

UNIVERZA V LJUBLJANI
FILOZOFSKA FAKULTETA
ODDELEK ZA SLOVENISTIKO

Študijski program: dvopredm. univ.

zgodovina in

slovenski jezik

ROK BEČAN

STILOMETRIČNA ANALIZA IDIOLEKTA

**ugotavljanje sloga posameznikovega govora s pomočjo statistike
besednih oblik v govornem korpusu**

diplomsko delo

Mentor: dr. Miran Hladnik, red. prof.

Somentor: dr. Primož Jakopin, doc.

LOVRENC NA POHORJU, AVGUST 2007

SAMOZALOŽBA

Opomba o oblikovanju: Po soglasju z mentorjema je besedilo zaradi velikega obsega diplome natisnjeno obojestransko in z enojnim vrstičnim razmakom.

Diploma in raziskovalno gradivo sta v elektronski obliki dosegljiva tudi na zgoščenki – na zadnji prilogi ter na avtorjevi osebni spletni strani <http://www.soncek.net/rok/>.

KAZALO VSEBINE

POVZETEK – ABSTRACT	6
ZAHVALE	7
I. IZBOR CILJEV, IZHODIŠČ, GRADIVA IN METOD DELA	8
1. CILJI DELA.....	8
1.1. Od raziskovalnih tradicij do opredelitev izhodišč dela	9
2. PRISTOP K JEZIKU KOT PREDMETU RAZISKOVANJA.....	10
2.1. Razčlemba jezikovnega sporazumevalnega produkta – od aktualizacije v jezikovni sistem 11	
2.1.1. Govorni delni sistem jezika v tradiciji jezikovnozvrstne teorije	11
2.1.2. Ena ali več slovnice?	13
2.2. Govorni znaki in pisava – semiotični vidik.....	17
2.2.1. Osnovna in pomožna izrazna sredstva	17
2.2.1.1. Pisna izrazna sredstva.....	19
2.2.1.2. Govorna izrazna sredstva	21
2.2.2. Izrazna sredstva po jezikovnih ravninah.....	23
2.2.3. Slovenski pravopis v funkciji znanstvenega zapisa govora.....	25
2.3. Izbor enot z besedne ravnine.....	27
2.3.1. Beseda in njen(i) sistem(i).....	28
2.3.1.1. Izrazno različne besede – pojavnica in različnica.....	30
2.4. Kvantitativni pretres jezikovnega gradiva s statistikami in pogostostmi.....	32
3. SPORAZUMEVALNE OKOLIŠČINE BESEDILA	35
3.1. Model sporazumevalnih dejavnikov	38
3.1.1. Neodvisne okoliščine sporazumevanja	39
3.1.2. Vpliv soudeležencev na sporazumevanje.....	40
3.1.2.1. Idiolekt – koncept diasistema jezika, vezan na posameznika	42
3.1.3. Sporazumevalna sredstva	43
3.2. Vrste besedil in meje med besedili	45
3.2.1. Meje med besedili v tej nalogi.....	45
3.2.2. Sobesedilo.....	46
4. SLOG RABE JEZIKA.....	47
4.1. Tradicionalni stilistični pogledi.....	48
4.1.1. Pomen sloga v literarni vedi.....	49
4.1.1.1. Govor v književnosti in publicistiki	50
4.1.2. O slogu v jezikoslovju.....	51
4.1.2.1. Jezikovne ravnine stilemov	52
4.1.2.2. Deskriptivna stilistika: slog kot učinek nekaterih izraznih sredstev.....	52
4.1.2.3. Vsebina sloga: konotativni pomen	53
4.1.2.4. Sistemska in strukturalna umestitev razlikovanja sloga.....	55
4.1.2.5. „Obstranskost“ sloga.....	57
4.2. Poskus strukturiranja sloga z modelom.....	58
4.3. Izbrane stilometrične metode	61
4.3.1. Umestitev pogostosti različnice v slogovni model	61
4.3.1.1. Pogostost različnice kot kvantitativno <i>izražanje</i> sloga.....	62
4.3.1.2. Slogovni pomen različnice in njene pogostosti	64
4.3.1.3. Ovinek k pogostostim leksemov	65
4.3.2. Kvocien različnic in pojavnic (R/P) v slogovnem modelu.....	66
4.3.3. Sistemi primerjav sloga	66
5. KORPUSNA ORGANIZACIJA IN RAČUNALNIŠKA OBDELAVA GRADIVA.....	68

5.1.	Zgledi govornih korpusov z govorom v ortografski transkribciji	68
5.1.1.	BNC.....	71
5.1.2.	Turdis-1 in UKGS.....	72
5.2.	Z računalnikom od transkribcij govora do njihove analize	73
5.2.1.	Obdelava zvočnih datotek.....	74
5.2.2.	Standardi računalniškega zapisa govornih besedil	75
5.2.3.	Programi za transkribiranje govora in obdelavo transkribcij	78
II.	TRANSKRIBCIJSKA KONVENCIJA STANID	81
6.	GRADIVO KOT ZBIRKA PODATKOV, NJEGOVA UREDITEV IN OBDELAVA.....	82
6.1.	Partitura in drugi formati transkribcije	82
6.2.	Druga mesta zapisa podatkov	85
6.3.	Šifriranje podatkov	88
7.	PRAVILA TRANSKRIBIRANJA TEKOČEGA GOVORA	88
7.1.	Razločevanje in zapis besedne (verbalne) strukture govora	89
7.1.1.	Razdelitev pisnih znamenj po njihovem pomenu v transkribciji	91
7.1.2.	Analitične izpostavitve podatkov	94
7.1.3.	Predizrazno nepopoln govor in nepopolno razumevanje	95
7.1.4.	Neproblematičen zapis	97
7.1.5.	Problematični zapisi: jasna povezljivost s standardno pisno obliko	98
7.1.6.	Problematični zapisi: ohlapna povezljivost s standardno pisno obliko	101
7.2.	Segmentacija – jezikovne enote tik nad besedno ravnino	103
7.2.1.	Izjave in premolki med njimi	103
7.2.2.	Meje povedi	104
7.3.	Zapis paraverbalnega.....	106
7.4.	Primeri zapisov govora iz drugih transkribcijskih konvencij	108
8.	ZAPIS SPORAZUMEVALNIH OKOLIŠČIN.....	111
8.1.	Tekoče spreminjanje okoliščin.....	111
8.2.	Metaoznake transkribcij.....	112
8.3.	Baza govorcev	112
8.4.	Besedila – poimenovanje in rezanje.....	114
8.5.	Besedila in njihove metaoznake.....	115
8.6.	Vloge govorcev v besedilih	118
8.7.	Primeri iz drugih korpusov	119
9.	DOKUMENTACIJA OBDELAVE GRADIVA IN KORPUSNE STRUKTURE	122
9.1.	Preverjanje tekočih zapisov v partituri	123
9.1.1.	Napake zapisa na verbalnih trakovih.....	123
9.1.2.	Napake zapisa na trakovih [pv] in [Ozadje].....	124
III.	POTEK OBDELAVE GRADIVA	125
10.	SNEMANJE IN PRIPRAVA ZVOČNIH DATOTEK.....	127
10.1.	Snemanje.....	128
10.2.	Predelava posnetkov – prenos na računalnik	129
11.	ZAPIS „ČISTEGA“ BESEDILA IN OSTALEGA V PARTITURI.....	130
12.	ZAPIS SPORAZUMEVALNIH OKOLIŠČIN.....	132
12.1.	Preglednica transkribcij – stopnja .3 <i>Transkribcije</i>	132
12.2.	Preglednica govorcev	133
12.3.	Rezanje gradiva na besedilne enote oz. besedila	134
12.3.1.	Zamišljanje rezov oz. besedil.....	135
12.3.2.	Rezanje posnetkov.....	136
12.3.3.	Rezanje transkribcij	137
12.3.4.	Uskladitev razrezanih posnetkov in transkribcij	138
12.3.5.	Zlepljanje mejnih delov materinskih transkribcij v besedila	138

12.4.	Metaoznake besedil – zapis okoliščin sporazumevanja.....	139
12.4.1.	Preglednica govorcev v besedilih	139
12.4.2.	Preglednica besedilnih enot.....	141
13.	PRIPRAVA ANALIZ	142
13.1.	Delo s transkribcijami – besedili	142
13.2.	Izkoriščanje podatkov o okoliščinah sporazumevanja.....	145
13.3.	Končna mapna in datotečna struktura	146
IV.	ANALIZE	147
14.	KORPUS KOT CELOTA.....	147
14.1.	Pregled jezikovnih elementov in transkripcijskih posebnosti	148
14.1.1.	Pisna znamenja.....	149
14.1.2.	Besede	150
14.1.3.	Povedi.....	152
14.1.4.	Besedila	153
14.2.	Slog govora v korpusu kot celoti.....	154
15.	OKOLIŠČINE SPORAZUMEVANJA MED SEBOJ	156
15.1.	Slogi sporazumevanja v pogostejših okoliščinah	157
16.	GOVORCI MED SEBOJ	161
16.1.	Slogi govorcev.....	163
16.2.	Vloge izbranega govorca proti sobesedilu	164
16.3.	Vloge izbranega govorca med seboj	165
17.	OVREDNOTENJE KVOCIENTA R/P	169
V.	ZAKLJUČKI	171
VI.	SEZNAMI.....	176
18.	KRAJŠAVE, POJASNILO OBLIKOVANJA TISKA.....	176
19.	KAZALA TABEL, SLIK IN PRIMEROV	177
20.	SEZNAM NAVEDENE LITERATURE.....	178
20.1.	Druga zanimiva literatura – transkribiranje	184
21.	ZANIMIVE SPLETNE POVEZAVE.....	184
21.1.	Seznami govornih korpusov	184
21.2.	Seznami orodij za urejanje transkripcij.....	184
VII.	PRILOGE	185
	PRILOGA 1: PREGLED BESEDIL V KORPUSU	185
	PRILOGA 2: BAZA GOVORCEV	205
	PRILOGA 3: ABECEDNI SEZNAM RAZLIČNIC KORPUSA	209
	PRILOGA 4: 1210 NAJPOGOSTEJŠIH RAZLIČNIC V PROSTEM GOVORU KORPUSA	227
	PRILOGA 5: PRIMERI KONKORDANC DVEH RAZLIČNIC.....	230
	PRILOGA 6: NAJPOGOSTEJŠE N-KRATNICE BESED V KORPUSU, N= [2, 3, 4, 5]	234
	PRILOGA 7: POVEDI V KORPUSU	235
	PRILOGA 8: XML-STRUKTURA TRANSKRIBCIJSKE DATOTEKE, EXMARALDA.....	237
	PRILOGA 9: XML-STRUKTURA TRANSKRIBCIJSKE DATOTEKE, ELAN	238
	PRILOGA 10: PRIMER PARTITURE IZ KORPUSA – BESEDILO 1NED19VG	242
	PRILOGA 11: IZJAVA O AVTORSTVU IN PODATKI AVTORJA	246
	PRILOGA 12: ZGOŠČENKA: PRIMERI POSNETKOV IDR.....	247

POVZETEK – ABSTRACT

Gradivo raziskave je govorni korpus slovenščine, obsegajoč 48.450 besed, v katerem je zajeto realno sporazumevanje v neprekinjenih posnetkih za dva dni – skupaj 28 ur. Snemalec je izbrani govorec – nosilec mikrofona, njegov govor ali idiolekt je v paleti sedmih primerjav govora iz korpusa stilistično opisan. Za potrebe opisa sloga je na podlagi strukturalnega pristopa razvit slogovni model. Kvantitativna metoda stilističnega opisa se poslužuje seznama najpogostejših besednih oblik ali različnic ter kvocienta različnic in pojavnic, kvantitativno pa so obdelani tudi pisna znamenja, besedne n-kratnice in povedi v korpusu. Zapis korpusa je določen v transkripcijski konvenciji StAnId, ki vključuje modificirano ortografsko transkripcijo v partiturnem formatu in teoretični model sporazumevalnih dejavnikov, s katerimi so podrobno označena besedila. Transkripcije so vzporejene z zvočnimi posnetki, možno pa jih je preko prosto dostopnih programov iz drugih virov (Elan, EXMARaLDA Partitur Editor in Eva) tudi dopolnjevati, iskati konkordance itd. Korpus StAnId je raziskovalcem na razpolago v elektronski obliki pri avtorju.

