatkov v kontekst (tako kontekst dramskega besedila kot tudi celotne uprizoritve ter zunanjih dejavnikov, npr. velikost dvorane, družbeni kontekst, umetniški kontekst oz. obdobja, dramski avtorji itd.). V nadaljevanju prispevka opredeljujem prednosti takšnega načina preučevanja in predstavljam interdisciplinarno strategijo raziskovanja, ki raziskovalne rezultate kontekstualizira ter tako osmišlja preučevanje umetniškega fenomena odrskega govora, ki povezuje znanost in umetnost.

3 Nastanek govornega jezika v gledališču – sodelovanje režiserja, dramaturga, lektorja in igralca

V gledališču v ustvarjalnem procesu, v katerem sodeluje skupina umetnikov različnih profilov, iz zapisane (navadno dramske) predloge nastane najprej uprizoritveno besedilo,¹ nato pa govorni odrski jezik,² ki skupaj z ostalimi uprizoritvenimi dejavniki (scena, mizanscena, kostumi, glasba, igra itd.) tvori uprizoritev.³

¹ Besedno zvezo 'uprizoritveno besedilo' v prispevku uporabljam v pomenu »besedil[a], ki ga v uprizoritvi govorijo nastopajoči in se včasih razlikuje od zapsanega dialoga v dramskem besedilu« (Humar idr. 2007: 193). Gre torej za neke vrste scenarij za gledališko uprizoritev, na podlagi katerega nastane odrski govor uprizoritve.

² Odrski govor »temelji na razločni izreki, ustreznosti slinosti, glasovni izraznosti, usklajen z besedilnimi in odrskimi okoliščinami, gledališko estetiko« (Humar idr. 2007: 193).

³ »Navadno na režijski, dramaturški, scenografski zamisli temelječa uresničitev dramskega besedila, druge predloge v obliki čutno dojemljivega dogajanja z nastopajočimi igralci, pevci, plesalci« (Humar idr. 2007: 135).

Strokovnjak za jezik in govor – gledališki lektor⁴ v prvi fazi nastajanja odrskega govora v sodelovanju z režiserjem in dramaturgom pripravlja besedilno predlogo za uprizoritev – uprizoritveno besedilo, pri čemer se dramska predloga primerno dramaturško sčrta, besedilo se lahko tudi dopiše, spreminja se vrstni red replik ipd. Nato se določi zvrst, ki se jo (glede na dramsko predlogo) pogosto spreminja – lahko v smeri pogovornega jezika, posamezne vloge se lahko zvrstno gradi glede na njihovo socialno, izobrazbeno idr. ozadje. Taka pripravljena besedilna predloga se na prvih bralnih vajah, ki potekajo za mizo, še ne v prostoru, z igralci prebere, pri čemer se deloma že določi besedilnofonetična struktura govora (premore, poudarki, intonacija, govorni ritem, glasovne modulacije itd.), dogovori se o jezikovnofonetični ravni – npr. o naglasih (predvsem je pomembna izbira med naglasnimi dvojnicami, upoštevanje/neupoštevanje pravorečne norme in določitev naglasnih mest, dolžine in kakovosti naglasov) ter opozarja se na različne glasovne premene, predvsem premene po zvanečnosti. Branje je na tej stopnji večinoma še neinterpretativno.

Po tem, ko se uprizoritev že postavi v prostor, se lektor ponovno vključuje v smislu preverjanja, ali dogovorjena govorna struktura deluje – tako igralce popravlja, jim svetuje. Na tej stopnji lahko ugotovi, da je treba določene dele spremeniti, popraviti. Manj se ukvarja s pravorečjem (s tem, na kar nakazuje ime gledališki lektor), bolj pa z ustvarjalno platjo odrskega govora, torej kako govor funkcionira v uprizoritvi. Glede na posamezno vlogo igralcem tudi svetuje, kako z glasovno ustvarjalnostjo doseči različne učinke.

⁴ Poimenovanje je precej problematično, saj beseda lektor pogosto asociira z lektoriranjem zapisanega jezika in se razume kot popravljanje napak – torej v smislu, da lektor pripravlja normativno neoporečno besedilo. Gledališki lektor pa pripravlja ustrezno govorjeno besedilo, ki ni nujno v zborni različici jezika, ampak se lahko uporablja zelo različne zvrsti in tako pogosto namenoma krši normo. Predlagajo se različne zamenjave za poimenovanje gledališki lektor, npr. govorni svetovalec (Stanič 2006) in oblikovalec govora (Vrtačnik 2012).

4 Raziskovalna strategija preučevanja govora v gledališču⁵

4.1 Raziskovanje dramskega in uprizoritvenega besedila – preučevanje jezika

Po prikazanem procesu ustvarjanja gledališkega govora je jasno, da raziskovanje ne more potekati le na ravni govora, ki je le rezultat vseh stopenj ustvarjalnega procesa. Najprej je treba raziskati jezik uprizoritvenega besedila, nastalega iz dramske predloge, se torej poglobiti v prvo stopnjo v procesu ugledališčenja. V zgodovini je razmerje med uprizoritvenim in dramskim besedilom zelo raznoliko, od tega, da uprizoritev čimbolj v popolnosti sledi dramskemu besedilu, do tega, da od njega radikalno odstopa (Žavbi 2016, 2019), kar dokazujejo predvsem uprizoritve v zadnjih desetletjih, pogosto že v naslovu opredeljene »po motivih« literarnega dela, »z mislijo na« literarno delo ipd. V sodobnem gledališču besedilo ni več sveto – premisa, da je edini avtor besedila dramatik, ki ga je treba pri nastanku uprizoritve v celoti upoštevati, danes nikakor ne drži. V sodobnem gledališču je avtorjev uprizoritvenega besedila več – prvi je avtor (dramske) predloge, drugi pa je uprizoritvena ekipa (dramaturg, lektor, igralci) z režiserjem na čelu. Ta ekipa dramsko besedilo bolj ali manj spremeni – gledališki lektor sodeluje že pri tem koraku, raziskovalec gledališkega govora pa tudi raziskuje jezik uprizoritvene predloge.

