

# Raziskovanje govorjenega jezika

*Jana Zemljarič Miklavčič*

The present chapter focuses on spoken language, or more precisely, spontaneous speech. Spoken language is a topic explored by a subfield of translation studies: in this context, it is necessary to use an established research and terminological framework. In the past, spoken language was neglected by traditional linguistic research which centred above all on written language, a language variety considered to be of high status and culturally prestigious. A change in the perspective in recent decades meant that the linguistic approach changed as well. Interest in spoken language has been growing and technological advances have greatly improved the research methodology. It is the purpose of this chapter to provide an introduction and a more detailed definition of the subject for those researchers who wish to deal with spoken language in the context of translation studies. A further purpose is to present the possibilities and limitations in spoken language data collection and various approaches used in transcribing speech and non-speech events.

**Ključne besede:** govorjeni jezik, spontani govor, govorni korpusi, zapisovanje govora, (ne)standardni jezik

V pričujočem poglavju se podrobneje ukvarjamo z govornim jezikom oziroma s spontanim govorom. Prevodoslovje se namreč v določenem delu svojih raziskav ukvarja tudi z govornim jezikom, za kar potrebuje izdelan znanstveni oz. terminološki aparat. Govorjeni jezik je bil tradicionalno zanemarjeno področja jezikoslovnega raziskovanja, v nasprotju s pisnim jezikom, ki je bil kulturno in statusno prestižna vrsta. Spremenjena jezikoslovna filozofija zadnjih desetletij je ta pogled spremenila, zanimanje za govorni jezik med raziskovalci raste, s tehnološkim razvojem pa so se bistveno izboljšale tudi možnosti za njegovo raziskovanje. Namen poglavja je raziskovalcem, ki se bodo v okviru prevodoslovnih študij ukvarjali z govornim jezikom, približati in podrobneje definirati predmet opazovanja, predstaviti možnosti in omejitve pri zbiranju gradiva ter različne načine zapisovanja govora in dogodkov, ki spremljajo govor.

V uvodnem delu je na kratko predstavljeno raziskovanje govornega jezika na Slovenskem; pozornost je usmerjena predvsem na raziskave, ki glede na obravnavano gradivo, metodologijo in filozofijo pomenijo nov pristop v raziskovanju govorne slovenščine. Predstavljen je pomen govornih korpusov in nekateri najbolj znani referenčni govorni korpusi. Sledi obravnava temeljnih pojmov in razmislek o možni taksonomiji govornih besedil. Nadaljevanje je posvečeno problematiki zapisovanja govornega jezika in s tem povezane avtentičnosti govora. V zadnjem delu razprave so predstavljene nekatere možnosti za analizo govornega jezika, ki jih je omogočil *Učni korpus govorne slovenščine*, skupaj s poskusom interpretacije izluščenih podatkov. Celovitejši opis govorne slovenščine zaenkrat ostaja še stvar prihodnosti.

## 1. RAZISKOVANJE GOVORJENEGA JEZIKA NA SLOVENSKEM

Za raziskave govornega jezika v preteklosti je bilo značilno, da so opazovale predvsem odstopne od knjižne norme (kadar ni šlo za dialektološke študije). Na tak način je govorni jezik obravnavan tudi v osrednjih jezikovnih priročnikih: opis »splošno- in knjižnopogovornega jezika« ter »pokrajinskih pogovornih jezikov« v *Slovenski slovnici* obsega osem strani (Toporišič 2000<sup>4</sup>: 16–23), kar predstavlja približno 1 odstotek cele slovnice; razmerje nazorno kaže, kakšen delež je bil govornemu jeziku, ki predstavlja večji del človeške komunikacije, dodeljen znotraj temeljnega jezikovnega opisa. V *Slovarju slovenskega knjižnega jezika* so besede, ki so jih leksikografi prepoznali kot neknjižne, označene s kvalifikatorji, ki »kažejo na normo knjižnega jezika ali opozarjajo na razmerje do nje« (SSKJ I, XVIII): npr. stilno-plastni (*nar.*, *pog.*, *nizje pog.*), ekspresivni (*nizko*, *vulg.*) in posebni normativni (*neprav.* in *neustal.*).<sup>36</sup> Tudi SP 2001 vrednoti besedišče

<sup>36</sup> Kvalifikator *pog.* ima 3344 izrazov ali posameznih pomenov (npr. a, adijo, aja, apezejevec, arduš, baby itd.), od tega 543 kvalifikator *nizje pog.* (med njimi afna, bolovati, britof, cagavec, džumbus, fentati, fris itd.). Z *nizko* je označenih okrog 200

s številnimi kvalifikatorji, ki so razvrščeni v naslednje skupine: *socialnozvrstni, funkcijskozvrstni, čustvenostni, časovni, pogostnostni*, uvedene pa so tudi splošne normativne oznake *prepovedano, nepravilno in odsvetovano*.<sup>37</sup>

Med starejšimi razpravami poskus celovitejše obravnave govornega jezika predstavlja razprava *Vprašanja govornega jezika* (Pogorelec 1965). Avtorica se dotakne vseh jezikovnih ravnin, pa tudi okoliščin oz. situacijskega konteksta ter odnosa med govorniki, zaradi omejenih tehničnih možnosti pa razprava ostaja bolj ali manj na ravni hipotetičnega razmisleka. Med odmevnejše razprave v zadnjem desetletju sodijo *Govor celjskega predmestja Gaberje* (Škofic-Guzej 1998), *Jezikovno načrtovanje govornega jezika pri Slovencih* (Pogorelec 1998), *Govorjeni knjižni jezik v televizijskih dnevnoinformativnih oddajah* (Verovnik 2004), *Nekateri vidiki zvrstnosti govornega diskurza s stališča poslušalca* (Vogel 2004), *Podoba in funkcija govornega knjižnega jezika glede na neknjižne zvrsti* (Tivadar 2004) in druge. Novejše raziskave izkoriščajo sodobne možnosti raziskovanja; tako npr. Vitez in Zwitter Vitez (2004) z računalniškimi orodji raziskujeta prozodične lastnosti spontanega govora. Korak naprej predstavljata tudi doktorski disertaciji, ki temeljita na gradivu, posnetem posebej za namen raziskav, pri transkribiranju posnetkov pa ohranjata specifične (tudi fonetične in deloma prozodične) lastnosti govora; to sta *Vpliv besedilne vrste na uresničitev skladenjskih struktur (primer narativnih besedil v vsakdanjem spontanem govoru)* Mojce Smolej (2006) in *Analiza diskurza kot podpora sistemom strojnega simultanege prevajanja govora* Darinke Verdonik (2006).<sup>38</sup> Uveljavitev korpusnega pristopa pri raziskovanju govornega jezika je predstavljena v monografiji *Govorni korpusi* (Zemljarič Miklavčič 2008a), s fonetičnimi značilnostmi govornega jezika se ukvarja disertacija *Kakovost in trajanje samoglasnikov v govornem knjižnem jeziku* (Tivadar 2008), širši humanistični pogled na govorni jezik (poleg jezikoslovnega) pa predstavljajo *Spisi o govoru* (Vitez (ur.) 2008).

Kljub naraščajočemu številu in vplivu razprav s področja govornega jezika pa o celostnem opisu govornega jezika, posebej spontanega govora, v slovenščini zaenkrat še ne moremo govoriti; ta bo predvidoma lahko nastal šele na podlagi referenčnega govornega korpusa.

izrazov oz. pomenov (aufbiks, baba, crkavati, debeloritnik itd.), *neprav.* 27 (tu gre predvsem za skladenjske zveze - brez da, izgubiti na privlačnosti/vrednosti, delovno mesto, za katerega je potrebno ..., upravljati s skladom itd.).

<sup>37</sup> *Prepovedano*, npr. Rebulin (nam. Rebulov), *nepravilno*, npr. toliko, gospej (nam. toliko, gospe) in *odsvetovano*, npr. dvomiti v ... (nam. dvomiti o ...).

<sup>38</sup> Kasneje monografija - *Jezikovni elementi spontanosti v pogovoru* (Verdonik 2007).