KLJUČNE BESEDE: stilometrika, stilistika, idiolekt, govorni jezik, pogostost besed, različnice in pojavnice, govorni korpus, transkripcijska konvencija

The research material is a spoken corpus of Slovenian (48.450 words), which includes natural spoken communication on continuous records of two days (28 hours). The idiolect (as spoken material of the person who carries microphone) is stylistically described inside the group of 7 comparisons of speech in the corpus. Stylistic description itself is based on original stylistic model, which is derived from structural grammar. The quantitative method of stylistic description uses list of the most frequent word forms (or types) and the type-token ratio. Documentation of material in the corpus is defined in the StAniD transcription convention and includes modified orthographic transcription in partitur format and theoretic model of communication factors; the latter are attributed to each text in corpus. Transcriptions are aligned with the records and through freely available programs (Elan, EXMARaLDA Partitur Editor and Eva) they can be updated, searched for concordances etc. The StAnId corpus is available free for researchers on authors address.

KEY WORDS: stylometry, stylistics, idiolect, spoken language, word frequency, types and tokens, spoken corpus, transcription convention

ZAHVALE

Rad omenjam naslednje osebe ali vire, katerim sem dolžan zahvalo za njihov prispevek pri tej nalogi:

- mojim domačim in bližnjim za podporo pri študijskem delu,
- mentorjema Miranu Hladniku in Primožu Jakopinu za potrpežljivo sprejemanje dolgotrajnega in obsežnega dela, posebej slednjemu za uvajanje v računalniško jezikoslovje, pomoč pri oblikovanju osnovnega cilja naloge in prosto uporabo njegovega programa Eva,
- predavateljema Petru Grzybku in Emmerichu Kelihu s slavistike v Gradcu za uvajanje v stilometrijo in posredovanje literature s tega področja,
- Jani Zemljarič Miklavčič za posredovanje računalniških prosojnic s predstavitvijo njenega korpusa,
- ustvarjalcem prosto dostopnih programov Audacity, Elan in EXMARaLDA (projekt s paketom več programov),
- rodnemu bratu Anžetu Bečanu za pomoč pri računalniških zagatah,
- dekanu Bernardu Geršaku za posojilo minidiska za potrebe snemanja
- ter navedenim raziskovalcem za dostopnost njihovih dognanj in gradiv.

I. IZBOR CILJEV, IZHODIŠČ, GRADIVA IN METOD DELA

*„Dan izreka dnevu govor, noč razglaša noči znanje.
To ni govor, to niso besede, njihov glas je neslišen;
po vsej zemlji gre njihov val, do konca sveta njihove besede.“
(Ps 19, 3–5)*

Je govor (jezik, sporočilo ...) stvaritev posameznika ali skupnosti? Kakšen slog jezika ima posameznik, kako ga prilagaja okoliščinam sporazumevanja? Kaj vse sploh vpliva na sporazumevanje? Kakšen vpliv na sporazumevanje ima sam govor kot delni sistem jezika? Katere slogovne značilnosti govora lahko zapišemo (transkribiramo), katerih ne? Kako lahko govor strojno oziroma računalniško obdelujemo? Kako lahko slog opišemo s pomočjo statističnega opisa jezika? Katere enote jezika lahko štejemo? Kaj nam pove beseda kot pregibna besedna oblika, kaj kot leksem ali lema (tj. beseda, kot jo najdemo v slovarju)? Kaj nam povedo funkcije besede v sporočilu in kaj slovnične kategorije gole besedne oblike? Na podlagi česa nastane slovar ali slovnica? Je gradivo jezikovnih priročnikov govorni ali pisni jezik, vsakdanji, strokovni ali umetnostni jezik? Kakšno je razmerje realnega, živega, naravnega govora do standard(*izžira*)nega jezika, *parole* do *langue*? V katerem sistemu jezika (diferenciaciji, zvrsti, prenosniku, kodu ...) mi, jezikovni uporabniki, *mislimo, iščemo izraz in svoj slog, pripovedujemo sporočilo*? Kje se rodi beseda, jezik?

V začaranem krogu takšnih globokih in tudi bolj konkretnih vprašanj se rojeva tudi ta diploma, seveda pa ne daje vseh odgovorov.

1. CILJI DELA

Osnovna ideja, iz katere se je razvilo pričujoče diplomsko delo, je bila na podlagi kvantitativne analize besedne jezikovne ravnine odkriti (statistično opisljive) posebnosti spontanega (po)govornega jezika. Cilji dela so bili:

1)) Širino dela omejiti z izborom opazovanja idiolekta (govora ene osebe) z vidika besed (leksikalno, skladijsko) in okoliščin (po)govora. Izbranih naj bi bilo nekaj (statističnih) stilometričnih analiz besedne jezikovne ravnine (pogostostne liste različnic, lem in oblikoskladijskih določitev), okoliščine govora pa enostavno kategorizirane.

2)) Delo utemeljiti in opredeliti na podlagi stilistične analize, teorije analize govornega diskurza (jezikoslovno, sociolingvistično), kvantitativne lingvistike in aktualnih jezikovno-tehnoloških raziskav (jezikovne tehnologije, govorni korpusi).

3)) Za jezikovni vir narediti enotedenski neprestani zvočni posnetek govorne (in slušne) dejavnosti ene osrednje opazovane osebe in njenih sogovorcev oz. govorno-slušnih ssporazumevalcev. Vir naj bi v ortografski transkripciji obsegal 100.000 besed, bil pa tudi oblikoslovno označen, lematiziran in razdvoumljen. Besedila naj bi bila oblikovana v preprost korpus za potrebe interpretativnega branja in kvantitativnih analiz.

4)) Obdelava gradiva naj bi v čim večji meri potekala strojno (program za transkribiranje, strojni določevalnik besednih oblik), s prosto dostopno programsko opremo, po enostavnih postopkih in s kratko časovno obdelavo, a s standardiziranimi rezultati.

Možnosti dela so seveda okrnjene glede na sposobnosti raziskovalca in obseg študijske obveznosti – diplome. In gradivo in rezultati naloge naj bi bili javno predstavljeni, knjižno (v okviru diplomske naloge) in na spletu, morda pa še kot članek v strokovni reviji.

Opomba: Od teh začetnih ciljev do konca diplome ni bil izpolnjen tretji in deloma prvi: skrčen je bil obseg dela in tako sta bila od posnetkov vsega tedna transkribirana le dva dni (48.450

besed), transkribcija pa ni lematizirana in oblikoslovno označena. Več o rezultatih dela v zaključkih, poglavje V.

1.1. Od raziskovalnih tradicij do opredelitev izhodišč dela

Pričujočo diplomu zaznamuje dvojnost: njen prvi konkretni cilj je priprava gradiva (besedila v širšem pomenu: z oznakami konteksta in jezikoslovnimi metaoznaki), drugi konkretni cilj pa izvedba analize jezika, ki o tem gradivu nekaj pove. Ta dvojnost ni toliko posledica uporabe računalnika in statističnih analiz; je posledica pretvorbe posnetega jezika preko transkribcije v *sistem informacij*, urejen tako, da iz primarnega vira odslikava značilnosti, ki jih želi zajeti raziskovalec. Ta sistem informacij seveda običajneje imenujemo **besedilni korpus**. „Odločitve, narejene v času transkribcije in označevanja, bodo vplivale na ves proces analize podatkov. Previdno premišljena transkribcija lahko odločilno pripomore k analizi; slabo odločanje v zgodnji fazi pa lahko naredi sledeče analize težavne ali brezvredne“ (Leech idr. 1995: 15–16). Delovne naloge raziskovanja so kljub dvojnosti diplome med seboj tesno povezane in vseskozi aktualne, čeprav jih lahko postavimo v idealno zaporedje: kaj preučujemo, o čem se sprašujemo, kako preučujemo, kaj lahko na koncu povemo in kako lahko ugotovljeno uporabimo.

Predmet preučevanja je realni jezik, celostno označen kot „naravni, človeški“, ožje izbrano pa gre za slovenščino v najrazličnejših situacijah, za njen govorno-slušni prenosniški kod. Opredelitev teoretičnih izhodišč, ki se tukaj začne, je odvisna od že omenjenih ciljev in od same dejanskosti objekta (kod, prenosnik, struktura, drugi sporazumevalni dejavniki ...). Če je raziskovalcu navidez lahko določiti cilje, pa se epistemološkim težavam pridruži še odvisnost njegovega znanja od tradicionalnih in dostopnih raziskovalnih koncepcij, metodologij. Naravni jezik je namreč tako kompleksen pojav, da zmore redko katera (če sploh katera) jezikoslovna šola ali disciplina zaobjeti vse vidike, ki ga celovito opisujejo, pojasnjujejo.

Osnovni pogled na preučevani jezik sledi „tradicionalnemu“ **strukturalističnemu jezikoslovju**, saj je jezik najpodrobneje analiziran kot *produkt* sporazumevalne dejavnosti, namreč besede kot njegove pomenske in oblikovne enote, določene s strukturnimi definicijami. Ker pa je v naravi družbenega ravnanja oz. sporazumevanja jezik tesno povezan s samim procesom svoje geneze in uporabe (Brinker/Sager 2001: 19–20), je upoštevan v nalogi še drug metodični pogled na jezik – ta, ki ga poudarjata **sociolingvistika** in **analiza govornega diskurza**. Priklic tega vidika precej zakomplicira pripravo gradiva (korpusov govornjenih besedil), saj se poveča število opisnih enot – poleg strukturnih enot jezika zajemamo še sporazumevalne okoliščine, le-te pa je zaradi pestrosti družbenih situacij še težje strukturirati za potrebe konsistentnih analiz.

Namera tega dela pa nikakor ni celovito opisati govorno slovenščino ali idiolekt nekega slovenskega govorca. Totalni opisi so v znanosti pravzaprav nezaželeni, saj so prisiljeni k uporabi modelov, ki so izrazito mimetični glede na objekt in zato izgubijo analitično distanco. (Npr. kritika iracionalizma in subjektivizma v „tradicionalni“ literarni znanosti pri Rolandu Barthesu govori o repoetizaciji literature; prim. Poniž 1974: 18–19.) Predmet te diplome je torej analitično *reduciran*, in sicer na podlagi statistične oz. kvantitativne metode, interpretacija rezultatov te analize pa je narejena v tradiciji **kvantitativnega jezikoslovja**. Redukcija objekta je izvedena že tudi na vseh stopnjah obdelave gradiva (kako, gl. Slika 12, str. 125), hkrati pa je z *označevanjem* jezikovnih enot na sintagmatski ravni in z označevanjem zunajjezikovnih okoliščin poskrbljeno za dokumentiranje in možnost procesiranja relevantnih jezikovnih lastnosti objekta. Pri tem delu se naslanjamo na izkušnje **računalniškega** in **korpusnega jezikoslovja** ter **jezikovnih tehnologij**.