V tej točki raziskovalca zanima, v kolikšni meri jezik uprizoritvenega besedila odstopa od dramskega – koliko je dramaturških črt in kakšne so, kakšno je morebitno dopisano besedilo (ali je blizu dramskemu ali je diametralno drugačno). Nato pa se precej ukvarja z jezikovno zvrstnostjo uprizoritvenega besedila – ali je ekipa spreminjala jezikovno zvrst in na kakšen način je to naredila, ali je posodabljala jezik ipd. V sodobnem gledališču je možnosti neskončno. Pogosto se zvrstnost niža v smeri pogovornih različic jezika, kar lahko služi tudi karakterizaciji, npr. različni liki govorijo v različnih socialnih zvrsteh, glede na to, iz kakšnega okolja prihajajo, koliko so stari, kakšna je njihova izobrazba, kakšna so hierarhična razmerja med njimi, različna čustvena stanja, govorni položaji itd. Raziskovalec odrskega govora zato ne more biti le jezikoslovec in popisati, kakšno je uprizoritveno besedilo, npr. na fonetični ravni, na ravni besedja, skladnje itd., ampak mora ta spoznanja analizirati

⁵ Raziskovalna strategija je podrobneje razčlenjena v doktorski disertaciji *Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja* (Žavbi 2016) in v prispevku z naslovom *Odrski govor – slušnozaznavna in računalniška fonetična analiza uprizoritev Cankarjevih Hlapcev* (Žavbi 2019).

in interpretirati v kontekstu, in sicer interdisciplinarno. Omenjene značilnosti je treba razumeti tudi v razmerju do dramskega besedila in dramatika, ki je besedilo napisal, v okviru režijske poetike režiserja ali celotne ekipe, ki besedilo postavlja na oder, tudi v okviru konkretnih gledaliških in družbenih okoliščin časa, v katerem uprizoritev nastane ipd.

4.2 Raziskovanje na odru izrečenega besedila – preučevanje govora

Na drugi točki se zapisano uprizoritveno besedilo z igralsko govorno interpretacijo pretvarja v govorno različico jezika. Odrski govorni interpreti (igralci) imajo prostor za lastno govorno kreativnost – govorno interpretacijo – zlasti v polju prozodičnih izraznih sredstev (intonacija, premori, tempo, jakost, register, barva glasu, glasovne modulacije itd.) (Žavbi 2013: 651, 652), pomemben pa je tudi dogovor o pravorečnih usmeritvah odrskega govora, pri čemer ima ključno vlogo gledališki lektor. Igralec pri govorni interpretaciji literarnega besedila »ustvarjalno uresničuje besedilo z govornimi sredstvi in pri tem usklajuje besedilne posebnosti z govorno podobo« (Podbevšek 2007: 20). »Nastajanje odrske govorne interpretacije dramskega besedila [...] je neločljivo povezano z drugimi uprizoritvenimi komponentami, kot so scena, kostumi, osvetljava, mizanscena, glasba, gib itd., bistveno pa je odvisno od režijskega koncepta« (Žavbi 2013: 651).

Ta del raziskave obsega analizo govora tako v smislu govora posameznega igralca (posamezna vloga) ter tudi na ravni uprizoritve kot celote. Odrski govor je analiziran s kombinacijo dveh metod: slušnozaznavne (raziskovalčeva slušna zaznava prozodičnih sredstev: glasnost, hitrost govora, intonacija in register, ritem, barva glasu, premori itd.) in računalniške fonetične (rezultati slušnozaznavne metode so preverjeni s pomočjo fonetičnega računalniškega programa Praat). »Fonetična analiza s Praatom omogoča znanstveno preučevanje, za katerega je ključna objektivnost, zanesljivost« (Žavbi 2019: 55). Fonetično raziskujemo s pomočjo posnetka uprizoritve. Poslušamo in analiziramo posamezne replike, si v transkripciji izgovorenega slišane besedila zaznamujemo slušne zaznave (glasnost, hitrost govora, premori, register in intonacija, barva glasu itd.) in vidno (mimika, geste, premikanje v prostoru) neverbalno komponento govora. Nato svoja opažanja raziskovalec preverja s fonetičnim računalniškim programom Praat, s katerim izmeri glasnost, register, izračuna hitrost govora, intonacijske razpone, dolžine premorov itd. Računalniško izriše tudi posamezne ilustrativne prikaze, npr. potek intonacije, glasnosti ipd., s čimer preveri in potrdi ali ovrže svoja slušna zaznavanja. Preverjanje

s fonetičnimi računalniškimi programi se je izkazalo kot izjemno uspešno zlasti pri nekaterih primerih, ko nas uho lahko vara oziroma so problematični za slišanje. Omenim lahko gibanje predvsem končne intonacije, zlasti pri čustveno obarvanih replikah, ki so izgovorjene glasno, hkrati pa imajo zadnje besede v povedi končni naglas (torej na zadnjem zlogu) – včasih nas naša slušna zaznava zaradi teh dejavnikov vara in slišimo rastočo intonacijo, v resnici pa se izkaže, da je intonacija padajoča. Prav tako je smiselno pri primerih t. i. ravne intonacije slušno zaznavo preverjati s Praatom, saj le objektivna potrditev da relevantne rezultate (pomembna je njihova preverljivost).