## 2. VLOGA KORPUSOV PRI RAZISKOVANJU GOVORJENEGA JEZIKA

Kljub temu, da je korpusni pristop v jezikoslovju dandanes najširše priznan, ga vsi raziskovalci ne razumejo na enak način. T. i. *delni korpusni pristop* metodo korpusne analize uporablja samo kot podporo tradicionalnemu strukturalističnemu jezikoslovju, korpusnega jezikoslovja pa ne vidi kot samostojne jezikoslovne vede z lastnim teoretičnim izhodiščem (Gorjanc 2005: 185). Nasprotno *popolni korpusni pristop* uporabo korpusa narekuje neodvisno od uveljavljenih jezikoslovnih teorij in interpretacij; pri tovrstnem razumevanju korpusnega jezikoslovja »opazovanje jezikovnih podatkov vodi v oblikovanje hipotez, nadalje v posploševanje in nazadnje v oblikovanje teoretičnih stališč oz. trditev« (Tognini-Bonelli 2001: 85). Vloga jezikoslovca v tem procesu je neizpodbitna: v vseh fazah sodeluje s svojim znanjem, izkušnjami in zmožnostjo interpretiranja (ibid.). Glede na to, da gre pri opisovanju govornega jezika za celostno dokaj neraziskano področje, je bil popolni korpusni pristop v številnih jezikovnih skupnostih prepoznan kot najboljša pot raziskovanja – v filozofskem in metodološkem pogledu. V tem smislu je treba razumeti tudi Teubertovo ugotovitev, da »je analiza govornega jezika tista, ki je prva zasidrala korpusno jezikoslovje kot disciplino, samostojno tudi v svoji teoretični težnji« (Teubert 1999). Tako je bil že najstarejši govorni korpus London-Lund (*London-Lund Corpus*), ki nastal na podlagi korpusa *Survey of English Usage* (SEU), namenjen za preučevanje govornega in pisane britanske angleščine odraslih govorcev in je služil kot vir za slovnico angleškega jezika *A comprehensive grammar of the English language* (Quirk et al. 1985).

*Govorni korpusi* so elektronske zbirke transkribiranih posnetkov (predvsem) spontanega govora; s specifično metodologijo gradnje se oblikujejo kot podkorpusi referenčnih korpusov (govorni komponenti *BNC*, *The Bank of English*), lahko pa tudi kot samostojni korpusi (*Nizozemski govorni korpus*, *Švedski govorni korpus*). Referenčni govorni korpusi so namenjeni predvsem opazovanju skladensko-leksikalnih jezikoslovne lastnosti, prozodičnih pa le, če je korpus ustrezno označen. Korpusi za potrebe fonetično-fonoloških raziskav nastajajo drugače, do nedavnega predvsem kot studijski posnetki, zajemajo pa največkrat samo izbrane (in prebrane) stavke; imenujemo jih korpusi govora ali *govorne zbirke* (Gorjanc 2005, 8).

Za razvoj govornih korpusov so bila odločilna osemdeseta in prva polovica devetdesetih let; v tem času so bili zgrajeni veliki govorni korpusi in oblikovane delovne skupine, ki so načrtovale enotno označevanje korpusov. Najprej so nastali korpusi za vse različice angleškega jezika (britansko, ameriško, mednarodno), nato pa so začeli nastajati govorni korpusi tudi za druge jezike, poleg že omenjenih tudi govorna komponenta *Češkega nacionalnega korpusa*, *C-ORAL-*

*ROM* (korpusni paket francoščine, italijanščine, španščine in portugalščine) in drugi. Med vsemi naštetimi je bil Britanski nacionalni korpus v devetdesetih letih najplivnejši referenčni vir za gradnjo korpusov; s svojimi načeli gradnje in industrijskimi razsežnostmi produkcije je postavil temelje novemu obdobju gradnje govornih korpusov. Vendar pa je mlajšim načrtovalcem korpusov časovni zaostanek petih ali desetih let prinesel tudi veliko tehnološko prednost: napredek jezikovnih tehnologij v tem obdobju je omogočil sinhronizacijo zvočka in transkripcij. V tem smislu za naprednejšega danes npr. velja Nizozemski govorni korpus: po velikosti je primerljiv z govorno komponento BNC (8-10 milijonov besed); raznovrstne oznake, od ortografske do fonetične in prozodične transkripcije, oblikoskladenjske in skladenjske oznake, poleg tega pa še sinhroni dostop do zvočnih posnetkov pa nizozemski korpus trenutno uvrščajo med vodilne govorne korpuse na svetu.

V zadnjem desetletju je bila potreba po izdelavi govornega korpusa za slovenščino večkrat izražena. Na to so opozarjali sestavljalci prvega referenčnega korpusa FIDA (1998), pobude za gradnjo pa so prihajale tudi s strani drugih jezikoslovnih in sorodnih vej, npr. dialektologije, analize diskurza in govornih tehnologij. Za raziskovanje govora je bil v letih 2004–2009 dostopen *Učni korpus govornjene slovenščine*, ki je nastal v okviru doktorskega študija na Univerzi v Bergnu na Norveškem.<sup>39</sup> Kljub omejenemu obsegu je bilo s pomočjo korpusa mogoče opazovati številne značilnosti govornjenega jezika, posebej tiste z visoko frekvenco pojavljanja. Rezultatov analiz zaradi neuravnoveženosti in relativno majhnega obsega korpusa ni mogoče posploševati na govor v slovenščini kot celoto, lahko pa na njih gradimo hipoteze, ki jih bo mogoče preverjati na referenčnem korpusu. Gradivo in opisi govornjenega jezika, ki bodo uporabljeni v nadaljevanju razprave, so sestavni del *Učnega korpusa govornjene slovenščine*.

Leta 2008 so bila v okviru projekta *Sporazumevanje v slovenskem jeziku* (2008–2013)<sup>40</sup> zagotovljena sredstva za izgradnjo referenčnega govornega korpusa slovenščine v obsegu 1 milijon besed, s čimer so bili natanko desetletje po prvem pozivu izpolnjeni pogoji za začetek gradnje. Trenutno poteka snemanje, transkribiranje in označevanje posnetkov na celotnem območju, kjer se strnjeno govori slovenski jezik. Gradnja korpusa ima tri osnovne cilje:

- zajeti vzorčne primere različnih govornjenih besedil v različnih situacijah,
- zajeti govornjena besedila demografsko reprezentativnega vzorca govorcev,
- zajeti predvsem tiste govorne situacije, v katerih so uporabniki jezika najpogosteje produktivno-receptivno udeleženi.

<sup>39</sup> *Učni korpus* je podrobneje opisan v Zemljarič Miklavčič 2008a; zaradi težav s strežnikom na Univerzi v Bergnu korpus od junija 2009 ni več dostopen.

<sup>40</sup> <http://www.slovenscina.eu/Vsebine/Sl/Domov/Domov.aspx>

Referenčni govorni korpus za slovenščino bo javnosti predvidoma dostopen konec leta 2010, predstavlja pa samo enega izmed delnih ciljev projekta; skupaj z nadgrajenim referenčnim korpusom pisnega jezika naj bi služil kot izhodišče za prenovo temeljnih jezikovnih priročnikov, in sicer slogovnega priročnika, leksikalne baze in pedagoške korpusne slovnice.

Čeprav so govorni korpusi postali temeljni vir za raziskovanje govornega jezika, pa pri analizi govornih besedil nikakor ne smemo pozabiti na pomen celih besedil. Posamezne pojave lahko namreč raziskujemo predvsem ali samo na ravni besedila, npr. nekatera kohezivna sredstva, ponavljanja, popraviljanja, prozodične lastnosti, prekrivni govor in drugo. V tem smislu je ob gradnji govornega korpusa smiselno načrtovati tudi zbirko besedil, dostopnih v avtentični in transkribirani obliki, saj se na ta način izognemo podvojenemu zbiranju gradiva. Razmerje med besedilom in korpusom na zanimiv način vzpostavlja Tognini Bonelli (2001: 3). Ugotavlja, da korpusno analizo lahko razumemo kot raziskovanje jezika, kakor je realiziran v besedilih. To pomeni, da korpusno jezikoslovje izhaja iz istih predpostavk kot besediloslovje (besedilo je glavni nosilec pomena); kljub temu gre za dva pristopa k raziskovanju, ki se razlikujeta v več pogledih:

BESEDILO	KORPUS
beremo kot celoto	beremo kot fragment
beremo horizontalno	beremo vertikalno
beremo zaradi vsebine	beremo zaradi vzorcev
beremo kot enkratni dogodek	iščemo ponavljajoče se dogodke
beremo kot individualno dejanje	beremo kot primere družbene prakse
je primer <i>parole</i>	omogoča vpogled v <i>langage</i>
je koherenten komunikacijski dogodek	je nekoherenten komunikacijski dogodek

### Slika 1: Besedilo in korpus (Tognini Bonelli 2001: 3)

Za raziskovanje govornega jezika sta, enako kot velja za pisni jezik, pomembna oba pristopa.