Po tej samosvoji raziskovalni poti prihajamo do predvidene „krone“ dela, ki vsa polja raziskave povezuje in jih skuša osmisлити: stilistična analiza gradiva, ki skuša ugotoviti, kaj je možno povedati o posebni rabi jezika oz. o edinstvenostih v tem gradivu – ali z eno besedo: kaj je možno povedati o *slogu* v gradivu. **Stilistika** je tu mišljena v literarnovednem smislu, posluhuje pa se seveda jezikoslovnih metod.

Povezanost vseh vidikov naloge opisuje naslov **Stilometrična analiza idiolekta**: *metrična* (kvantitativna) analiza *idiolekta* (govornega jezikovnega sistema posameznega uporabnika jezika v sociolingvističnem smislu) nas pripelje do stilističnih ugotovitev (*slog* kot *idiolekt* v stilističnem smislu – značilna raba jezika, ki jo izvede določena oseba).

Sledeča uvodna poglavja želijo ozavestiti ozadje konceptov in pojmov, uporabljanih v nalogi. S tem naj bi bila dosežena enoznačnost terminologije in jasna določenost metod. Prav tako želijo podrobneje navezati diplomu na konkretna relevantna in aktualna raziskovanja, nakazana že zgoraj, ki so neposredna teoretična podlaga naloge in zato vir pogledov, metod, pojmov, terminologije, podatkov in idej.

2. PRISTOP K JEZIKU KOT PREDMETU RAZISKOVANJA

Pri pričujoči raziskavi jezika, ko z vedno podrobnejšo razčlemba in interpretacijo želimo povedati nekaj konkretnega o jeziku, je potrebna kopica teoretičnih odločitev. Posneli smo vse, kar mikrofoni in snemalna oprema zaznata kot zvočno dogajanje, osredotočimo se na antropogeno dogajanje in dalje na jezikovne znake. Da bi le-te lažje obdelali, jih transkribiramo v pisni prenosnik, zaznamujemo pa tudi okoliščine sporazumevanja. Pri obravnavi se držimo golih načel znanstvenosti, konsistentnost in celostnost pri razčlembi pa sta potrebni tudi zaradi nadaljnje strojne obdelave. Razbitje jezika na sistem enot (besede, segmenti, besedila, dejavniki sporazumevanja itd.) je posledica izbora vidikov in kriterijev (za besednovrstne, oblikoskladenjske in besedilne kategorije, pogostost itd.), ki jih je potrebno upoštevati tudi pri ugotavljanju posplošitev o jeziku, pri primerjavah z drugim gradivom, pri sklepanju o zunajjezikovnem svetu in pri uporabi jezika v prihodnjih namenih.

Naj naštejemo še nekaj **dilem**, nastajajočih pri delu, na katere skušam odgovoriti s pomočjo drugih teoretikov in z izborom lastnih rešitev v pričujočem delu. (Po)govorno sporazumevanje, ki se mu posvečamo – za kakšen delni sistem jezika gre, nas zanima v tradiciji razlikovanja prenosniških, socialnih ali funkcijskih zvrsti jezika? Ali ga sploh lahko opisujem s pomočjo slovničnih koncepcij, nastalih (predvsem) na podlagi pisnih virov (slovenski *knjižni* jezik, Slovar slovenskega *knjižnega* jezika, *pisni* besedilni korpusi, *pisna* besedila v učnih korpusih strojnih oblikoskladenjskih in besednovrstnih določevalnikov in lematizatorjev)? Na katere jezikovne ravnini opisovanja govornega prenosnika se je smiselno osredotočiti, kako segmentirati govorni jezik? Kako naj govor transkribiramo oziroma prenesemo v pisno obliko in ga tu analiziramo, da se ohranijo vsaj nekatere poteze, ki so posledica govornega prenosnika? V čem je smisel izbora ortografske transkripcije in katere so posledice pri nadaljnji obdelavi?

Opazovanje govornega diskurza (predvsem t. i. spontanega ali prostega) je še posebno zanimivo, ker se pri njem razkriva meja med paradigmatskim in sintagmatskim v jeziku (Smolej 2004), pa tudi odnos med jezikovnim in zunajjezikovnim (predvsem družbena razsežnost sporazumevanja). Z vidika upoštevanja sporazumevalnih dejavnikov se na kompleksnejši ravnini ponovijo vprašanja zvrstnosti in enot (taksonomija besedil): Kakšno je razmerje posameznika in njegovega idiolekta do jezikovnega sistema skupnosti? Kako upoštevati tudi slušno aktivnost posameznega „govorca“ (tvorca, sporazumevalca)? Kako zajeti sporazumevalne dejavnike realnega sporazumevanja v uporaben model, jih opisati z enostavnimi in konsistentnimi kriteriji? Kateri dejavniki nas sploh zanimajo? Kje so meje sporazumevalnih enot oziroma besedil? Ali so ta primerljiva s taksonomijo pisnih besedil?

2.1. Razčlemba jezikovnega sporazumevalnega produkta – od aktualizacije v jezikovni sistem

Kot že rečeno, se najprej osredotočimo v nalogi na vse govorjeno gradivo kot na jezikovni kod, aktualizirano (kot *parole*¹) in fiksirano v posnetku in transkripciji, vzeto *sinbrono* in kot jezikovna celota (čeprav je od prvega do zadnjega posnetka preteklo 7 dni in se menjala paleta sporazumevalnih situacij, funkcij, tvorcev in naslovnikov itd.), in ga umestimo v kontekst slovenskega jezika. Ko ga kot takega nadalje analiziramo, primerjamo s temeljnimi sistemskimi možnostmi (*langue*), sledimo „tradicionalnemu“ oz. struktural(istič)nemu² in (pre)tvorbenemu jezikoslovju. Da pa teh vidikov ni lahko obdržati, ugotovimo že, ko se soočimo z jezikovnozvrstno teorijo.

2.1.1. Govorni delni sistem jezika v tradiciji jezikovnozvrstne teorije

Na podlagi govornega gradiva te naloge želimo govoriti o **govorni prenosniški zvrsti jezika**, kar je očitno predvsem na podlagi situacij, v katerih je bilo posneto gradivo. Situacij in tem se namreč ni izbiralo, ampak so bile posnete kar vse, na katere je nosilec mikrofona naletel. (Če bi bili posneti le npr. pogovori o vsakdanjih stvareh, bi lahko govorili o praktičnosporazumevalni funkcijski zvrsti jezika, odvisno od natančnejše družbene situacije pa še o socialni zvrsti jezika, verjetno o pogovornem jeziku ali o narečju.)

Da ne govorimo o socialni ali funkcijski zvrsti jezika,³ bi rad poudaril zato, ker je v slovenski tradiciji jezikovnozvrstne teorije dokaj zmede⁴ ob opozicijskih oznakah govorni (pogovorni, govorjeni, spontani govorni) – pisni (pisani, knjižni) jezik. Stična točka zamenjav zvrstnih vidikov je prav gotovo nedodelana predstava o „**pogovornem jeziku**“, ki da je (karikiram) „prisoten le v govoru, najpogosteje obsega teme iz vsakdanjega življenja, je uporabljan v zasebnih situacijah, je ena najpogostejših oblik sporazumevanja, v njem najobičajneje sodelujeta 2 govorca“. Gotovo drži le to, da ga

ni mogoče opisati „enkrat za vselej“, ampak ga je bolj kot ne mogoče prikazati le v njegovem življenju, torej v okviru posameznih govornih dejanj – preko teh uresničitvev pa predvideti njegovo zgradbo in rabo tudi v drugačnih govornih okoljih oziroma drugih govornih dejanjih (Škofic 1991: 9–10).

Zdi se torej, da je za tradicionalni strukturalni (normirajoči, „statični“; prav tam: 14–15) opis jezikovnih zvrsti dosti bolj primeren koncept prenosniških zvrsti, s katerim je na nižjih jezikovnih ravneh lažje predvideti s prenosnikom določena izrazila.

S tem ko smo izbrali gradivo na podlagi prenosniške zvrsti, smo zajeli še vedno dokaj nepredvideno množico govornih situacij, ki jo hočemo še dalje razčleniti. Taksonomija besedil govornega jezika z doslej omenjenimi „**zvrstnimi snopi**“ (Toporišič 1992: ESJ: 381) ni prav podrobno razdelana, sploh pa ne v primerjavi s taksonomijo pisnih besedil (prim. tudi mnenje v Zemljarič Miklavčič 2004: 503). Da ne gre le za situacijo v slovenskem jezikoslovju, kažejo tudi

¹ Običajni prevod tega Saussurjevega termina, *govor*, nima nič opraviti z nakazovanjem na govorni prenosnik v tej nalogi; Saussurjeva *parole* je lahko tudi aktualizacija sistema v pisnem besedilu. Prim. Vitez 1999: 25, opomba 7.

² „V tem in onem je bilo jezikoslovje vedno strukturalno, tj. je stvari obravnavalo med seboj povezano“ (Toporišič 1992: ESJ: 316: **strukturalno jezikoslovje**). V tem širokem smislu tudi tukaj jemljemo oznako strukturalizem; ne pomeni nam le npr. postavljanja razlikovalnih opozicij med preprostimi jezikovnimi elementi po Trubeckojevem zgledu.

³ Strokovni izrazi za zvrsti so iz Toporišič SS²⁰⁰⁰: 11–34, poglavje Zvrstnost slovenskega jezika.

⁴ Jožica Škofic je v magistrski razpravi ob podrobnem kronološkem pregledu slovenskih jezikoslovcov o pogovornem jeziku v 20. st. ugotavljala, „da so mnogi jezikoslovci, katerih dela sem pregledovala, povezovali (nekateri pa celo enačili) npr. govorjeni jezik (prenosniška zvrst) s pogovornim jezikom ali narečjem (socialni zvrsti), knjižni jezik (socialna zvrst) s pisanim jezikom (prenosniška zvrst) ipd.“ Škofic 1991: 15.

Priporočila za elektronsko kodiranje in izmenjavo tekstov konzorcija TEI: transkribicijam govora je namenjen en sam skupen nabor oznak, posebni nabori oznak pa so za prozo, poezijo, dramatiko, slovarje ipd. (gl. Sperberg-McQueen/Burnard 2001).

Zakaj so take razlike v taksonomijah glede na prenosnik? Najširše polje jezikoslovnih zanimanj govornega prenosnika zavzema prav spontani (po)govor, pisnemu prenosniku sorodnejša področja so obrobna. Največji problem je težavnost zajema govornega prenosnika prav v spontanah sporazumevalnih situacijah, ne pa v nepredvidljivosti teh situacij (Kranjc 1998 opozori še na razvoj tehnike za zbiranje gradiva). Zajem jezika je namreč lažji tam, kjer materialno naravo sporazumevanja *fiksirajo mediji*⁵ ali posredniki, in sicer tako, da zmanjšajo možnosti spreminjanja prenosnika in družbenega razmerja ssporazumevalcev. To nasploh velja tako vizualne kot zvočne množične medije, v katerih je sporazumevanje v odnosu do naslovnikov enosmerno – prim. dejstvo, da k velikosti pisnih korpusov največ pripomorejo časopisni in knjižni opusi, k velikosti govornih korpusov pa RTV objave (npr. t. i. govorne zbirke Broadcast News v govornih tehnologijah). Razkorak tipologij med pisnim in govornim jezikom pa je bolj viden na področju zasebnega oz. nejavnega sporazumevanja, kjer je zajem jezikovnega gradiva spet lažji za vse medije pisnega prenosnika: razglednice, sms-i, različne vrste pisem, dnevniki in kronike, poročila, dopisne seje itd.