4.3 Interdisciplinarni pristop – interpretacija raziskovalnih rezultatov

Po podrobnem preučevanju tako jezika uprizoritvenega besedila kot odrskega govora v dveh korakih raziskovalec ugotavlja, katero jezikovno zvrst je uporabila umetniška ekipa ter katera govorna sredstva so uporabljali posamezni igralci. Vendar pa rezultati sami na sebi ne povedo zadosti. »Cilj pr[e]učevanja ni le pridobiti fonetične informacije o rabi prozodičnih sredstev v posamezni uprizoritvi [...], ampak interpretirati, kako odrski govor igralca učinkuje v konkretnem odlomku« (Žavbi 2019: 56). Raziskovalca ne zanima le, katera sredstva je uporabil igralec, ampak predvsem, zakaj jih je uporabil in zakaj v točno določenem odlomku, v konkretni repliki, zato rezultate interpretira interdisciplinarno – spoznanja jezikoslovne vede (fonetika, besedilna fonetika itd.) postavlja v kontekst celotne uprizoritve (teatrologija) in poskuša govor razumeti kot le enega od uprizoritvenih dejavnikov. Upošteva tudi druge vede (sociologija, psihologija, filozofija), razmišlja o vplivu različnih zunajjezikovnih okoliščin (družbenih, političnih, kulturnih) na govor.

5 Primer raziskovanja konkretne uprizoritve

V nadaljevanju poskušam predstavljeno raziskovalno strategijo prikazati na primeru uprizoritve Cankarjevih *Hlapcev* v režiji Sebastijana Horvata leta 2015 v Slovenskem stalnem gledališču Trst.⁶

⁶ Ivan Cankar: *Hlapci*. Slovensko stalno gledališče Trst. Premiera: marec 2015. Režiser: Sebastijan Horvat. Avtor priredbe in dramaturg: Milan Marković Matis. Scenograf: Jürgen Kirner. Kostumografka: Belinda Radulović. Skladatelj: Drago Ivanuša. Lektorica: Tatjana Stanič. Asistent režije: Žiga Divjak. Igrajo: Radko Polič – Jerman, Romeo Grebenšek – mladi Jerman, Jure Kopušar – župnik, Nikla Patruška Panizon – Lojzka, Iztok Drabik Jug – nadučitelj, Primož Forte – Kalander, Maja Blagović – mati, Matija Rupel – Komar, Patrizia Jurinčič – Minka, Tina Gunzek – Anka, Luka Cimprič – Pisek.

5.1 Analiza uprizoritvenega besedila (jezik)

Uprizoritev je precej sodobna, saj gre za adaptacijo drame *Hlapci*, ki je le »osnova, na kateri je grajena uprizoritev« (Žavbi 2016: 224). Uprizoritveno besedilo se v veliki meri razlikuje od drame Ivana Cankarja. Del, ki se dogaja v preteklosti, sledi besedilu drame – vendar s številnimi in obsežnimi dramaturškimi črtami, tudi s spremenjenim vrstnim redom replik in različnih delov dramskega besedila, nekatere osebe iz drame v uprizoritvi sploh niso prisotne. Drugi del uprizoritvenega besedila je dopisan in se dogaja v sedanjosti ostarelega Jermana (Radko Polič). Tudi jezikovne zvrsti uprizoritvenega besedila zato ne sledijo v celoti jezikovnim zvrstem dramskega besedila. Včasih se besedilo približuje pogovornim različicam, v nekaterih delih pa je jezik posodobljen tudi na ravni besedja, besednih zvez, stavkov in povedi. Besedni vrstni red je pogosto spremenjen v smeri danes slogovno bolj nezaznamovanega. Besedilo je manj spremenjeno v delih, ki se dogajajo v preteklosti, ki predstavljajo spomine ostarelega Jermana – v teh delih jezikovnozvrstno večinoma sledi drami. V drugem delu so iz drame prisotni le fragmenti, v zelo različnem zaporedju, večina besedila je na novo napisana; govori večinoma Jerman, ki je zaradi življenjskih izkušenj zelo drugačna oseba. Jezik je precej pogovoren, tudi približan sodobnemu – takšna je seveda tudi vsebina – ostareli Jerman preizprašuje svoje življenjske odločitve.

5.2 Analiza odrskega govora v dveh korakih⁷

V uprizoritvi *Hlapcev* v režiji Sebastijana Horvata se je izkazalo, da so prozodična sredstva rabljena precej različno, predvsem pa večkrat nepričakovano in zelo ustvarjalno – npr. igralci pogosto »kršijo« pravila intonacijskih potekov in (dolžine) premorov pri ločilih, npr. pri vejici in vprašajih. Značilnosti ustvarjalne rabe prozodičnih sredstev se namreč v interpretaciji kažejo predvsem v njihovi neustaljeni rabi. Premore v besedilu drame označene z ločili, igralci upoštevajo le v določeni meri. Glede na interpretacijo lahko vejice izgovarjajo različno, tako glede premorov (jih sploh ni ali pa so zelo dolgi) kot tudi glede intonacijskih potekov (rastoča, padajoča intonacija, končna, nekončna ...). Tudi hitrost govora ni odvisna le od jezikovnih znakov, ampak predvsem od vsebinskih poudarkov, ki so nastali v

⁷ Primer navajam iz doktorske disertacije *Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja* (Žavbi 2016: 224–278).

kontekstu uprizoritve. To pa lahko razložimo šele, ko prestopimo meje strogo fonetične analize, torej v fazi interpretacije rezultatov.