Zbiranje gradiva pri raziskovanju govornega jezika poteka v dveh korakih: najprej snemanje, nato zapis posnetega gradiva. Ob tem se zastavlja vprašanje avtentičnosti: zavedanje govorcev, da se njihovo govorenje snema, zagotovo vpliva na njihovo produkcijo; najbolj avtentične posnetke govora bi dobili, če bi snemali brez vednosti govorcev. To žal ni mogoče, saj je tajno snemanje pri nas (in podobno tudi drugod) prepovedano in kaznivo.<sup>41</sup> Na podlagi posnetkov spontanega govora, narejenih za Učni korpus govornega slovenščine, in nekaterih tujih

<sup>41</sup> Kazenski zakonik RS (KZ-UPB1), <http://www.uradni-list.si/1/objava.jsp?urlid=200495&stevilka=4208>.

izkušenj lahko sklepamo, da govorniki dokaj hitro pozabijo na mikrofonski in govorniški popolnoma neobremenjeno, sem ter tja pa se na mikrofonski spet spomnijo, kar je razvidno iz njihovih reakcij. Zaradi zakonskih določil moramo biti pripravljeni na to, da pri analizi govornega jezika nikoli (vsaj zaenkrat) ne bomo razpolagali s stoddostno avtentičnim gradivom – v tem smislu je torej spontani govor za raziskovalce neulovljiv.

### 3. VRSTE GOVORJENIH BESEDIL

Z naraščanjem zanimanja za raziskovanje govornega jezika se utrjuje, deloma pa tudi na novo vzpostavlja izrazoslovje obravnavanega področja. *Govorjeni jezik* nam pomeni vrsto jezika, ki jo definira prenosnik, torej tisto, kar govorimo oz. slišimo, v nasprotju s pisnim jezikom. Vendar ta definicija ni dovolj natančna: opredeliti se je treba tudi do stopnje spontanosti govora. Kadar govorimo o govornem jeziku, pogosto mislimo na *spontani govor*, pri čemer gre lahko za različne stopnje spontanosti, od povsem nepripravljenega do pripravljenega govora. Branjen besedil v tem smislu ne razumemo kot govorni jezik, saj gre samo za oralizacijo pisnega jezika. Kot enoto govornega jezika pogosto uporabljamo izraz *govorjeno besedilo*, nekateri tudi *govorjeni diskurz* (npr. Verdonik 2006); govorno besedilo mora, enako kot pisno besedilo, zadostiti sedmim merilom tekstualnosti (de Beaugrande in Dressler 1981, 3).

V nasprotju z razmeroma enotnim knjižnim jezikom obstajajo številne podobe govornega jezika. Razlike nastanejo po eni strani zaradi demografskih značilnosti govorcev, npr. regijske pripadnosti, starosti in izobrazbe; ti dejavniki dokazano vplivajo na razlike v govoru. Vsak posameznik lahko znotraj svojih demografskih okvirov producira različne oblike govornega jezika – glede na okoliščine sporočanja, odnos do sogorcev, prenosnik, namen in drugo. Govorjena besedila se tako razlikujejo vsaj glede na:

- stopnjo spontanosti: spontani, pripravljeni, brani govor,
- število govorcev: monolog, dialog, multilog,
- okoliščine: javno, zasebno, uradno, neuradno
- govorni položaj: formalni, neformalni,
- prenosnik: osebni stik, telefon, mediji, internet
- namen besedila.

Posameznih besedil ni vedno lahko uvrstiti v posamezno kategorijo, saj so meje med njimi zabrisane. Kategorija *uradno/neuradno* se ne prekriva s pojmom *javno/zasebno*, saj imamo uradna in neuradna besedila tako znotraj javnih kot znotraj zasebnih besedil. V slovenskem jezikoslovju na podoben razmislek nalletimo pri novejši klasifikaciji zvrsti govornih besedil (Vogel 2004: 461), kjer

avtorica razlikuje med “javnimi govornimi nastopi, javnimi pogovori (pogovori pred javnostjo), uradnimi zasebnimi pogovori in neuradnimi zasebnimi pogovori.” Da je definicija javnega/zasebnega v govoru nujno potrebna, ne pa tudi zadostna, potrjuje tudi Kranjc (1996/97: 308): »Poleg javnosti/zasebnosti je treba upoštevati tudi družbeno distanco med sogovorcema, njuno starost in to, ali sta prijatelja ali le bežna znanca, skratka, pozorni moramo biti na sociolingvistične danosti /.../.« Avtorica v bistvu govori o formalnem in neformalnem govornem položaju znotraj javnega in zasebnega govora, z drugimi besedami: »Govorne položaje ločujemo na formalne in neformalne. V obeh vrstah govornih položajev pa se lahko odvija javni ali zasebni diskurz« (ibid.).

Razmerje med govorcema pogosto označujemo z izrazoma *formalno/neformalno*. Formalnost/neformalnost govornega položaja po Kranjc (1996/1997: 309) določa “socialna razdalja med govorcema ter njun status in vloga v družbi.” Sigley v študiji o formalnosti besedil (1997: 206–208) pojasnjuje, da pojem formalnosti lahko razumemo kot situacijski kontekst (tu lahko opazujemo vzrok), lahko pa gre za jezikovno dejstvo (tu opazujemo jezikovno izbiro, ki je posledica situacijske formalnosti, pri nas npr. *vikanje*). Obenem velja, da jezikovni podatki ne merijo oz. ne izkazujejo vedno ustrezno formalnosti situacije, in obratno, na podlagi formalnosti situacije ne moremo zanesljivo predvidevati, kakšne učinke bo imela na jezik govorcev. Sigley nadalje navaja situacijske dejavnike, ki vplivajo na stopnjo formalnosti:

Formalno		Neformalno
visoka, pomembna	<i>družbena vrednost dogodka</i>	nizka, nepomembna
vnaprej določena	<i>struktura govornega dogodka</i>	prosta
javno, institucionalno	<i>okolje</i>	domače, zasebno
tujci socialne vloge	<i>udeleženci, odnos, vloge</i>	intimen odnos, individualne osebe
informativni	<i>namen dogodka</i>	interakcija
abstraktna, specifična, vnaprej določena	<i>tema</i>	zasebna, splošna, nedoločena

### Slika 2: Situacijski dejavniki formalnosti (Sigley 1997: 209)

Gre za zapleten preplet mehanizmov, ki jih nezavedno pregledujejo človeški možgani, ko se človek odloča za izbiro jezikovnih sredstev. Kar zadeva slovensko jezikovno situacijo, lahko k Sigleyjevi shemi dodamo vsaj še en demografski dejavnik – starost udeležencev; ta v še tako neformalnih okoliščinah lahko vpliva na izbiro formalnega jezikovnega sredstva – vikanja. Vprašanje je, v kolikšni



meri se med seboj prekrivajo pojmi *uradno/neuradno* in *formalno/neformalno*; v slovenski jezikoslovni teoriji, kot smo videli, ni enotnega pojmovanja teh izrazov. Načeloma naj bi uradne okoliščine pogojevale večjo mero formalnosti govora, neuradne okoliščine pa ravno nasprotno. Obenem pa vemo, da v realnem govoru ni vedno tako: analize empiričnega gradiva bodo lahko pokazale, kdaj in kakšna so odstopanja, morda pa tudi, na kakšen način odstopanja vplivajo na sporazumevanje.

Posebno problematiko predstavlja *segmentiranje* govornih besedil. Večina pisnih (knjižnih) besedil je strukturiranih po standardiziranih načelih in jih lahko razdelimo na manjše enote, kot so poglavja, odstavki, povedi in stavki. Naslovi, ločila in različna tipografska orodja še nadalje strukturirajo pisna besedila in jim dajejo prepoznavno obliko. Popolnoma drugače pa je pri govornih besedilih: tu gre za bolj ali manj strnjen dogodek v času, ki ga je težko razdeliti na manjše enote. Segmentacija govornih besedil poteka na podlagi prozodičnih lastnosti govora, bodisi na podlagi tonskih enot ali pa zgolj premorov v govoru in menjav govorcev. V zvezi z enoto govora najpogosteje govorimo o *izjavi*: »Praviloma gre za enoto govora, izgovorjeno med premoroma, običajno tudi med vdihom govornika« (Verdonik 2006: 50). Tako definirana izjava je predvsem prozodična enota govora, kljub temu pa je pogosto tudi »skladenjsko-semantično zaokrožena enota« (ibid.), čeprav to ni pogoj. Vitez in Zwitter Vitez (2004: 8) ob analizi spontanega govora ugotavljata, da je »opredelitev analiziranih govornih enot v osnovi usklajena z negramatikalnimi.« *Izjava* kot najmanjša enota spontanega govora je torej prozodična in ne nujno skladenjsko-semantična enota.