Kljub različnim zvrstnim vidikom in njihovim zamenjavam pa se zdi, da lahko pri navidez vzporednih prenosniških zvrsteh (pisni in govorni) nanju enakomerno prenesemo besedilne taksonomske kriterije, ki izhajajo iz delitve jezika na **funkcijske in na socialne zvrsti**. Namreč, tako kot poznamo knjižni in govor(je)ni zborni jezik, tako bi lahko socialno zvrst pogovornega jezika dosledno delili na govorno in pisno (čeprav je slednje manj, pa vendar npr. v zasebnih pismih, spletnih klepetalnicah). Enako pri funkcijskozvrstnih kriterijih. Pravzaprav je problem kvečjemu subjektivna ocena, da značilnosti vseh teh besedilnih vrst veljajo za pisni prenosnik; ta velja le toliko, kolikor smo opise teh vrst zgradili predvsem na podlagi pisnih gradiv in jih naknadno prenesli še v normativni vidik (npr. pri razvoju standardnega jezika ali v pedagoškem procesu). Ta ocena je gotovo veljala nekdaj, dandanes pa sta se z diferenciacijo družbe in tehničnim razvojem sporazumevalnih sredstev funkcijska in socialna raba obeh prenosnikov zelo izenačili.

Pa se vrnimo k pazljivem poimenovanju (oz. koncipiranju) samega **prenosnika**. Za pričujoče gradivo je bistveno, da so jezikovni znaki slušno uresničeni prav z govorom/govorjenjem, od tod oznaki *govorni* in *govorjeni* (jezik).⁶ Tukaj menim, da ju je smiselno uporabljati kot sopomenki, kot sta sopomenki tudi govor in govorjenje, zgolj zaradi analogije z večino vrstnih pridevnikov pa raje uporabljam nedeležniško obliko, torej **govorni jezik**. So pa še druga razlikovanja: nekateri zaradi kalkiranja iz angleščine (*spoken language*) uporabljajo deležniško obliko *govorjeni*. Toporišič v našem pomenu uporablja le oznako *govorjeni jezik*, medtem ko oznako *govorni jezik* pridrži za prosto/spontano govorjenje (brez branja, recitiranja ipd.).⁷

⁵ Zaradi jasnosti poudarjam nesopomenskost terminov medij in prenosnik v strokovnem diskurzu o sporazumevanju, posebno v sporazumevalnih modelih. **Prenosnik/kanal** (govorni, pisni, otipni ipd.) govori o *fizični* naravi človeških organov, s katerimi človek tvori, modificira in sprejema sporazumevanje, s tem pa tudi *materialno* naravo jezikovnih znakov (pogostost, jakost, trajanje, višina zvoka; grafično in otipno obliko itd.). **Medij/posrednik** (neposredni ali posredni medosebni stik) pa predvsem omejuje sporazumevalce pri večsmernosti in množičnosti sporazumevanja in s tem pri spreminjanju družbenih odnosov med sporazumevanjem: *posredni mediji* so radio, časopis, knjiga, kinopredstava, *neposredni* pa predvajanje in pogovor v živo (z vidnim in/ali zvočnim stikom), telefonski ali videokonferenčni pogovor, deloma tudi dopisovanje. Primarnejši koncept oz. podpomen besede medij pa je 'tehnična sporazumevalna naprava'. Mediji so kot taki poimenovani po *materialni/tehnični* naravi prenosnika, v katerem deluje naprava: tiskani medij, zvočni, vidni, avdiovizualni.

⁶ Še posebej gre pri govornem prenosniku vedno za primaren način človeške tvorbe, ne glede na materialno naravo medija oz. tehnične naprave, ki lahko jezikovne znake zajema in kasneje posreduje: neposredni stik, v avdiosnemalnih napravah mikrofona, v avdiopredvajalnikih zvočnik.

⁷ Toporišič 1992: ESJ: 53, 54, gesli **govorjeni jezik in govorni jezik**; enači pa **govor** in **govorjenje** – gl. še ti gesli, prav tam.

Glede na samo materialno naravo prenosnika je možna še širša oznaka, namreč *slušni, zvočni* (tudi *glasovni*) jezik, ki poleg govora zajame še druge prenosniške kode, kot so npr. zvočni jezikovni⁸ znaki siren, zvonjenje telefona, pozivnikov, ur, različne vrste avizov v avdio(vizualnih) medijih, nabori zvočnih znakov v vojski itd. (prim. Korošec 1986: 393, opomba 1). Ti kodi (nabori zvočnih znakov) običajno niso tvorjeni z govornimi, zato se ne pokrivajo z oznako *govorni jezik*. Zanimiv je tudi vidik **besednosti** (verbalnosti; gl. še opomba 22 na str. 18): pod govornimi jezikom spadajo tako besede kot parabesedni (prozodija) in nebesedni (z glasom izražen smeh, vzkliki, namerni kašelj ipd.) jezikovni znaki, prej omenjeni kodi pa običajno niso sposobni oblikovati besed – izjema je npr. Morsejeva abeceda v zvočni izvedbi.

Poleg govornega in pisnega (zapisanega, vidnega, bralnega ...) prenosnika obstaja vsaj še otipni, pri opredelitvah pa je z ozirom na tvorbo, sprejemanje, fiksiranje, verbalnost in jezikovnost nabora znakov potrebno biti podobno previden kot pri govornem jeziku. Npr. oznaka pisni jezik v evropski tradiciji zajema abecedni besedni prikaz pomena, v npr. kitajski pisavi pa ideografski nebesedni prikaz pomena, ki je po strukturi izrazil bližje kretalnemu jeziku (znotraj koncepta *vidni jezik*) za sporazumevanje gluhih. V nadaljevanju se z oznako **pisni (jezik)** nanašam predvsem na abecedni črkovni zapis kot temelj slovenskega pisnega sistema, kar je bistveno pri primerjavah govornih in pisnih jezikovnih sredstev.

Model delitve jezika na prenosniške zvrsti je preprost, dejanske sporazumevalne situacije pa vključujejo bodisi enega bodisi več prenosnikov: npr. v neposrednem didaktičnem diskurzu je kombiniranje govora in vizualnih vsebin stalna praksa. Če pri analizi takega konkretnega večprenosniškega sporazumevanja nato obdelamo le en prenosnik, izgubimo del informacij. Ko se skušam v pričujoči diplomii omejiti zgolj na govorni jezik, kot sem ga posneli, izgubljam informacije o vsej vidni komunikaciji, ki se je morda hkrati dogajala. Ko pa posnetek še transkribiram (*zapišem* – sic!), se izgubijo še informacije o parabesednem in nebesednem oblikovanju in govornega in zvočnega prenosnika, če jih ne zabeležimo ali ubesedimo na dodaten način v novem prenosniku. Vse te izgube (zmanjšanje obvestilnosti, povečanje entropije) lahko poimenujemo tudi **šum v pretvorniku** (Korošec 1986: 394).

2.1.2. Ena ali več slovnice?

Jezikovna sredstva v izbranem prenosniku želimo obravnavati sistemsko, torej v odnosu do vseh izrazil jezika in njihovih medsebojnih odnosov – slovnice. Že zgoraj smo se naslonili na tezo, da večina slovnicih opisov jezika temelji na opisih pisnega gradiva oz. jezika, govornimi jezikom pa je nekoliko zapostavljen (Krajnc 1998). Podrobneje to poskuša razložiti naslednji navedek:

Iz zgodovine opisnih in normodajnih jezikoslovnih pristopov večine standardnih jezikov se da razbrati dolgo kulturno in statusno prevlado jezika pisnih besedil.⁹ Prestižna pisna besedila (sprva religiozna, nato dolgo časa umetniška) so bila temelj za opis in normiranje standardnega jezika. Govor oz. jezik govornih besedil je šele pozno postal legitimen predmet jezikoslovnega raziskovanja (razen pri nekaterih vejah jezikoslovja, npr. dialektologiji /.../). Zaradi psiholoških oz. kognitivnih in komunikacijskih razlogov, ki so danes v precejšnji meri pojasnjeni, je govor veljal za manj popolno obliko sporazumevanja, torej tudi manj prestižno, zaradi česar je bil tudi neprimeren kot vir analize oz. spoznavanja jezika. (Stabej/Vitez 2000)

In

⁸ Govorimo samo o **jezikovnih znakih**, tj. (antropogenih) znakih, ki nosijo pomen in so s tem namenom tudi bili oblikovani, aktualizirani.

⁹ Še mnenje Toma Korošca: „Zapis jezika (jezikovnega sporočila) je ena najpomembnejših sestavin civilizacijskega stremeljenja v razvoju človeštva. Zapisano sporočilo je sposobno premostiti časovno vrzel, premagati enkratnost govornega sporočanja; zato je od nekdanj kulturna in zgodovinska vrednota. Pisava je iz prakture gradila civilizacijo, katere definicijska prvina je prizadevanje po stalnosti (pustiti za seboj pričevanje o obstoju).“ Korošec 1986: 393–394.

/z/lasti premalo pa vemo o razmerju obeh osnovnih jezikovnih plasti, o plasti govorjenega jezika in o pisanem jeziku. Obe plasti se med seboj močno razlikujeta, saj je pisanje pogojeno z drugačnimi zakonitostmi kot govorjenje. (Pogorelec 1965: 132)

Dandanes, ko je empirični pristop k jeziku korpusno jezikoslovje prignal spet k višku popularnosti (po prvem valu v 60. letih 20. st.), pisni korpusi pa so generacijo starejši in „za velikostni razred večji“ od govornih, bi lahko gornji citat razumeli v eksaktnem empiričnem duhu, da so bila pisna besedila *edini* vir občnih opisov jezika. Tega pa nihče ne bo trdil, niti ne za slovnico slovenskega standardnega („knjižnega“) jezika. Pisce slovnice je gotovo vodil tudi jezikovni instinkt oz. zmožnost in drugo v protipolu empirizmu, npr. v valu tvorbene (generativne) slovnice, in to k zajemanju iz govornega jezikovnega prenosnika, primarnejšega in priročajšega v osebni jezikovni izkušnji. Morda sicer ne poznamo razmerja med eno in drugo plastjo jezika, bodisi v jeziku kot *sredstvu za sporazumevanje* bodisi v podstatih slovnice. Zato pa nas dejstvo, da je jezik „preplet mnogih izraznih sistemov“ (Pogorelec prav tam) in „da različne oblike človeškega jezikovnega vedenja vendarle ustrezneje pojasnjuje *izhodiščna* jezikovna zmožnost /kot dve oz. več jezikovnih zmožnosti nekega jezika/“ (Stabej/Vitez 2000), napeljujeta k priznavanju zgolj ene *temeljne slovnice jezika*.

V prvi meri nam temeljna slovnica pomeni **temeljni „sistem jezikovnih sredstev in njihovih medsebojnih odnosov“** (SSKJ.: slovnica). (Odnosom v tem sistemu lahko rečemo tudi strukture.) Lahko pa temeljnost opišemo še drugače, ne le z vidika izraznega prepleta:

- k temeljnosti, izhodiščnosti oz. naravnosti nas vodi psihološka narava človeškega uma, ki teži k poenostavljanju, tipiziranju, partikularizaciji izraznih možnosti in s tem k strukturiranju jezikovnih sredstev (prim. t. i. načelo jezikovne gospodarnosti), in ki je hkrati te tipizacije sposoben prožno in konkretno uporabljati (Vitez 1999: 40);
- temeljen (nekonkreten) je predvsem splošen opis jezika, ki v abstrakciji preide od rabe (parole) k sistemu (langue);¹⁰
- temeljen (skupen vsem uporabnikom) je jezikovni kod zaradi interakcijske oz. socialne vloge v sporazumevanju, takšen kot je najbolj razširjen (gl. še opombo 53, str. 36);
- temeljna postane z normiranjem oz. standardiziranjem neka slovnica jezika v različne namene: socialne; ožje določeno identitetne, kulturne ...