V prizoru, v katerem je izjemno zanimiv govor ostarelega Jermana, ki jemlje nekatere replike iz Jermanovega govora v krčmi (besedilo Ivana Cankarja), večinoma pa je tekst na novo zapisan ali so replike celo obrnjene, njihov pomen pa diametralno nasproten, je zelo inovativna igralska govorna interpretacija Radka Poliča v vlogi ostarelega Jermana.⁸ Za interpretacijo rezultatov analize je pomembna vsebina povedanega oziroma sporočilo ostarelega Jermana (logičnopomenska govorna realizacija).

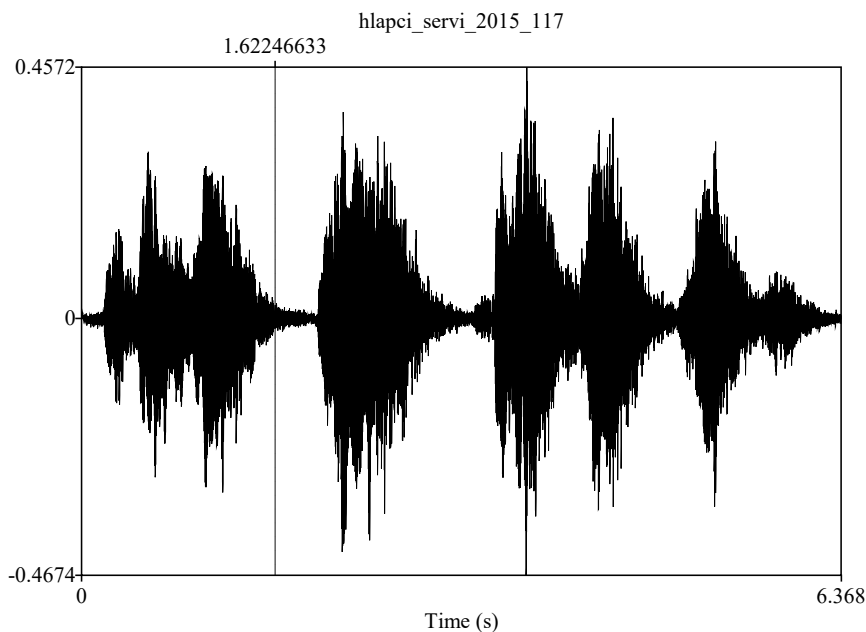
S slušnozaznavno analizo, pri kateri raziskovalec zapiše slišane premore, register, intonacijske poteke, hitrost, glasnost govora, govorni ritem, morebiti tudi govorne modulacije, je bilo ugotovljeno, da je govor precej čustveno obarvan. Na besedilnofonetični ravni je to razvidno iz glasnosti govora, ki je na nekaterih mestih upočasnjen ter vsebuje dolge premore, s katerimi je pogosto nakazana pomembnost povedanega (poudarjanje). Čustvenost je razvidna tudi iz velikih intonacijskih razponov (vzklik, krik) in na nekaterih mestih zelo povišanega registra. Pri preučevanju umetniškega govora je pomembna tudi neverbalna komunikacija, ki se jo opazuje hkrati s poslušanjem zvočne plati. V tem odlomku je takšno preučevanje oteženo – mimika je zelo slabo vidna, saj je Jerman sneman od zadaj. Na raziskovanje torej vpliva tudi zorni kot snemalca, saj govor uprizoritve lahko preučujemo le s pomočjo videoposnetka. Po celotni slušnozaznavni analizi, pri kateri raziskovalec zapiše slišane premore, register, intonacijske poteke, hitrost, glasnost govora, govorni ritem, morebiti tudi govorne modulacije, svojo slušnozaznavno analizo preverja z računalniškim programom Praat. Pri interpretaciji rezultatov je treba biti previden in upoštevati veliko zunanjih dejavnikov – tudi način in pogoje snemanja, ki se razlikujejo pri posameznih uprizoritvah, prav tako je pomembno upoštevati, kako so postavljeni mikrofoni ipd., kar je razvidno iz spodnjega primera.

Na primerih, ki jih predstavljam v nadaljevanju (akustična analiza s programom Praat), se pokaže pomembnost kombiniranja obeh analiz – slušnozaznavne in računalniške. Govorni način oziroma govorno interpretacijo Radka Poliča prikazujem kot vzorčni primer ustvarjalne igralske uporabe prozodičnih sredstev.

⁸ Zanimivo je, da je Radko Polič v vlogi Jermana nastopal že tri desetletja pred tem, in sicer leta 1980 v *Hlapcih* v režiji Mileta Koruna v produkciji SNG Drama Ljubljana. Predstava je bila po kritiških odzivih izjemna in v uprizorjanju *Hlapcev* tudi prelomna.