#### 4. ZAPISOVANJE GOVORJENIH BESEDIL

Tudi naslednji korak pri zbiranju gradiva za analizo govornih besedil, zapis govora, predstavlja svojevrstno problematiko na različnih ravneh. Temeljno vodilo zapisovalcev govora bi moralo biti ohranitev čim več značilnosti govornega jezika. Obstajajo različne vrste transkripcij in raziskovalci se med njimi odločajo glede na vrsto in namen raziskave: kogar zanimajo predvsem ponavljajoči se (skladenjski) vzorci v jeziku, pogostnost pojavljanja ter pomeni besed, ne bo potreboval fonetične transkripcije, ampak bo zanj bolj pomembna (velika) količina transkribiranega gradiva. Fonetik nasprotno potrebuje manjšo količino podatkov, vendar morajo biti ti mnogo bolj podrobno transkribirani v smislu prozodije in akustične realizacije. Za sociolingvistiko in analizo diskurza so spet bistvene informacije o okoliščinah in sobesedilu. Večje količine gradiva je najlažje transkribirati po načelih t. i. ortografske transkripcije, ki za zapisovanje govornih besedil uporablja standardni (knjižni) zapis besed, kadar ta obstaja. Fonemska transkripcija temelji na zapisu izgovorjenih fonemov, pri čemer upošteva izgovor besed v

kontekstu ter tudi morebitno izpuščanje ali dodajanje fonemov; še podrobnejša, fonetična transkripcija, si prizadeva natančno zapisati, kaj in kako je govorec izgovoril, tudi z variantami fonemov;<sup>42</sup> ločuje vse segmente govora, ki so prepoznavni kot samostojne enote na sliki akustičnega valovanja ali na spektrogramu.

Spodaj so prikazani različni načini transkribiranja – najprej ortografska (poknjžena) transkripcija, nato fonemska in nazadnje fonetična transkripcija po sistemu SAMPA:

**G19:** sem pa danes spila že kakšne tri kofete

**G20:** jaz sem ga dopoldne enega sem kuhal potem sem pa nič potem pa ko je enkrat ko imaš tukaj prištmano ko je enkrat potem {neraz} ne kaj pa križanka pa kavica pa cigaret pa vse sorte

**G19:** sɛm pa dɔns spila že êne tri kofête

**G20:** jəs sɛm ga dopɔudne ênga sɛm kuhou po sɛm pa nɛç po pa kə je ênkat kə maš tlê prštman kə je ênkat pol {neraz} ne ka pa križanka pa kavica pa cigarét pa vse sôrte

**G19:** s@m pa "dO:ns "spi:la Ze "E:ne "tri: ko"fe:te

**G20:** j@s s@m ga do"po:Udne "E:nga s@m "ku:hoU po s@m pa "n@tS po pa k@ je "E:Nkat kə "ma:S "t\_IE: p@r"Sti:man kə je "E:Nkat "pO:l' {neraz} ne ka pa "kri:ZaNka pa "ka:vica pa tsigà"re:t pa WsE "so:rte

Prvi način je za zapisovanje najlažji, saj je uporabljen standardni črkopis; slaba stran takega zapisa je, da so izgubljene številne imanentne lastnosti govornega jezika; druga dva zapisa govoru bolj dosledno ustrezata, vendar sta zaradi posebnih znakov zahtevna za zapisovanje. Nobeden izmed navedenih zapisov se torej ne zdi vsestransko primeren za zapisovanje spontanega govora, zato je bilo potrebno poiskati ustrežnejšo rešitev.

V nekaterih novejših razpravah o govornem jeziku se zato srečujemo s problematiko zapisa govora. Verdonik (2006: 69) v analizi telefonskih pogovorov, opravljenih med turistično agencijo in strankami,<sup>43</sup> pri zapisu govornih besedil sledi priporočilom EAGLES.<sup>44</sup> Besedilo je zapisano »ortografsko, skladno s knjižnim standardom – to pomeni, da so tudi pogovorno, narečno, površno ipd. izgovorjene besede zapisane tako, kot je predvideno v knjižnem standardu, ne tako, kot so dejansko izgovorjene.« Izjema pri tem so naslednje besede:

<sup>42</sup> Najpogosteje v mednarodni računalniški fonetični abecedi SAMPA, <http://www.phon.ucl.ac.uk/home/sampa/index.html>.

<sup>43</sup> Turistična agencija se nahaja v Mariboru.

<sup>44</sup> Expert Advisory Group on Language Engineering Standards, <http://www.ilc.cnr.it/EAGLES/home.html>.

- pogovorni nedoločnik je prepisan brez končnega –i, če je tako izgovorjen,
- *mogli* v pomenu morali,
- *najdli* (od najti),
- *taki* (v pomenu tak, takšen),
- *pol* (v pomenu potem),
- *more* (namesto oblike mora),
- *večih, večim* (sklanjanje nesklonljive besede *več*).

Vidimo lahko, kako raziskovalka v določenih primerih (zdi se, da gre v primerih z veliko frekvenco pojavljanja) odstopa od siceršnjega standardiziranega zapisa v prid ohranjanja avtentičnosti govora, saj predvideva, da s tem ohranja pomembne podatke za raziskovanje. Podobno Smolej za zapis govora uporablja »/.../ knjižni oz. ortografski zapis. Kljub odločitvi za knjižni zapis pa so pri transkribiranju upošteevane nekatere izjeme, ki niso skladne s standardnim slovarjem (SSKJ) – odločili smo se za tiste izjeme, ki so skupne vsem govorcem posnetih besedil<sup>45</sup> in kljub drugačnosti od knjižnega zapisa ne otežujejo berljivosti« (Smolej 2006: 13).

Izjeme v tem primeru so:

- izguba nenagl. končnice –o pri samostalnikih srednjega spola (*mlek*),
- izguba glasu i pri deležnikih množinske oblike moškega sploa (*igral, spal*) in izguba glasu o pri deležnikih srednjega spola (*dogajal*),
- izguba i-ja in u-ja v predponah in v osnovi (*prpravt, prjatu, drgač*),
- izguba končnice –i v dajalniku in mestniku (*men, pr men*),
- izguba nenaglašene e in a v besedah (*clo, zlo*),
- ohranjeno ukanje v prednaglasnem zlogu (*utrok, purudnišnica*),
- deležnik na –l za moško obliko se razen izjemoma zapisuje z –u (*naredu*) (Smolej 2006: 13–14).<sup>46</sup>

V tem primeru gre za nekakšen hibrid med ortografski in fonetičnim zapisom, ki pa je ustrezal namenu raziskave (skladenjske strukture v govoru). Smolej navaja tudi seznam t. i. *besedilnih aktualizatorjev*, ki prav tako niso del knjižnega jezika, v govoru pa so nepogrešljivo sredstvo vzpostavljanja kohezije in koherence ter medbesedilnosti. To so:

- *en* (sredstvo nedoločnega uvrščanja v besedilni svet),
- *ta* (v funkciji določnega člana in v deiktični vlogi),
- *tist* (v deiktični vlogi in drugo),
- *un* (splošno znana referenca in drugo).

Ko zapisujemo govorni jezik, prestopamo v kod, ki so ga do nedavnega strogo obvladovala pravila, kar v zapisovalcih še vedno prikrito deluje na intuitivni ravni. Zapisovanje (za knjižni jezik) nestandardnih besed in oblik od nas pravzaprav

<sup>45</sup> Posnetki so nastali v Ljubljani.

<sup>46</sup> V nadaljevanju so navedene besede, ki se kljub drugačnemu izgovoru zapisujejo knjižno, npr. *jaz, brat, zdaj*, predlog *v* itd.

terja, da najprej definiramo, kaj je v jeziku standardno. S pojmom *jezikovnega standarda* se je pri nas mdr. ukvarjal Skubic, ki pojem definira na dva načina:

“Izraz ima potencialno dva pomena: (1) jezik določene *kakovosti*, ki jo je treba doseči (jezik kot merilo), ter (2) najbolj *običajna*, splošno uveljavljena različica jezika jezika (standard kot običajnost). V prvem pomenu, ki je v literaturi najpogostejši, je ta izraz najbliže v slovenščini uveljavljenemu izrazu *knjižni jezik* /.../ Večkrat pa se v slovenskem jezikoslovju ta izraz uporablja v tem drugem smislu (običajni jezik, lahko npr. določene družbene skupine ali občila (Pogorelec 1967, 1982), in takrat z izrazom *knjižni jezik* seveda ni prekriven. Zdi se, da v zadnjem času pojem pravopisne kodifikacije pri mlajših, korpusno usmerjenih slovenskih jezikoslovcih implicira nekakšno sintezo obeh pomenov – ko naj bi jezikoslovec na podlagi dejansko najbolj razširjene rabe posplošil pravilo in ga povzdignil v kakovostno normo, torej iz standardnega jezika v drugem v standardni jezik v prvem pomenu” (Skubic 2004: 211).