Ko priznavamo eno temeljno slovnico jezika, lahko s pomočjo opisov¹¹ njene temeljne strukture izvajamo tako enega kot druge izrazne sisteme. Netemeljna jezikovna sredstva in strukture, lastne prav tem delnim (izraznim) sistemom, pa bi lahko imenovali in opisovali tudi kot diasisteme ali *delne slovnice*, „delne jezike, jezikovne sloge ali jezikovne zvrsti“, odvisno od predstavnih izhodišč (Pogorelec 2003: 26). V pričujoči raziskavi govornega jezika nas seveda zanima pozicija *izraznega* (prenosniškega) delnega sistema v odnosu do temeljne slovnice jezika, in ne delni sistemi za posamezne „oblike in teme glede na priložnosti komunikacije“. ¹² Če jezik pojmuje kot sistem jezikovnih sredstev, v katerem velja načelo funkcijske (komunikacijske)

¹⁰ Kot pravi druga razprava: „Opis fonološkega sistema je možen le na podlagi abstrakcije konkretnih spoznanj, ki jih omogoča opazovanje stvarnih pojavov v govornih uresničitvah. Samo vsled tovrstne idealizacije fonološkega sestava je možna natančna določitev nekontinuiranih, jasno zamejenih oblik, ki stopajo v opozicijski odnos z drugimi enakovrednimi prvini znotraj sistema. Vsakršna podrobna fonetična razčlemba pojavov v izvedbah govornih verig in njihovega delovanja v govorni komunikaciji pa spregovori v prid dejstvu, da je narava teh stvarnih izvedb že v temelju nasprotna shematizmu fonoloških struktur.“ Vitez 1999: 31. O tem bo še govora v naslednjih podpoglavjih.

¹¹ Kot opis jezika tu ne mislimo samo slovnice 'nabor urejevalnih možnosti jezika', ampak tudi slovar-tezaver 'nabor poimenovalnih možnosti jezika'.

¹² Le-take notranje diferenciacije jezika (socialne, funkcijske zvrsti) Pogorelčeva v slednji razpravi ovrže kot delne sisteme jih in razloži raje s teorijo zvrstnosti osmoze, „zlasti v govornih zvrsteh“. Po njenem se „notranja diferenciacija potemtakem odvija v samem procesu komunikacije in ob nastajanju konkretnega besedila.“ (prav tam). Z drugimi besedami, če netemeljne delne sisteme (ali notranje diferenciacije) jezika hierarhiziramo, so prenosniške zvrsti bližje temeljni slovnici jezika; sledijo jim socialne (zaradi relativne ustaljenosti družbenih institucij in odnosov, v katerih prihaja do sporazumevanja; prim. koncept „obrazov jezika“ Skubic 2005) in nazadnje funkcijske zvrsti.

raznovrstnosti, je koncipiranje socialnih in funkcijskih zvrsti kot delnih sistemov neustrezno (Pogorelec prav tam), prenosniške zvrsti pa pri tem pojmovanju obstanejo.

Če se torej osredotočimo na **delna izrazna sistema**, govorni in pisni jezik, nas zanima, katere *jezikovne strukture* so vsakemu sistemu lastne, tj. različne od lastnih jezikovnih struktur drugega sistema in hkrati različne od skupnih jezikovnih struktur. Kaj je različnega, pokaže le primerjanje posameznih strukturnih enot z unijo obeh izraznih sistemov, kar bomo pokazali v naslednjih poglavjih. Kaj je skupnega, pa med drugim razkrivajo podatki o tem, kaj je bilo iz (po)govornega jezika sprejetega v slovnice in slovarje slovenskega knjižnega (standardnega) jezika in tako postalo tudi struktura pisnega jezika (diahroni vidik).

Ta zgodba se ne začne šele pri tiskanih knjigah kot množičnih medijih, ampak že pri izboru **abecedne črkovne pisave** za slovenski jezik po splošni evropski kulturni tradiciji. „/M/odel korespondenc med fonemom in njegovim potencialnim zapisom (črko)“ namreč postavlja črko v položaj drugotnega znaka glede na jezikovne znake v govoru (Grgič 2006: 14). Kjer so se pisave razvile neposredno iz govora, odražajo fonetični sistem jezika – tako se razlikujejo *konzonantne* (npr. arabska) in *fonetične* (npr. cirilica) abecede. Kulturni vplivi pisav pa do neke mere v zapis govora vnašajo standarde – kljub različnim fonetičnim in prozodičnim sistemom med današnjimi jeziki je prevladujoče zapisovanje v latiničnem črkopisu.

T. i. alfabetska pisava pomeni povezan odnos govornega in pisnega sporazumevalnega koda na jezikovni ravni, ki jo običajno razvrščamo nižje od drugih sistemov pisav. Pisave z višjo osnovno ravni so: zlogovna (npr. japonska kana) in logo-zlogovna (kitajska pisava), ki ju še štejemo med *popolne pisave*, in *protopisave*: piktogrami (npr. eskimska pisava), rebusni simboli, glasbena notacija, matematični in znanstveni simboli ...¹³ Pisava kot celoten *nabor pisnih znamenj* pa ne obsega le črk, ampak tudi druga pisna znamenja, lastna le pisnemu jeziku. Ta se pojavljajo kot jezikovna sredstva na različnih jezikovnih ravneh, posamezno pisno znamenje v svoji konvencionalnosti in večpomenskosti lahko tudi na več ravneh¹⁴ – prim. npr. različne skladijske in neskladijske rabe ločil, številke in številke, oznake za odstavke in poglavja, oznake za razmerja \leq $<$ $|$ \Rightarrow \Leftarrow , tudi sestavljene $\leq \pm \geq | \neq | \subseteq \supset$ itd.

Podvržene osnovni znakovni strukturi pa prehajajo iz govornega v pisni sistem tudi enote jezikovnih ravnin, kompleksnejših od ravnin fonema in črke. Anton Breznik na začetku slovnice iz leta 1924 opisuje, da je „slovenski pismeni jezik nastal iz več narečij in je dobil sedanjo obliko v teku stoletij“. Najprej razlaga, da je osnova dolensko narečje 16. st., nato pa kronološko našteva sprejemanje „oblik“ (gre za različne razvoje vokalov, predpone in končaje, njihov zapis) iz gorenjskega, štajersko-koroških in vzhodnoštajerskega narečja, oblikovanje po zgledu srbohrvaščine, in zaključí: „Pismeni jezik sloni torej na glasovju in oblikovju, ki živi od 16. stoletja dalje, t. j. od začetka našega slovstva, in se nahaja v vseh glavnih slovenskih narečjih.“ Pred tem pa: „Ker je naš pismeni jezik proizvod dolgotrajnega razvoja, ni enovit in dosleden, temveč je sestavljen iz različnih sestavin in se v celoti kot tak ne govori v nobenem našem narečju in se v tej obliki tudi v zgodovinski dobi, t. j. v 16. veku, ni govoril.“ (Breznik 1924: 1–5, poudarka sta njegova).

O skupnem obeh prenosniških zvrsti na leksikalni ravni lahko po podatkih v zadnjih dveh večjih slovenskih slovarjih sklepamo samo posredno, saj sta oba zajemala iz pisnega prenosnika.¹⁵ SP²⁰⁰¹ v Pravilih sicer omenja prenosniške zvrsti (§ 1039 oz. 1044), a na njihovi podlagi v Slovarju ne prisojajo posebnih oznak; da bi na podlagi socialnozvrstnih oznak presojali govorni izvor, pa se zaradi neporočanja o izpisovanju zdi nesmiselno. SSKJ pa pojasnjuje, da so „posebnosti pogovornega jezika in žargonov ter narečno besedišče v slovar vključene, če so močnejše zastopane v pisanem knjižnem jeziku“ (Uvod, § 2). SSKJ je torej na prvi plati izbora

¹³ Vzeto po Robinson 1999. T. i. popolne pisave avtor klasificira tako: zlogovni sistemi (linearna B, kana – japonska, čerokeška), logo-zlogovni sistemi (sumerska, kitajska, majska), logo-konzonantni sistemi (egipčanska), konzonantne abecede (feničanska, hebrejska, arabska), fonetične abecede (grška, latinska, finska), logo-fonetične abecede (angleška, francoska). Slovenska je verjetno tu nekje med zadnjima dvema razredoma.

¹⁴ V primeru posamezne aktualizacije seveda le na eni ravni.

¹⁵ Le SSKJ pravi, da so zelo redke besede, za katere ni izpisov, vzete iz Pleteršnikovega slovarja, Slovenskega pravopisa (1962) ali zapisov urednikov (Uvod, §11) – med njimi bi utegnil kak vir biti govorjen.

gradiva zasnovan prenosniško (prav tako kot gradivo te diplome) in s tem zajema leksiko iz vseh socialnih in funkcijskih zvrsti, a z izborom medija (knjige oz. tiska kot množičnega medija) zajame več leksike, ki je uresničena v standardni socialni zvrsti jezika. Pri SSKJ moramo tudi paziti, da 5 „stilno-plastnih kvalifikatorjev“ (nar., otr., pog., nižje pog. in žarg.; Uvod, § 137–141), za katere moremo sklepati, da naznačujejo izvor besed v govornem prenosniku, vzamemo zgolj kot oznake za socialno zvrst. Vsa ostala leksika, ki nima katerega teh petih kvalifikatorjev, se namreč prav tako lahko pojavi v govornem prenosniku.

Diahroni vidik nam pokaže, kako je nastal slovenski „knjižni/pismeni/standardni“ jezik kot neka vrsta temeljne slovnice: s prepletom izraznih sistemov. Ne obstaja torej temeljna slovnica sama po sebi, saj brez izraza sploh ne more obstajati; če obstaja v enem jeziku en sam izrazni sistem, je le-ta hkrati temeljna slovnica, če pa je izraznih sistemov več, se temeljna slovnica razvije postopno kot njihov preplet, izrazljiv tako v enem kot drugem delnem sistemu, ki živita hkrati sama zase v svojem izrazu (prenosniku) in povezana preko temeljnega sistema. Če pogledamo konkretno pri slovenskem primeru, najprej govorno plat: pokrajinski jezik je preplet narečnih izraznih sistemov. Dokler nastaja le v govoru, je sproten in razpada z zaključkom vsake situacije.¹⁶ Knjižni zapisi so nastajali bodisi narečno bodisi pokrajinsko, bodisi pa umetno nadpokrajinsko, kar je pripeljalo do fiksirane (vse)slovenske slovnice. Preplet govornega (ki je sam preplet več narečnih sistemov) s pisnim jezikom (ki je prav tako preplet več pokrajinskih knjižnih jezikov), je pripeljal do nastanka standardne temeljne slovnice, ki jo imenujemo **slovenski knjižni jezik**, predvideva pa svoj *pisni* in *govorni izraz*.¹⁷ Njena struktura je odraz različnih zgodovinskih in sodobnih izraznih sistemov (govornih narečnih in pokrajinskih; pisnih – tudi različnih) in ni enovita.

Primer 1: Mešanje vplivov pisnega in govornega izraznega sistema ob prevzemanju besed

Še dandanes se kaže mešanje vplivov pisnega in govornega izraznega sistema – lep primer je prevzemanje besed iz tujih jezikov, njihovih govornih in pisnih izraznih sistemov s slovenskimi, npr. zapisa *anžambmá* ali *enjambement*, pri obeh standardni izgovor [anžambmá] ali pri drugem nestandardni [enjambement], ki ga izpričujem iz lastnih izkušenj. Prim. še iztočnici obe v SP²⁰⁰¹.