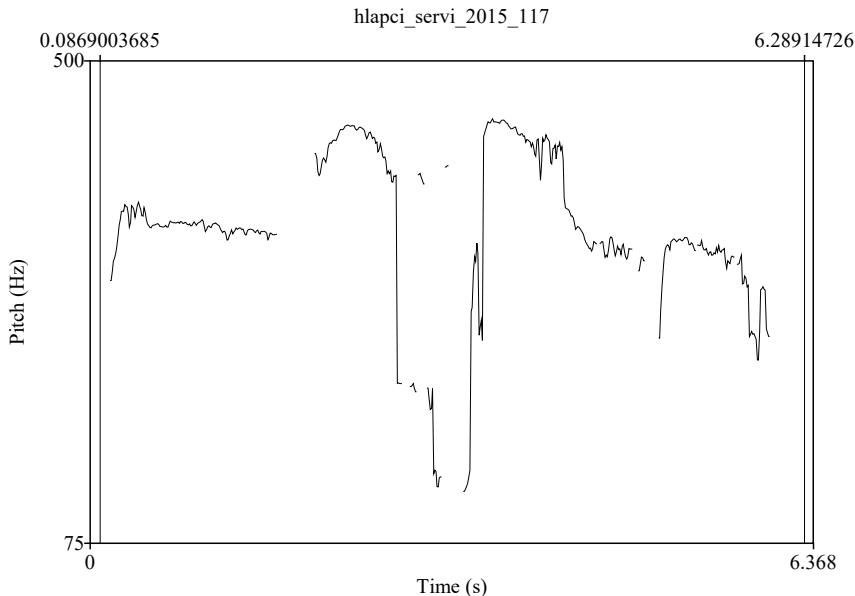
Prva prikazana replika Radka Poliča (graf 1 in graf 2): ***Skupaj smo in nismo hlapci!***

Iz slušne analize sem zaključila, da je replika izgovorjena zelo glasno, povprečno hitro (glede na ostali del govora), intonacije ob premorih (ki so trije med repliko in eden na koncu) so padajoče, register je visok. Kljub temu da je replika po slušnozaznavni analizi izgovorjena zelo glasno, je Praat izmeril povprečno glasnost 71,2 dB (graf 1). Razlog je v spremenjenih pogojih snemanja. Tako po slušni zaznavi kot po meritvah so intonacijski loki veliki – 328,2 Hz (120,8–449,0 Hz). To pomeni, da gre za vzklík, za veliko čustvenost govorca (graf 2). Izmerjena hitrost govora je 1,3 zloga v sekundi, kar pomeni, da Jerman govori zelo počasi. »V repliki sta dva premora, in sicer pred in za besedico *in* – drugi premor poudarja pomembnost besede, ki se nahaja za njim. Analiza s Praatom je (z izjemo glasnosti izgovorjene replike) potrdila slušnozaznavno analizo govora« (Žavbi 2016: 270).



Graf 1: Oscilogram replike ostarelega Jermanas

Vir: Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja (Žavbi 2016: 269)

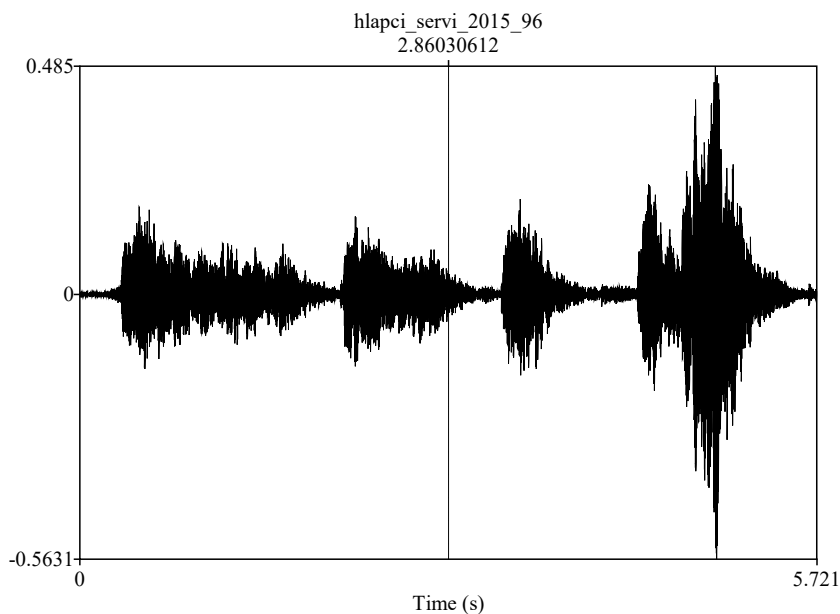


Graf 2: Intonacijski potek replike ostarelega Jermana

Vir: Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja (Žavbi 2016: 270)

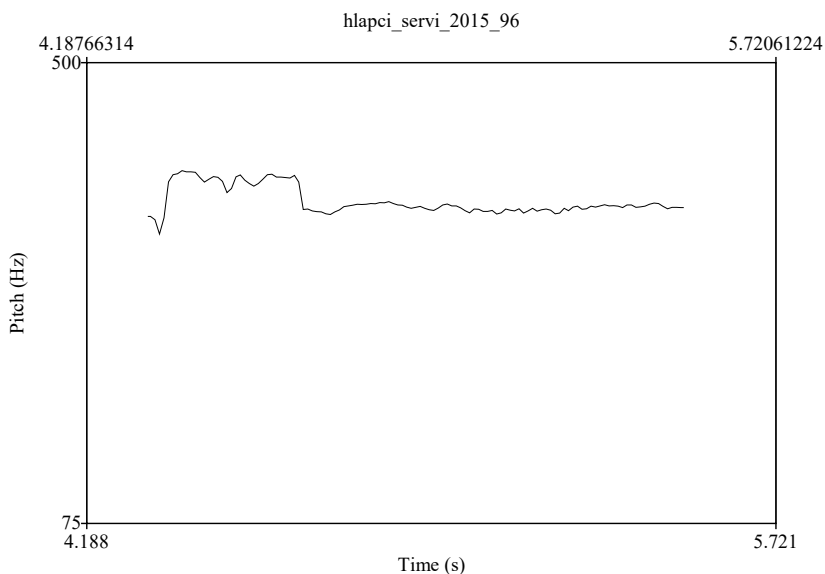
Druga prikazana replika Radka Poliča (graf 3 in 4): *Vsi smo pozabili, kaj pomeni bit skupaj.*