Razumevanje standarda v prvem pomenu za govorni jezik pride v poštev samo v določenih govornih položajih. Za pisni jezik je še do nedavnega veljalo, da imamo en sam standard; z razvojem novih tehnologij in vzpostavitvijo elektronskih komunikacijskih kanalov (SMS, elektronska pošta, forumi, klepetalnice, blogi itd.) se je to razumevanje močno relativiziralo. Sedaj se tako v govornem kot v pisnem jeziku standard v pomenu zahtevane kakovosti lahko pričakuje samo v določenih okoliščinah in govornih položajih, v vseh drugih okoliščinah in govornih položajih pa je raba določene oblike, leksema ali skladenjske strukture lahko povsem *običajna* (standardna v drugem pomenu besede), odstopa pa od meril, ki veljajo za standard v prvem pomenu besede, to je kodificirani knjižni jezik. Odstopi od idealne (predpisane) normativne izreke in zapisa v teh položajih predstavljajo poglobitve prvine jezikovne realnosti (Stabej in Vitez 2000: 80), in jih je kot take treba tudi razumeti, brez nenehnega sopostavljanja nasproti normativnemu predpisu.

Kolikor lahko sklepamo iz dosedanjih razprav, se odločitev o tem, kaj je “nestandardno”, raziskovalcem vsiljuje na intuitivni ravni, kar seveda ni v skladu s filozofijo korpusnega jezikoslovja. Verdonik v svoji razpravi označuje nekatere besede in oblike kot “pogovorna oblikoslovno-skladenjska raba oz. pogovorni izraz”, npr. *boste mi dal, tam en hotelček, pa morate vedet, lahko pokličete čez njih pet minut, če slučajno vejo na ministrstvu, pa se mogoče pol čujeva, te pa dorečema pozneje, se pomenima, toto animacijo, šalter, ziher, fajm, glih, garantiram, duplih rezervacij, brezveze, nonstop, gor na spletni strani* itd. (Verdonik 2006: 180–223). Kot lahko vidimo, so označene besede in oblike nestandardni izrazi, seveda v odnosu do knjižnega standarda; v pogovornem jeziku so nekatere navedene besede in oblike

povsem nevtralne, nekatere pa občutimo kot pokrajinsko (narečno) obarvane. Verdonik ne pojasnjuje kriterija, po katerem je nastal ta seznam; predvidevam, da na podlagi avtoričine jezikovne intuicije. Podobno so bile pri transkribiranju Učnega korpusa govornega jezika v prvi fazi nekatere besede označene kot nestandardne, npr. *angažma, benz, cajt, carsko, dek, dila, dofilati, faks, fila, fora, frej, ful, furajo, gužva, v iber hudih, jabki, jabčki, jebala ...* (Zemljarič Miklavčič 2008a: 157), kar kaže, kako težko je pravzaprav sprejeti nov zorni kot opazovanja jezika – ne več v odnosu do knjižnega jezika, ampak neodvisno od njega.

Zdi se, da so bili ti nekoliko obotavljajoči se koraki zapisa govora potrebni, da se je postopno izoblikovala zavest o neodvisnem zapisovanju pogovornega jezika. Pri transkribiranju za referenčni govorni korpus, ki nastaja v okviru projekta *Sporazumevanje v slovenskem jeziku* (prim. op. 5), je tovrstni način označevanja nestandardnih besed v celoti presežen oz. ukinjen. Na prvem nivoju zapisa se poskuša čim bolj ohraniti glasovno podobo govora, in sicer z uporabo veljavnega slovenskega črkopisa. To pomeni, da se npr. v zapisu ohranjajo redukcije (glasovi, ki niso izgovorjeni, se ne zapisujejo) – *tud, tko, mam, mislm, prijatci, navm (ne bom)* oz. se zapisujejo glasovi, kot so izgovorjeni – *jest, pršu, vidu*, pokrajinsko specifični fonemi pa se zapisujejo z najbližjimi ustreznimi črkami – *šiest, šuola, hriem* (primorsko *grem*); podrobneje je transkribiranje za referenčni govorni korpus predstavljeno v Zwitter Vitez et al. (2009).

Poleg odločitve o vrsti transkripcije se je glede na namen raziskave potrebno opredeliti tudi do zapisa besedilnih delov, ki so v pisnem jeziku nestandardizirani, vendar izrazito značilni za govorni jezik, npr. polverbalni (*mhm, polglasnik*) in neverbalni zvoki (*smeh, kašelj*), nedokončane besede, popravljanja, prekrivni govor, premori in drugo; pri tem je smiselno upoštevati mednarodne standarde in priporočila,<sup>47</sup> znotraj teh pa iskati najprimernejše rešitve in prilagoditve. Vedno pa je treba imeti pred očmi predvsem ohranjanje avtentičnosti govora.

## 5. NEKATERE IMANENTNE ZNAČILNOSTI GOVORJENEGA JEZIKA

Med govorenjem zaznavamo tudi glasove in glasovne sklope, ki lahko imajo komunikacijsko funkcijo, pa niso besede ali vsaj ne polnopomenske besede; to so medmeti kot *mhm, uaa, oh*, pa tudi drugi glasovi, ki jih proizvaja človek med govorenjem – smeh, tleskanje, vzdihovanje, kašljanje ipd. Razlikujemo polverbalne glasovne dogodke (*mhm*)<sup>48</sup> od neverbalnih dogodkov (vzdih). Kot smo vi-

<sup>47</sup> Npr. TEI (Text Encoding Initiative) in EAGLES (Expert Advisory Group on Language Engineering Standards).

<sup>48</sup> V Učnem korpusu govorne slovenščine so se pojavili *ajd, ah, aha, aja, ane, ej, evo, hjah, ə, əm, jah, joj, m, ma, mah, mhm, no, oh in oja*.

deli, večino glasovnih sklopov, ki se kot besedna vrsta uvrščajo med medmete (ali členke), v okviru sodobnih tokov analize diskurza opazujemo v vlogi diskurznihih označevalcev (prim. npr. Verdonik 2006); gre za izrazito in pogosto lastnost govorjenega jezika, ki je v slovenščini zaenkrat še slabo raziskana. Spodaj je prikazan primer rabe diskurznehih označevalca – besede *ne* v Učnem korpusu:

mesece ma se je <neraz>] per <lz>	ne	+G01+	ampak oni bojo ziher vedeli
ga je <lz> ko je vozil se s kolesom	ne	+G01+	pa nima blatnika +G01+ ima
je vpliv veš to je ob Adiži v bistvu	ne	+G01+	skozi od morja +G10+
glej Oslo Bergen je v bistvu daleč	ne	+G03+	[tako da <pavza> ja zdaj
je v +G03+ je v <nst> penziji </nst>	ne	+G03+	[<shift=vpr> a to je domači
no glej <nst> klele </nst> je Bergen	ne	+G03+	[čisto tukaj a <neraz> zato
obilo dežja <pavza> [mhm] ne v Bergnu	ne	+G03+	a je <nst> scalo </nst> <n
če bi bila tukaj kakšna karta Evrope	ne	+G03+	da je od Ljubljane do
izhodišče za vse te taglavne fjorde	ne	+G09+	[kako dobro] +G09+ kako
+G01+ Adiža <pavza> a to je tole	ne	+G10+	<pavza> ne jaz mislim da je
aja to pa ja <nst> valjda </nst>	ne	+G11+	na <nst> konteste </nst> se
koliko jih <nst> sfuraš </nst> koliko	ne	+G11+	večinoma se <?> park </?>
[ali pa] karkoli je kaj takega	ne	+G13+	[ja] +G11+ ker je to v bis
pod resno misliš <neraz> literaturo	ne	+G16+	[<neraz>] se štejejo pač ti
<shift=vpr> kako lahko </shift=vpr>	ne	+G17+	[ja ja zastopim] ki se itak
sem gledal zdaj ko ta ðm Vega	ne	+G17+	toži Mobitel
[zakaj ne] zato ker računalnik	ne		[dela ali je kaj drugega
se <?> park </?> <nst> fura </nst>	ne	[na kontestih] +G12+	[ja v
Američan] a veš <nst> pizda </nst>	ne		[oni ki imajo] <pavza> [ne

### Slika 3: Delni prikaz kolokacij besede *ne* iz neformalnih besedil UKGS<sup>49</sup>

Prikazanih je prvih 19 vrstic (10 %) od skupno 199, kolikor je pojavitev te besede v neformalnih besedilih korpusa. Vidimo lahko, da je *ne* 18-krat v funkciji diskurznehih označevalca (po katerem skoraj brez izjeme sledi predaja besede), in samo enkrat v vlogi klasične nikalnice pred glagolom;<sup>50</sup> to pomeni precejšnje prevrednotenje pomske napolnjenosti obravnavane besede v neformalnih govorjenih besedilih.