Ob dveh temeljnih izrazih slovenskega knjižnega jezika kot temeljne slovnice, pisnem in govornem, ne smemo pozabiti na druge znakovne sisteme prenosnikov, ki običajno spremljajo sporazumevanje. *Kretalni jezik* (ne le kod za sporazumevanje gluhih, ampak vsa mimika, gestika in kinezika), verjetno najprvotnejši izmed vseh jezikovnih izraznih sistemov, spremlja govor v neposrednem vidnem stiku. *Slikovni jezik*, običajneje poimenovan grafično ali ilustrativno gradivo, pa je podobno neločljiv od pisnega jezika. Vendar ugotovimo, če vse štiri doslej naštete izrazne oz. znakovne sisteme primerjamo z vidika *izrazne moči strukturiranja pomena*, da najabstraktnejše in najbolj nedvoumno sporazumevanje omogočata ravno govorni in pisni jezik, najbolj prepletena in strukturirana izrazna sistema.

V tem poglavju smo torej ugotovili, da je govorni jezikovni sistem odražen v pisnem jeziku in da se to da razbrati tudi v splošnih sistemizacijah jezika oz. temeljnih slovnica; navedli

¹⁶ „Menim namreč, da je pogovorni jezik zvrst, ki jo je sicer mogoče – a zaradi njene spremenljivosti in velike raznolikosti zelo težko – opisati, vsaj kako njegovo varianto, nikakor pa je ni mogoče tudi predpisati. Tako kot narečje se namreč tudi pogovorni jezik oblikuje in predaja drugim govorcem le s sporočanjem, s sporazumevanjem v govornih okoljih, ki to jezikovno zvrst zahtevajo oziroma dopuščajo, je zvrst, ki se realizira skozi posamezna govorna dejanja oziroma se (sicer po določenih pravilih) oblikuje vedno na novo. Prav to pa je gotovo vzrok dvema skrajnima pogledoma na pogovorni jezik v slovenskem jezikoslovju – od (začetnega) prepričanja, da slovenskega pogovornega jezika sploh ni, do hotenja, natančno popisati in celo predpisati ne le pravila rabe pogovornega jezika, ampak tudi njegov sistem, njegovo izrazno plat na vseh jezikovni ravninah.“ Škofic 1991: 9.

¹⁷ Tudi prenosniški vidik jezika je torej standardiziran. Pravzaprav se standardizacija tiče le t. i. „zbornega“ jezika; ostale (socialne) zvrsti so v slovnica in jezikoslovnih delih zgolj opisovane, tudi kot jeziki dajalniki, le izjemoma pa tudi poučevane v didaktičnih priročnikih (npr. za učence slovenščine kot drugega/tujega jezika).

smo nekaj primerov. V naslednjem poglavju bomo skušali pregledati, katera govorna izrazila lahko v slovenščini transkribiramo s slovensko ortografijo, kje pa nam to spodleti.

2.2. Govorni znaki in pisava – semiotični vidik

Govorno gradivo te naloge želimo zapisati predvsem zato, da si olajšamo dve stvari: analizo samega jezikovnega koda in primerjavo s pisnim jezikom. Prvo bi bilo namreč težko, če bi ostali pri zvočnih posnetkih ali celo pri avdiovizualnih posnetkih, drugo pa sploh nemogoče. Pa tudi metajezikovno ukvarjanje z jezikom potrebuje en svoj izraz; za znanost je ustaljen¹⁸ in funkcionalen prav pisni izraz (iz razlogov, podobnih onim, po katerih se je razvila statusna prevlada pisnega prenosnika v standardnih slovnica, gl. opombo 9, str. 13). Prva prednost je preprostost fizičnih plati pisnega prenosnika: gostejše posredovanje informacij v trenutku in časovna večrazsežnost reprodukcije (pri zvočnem prenosniku je reprodukcija kot pri produkciji časovno linearna), možnost obdelave brez strojne opreme, pri računalniški strojni opremi enostavnejše elektronsko kodiranje, manj zasedanja pomnilnika, enostavnejše iskanje in premikanje po jezikovnih enotah ...

Oboje nas vodi k izboru take vrste zapisa (transkribcije), ki izhaja iz temeljne slovnice, – temeljnega ali „primerjalnega izraza“.¹⁹ Toda kako moramo uporabiti slovensko ortografijo kot primerjalni izraz, jo razločiti od vloge, v kateri določa zgolj in samo pisni izraz temeljne slovnice? Jo naj uporabimo ožje kot fonetični črkopis 25-ih znakov ali širše kot pravopis, ki definira rabo črk in drugih pisnih znamenj, zapis glasov in oblik vse do nedvoumnega izražanja pomena leksema bodisi v besedi bodisi v besedni zvezi? Ali še širše, skladenjsko, ker to vpliva na oblikoskladenjsko določitev vsake aktualizirane besede oz. pojavnice v gradivu?

Ker je naše izhodišče prenosniški vidik jezika in izrazno različno koncipiranih sistemov ne moremo primerjati neposredno, bomo „primerjalni izraz“ izvedli iz abstrakcije in strukturiranja obeh sistemov: iz primerjave izraznih enot govornega in pisnega sistema. Strukturalna umestitev enot mora pri tem izhajati iz narave sistema samega in ne sme biti iznajdena zaradi olajšanja jezikoslovnega opisa.

2.2.1. Osnovna in pomožna izrazna sredstva

Osnovne izrazne enote (oz. sredstva) v semiotičnem duhu imenujemo *pomenskorazločevalne*, kar pomeni, da jim s pomočjo opisa izraznorazlikovalnih lastnosti (angl. distinctive features) v razmerju do jezikovnega sistema ne moremo (ne želimo) pripisati manjše vloge pri oblikovanju pomena jezikovnega sredstva. Enostavno rečeno: koncipiramo jih kot (pomensko) nedeljive, zato nas ne zanima nadaljnja *sestavljenost* njihovega izraza (oblike), temveč samo njihova *razlikovnost*. Taki so fonemi oz. črke in glasovi kot relativno idealne enote, ki jih lahko enostavno sestavljamo v višjo raven, besedno.²⁰ Strukturalizem s pomočjo koncipiranja *osnovnih* enot (idealno enotnih v prekrivanju razlikovanja pomena in oblike) tako gradi jezikovne

¹⁸ Poleg tega smo v zapis prisiljeni zaradi ustaljene oblike diplomskega dela – pisne oz. knjižne.

¹⁹ Tudi če želimo kasneje besede transkribiranega gradiva besednovrstno in oblikoskladenjsko določiti s pomočjo strojnega določevalnika, ki določa na podlagi učnega korpusa tako knjižnih kot transkribiranih besedil, smo obvezani izbrati slovensko ortografijo, pravopis.

²⁰ **Beseda** pa je idealna, središčna enota strukturalnega pogleda na jezik, idealni *jezikovni znak*, v svoji *virtualnosti* (uresničeni v za-vesti jezikovnega uporabnika in v slovarju jezika) sestavljena iz oblike/zaznamujočega in pomena/zaznamovanega, po teh dveh kriterijih pa tudi v *aktualizirani* uresničitvi umeščena v sistem: po prvem sintagmatsko in po drugem paradigmatsko. Natančnejše opredelitve aktualizacij tega zbirnega pojava so recimo: leksem, besedna zveza, izrazna beseda, grafična beseda, naglasna enota, besedna oblika, (prosti besedni) morfem ... (Tudi vsaj nekatere morfeme bi lahko uvrstili pod besedo, čeprav le na njeno „spodnjo“ mejo; nenazadnje SSKJ in SP²⁰⁰¹ med iztočnice uvrščata tudi nekatere predpone.)

ravnine (fonemska, morfemska, leksemska, sintaksemenska, tekstemska), kot preplet oblike in pomena.²¹

Tukaj izhajamo iz oblike (oz. izraza, zakodiranja prenosnika) in *vseh* njenih aktualiziranih enot, ne le idealno koncipiranih, in jih poskušamo razvrstiti na jezikovne ravnine glede na raven oblikovanja pomena. Strukturalistični model razvrstitve jezikovnih enot pa omaga pri razvrstitvi izraznih pojavov, kot so:

1)) *enote* ali *vrednosti* prozodijske in stavčne fonetike (npr. naglasi, hitrost govora, izdišna enota) v govornem prenosniku ter

2)) nečrkovna pisna znamenja, grafični sklopi ali grafične vrednosti (npr. ločilo, vrstica, velikost pisave, smer pisanja, animacija napisa) v pisnem prenosniku.

Glede na osnovne enote bi ta izrazna sredstva lahko imenovali tudi **pomožna izrazna sredstva** (ne moremo govoriti le o enotah, ampak tudi o relativnih vrednostih; tako uporabimo zbirni pojem sredstva).

Lahko bi pomožna izrazna sredstva označili tudi za parajezikovna²² oziroma natančneje **paraprenosniška**: niso sposobna strukturalizacije kot osnovni jezikovni znaki v posameznem kod, hkrati pa vendarle prispevajo k oblikovanju pomena in so del prenosnika/kanala.

Tabela 1: Umestitev jezikovnega sredstva z vidika prenosnika in koda (pomenskorazločevalna raven)

Je znak <i>jezikovno</i> sredstvo?	jezikovno (alo)fon, črka, vzključna intonacija, velika začetnica, mašilo, skomig z rameni, zvočni signal za črko j, reliefni del znaka vsote na bankovcu		nejezikovno (nepomenljivi) šum
Pripada posameznemu konceptu <i>prenosnika/kanala</i> ?	prenosniško npr. govorni prenosnik: (alo)fon, tonem, premolk, ritem, dolžina		neprenosniško (pripada drug. prenosniku) ločilo
Ga lahko umestimo v <i>posamezen znakovni kod</i> ? ²³	kodno OSNOVNA IZRAZNA SREDSTVA (ENOTE)	paraprenosniško in parakodno POMOŽNA IZRAZNA SREDSTVA	nekodno (pripada drugemu naboru znakov)

Relevantno je govoriti tudi o **jezikovni redundanci**, saj se zdi, da prav paraprenosniška jezikovna sredstva tvorijo v sporočilu, kar se z vidika osnovnih enot zdi *dodatna, ne nujno potrebna* (pomen besede redundanten) informacija. Nikakor pa niso odvečna, saj pripomorejo k „celostni koherenci sporazumevanja“, h kateri težijo s prehajanjem med jezikovnimi ravninami (gl. Vitez 1999: 28–29, ki te vidike povezuje malo drugače). Odvečen je morda zapis stavčne fonetike s

²¹ Dihotomnost oblike in pomena velja seveda le v **virtualnosti**, abstraktnosti (v procesu sporočanja v zavesti tvorca in naslovnika), v **aktualiziranosti** (rekli bi, ko je vzpostavljeno referenčno razmerje) pa ne. Prim. mnenje, ki želi prestopiti ta strukturalistični vidik: „Fonetika in fonologija si prizadevata znanstveno dosledno pokazati, da govorni prehod od fonemskega ideala h glasovni stvarnosti – in nazaj – pravzaprav ni posledica čudežnega prepleta duše in telesa, v katerega lahko verjamemo, razumu pa ni dosegljiv. Razložiti ga skušata prav tako, kakor fizika, fiziologija in nevrologija pojasnjujejo vse druge pojavnosti človekovega gibanja.“ Vitez 1999: 31. To počnemo tudi tukaj, kasneje, ob prehodu v pragmatični pogled na jezik (poglavje I.3), ki vključuje relevantno zunajjezikovno stvarnost – sporazumevalne okoliščine oz. dejavnike.