Igralec Radko Polič svojo govorno interpretacijo oblikuje prepoznavno individualno, zlasti v rabi višjega registra, zelo raznolikih premorov (tako po dolžini kot po njihovem mestu), glasnosti ter v nepričakovanih intonacijskih potekih, ki jih računalniški program zazna kot ravne, slišimo pa jih kot zelo rahlo padajoče (graf 4). Iz grafa 3 je razvidna ustvarjalna raba premorov, ki jih igralec izvede na skladijsko nepričakovanih mestih, in poudarek na zadnji besedi (glasneje izgovorjeni del), ki je vsebinsko utemeljen.



Graf 3: Oscilogram specifične govorne realizacije Radka Poliča

Vir: Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja (Žavbi 2016: 272)



Graf 4: Intonacijski posnetek specifične govorne realizacije igralca Radka Poliča

Vir: Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja (Žavbi 2016: 273)

5.3 Interpretacija rezultatov z interdisciplinarnim pristopom ter umestitev odrskega govora v uprizoritveno celoto

Iz uprizoritve *Hlapcev* v režiji Sebastijana Horvata je razvidna velika razlika med dramskim in uprizoritvenim besedilom ter velika govorna ustvarjalnost igralcev, ki se kaže v izvirni rabi prozodičnih in vidnih neverbalnih govornih sredstev. Uprizoritev je nastala leta 2015, v času, ko dramskemu besedilu ni več treba slepo slediti in je ustvarjalna ekipa v enaki meri avtor uprizoritve, kot je dramatik avtor dramskega besedila. Odrski govor funkcionira kot eden od elementov uprizoritve (skupaj s sceno, mizansceno, kostumi, glasbo itd.), usklajen mora biti z režijskim konceptom in je umetniška kreacija igralca. Iz povedanega je jasno, da je govor posameznih igralcev treba razumeti znotraj uprizoritve, ki je izrazito avtorska in precej odstopa od drame Ivana Cankarja. Tudi samo uprizoritveno besedilo se prilagaja temu konceptu, zato je pogosto črtano, dopisano, jezikovna zvrst je spremenjena. Igralec (kot je nakazano na primeru govorne realizacije Radka Poliča) uporablja ustvarjalne pristope in lahko precej odstopa od pričakovane govorne izvedbe – kar se z novimi metodami da tudi znanstveno dokazati.

V uprizoritvi se prepletata dva različna tipa odrskega govora: prvi je govor preteklosti v delu uprizoritve, ki odraža duha Cankarjevega časa, zato je v tem delu govor arhaičen, zvest dramski predlogi; drugi tip govora je sodoben, zato tudi jezikovnozvrstno spremenjen. Ta del močno odstopa od Cankarjevega jezika – besedilo uprizoritve je adaptacija dramskega besedila (replike so izdatno črtane in se pojavljajo le v fragmentih, ki skupaj z dopisanim besedilom tvorijo celoto). Govor »je z rabo različnih prozodičnih sredstev v različnih razmerjih in z različno intenzivnostjo smiselno zgrajen, učinkovit, hkrati pa s tem dosega visoko estetsko raven« (Žavbi 2016: 277). V obeh delih se govor ujema s kostumi, lučjo, glasbo, sceno, mizansceno; tako z ostalimi uprizoritvenimi dejavniki skupaj izvrstno funkcionira in v tem sledi režijskemu konceptu, s čimer uprizoritev kot celota odlično deluje.

V uprizoritvi h govoru pristopajo izrazito ustvarjalno. Cankarjev jezik že dolgo v gledališču ni več nedotakljiv, ta uprizoritev pa ga prva tako radikalno spreminja. »Ustvarjalna ekipa je po režiserjevi zamisli ustvarila celostno podobo odrskega govora, ki jo sestavljajo ustvarjalno oblikovani govori posameznih interpretov. Med

njimi po ustvarjalni moči izstopa kreacija odrskega govora ostarelega Jermana (Radko Polič)« (Žavbi 2016: 277, 278).