Naslednja lastnost govorjenega jezika, ki je knjižni (niti pisni) jezik ne pozna, so *ponavljanja*.<sup>51</sup> Ta imajo v spontanem govoru lahko več funkcij. Poleg tega, da so sredstvo vzpostavljanja besedilne koherence (kadar npr. govorec ponavlja dele besedila v znak razumevanja ali strinjanja), imajo hkrati tudi vlogo kohezivnega sredstva. Eden izmed vzrokov, ki v spontanem govoru pogosto privede do nenačrtovanega ponavljanja, je pridobivanje časa: govorniki so pogosto prisil-

<sup>49</sup> Označke G01-G17 so identifikacijske oznake govorcev in stojijo pred njihovimi izjavami; v zgornjem izpisu kažejo, da je po besedici *ne* v vlogi diskurznehih označevalca pogosto sledila predaja vloge.

<sup>50</sup> Če pogledamo vse pojavitve v neformalnih besedilih, je približno ena četrtna pojavitev v funkciji nikalnice.

<sup>51</sup> Oz. tam delujejo redundantno, če niso v ekspresivni funkciji.

jeni proizvajati svoje prispevke z minimalnim načrtovalnim časom; tega si lahko podaljšajo s ponavljanjem. Nadaljnje funkcije ponavljanj so še poudarek pomena leksema, poudarjanje trajanja dogodka, odprava nejasnosti, potrditev pravilnosti izjave, čustvena angažiranost govorca in druge; navedeni so nekateri primeri ponavljanj iz Učnega korpusa:

- (1) kar fajn dala sva za vsako sva dala petnajst jurjev
- (2) razlika je za najširšo javnost težko <pavza> težko natančno dojemljiva
- (3) mora biti to kar udeleženec opazi mora biti pa veliko bolj lahkotno
- (4) [ja zdaj mogoče bi res lahko] mogoče ločili
- (5) kar se tiče kar se tiče tega ne kar se tiče ne pa da da na rezervo dudlajo vzorce  
i= in in skladen= in in in ne vem besede
- (6) ful je veter pihal ful ful ful
- (7) in se zelo zelo intenzivno učijo
- (8) to si mislila ja bravo bravo

Poleg ponavljanj v skladnji spontanega govora lahko opazujemo tudi *napačne začetke*. Ti so eden izmed najočitnejših razlikovalnih elementov glede na pisni jezik. Termin označuje dogodek, ki se zgodi na ravni besede; govorec besedo začne, pa je ne dokonča – je prekinjen, si premisli, se zmoti itd. V slovenščini je pojav zaenkrat slabo raziskan, manjša raziskava na učnem korpusu pa je pokazala, da do prekinitev na ravni besede pogosteje prihaja v formalnih govornih položajih in da je delež napačnih začetkov v neformalnem govoru zanemarljiv (Zemljarič Miklavčič 2008b); eden izmed vzrokov za to je morda večja sproščenost in tekočnost neformalnega govora. V formalnem govoru govorniki tvorijo daljše in zahtevnejše sintaktične strukture, zato potrebujejo več časa za načrtovanje govora; zaradi zahtevnosti formulacij so tudi napačni začetki pogostejši. V neformalnih besedilih je do napačnih začetkov prišlo v primerih, ko so nastopile zunanje motnje pogovora in v primerih, ko so se govorniki pogovarjali o vsebinsko zahtevnejših temah (konkretno o umetnosti, politiki itd.). Za potrditev te hipoteze bi bilo potrebno analizirati večjo količino gradiva.

*Popravljanja* so razširjena kategorija napačnih začetkov. Termin se nanaša na besede in besedne zveze, ki so do konca izgovorjene, vendar jih govorec zaradi določenega razloga popravi oz. ponovno izreče – v drugačni obliki, drugačnem zaporedju, z drugo besedo itd. S popravljanji se v okviru analize diskurza podrobneje ukvarjata Smolej (2006) in Verdonik (2006); slednja raziskuje težave, ki jih ta značilnost spontanega govora povzroča pri razvoju sistemov strojnega prevajanja govora; vsekakor govorec s popravljanji delno razkriva procese psihološke in sociološke narave, ki potekajo ob tvorjenju govora. Prikazani so nekateri primeri iz Učnega korpusa:

- (1) v začetku že skušaš poskusiš govoriti spregovoriti
- (2) kot danes ko je bila dosežena dosežen konsenz

- (3) so tukaj navade kakšni *əm* ne vem kakšen je delovni čas  
 (4) hočeš vse imeti iz ne vem od iz Obsessiona  
 (5) mislim *və* primerljiv seveda z ne vem s francoskim inštitutom.

Izrazita značilnost govornega besedila je tudi *prekrivni* ali *hkratni govor* več govorcev. Gre za pojav, ki ga zaznamo praktično v vseh dialoških ali multiloških besedilih, ne glede na okoliščine ali stopnjo formalnosti, čeprav se količina prekrivnega govora v bolj formalnih okoliščinah predvidoma znižuje (kar bi bilo potrebno še dokazati), ker hkratno govorenje več govorcev v formalnih okoliščinah velja za manj sprejemljivo. Tudi razloge za nastanek prekrivnega govora bo treba še natančneje raziskati, znana pa je npr. ena izmed funkcij prekrivnega govora, to je nakazovanje, da govorec sledi diskurzu, ne da bi želel prevzeti vlogo; prikazan je primer iz Učnega korpusa:<sup>52</sup>

ne] +G01+ [mhm] +G09+	[mhm] +G10+ [mhm] +G15+ [mhm]
[mhm] +G09+ [mhm] +G10+	[mhm] +G15+ [mhm] +G03+ obilica
časa za [druženje] +G04+	[mhm] [ampak tista prioriteta]
[s Slovenci] +G07+	[mhm] [mhm] <i>əm</i> kaj vam je to
usposobljena za slovenščino +??+	[mhm] [mhm] <i>əm</i> <ime>
[da] +G02+ [mhm]	[mhm] [ə doktorica <ime>
na seminarju [ane] +G03+	[mhm] ampak res predvsem v
kaj bom zdaj jaz +G04+	[mhm] bo mogoče <smeh> še ona
čez [štirideset] let +G04+	[mhm] bog ve kaj bo ne jaz
funkcijo] čeprav je res +G04+	[mhm] da je potreba po [nečem
dejansko je pa res +G04+	[mhm] da je- <i>ə</i> bo treba po
mi včasih smo to +G04+	[mhm] doživljali tako ne dosti
[dežela] za nas +G04+	[mhm] drugače jaz sem študiral
<ime> +G15+ je pa obilo dežja <pavza>	[mhm] ne v Bergnu ne +G03+ a je
in slovenščini [ne] +G04+	[mhm] no kot je bilo že prej
<repet/> naj bi bil razlog	[mhm] potem <i>ə</i> če lahko govorim
[drugi] del publike +G04+	[mhm] recimo ki študira
govo- n- [govorcev] +G04+	[mhm] saj v končni fazi ne gre
posebej [dopovedovati] +G04+	[mhm] torej seminar je zasnovan
[ljudi] v +G04+	[mhm] tujih državah da bodo
od ponedeljka do petka ne ja	]mhm] <neraz> pa naši semenir-
življenje] [<neraz>] +G04+	[mhm] <pavza> [Slovenija] in to
kot za eno državo +??+ [mhm]	mhm] <pavza> mislim in to
precej na delovnem mestu +G04+	[mhm] <i>ə</i> doktor <ime> <priimek>
razsežnosti <i>ə</i> [Slovenije] +G04+	[mhm] <i>ə</i> je pa razlika
kulture +G07+ [seminarja]	[mhm] <i>ə</i> ki prihaja iz
oziroma] morajo biti +G04+	[mhm] <i>ə</i> na razpolago tako da se
se mi zdi] ne +G04+	[mhm] <i>ə</i> no morda v tem
</nv>] [ne] +G04+	[mhm] <i>ə</i> no saj <i>ə</i> zanimivo bi

<sup>52</sup> Besedilo v oglatem oklepaju je izgovorjeno hkrati z drugim besedilom, kar je mednarodno uveljavljen zapis hkratnega govora; G01-G10 so oznake govorcev in v transkripciji označujejo, da je vstopil nov govorec; več o označevanju govornih besedil za potrebe referenčnega korpusa v Zemljarič Miklavčič 2008b.



pa isto misel ne +G04+ [mhm	<b>mhm]</b>	ə to da je na Slovenstvu
[da ga ima] +G04+	<b>[mhm]</b>	ə to je mogoče se slišijo
od ponedeljka do petka ne ja	<b>mhm]</b>	<neraz> pa naši semenir-
življenje [<neraz>] +G04+	<b>[mhm]</b>	<pavza> [Slovenija] in to
kot za eno državo +??+ [mhm	<b>mhm]</b>	<pavza> mislim in to
precej na delovnem mestu +G04+	<b>[mhm]</b>	ə doktor <ime> <priimek>
razsežnosti ə [Slovenije] +G04+	<b>[mhm]</b>	ə je pa razlika
kulture +G07+ [seminarja]	<b>[mhm]</b>	ə ki prihaja i

#### Slika 4: Primeri prekrivnega govora iz UKGS

Prekrivanja v teh primerih s stališča organizacije diskurza niso moteča, ampak pričakovana, za govorca je lahko moteča pravzaprav njihova morebitna odsotnost.