²² Paziti moramo pri podomačevanju tujk **verbalnost/paraverbalnost/neverbalnost** (tudi nonverbalnost). Podomačevanje lahko izvajamo iz verbalen=jezikoven ali verbalen=beseden. Predpona *para...* pri tem pomeni 'nepravi, podoben, približen' (SSKJ), torej je točna opredelitev odvisna tudi od znanstvenosistemskega konteksta. Za primere prvega izvajanja podomačevanja gl. Tabela 1, str. 18, za primere drugega pa poglavje I.2.1.1.

²³ Toporišič 1992: ESJ: 349–350: **vez**: „Sestav sredstev, s pomočjo katerih lahko kaj odrazim, s čimer čemu lahko damo drugotno podobo, podobo sporočila. Tak vez je tudi jezik, saj v njem vsemu, kar je, lahko damo posebno, sporočilno obliko obstajanja, tj. obliko besedila. – Sopom. kod.“ Kaj vse sestavlja nabor, pa ni odvisno od prenosnika, marveč je antropogeno (kulturno) pogojeno oz. dogovorjeno – v tem smislu lahko kodu damo sopomenko jezik (nor. slovenski vs. nemški, ena vs. druga socialna zvrst). S podreditvijo prenosniku je koncept bistveno omejen; lahko bi ga imenovali tudi *nabor (jezikovnih) znakov* ali *govorica* (v prenesenem pomenu, z obveznim prilastkom: *govorica kretenj*; ne pa tudi širše: *govorica arhitekture*).

posebnimi pisnimi oznakami, ki jih v ortografski transkripciji nismo vajeni, smo pa vajeni raznih ločil (če bi recimo slušni predlogi odvezli vse naglasne razlike in intonacijo oz. informacijo o njih, bi to izgledalo podobno kot tekst brez presledkov in ločil).

Preden gremo na umeščanje govornih in pisnih med bodisi osnovna bodisi pomožna izrazna sredstva, še razmislek, kako obsežna je pravzaprav vsaka posamezna množica, iz katere izhajamo. Oznaka *govorna izrazna sredstva* nam pomeni veliko nehomogeno množico jezikovnih sredstev (nehomogeno s socialno- in funkcijskosporazumevalnega vidika notranje diferenciacije jezika). V različnih govornih situacijah so govorna izrazna sredstva različno popolnena, bližje ali dlje od temeljne slovnice, jezik pa bolj ali manj notranje diferenciran. Prim. recitiranje po spominu, branje ali petje, govor po radiu ali televiziji, govor simultane prevajalca, razlaganje oz. predavanje (po predlogi ali brez nje), pogovor po video zvezi ali telefonu, pogovor v ustaljenih situacijah (med sodelavci, med uradniki in njihovimi strankami, med znanci), spontani govor v manj predvidenih ali v nepričakovanih situacijah (na izpitu, med neznanci), samogovor med glasnim učenjem, pogovor človeka samega s sabo itd. Nikakor ne menimo le standardizirane „enotne izreke“ (Rupel 1946: 4), kot jo opisuje pravorečje ali pravopis (nabor razločevalnih glasovnih enot in njihovih variant, jakostni in tonemski naglas, posebne glasovne zveze, SP²⁰⁰¹ § 615–745; oblikoslovni naglas § 896–920). V opisovanju realnega jezika lahko približke zborni izreki sicer najdemo (v situacijah, naštetih zgoraj na začetku), najdemo pa tudi nezborna izreka in kompleksnejše govorne enote, ki jih normodajni priročniki prepuščajo retoriki (Rupel prav tam) ali previdno nakazujejo (SP²⁰⁰¹ § 484).²⁴

Še manj običajno pomislimo na odvisnost *pisnih izraznih sredstev* od različnih pisnih situacij, saj pisno sporazumevanje običajno enačimo s standardnim sporazumevanjem (tu ima slab vpliv poimenovanje slovenskega standardnega jezika „knjižni“ jezik). Še v tiskanih ali tele-*vizijskih* množičnih medijih (tudi internet) zadnje čase ni prisotno le pisanje po pravopisu.²⁵ V *rokopisih* pa že od nekdaj ni tolikšne bližine standardnemu sporazumevanju in so zaradi neomejenosti pri oblikovanju tako sporazumevalno najpestrejša množica pisnih besedil (čeprav so tudi tu pogoji, ki standardizirajo pisanje, npr. vrstice v zvezkih, velikost listov, prepisovanje, omejitve glede debeline in barve pisala, materialov podlage in pisne sledi itd.).

2.2.1.1. Pisna izrazna sredstva

S primerjanjem različnih pisnih in govornih sporazumevalnih situacij želim opozoriti na razmerje vsakega delnega izraznega sistema do temeljnega sistema standardiziranih izrazil. Le *na prvi pogled* se namreč zdi, da je pisava nasploh kot naravni (realni) znakovni sistem preprostejša od govora; npr. pri osnovnih izraznih enotah: da je sistem črk enostavnejši kot sistem fonemov, le-ta pa enostavnejši od množice (alo)fonov nekega jezika (oz. katere njegove govornice). Povedano v terminologiji teorije informacij, zdi se, da je **entropija** črkovnega koda nižja kot entropija glasovnega koda (znakovnega sistema). Podobno naj bi verjetno veljalo tudi za ostala pisna znamenja (ločila itd.), torej da z vsakim *zapisovanjem* govora (ne le z ortografskim) kodifikacijo poenostavljamo.

²⁴ Rupel 1946: 4: „ne bi mogli pohvaliti predavatelja ali govornika, ki ne bi obvladal tehnike glasu, ne bi pazil na dinamiko in melodijo, na kadenco stavka. Vse to pa ne sodi v pravorečje, temveč je del retorike“. SP²⁰⁰¹ § 484: „Stavočnofonetični ustrezniki ločil: Gledano z druge strani /kot pri nadrejenem naslovu Izražanje slušnih pojavov z ločili, § 476/, lahko rečemo, da posameznim ločilom v skladijski vlogi *dokaj* ustrezajo značilni tonski poteki.“ (Poudaril R. B.) Dalje se piše o tonskih potekih, ki jih zaznamujejo posamezna ločila: kadenci, padajoči in rastoči polkadenci, antikadenci. Prim še Toporišič 1992: ESJ: 308: **stavčna fonetika**: „Stavočnofonetična dejstva *deloma* nakazuje pravopis, zlasti z rabo ločil.“ (Poudaril R. B.)

²⁵ Pa vendar se je delež strojno obdelanih besedil zaradi osebnih računalnikov in njihovih programov za obdelavo besedil močno povečal, tudi črkovalniki in samopopravki v zadnjem času popravljajo stopnjo spontano tvorjene standardnosti besedil (Stabej 2003: 7).

To pa ne drži. Le temeljni sistem standardiziranih, osnovnih izrazil (pri transkribiciji pisnih) je enostaven: ker je abstrahiran oz. idealiziran oz. tipiziran in kot tak v *absolutni* poziciji nasproti *relativnim* konkretizacijam v različnih sporazumevalnih situacijah (našteli smo jih zgoraj).²⁶ To razmerje pa ne velja le za višje jezikovne ravnine (npr. idealna enota – poved v pisavi ali v govoru ujemanje popolne skladijske enote s kadenco in zamejenost z daljšimi premori), ampak se notranja diferenciacija dogaja tudi na ravnini pomenskorazločevalnih izrazil. Saj standardni opis ne predvideva, ali bo npr. črka <a> pisana ali tiskana, krepka ali ležeča, temveč da pomeni določeni znak za razločevalni glas in da je lahko v določenih situacijah velika, v drugih pa mala (začetnica) itd.

Kvantitativni vidik je odvisen od tega, katere znake in vrednosti vse „vzamemo v račun“ kot *enote* jezikovnega znakovnega sistema, saj se s tem izračun entropije spreminja. Možni pisni nabori: 23 črk klasične latinice (brez w, v in j), 25 črk slovenice, lahko jim dodamo 10 arabskih števk, pa ločila, ki imajo možnost skladijske rabe (teh je v slovenščini vsaj 11, če ne štejemo različnih vrst oklepajev in narekovajev, SP²⁰⁰¹ § 228). Nadaljnja diakritična znamenja lahko definiramo kot znake zase ali pa kot kombinacije z osnovnimi črkami (drugo je pravilneje),²⁷ podobno lahko naredimo s tipografskimi variantami (majuskula, minuskula, debelo, ležeče in druge, pa še rokopisne variante). Ko prištejemo nabore znamenj iz raznih jezikoslovnih tradicij in drugih strok, postaja množica multifunkcijska in neenovita.

Podobno je v mednarodnih kodiranih naborih znakov, ob katerih se nam množica fonemov ali glasov pravzaprav zdi preprosta: ASCII (American Standard Code for Information Interchange, jugoslovanska različica je bila YUSCII) je imel samo 128 znakov – po 26 velikih in malih angleških črk, 10 števk, 33 izpisljivih in 33 neizpisljivih krmilnih znakov (Peterlin idr. 1998). Trenutno še aktualni srednjeevropski ISO 8859-2 je velik dvakrat toliko (8-biten, $2^8 = 256$ znakov), ISO 10646, ki ima načrt, da globalno zajame pisne znake, je pa že nepredstavljivo velik (32-biten, $2^{32} = 4.294.967.296$ znakov).

Ker želimo govor transkribirati z ortografsko pisavo in na računalniku, smo seveda vezani na določen kodirni nabor znakov, kar pa v primerjavi z rokopisom pomaga pri standardizaciji in poenostavljanju zapisa. Seveda ne bomo uporabili vseh znakov v naboru, ampak le izbrane, ki nam bodo služili kot že omenjeni temeljni oz. primerjalni izraz (katere znake in kateri mednarodni kodirni nabor, glej Tabela 12, str. 89).

Važna torej ni le kvantiteta, ampak tudi *strukturiranost nabora* za potrebe konkretnega jezika in znanstvenega koncepta jezikovnega opisa; odnosi, s katerimi dodatno opišemo znake in njihove vloge pri sporazumevanju. Še vedno smo pri pisnih izraznih sredstvih. Zaradi drugega prenosnika osnovna izrazna enota pomenskorazločevalne ravnine očitno ni *fonem*, bi pa lahko to enoto vzporedno fonemu poimenovali **grafem**, seveda z na novo koncipiranim podpomenom 'črka ali pisno znamenje, ki v danem jeziku razločuje pomen besed ali morfemov'; ne gre namreč samo za črke kot znake za foneme.²⁸

²⁶ Podobno tudi npr. digitalni sistemi v primerjavi z analognimi, binarni številski sistem v primerjavi s šestnajstiškim, 7-bitno kodiranje nabora znakov v primerjavi s 16-bitnim itd.

²⁷ Dejstvo je, da so pisave, ne glede na stopnjo fonetičnosti, podrejene tradiciji oz. kulturnemu vplivu izhodiščnega (primarnega) črkopisa in njegovega fonetičnega sistema. Morda je banalno dejstvo, ampak ni nujno, da bo fonem drugega jezika s sekundarno uporabljeno fonetično pisavo zapisan z enim (svojim) znakom. Primer slovenščine – latinica v sekundarni rabi kot slovenica: fonem nima zapisa oz. je ga je mogoče razbrati iz neenovitega zapisa z drugimi znaki (npr. /ə/), lahko je zapisan kot dvočrkje (npr. <dz> za /dʒ/) ali mu je prirejen drug znak s pomočjo diakritičnega dodatka (npr. s + ˇ = <š> za /ʃ/). Mednarodna fonetična abeceda, IPA, tako izhaja iz t. i. središčnega vokalnega sistema in pri konsonantih iz razlikovanj v latinici (in romanskih jezikih), za artikulacijske variante pa uporablja diakritične znake (Matthews 1997: 183). Fonetična OLA-abeceda pa ima drugačen, izhodiščno bolj slovanske fonetične posebnosti razločevalni črkopis.