6 Zaključek

V prispevku poskušam prikazati raziskovalno strategijo (v prvem koraku usmerjeno predvsem na jezik uprizoritvenega besedila, v drugem pa na realizacijo tega jezika v govoru igralcev), učinkovito in primerno za raziskovanje odrskega govora. Pomembno povezuje slušnozaznavno in akustično analizo, kar se kot ključno izkaže predvsem pri nekaterih problematičnih primerih, npr. kadar raziskovalec ni prepričan o svoji slušni zaznavi in jo želi preveriti z objektivnimi metodami. Na prikazanem primeru govora Radka Poliča se je to kot pomembno izkazalo predvsem pri posebnih oblikah intonacijskih potekov. V nekaterih primerih pa nam akustična analiza tudi ne more koristiti (npr. prekrivni govor) ali pa nas lahko celo zavede (različni pogoji snemanja tudi znotraj ene predstave). Kombinacija obeh metod prispeva k večji preverljivosti in sledljivosti raziskovalnih rezultatov, za kar se zavzemajo tudi v ostalih raziskavah govora (npr. Tivadar 2009: 366). Pomembna je tudi interdisciplinarnost raziskovalnega pristopa: »Raziskovalec odrskega govora ne sme dojemati le kot jezikoslovno (fonetično) kategorijo, pač pa kot umetniško delo, ki ga je treba razumeti in interpretirati s pomočjo teatroloških, socioloških, sociolingvističnih, psiholoških in drugih znanj« (Žavbi 2016: 280). Odrski govor je nujno razumeti znotraj določene predstave in ga kot takega tudi umeščati.

Literatura

- Ivan CANKAR, 1969: *Hlapci*. Ljubljana: Državna založba Slovenije. (Zbrano delo 5).
Hlapci, videoposnetek uprizoritve v režiji Sebastijana Horvata. 2015. SSG Trst. Avtor posnetka: Luca Quaia. Pridobljeno iz Arhiva SSG Trst.
- Damjan HUBER, 2017: Raziskovalna izhodišča besedilnofonetične analize slovenskega medijskega govora (na primeru besednovernstne predvidljivosti poudarkov). *Slavistična revija* 65/2, 281–299. https://srl.si/sql_pdf/SRI_2017_2_05.pdf
- Marjeta HUMAR idr., 2007: *Gledališki terminološki slovar*. Ljubljana: Založba ZRC. <http://bos.zrc-sazu.si/c/term/gledaliski/>
- Klasja Zala KOVAČIČ, 2019: Socialne zvrsti odrskega govora treh najbolj gledanih predstav sezone 2015/2016 v osrednjeslovenskih gledališčih. *Slovenski javni govor in jezikovno-kulturna (samo)zavest*. Ur. Hotimir Tivadar. Ljubljana: Znanstvena založba Filozofske fakultete. (Obdobja 38). 297–304. https://centerslo.si/wp-content/uploads/2019/10/Obdobja-38_zbornik.pdf
- Mateja PEZDIRC BARTOL, 2010: *Najdeni pomeni: empirične raziskave recepcije literarnega dela*. Ljubljana: Zveza društev Slavistično društvo Slovenije. (Slavistična knjižnica 15).

- Mateja PEZDIRC BARTOL, 2011: Branje dramskega besedila: primer empirične raziskave. *Primerjalna književnost* 34/2, 125–135. <https://www.dlib.si/details/URN:NBN:SI:doc-WRLDUHBO>
- Katarina PODBEVŠEK, 2007: *Govorna interpretacija literarnih besedil v pedagoški in umetniški praksi*. Ljubljana: Slavistično društvo Slovenije. (Slavistična knjižnica 11).
- Katarina PODBEVŠEK, 2008: Govor kot gledališko izrazilo (Na primeru Flisarjevega Akvarija). *Slovenski jezik, literatura, kultura in mediji: zbornik predavanj*. 44. seminar slovenskega jezika, literature in kulture. Ur. Mateja Pezdirc Bartol. Ljubljana: Filozofska fakulteta. 51–59. https://centerslo.si/wp-content/uploads/2015/10/ssjlk_44_zbornik.pdf
- Tatjana STANIČ, 2006: V primežu norme. *Kolokvij o umetniškem govoru II*. Ur. Katarina Podbevšek, Tomaž Gubenšek. Ljubljana: Akademija za gledališče, radio, film in televizijo, Katedra za govor. 65–68.
- Hotimir TIVADAR, 2004: Priprava, izvedba in pomen perceptivnih testov za tehnično-fonološke raziskave (na primeru analize fonoloških parov). *Jezik in slovnost* 49/2, 17–36. <https://www.dlib.si/stream/URN:NBN:SI:doc-TMTVZP16/4553f87e-4c38-484e-b6f8-d5a8c0106090/PDF>
- Tomaž TOPORIŠIČ, 2004: *Med zapeljevanjem in sumničavostjo: razmerje med tekstom in uprizoritvijo v slovenskem gledališču druge polovice 20. stoletja*. Ljubljana: Maska.
- Gordana VAROŠANEC-ŠKARIČ, 2005: *Timbar*. Zagreb: FF press.
- Karolina VRBAN ZRINSKI, 2022: *Kazališče i govor – dvojbe i ideali: prozodija scenskoga govora*. Zagreb: Fakultet hrvatskih studija.
- Martin VRTAČNIK, 2012: Gledališki lektor – njegova funkcija in namen v sodobnosti. *Jezik in slovnost* 57/3–4, 101–114. <https://www.dlib.si/details/URN:NBN:SI:DOC-8XJJAQWTA>
- Nina ŽAVBI, 2013: Analiza odrskega govora – primer Bergerjeve uprizoritve Hlapcev (Komentirana izdaja). *Slavistična revija* 61/4, 651–664. https://srl.si/ojs/srl/article/view/COBISS_ID-3626331
- Nina ŽAVBI, 2016: *Govorna interpretacija dramskih besedil na primeru uprizoritev Hlapcev Ivana Cankarja. Doktorska disertacija*. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko. <https://repozitorij.uni-lj.si/Dokument.php?id=109535&lang=slv>
- Nina ŽAVBI, 2019: Odrski govor – slušnozaznavna in računalniška fonetična analiza uprizoritev Cankarjevih Hlapcev. *Slavistična revija* 67/1, 51–67. <https://srl.si/ojs/srl/article/view/2019-1-1-4>

STANJE IN PERSPEKTIVE UPORABE GOVORNIH VIROV V RAZISKAVAH GOVORA

MIRA KRAJNC IVIČ (UR.)