Zadnja specifična lastnost govornega jezika, ki si jo bomo ogledali v okviru te razprave, so *prozodična sredstva*, ki so prav tako immanentna lastnost govornega jezika.<sup>53</sup> Omenjena je že bila njihova funkcija pri segmentiranju govora – tu so pomembni premori, stavčna intonacija, pa tudi višina in spremembe višine osnovnega tona, jakost ter hitrost govora. Sodobne raziskave prozodije govornega jezika temeljijo na meritvah, saj so orodja splošno dostopna, in ne več na posluhu raziskovalcev. Med dokaj neraziskana prozodična sredstva v govoru sodijo *premor*. Nekatere njihove funkcije so vsaj približno znane, npr. poudarjanje ali pridobivanje časa za načrtovanje govora. Analiza premorov na učnem govornem korpusu je pokazala naslednje: v korpusu je bistveno večji delež premorov v neformalnih besedilih (Zemljarič Miklavčič 2008a); v formalnem govoru so vsi premori brez izjeme dolgi največ 1 sekundo, v neformalnem govoru pa je ena četrtina premorov daljših (od 2 do 11 sekund). Gradivo torej kaže, da so premori manj moteči oz. bolj običajni v neformalnem govoru. V formalnem govoru se 93 % premorov zgodi v okviru monologa, naredi jih govorec sam od sebe, ne da bi ga kaj zmotilo; v 10 odstotkih premor podaljša še z zapolnjenim premorom (polglasnik). V neformalnem govoru so premori popolnoma drugačne narave, okoli premora se vedno »nekaj dogaja«: 24-krat po premoru spregovorita oba govorca hkrati, kot da se nista dobro razumela, komu pripada vloga govorca; 13-krat sta hkrati govorila pred premorom, očitno sta se oba ustavila v puščanju prednosti; 6-krat premoru sledi vprašanje, ki načenja novo temo; 5-krat je premor posledica motenj v okolici (ropot) ali pa glasovni zvoki iz okolice nakazujejo, da so govorniki prekinili govorjenje zaradi zunanjih okoliščin (gledajo televizijo, listajo knjigo). Že ta površna analiza gradiva je pokazala, da se premori v govoru zelo razlikujejo med seboj in da niso nekaj redundantnega in motečega, ampak imajo vedno svoj vzrok in pogosto tudi svojo funkcijo, kar je vse treba še natančneje raziskati.

<sup>53</sup> Prozodične lastnosti govora se po tradicionalnem pojmovanju prenašajo v pisavo predvsem v obliki ločil (kar potrjuje primarnost govora v primerjavi s pisanjem); v nekaterih elektronskih medijih se prozodične lastnosti, pa tudi splošnejši situacijski kontekst, še dodatno prenašajo v pisni jezik, npr. KVA? (kričimo), © (ob povedanem se smehljamo), ☺ (smo jezni).

</nst> ne [oni ki imajo	<pavza> [ne <neraz>] potem pa oni ves
[ja no] <neraz> +G01+ [<neraz>]	<pavza> [nič ne dela] +G15+ [ja mreža
fila </nst> [skozi] +G17+ <neraz>	<pavza> [pa] Irak <nv> pihne skozi
[ja jaz tudi nimam več] +G15+	<pavza> [samo na kakšno staro če se
[tukajle] +G10+ [mhm] +G03+	<pavza> [samo padeš dol] +G01+
+G10+ [<neraz> ja] +G15+	<pavza> [tudi letiš] +G01+ <pavza>
[<ime> ne <ime> <neraz>] +G01+	<pavza> [<govori s polnimi usti>
pet procentov pa [za Busha]	<pavza> [<neraz [ja] +G16+ [ja
tako no danes je Himalaja problem je	<pavza> a film ja aha <shift=vpr> a
napravo in drugi šumi] +G01+ Adiža	<pavza> a to je tole ne +G10+ <pavza>
[<neraz>] +G16+ [<neraz>]	<pavza> a več <neraz> ni jasno pa
doktorat dela pa rabi posnetke ne	<pavza> a več <pavza> pa je prosila
(ostro šumenje papirja) +G03+	<pavza> aja daj [preden] začnem ej
[kar] </smeh> +G16+ [aja]	<pavza> bo prišel [iskat] +G17+ [ja]
sigurno <shift=vpr> ane </shift=vpr>	<pavza> imate <shift=vpr> kako se
ja <shift=vpr> aja delaš </shift=vpr>	<pavza> itak pa ni ne petek je no- je
Avstrija] +G10+ [Italija <pavza> med	<pavza> Italija ja] +G15+ [v Italiji]
se lotili] +G01+ [domov pojdi	<pavza> ja +G10+ <??> <nv> smeh </nv>
</shift=vpr> +G11+ [ameriške firme]	<pavza> ja +G13+ [ti <nst> skejterski
v Prištini] +G19+ ja [<pavza> ja	<pavza> ja seveda ja] +G20+ ja ti si
je v bistvu daleč ne +G03+ [tako da	<pavza> ja zdaj grem jaz z z <repet/>
aja iz <lz> v Prištini] +G19+ ja	<pavza> ja <pavza> ja seveda ja]
(telefon zazvoni) a veš ja	<pavza> ja <shift=vpr> kaj je
spet <tj : ang> message </tj> +G20+	<pavza> tri mesece je mimo <nv> smeh
bo zmagal spet malo pogoljufali a	<pavza> vprašanje tudi če ne ne ja
ane potem pa še tam neke formalnosti	<pavza> zdajle dvanajstega imam enega
če ne dela +G10+ [<??> no </??>]	<pavza> <??> aja </??> +back sounds+
v Sloveniji tako ime recimo </> +G11+	<pavza> <ime> <priimek> +G01+ ja
skozi usta </nv> to samo gledaš	<pavza> <neraz> (ropot avtomobila)
+G15+ <pavza> [tudi letiš] +G01+	<pavza> <neraz> <?> telemarkt </??>
ja] +G15+ [v Italiji] +G09+	<pavza> <nv> smeh </nv> +G09+ <pavza>
da je kvečjemu tukaj bila kje +G09+	<pavza> <nv> smeh </nv> <ime> pomoč
potem je pa malo dolgčas gledati ne	<pavza> <shift=vpr> a boš šel gledat
kje pa je to </shift=vpr> ø na Viču	<pavza> <shift=vpr> a je to gostilna
pa tudi drugače kaj pa vem no	<pavza> <shift=vpr> ja [a veš] tisto
a veš <pavza> pa je prosila če lahko	<pavza> <shift=vpr> kaj pa dela za en
</nst> a veš <pavza> tako je ne	<pavza> <shift=vpr> kaj sta bila kaj
<neraz> <pavza> (5) film ja	<pavza> <shift=vpr> si slišal kaj je
potem v bližini] +G01+ no in a veš	<pavza> ø +G01+ tja grede smo šli
do konca septembra +G15+ a svojo	<pavza> ø po mo <lz> +G15+ b <lz>
<pavza> (3) samo da vidim +G19+	<pavza> ø <pavza> +G19+ <pavza> (11)
kamorkoli] si klicala +G19+ ja +G19+	<pavza> čakaj kako je Ie to +G19+ če
ja +G16+ [<nst> frej </nst>]	<pavza> če ne bo spet kaj

### Slika 5: Nekateri premori v neformalnem govoru UKGS

Sestavni del govora so poleg premorov, intonacije in tonske višine še drugi akustični dogodki, ki lahko odločilno vplivajo na razumevanje vsebine, pomena

ali namena govora. To so tudi različne oblike deformiranja glasu ali oponašanja – *smeh, kričanje, šepetanje, počasno govorjenje, odsekano govorjenje, oponašanje narečja* imajo v besedilu svojo funkcijo; so neke vrste govorni metajezik oz. diskurzni označevalci, saj označujejo odnos govorca do vsebine ali so odraz situacijskega konteksta. Podobno kot ostale prvine govora še potrebujejo podrobnejšo obravnavo.