Univerza v Mariboru, Filozofska fakulteta, Maribor, Slovenija
mira.krajnc@um.si

Znanstvena monografija *Stanje in perspektive uporabe govornih virov v raziskavah govora* predstavlja rezultate prvega leta dela v raziskovalnem projektu *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik* (J7-4642) kot tudi rezultate raziskovalcev, ki se ukvarjajo z govorom v drugih raziskovalnih projektih, s skupnim glavnim ciljem strateškega in učinkovitega razvoja prostodostopnih govornih virov. Prispevki izpostavljajo načrte, cilje in izzive ob gradnji govornih virov, in sicer na leksikalni in skladenjski ravni, iščejo rešitve za posebnosti govora na fonetično-fonološki ravni tudi v povezavi s standardi zapisovanja (narečnega) govora. Govor in jezikovna variantnost sta raziskovana kot manifestacija socialne slojevitosti, govorjeni diskurz pa tudi z vidika součinkovanja različnih semiotskih kodov v smislu vzajemnega učinkovanja za tvorjenje smisla sporočila. Z vidika prepletanja prvin govorenega in pisnega jezika so analizirana izbrana spletna besedila, upoštevana je tudi vloga govora v gledališki umetnosti.

DOI
[https://doi.org/
10.18690/um.ff.4.2024](https://doi.org/10.18690/um.ff.4.2024)

ISBN
978-961-286-882-6

Ključne besede:
govorni viri,
jezikovne tehnologije,
govor,
govorjeni diskurz,
slovenščina



Univerzitetna založba
Univerze v Mariboru

DOI

[https://doi.org/
10.18690/um.ff.4.2024](https://doi.org/10.18690/um.ff.4.2024)

ISBN

978-961-286-882-6

Keywords:

speech resources,
language technologies,
speech,
spoken discourse,
Slovene

STATUS AND PERSPECTIVES OF THE USE OF SPEECH RESOURCES IN SPEECH RESEARCH

MIRA KRAJNC IVIČ (ED.)

University of Maribor, Faculty of Arts, Maribor, Slovenia
mira.krajnc@um.si

The scientific monograph “*Status and Perspectives of the Use of Speech Resources in Speech Research*”, which emerged from the research project *Basic Research for the Development of Speech Resources and Technologies for the Slovenian Language* (J7-4642), presents 22 contributions on research into speech and spoken discourse, particularly in Slovene studies. The contributions highlight plans, objectives, and challenges in building speech resources, especially at the lexical and syntactic levels and seek solutions for the peculiarities of speech at the phonetic-phonological level, also in the context of recording standards of (dialectal) speech. Speech and language variability are explored as manifestations of social complexity, while spoken discourse is also analysed from the perspective of the interaction and interplay of different semiotic codes with regard to their mutual effect in generating the meaning of the message. Selected online texts are analysed from the perspective of the intertwining elements of spoken and written language, including the role of speech in theatre arts.



Potrebnost tovrstne znanstvene monografije za naš prostor, kot je ta z naslovom *Stanje in perspektive uporabe govornih virov v raziskavah govora*, dokazujeta že dejstvi, da so težnje po intenzivnejši raziskavi govorjene slovenščine stare vsaj dobrega pol stoletja in da je ta monografska obravnava nastala kot eden izmed rezultatov širše projektne raziskave z naslovom *Temeljne raziskave za razvoj govornih virov in tehnologij za slovenski jezik (J7-4642)*. Govor je tisti, ki nas označuje in določa, in kot najosnovnejša uresničitev jezikovnega sistema mora biti tudi ustrezno in kontinuirano opisan in preučevan. Stalno zavedanje, da za sprotno beleženje potrebujemo precej človeškega potenciala, ki zbira in to opisuje in preučuje, pa predstavlja tudi ta dovolj celovita študija, ki je predstavitvena in problemska hkrati.

Prof. dr. Andreja Žele
FF UL, ISJ ZRC SAZU

Vrednost monografije je prav v konkretnih predlogih in celoviti obrnavi govora, ki še vedno ni našel dovolj prostora v znanstvenih raziskavah, čeprav je prav govor primarna človekova dejavnost, s katero vzpostavljamo in uresničujemo jezikovno sporazumevanje. Pomen govora je priznaval že Aristotel, ki je človeka opredelil kot bitje, ki ima um in govor. Ne nazadnje je pomen te monografije še, da na enem mestu prinaša sistematičen pregled pomembnih tem, povezanih z raziskovanjem govora, z raziskovanjem njegove družabne in družbene ter kulturne razsežnosti. Prinaša pa tudi pregled programskih orodij za njegovo analizo, kritični pregled učnih načrtov in katalogov znanja za slovenski jezik ter ugotovitve o učenju slovenščine kot drugega ali tujega jezika, pri katerem je zaradi komunikacijskega pristopa, tudi ko gre za govor, fonetično-fonološka raven najmanj zastopana.

Prof. dr. Tamara Gazdić-Alerić
Pedagoška fakulteta Univerze v Zagrebu