## 6. ZAKLJUČEK

V okviru prevodoslovja je mogoče predvideti delež raziskav, ki se ukvarjajo z govorjenim jezikom, bodisi da gre za raziskave, ki preučujejo proces in značilnosti tolmačenja, bodisi strojno prevajanje govora, ali pa prevajanje pogovornega jezika – lahko tudi v zapisani obliki v literarnih delih. Govorjeni jezik na splošno velja za manj raziskovano in raziskano področje jezika, ki pa v zadnjih desetletjih, tudi ali predvsem v okviru korpusnih raziskav, pridobiva na pomenu. V prispevku so bile na kratko predstavljene nekatere sodobnejše raziskave govorjenega jezika in njihovi izsledki. Bistvo spremenjene filozofije je, da se govorjeni jezik opazuje kot samostojno jezikovno pojavnost s svojimi zakonitostmi in ne kot odstop od knjižne norme.<sup>54</sup> Predstavljeni so bili pogoji za zbiranje gradiva; za opis in analizo leksike, morfologije, skladnje in drugih značilnosti govorjenega jezika so običajno potrebna cela besedila in govorni korpusi, besedila pa morajo biti dostopna tako v zvočnih posnetkih kot v transkripcijah. Zapisovanje govora lahko poteka na različne načine, odvisno od ciljev raziskave, zato so bili predstavljeni različni načini transkribiranja. Tudi tu velja, da se vzpostavljajo novi standardi zapisovanja govora, pogosto neodvisni od knjižnega zapisa.

Iz predstavljenega je mogoče razbrati, da gre pri opisu značilnosti govorjenega jezika vendarle šele za fragmentarne izsledke posameznih raziskav. Pričakujemo lahko, da bo celovitejši opis govorjenega jezika nastal v naslednjih letih, ko bo raziskovalni javnosti dostopen referenčni govorni korpus za slovenščino.

### Bibliografija

- De Beaugrande, Robert Alain, in Dressler, Wolfgang Ulrich, 1981: Uvod v besediloslovje (prev. A. Derganc in T. Miklič 1992). Ljubljana: Park.
- Fillmore, Charles, 1992: Corpus linguistics or Computer-aided armchair linguistics. Jan Svartvik (ur.): *Directions in corpus linguistics*. Berlin: Mouton. 35–60.

<sup>54</sup> Celo nasprotno, v nekaterih okoliščinah govorjeni jezik prevzema vlogo "standarda" (v pomenu *bolj običajno*), pisni jezik se mu začne prilagajati (SMS-sporočila, blogi, forumi).

- Gorjanc, Vojko, 2005: *Uvod v korpusno jezikoslovje*. Zbirka Zrenja. Domžale: Založba Izolit.
- Kranjc, Simona, 1996/97: Govorjeni diskurz. *Jezik in slovtstvo* 42, 7. 307–320.
- McCarthy, Michael, 1998: *Spoken Language & Applied Linguistics*. Cambridge University Press.
- Pogorelec, Breda, 1965: Vprašanja govornega jezika. *Jezikovni pogovori*. Ljubljana: Cankarjeva založba.
- Pogorelec, Breda, 1998: Jezikovno načrtovanje govornega jezika pri Slovenih. Štrukelj, I. (ur.): *Jezik za danes in jutri*. Ljubljana: Društvo za uporabno jezikoslovje. 56–64.
- Quirk, R., Greenbaum, S., Leech, G. in Svartvik, J., 1985: *A comprehensive grammar of the English language*. London, New York: Longman.
- Sigley, Robert, 1997: Text Categories and Where You Can Stick Them: A Crude Formality Index. *International Journal of Corpus Linguistics* 2/2. 199–238.
- Skubic, Andrej E., 2004: Sociolekti od izraza do pomena: kultiviranost, obrobje in eksces. Kržišnik, E. (ur.): *Obdobja* 22. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik. 297–320.
- Slovar slovenskega knjižnega jezika. A–H*: 1980. Ljubljana: SAZU, Inštitut za slovenski jezik Frana Ramovša in DZS.
- Slovenski pravopis*, 2003. Elektronska izdaja. Ljubljana, ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša, Amebis. d. o. o.
- Smolej, Mojca, 2006: *Vpliv besedilne vrste na uresničitev skladenjskih struktur (primer narativnih besedil v vsakdanjem spontanem govoru)*. Doktorska disertacija. Mentor: M. Stabej. Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko.
- Stabej, Marko in Vitez, Primož, 2000: KGB (korpus govornih besedil) v slovenščini. Erjavec, T. in Gros, J. (ur.): *Jezikovne tehnologije za slovenski jezik*. 79–81.
- Stabej, Marko, 2008: Kako pa govoriš!? (Spis o normi in normalnem). Vitez, P. (ur.): *Spisi o govoru Ljubljana : Znanstvenoraziskovalni inštitut Filozofske fakultete*. 87–100.
- Svartvik, Jan, 2007: Corpus linguistics 25+ years on. Roberta Facchinetti (ur.): *Corpus Linguistics 25 Years on*. Amsterdam – New York, Editions Rodopi B. V. (Language and Computers: Studies in practical linguistics, No 62). 11–26.
- Škofic-Gujez, Jožica, 1998: Govor celjskega predmestja Gaberje. *Jezikoslovni zapiski* 4. Ljubljana: Inštitut za slovenski jezik Frana Ramovša. 89–98.
- Teubert, Wolfgang, 1999: Korpusno jezikoslovje in leksikografija. Gorjanc, V. in Krek, S. (ur.), 2005: *Študije o korpusnem jezikoslovju*. Ljubljana: Krtina.
- Tivadar, Hotimir, 2008: *Kakovost in trajanje samoglasnikov v govornem knjižnem jeziku*. Doktorska disertacija. Mentorica: A. Vidovič Muha. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko.

- Tognini Bonelli, Elena, 2001: *Corpus Linguistics at Work*. Studies in Corpus Linguistics 6. Amsterdam/Philadelphia: John Benjamin's Publishing Company.
- Toporišič, Jože, 1976: *Slovenska slovnica*. Založba Obzorja Maribor.
- Verovnik, Tina, 2004: Govorjeni knjižni jezik v televizijskih dnevnoinformativnih oddajah: študija primera. Poler Kovačič, M. in Kalin Golob, M. (ur.): *Poti slovenskega novinarstva – danes in jutri*. Ljubljana: Fakulteta za družbene vede. 157–173.
- Verdonik, Darinka, 2006: *Vpliv besedilne vrste na uresničitev skladijskih struktur (primer narativnih besedil v vsakdanjem spontanem govoru)*. Doktorska disertacija. Mentor: M. Stabej. Ljubljana: Filozofska fakulteta Univerze v Ljubljani. Oddelek za slovenistiko.
- Vitez, Primož in Zwitter Vitez, Ana, 2004: Problem prozodične analize spontanega govora. *Jezik in slovtvo* 49, 6. [3]–24.
- Vitez, Primož (ur.), 2008: *Spisi o govoru*. Ljubljana: Znanstvenoraziskovalni inštitut FF.
- Vogel, Jerca, 2004: Nekateri vidiki zvrstnosti govorenega diskurza s stališča poslušalca. Kržišnik, E. (ur.): *Obdobja* 22. Ljubljana: Filozofska fakulteta Univerze v Ljubljani, Oddelek za slovenistiko, Center za slovenščino kot drugi/tuji jezik. 453–468.
- Zemljarič Miklavčič, Jana, 2008a: Iskanje odgovorov na *Vprašanja govorenega jezika*. *Jezik in slovtvo* 53, 2. [89]–106.
- Zemljarič Miklavčič, Jana, 2008b: *Govorni korpusi*. Znanstvena založba FF, Oddelk za prevajalstvo (Zbirka prevodoslovje in uporabno jezikoslovje).
- Zwitter Vitez, Ana, Zemljarič Miklavčič, Jana, Stabej, Marko, Krek, Simon, 2009: Načela transkribiranja in označevanja posnetkov v referenčnem govornem korpusu slovenščine. Stabej, M. (ur.): *Infrastruktura slovenščine in slovenistike*. Simpozij Obdobja 28. Ljubljana, Znanstvena založba FF UL. 437–442