²⁸ V kolikor se ta definicija tiče črk kot znakov za foneme v primeru standardne slovenščine, ki ima 25 črk za 29 fonemov, lahko v njo podvomimo. Če je namreč razločevalna moč za pomene morfemov in besed primarno definirana za foneme, se razločevalnost pri prenosu v nesorazmeren sistem znakov (črk) podre. Te probleme mora reševati pravopis. „Knjižne izreke pa nam naš pravopis ne more dati, ker je, kakor smo videli, zgodovinski in etimološki, vrhu tega pa s svojim črkopisom *pomanjkljiv*.“ (Rupel 1946: 18; poudaril R. B.) Prim. še Toporišič 1992:

Merilo sistemske umestitve para-grafemskih znamenj, ki so bolj problematična za umeščanje, pa ni le *razločevalnost izraza*: na eni strani, da bi bila vsaka najmanjša grafična poteza svoj znak (kljukica, pika, tipografska inačica, soroden znak v drugače konceptualiziranem fonetičnem črkopisu, npr. <dž> po OLA ali <ɟ> po IPA) ali na drugi strani, da bi vsak element imel le eno pomenskorazločevalno funkcijo (npr. skladijska ločila). Merilo ravninskega strukturiranja teh sredstev je prav tako **razločevalnost pomena**, po njem pa jih lahko umestimo glede na njihovo razločevalno in sintagmatsko moč višje ali nižje od osnovne ravnine jezikovnega sistema.

Na osnovni grafemski ravnini (in hkrati *osnovne izrazne enote*) so tako črke in nekatere neskladijske rabe ločil.²⁹ Nižje od osnovne ravnine so znaki, ki se vežejo na posamezna znamenja (diakritična znamenja kot ločeno znamenje), še nižje tipografske variante in posebna prostorska razporeditev znakov (npr. nadpis). Malo višje od osnovne ravnine so ligature in znamenja za skupine znakov (spet diakritična znamenja), višje nato logografska znamenja (med njimi najvišje tista, ki se ne zmorejo povezovati), morda še višje slovničnorazmerna znamenja, ki lahko stojijo le na določenih mestih (velike začetnice pri večbesednih lastnih imenih in na začetku povedi, ločila v skladijski rabi).

Med vsemi grafemi imajo črke največjo zmožnost povezovanja (sintagem) in tako izražanja pomena na višjih ravninah (od morfemov oz. besed dalje), je pa ta lastnost odraz prvotnejšega glasovnega strukturiranja pomena, kot smo že povedali zgoraj. Nečrkovnim znamenjem na višjih jezikovnih ravninah, od besede dalje, vendarle ne moremo pripisati besednosti (verbalnosti), saj njihova pretvorba v pomen oz. besede, ubesedenje, ni enoznačna; takšna so kot druga *pomožna izrazna sredstva*. Tako tudi Toporišič pravi, da „/s/tavčna fonetika in ločila so v precejšnji medsebojni zvezi,“ ni pa popolnega vzporednostnega enačaja. In dalje: „Kakor torej ni res, da bi vprašaj zmerja izražal vprašalno intonacijo /.../, prav tako ni res, da bi klicaj vedno kazal velelno glasovno valovanje.“ (Toporišič 1967: 172).

2.2.1.2. Govorna izrazna sredstva

Enak strukturalni pristop kot za pisna izrazna nas zanima tudi za izrazila govornega prenosnika. Realen opis govornih enot je odvisen od gradiva in mere, do katere jih raziskovalec želi razlikovati od ostalih glasov (bodisi preko opisa artikulacije, bodisi sistemsko na podlagi raznih lastnosti). Opis fonemov je vedno „temeljen“ – odmaknjen od konkretnega (gl. navedka v opombi 10 na str. 14 o idealizaciji fonološkega sestava in v opombi 28 na str. 20 o fonemu).

Za lažjo predstavo vzemimo metodo izdelave t. i. fonološkega opisa (narečja) v strukturalni dialektologiji:

- 1) Iz zbranega gradiva se sestavi maksimalen fonemski inventar; to pomeni, da se popiše vse glasove, ki so fonemi (imajo pomensko razločevalno vrednost), opiše se tudi njihovo fonetično vrednost;
- 2) foneme se razvrsti v grafikone, tj. samoglasniške tri-/pravokotnike, soglasniške preglednice, doda še fonetične variante /.../, shematsko se prikaže naglas (vrsta in razporeditev);
- 3) sledi distribucija (razvrščenost) fonemov in prozodemov, opis omejitev pri distribuciji in navedba fakultativnih (prostih) in pozicijskih (položajnih) variant;

ESJ: 41: **fonem**: „Glas (ali skupina glasovnih različic/variant), ki nasproti drugim glasovom (skupinam različic) v danem jeziku razločujejo pomen besed ali morfemov“. Glej še primere v naslednji opombi.

²⁹ Termin „(ne)skladijska raba /ločil/“ povzemam po SP²⁰⁰¹, Pravila § 226–229. Osnovno raven (grafem) tako določa raba „sredi“ besede, višje ravni pa neskladijske rabe „na koncu povedi ali besede, vprašaj ali klicaj pa tudi ob robu besedila“ (prav tam, § 229). Po tem načelu razločimo npr. **neskladijske rabe pike**: Namesto vsake izpuščene črke v besedi ali številke v številki (§ 258) in med deli številke, da loči enote različnih stopenj (§ 255, 261) je pika kot osnovna raven grafema. Na višjih ravneh je pika za okrajšavami posameznih besed (npr., str., dr., gl.; § 249) in za številkami za zaznamovanje vrstilnih številnikov (§ 251, 252, 254) – ti dve rabi sta konkretno morfološki; na besedno raven pa lahko uvrstimo rabo pike kot znamenje množenja „krat“ – v tej rabi je pika celo nestična (§ 255), podobno kot grafične besede. Na nižji ravni je pika kot diakritično znamenje: npr. nad črkama i in j ali pod e-jevskimi in o-jevskimi vokali za zaznamovanje ozkosti po OLA-transkribiciji.

4) poda se izvor fonemov in prozodemov /.../;

5) poda se tudi izguba, izpad, popolna redukcija glasov (V in K). (Smole 2000: 6–7.)

Prozodične lastnosti v fonoloških opisih so:

- za zlogotvoren *glas* (običajno samoglasnik) oz. v primerjavi *zlogov* v eni *besedi* ali *med besedami* (lahko v besedni zvezi) dolžina ali kolikost (enota kronem),³⁰ ne-/naglašenost, jakost in ton oz. vrsta naglasa (jakostni, tonemski), in količina naglasov (na besedo), mesto in okoliščine naglasa;

- za *stavek* oz. *poved* in dele besedila (stavčna fonetika) členitev s premori, poudarek ali jakost, intonacija ali tonska višina, register (tonski pas), hitrost ali tempo, glasovno barvanje (Toporošič 1992: ESJ: **fonološki sestavi, prozodem, prozodija, stavčna fonetika**).

Razporeditev govornih enot povzamemo po fonologiji, ki ima tako v slovenskem jezikoslovju kot v vseh strukturalističnih šolah mnogo večjo tradicijo kot strukturalna zavest o pisnih znamenjih.³¹ Govorne enote oz. fonetične pojave v pomenskorazločevalni vlogi fonologija običajno razdeli po glasovnih sestavih kot **foneme**, ki ji pomenijo *osnovne segmentne enote*, sposobne razločevati pomen morfemov in besed s tem, da jih gradijo, in po prozodičnih sestavih kot **prozodeme**, ki ji pomenijo dodatne, neosnovne lastnosti, dodane raznim jezikovnim enotam/sintagmam: glasovom, zlogom, besedam, besednim zvezam, stavkom, povedim.³² Prozodične lastnosti so tradicionalno pripisovane enotam, večjim od segmentov (osnovnih enot, običajno glasov), zato tudi *nadsegmentne* lastnosti, bolje pa bi bilo poimenovanje „*parasegmentne*“ zaradi nekaterih lastnosti, pripisanih že zlogotvornim glasovom. Zanimivo z vidika ravninske umestitve je tudi vprašanje, katere fonetične/prozodične pojave je možno uslovariti: očitno vse, kar se tiče naglasa in posameznih glasov, in nič, kar se tiče stavčne fonetike. So pa fonetična sredstva oz. vrednosti, ki imajo vlogo na obeh od pravkar razločenih stopenj: jakost (jakostni naglas vs. poudarjanje besed v stavku), tonskost (tonemski naglas vs. kadenca in register), dolžina (glasov vs. hitrost ali tempo).

Pravzaprav se koncipiranje pomenskorazločevalnih govornih lastnosti in s tem gradbenih enot na različnih ravneh razlikuje od jezika do jezika, zato meja med osnovnimi in pomožnimi govornimi izraznimi sredstvi ni tako jasna, če fonološkega strukturiranja ne izvajamo dosledno samo za npr. slovenščino.³³

Problem sistemiziranja izraznih sredstev (tako osnovnih kot pomožnih) je v tem, da neka fizično opisljiva razlikovalna lastnost velja (se pojavlja) samo pri nekaterih sredstvih od teh, ki jih uvrščamo v isto ravnino – ki jim pripisujemo enako vlogo razločevanja/oblikovanja pomena. Gre

³⁰ Toporošič (SS²⁰⁰⁰: 59) pripisuje samoglasnikom poleg „osnovnih (segmentnih) lastnosti, izhajajočih iz položaja govornil v ustni votlini“ še 3 „t. i. prozodične (tudi nadsegmentne) lastnosti, tj. trajanje, jakost in ton.“ Menim, da dolžina (kolikost) ni nadsegmentna lastnost samoglasnika, namreč ne v smislu kot pri naglasu (koncipiranem na podlagi jakosti ali tonemskosti); npr. v tem, „da posamezne zloge povezuje v besedo“ (Toporošič 2000: 64). Prav tako oblikovanja te lastnosti ni mogoče umestiti dosledno izven ustne votline, v kolikor je to merilo strukturiranja različnih fonetičnih lastnosti. Pri dolžini gre pravzaprav za aktualizirano fonetično značilnost samoglasnika, za opis njegove artikulacije, podobno kot lahko opišemo dolžino z okoljem pogojenih artikulacijskih variant zvočnikov in nezvočnikov.

³¹ Ima pa slovensko jezikoslovje obilo tradicije v pisno-pravopisnih razpravah in polemikah.

³² V ožjem podpomenu pa je prozodija zgolj naglas (le-ta omejen zgolj na zlog). Vendar novejša raziskava govora prepoznavajo in koncipirajo gradbene enote na višjih ravneh (podobnih besednozvezni, stavčni) prav na podlagi prozodičnih lastnosti. Primer obrazložitve iz metodologije prozodične analize spontanega govora, razvite za francoščino in uporabljene tudi na slovenščini: „Osnovna definicija govornih enot temelji na komunikacijskih načelih in specifičnih (tudi prozodičnih) sredstvih izrekanja. Opisovanje mikrostrukture posameznih govornih enot se v podrobnejših analizah delno naslanja tudi na skladenjske koncepte.“ (Vitez/Zwitter Vitez 2004: 8). Navedena metodologija proučuje zgradbo spontanega govorne izmenjave in vlogo prozodije kot strateškega zvočnega sredstva za doseganje komunikacijskih namenov v spontanem govoru. Prim. tudi Toporošič 1992: ESJ: 309: **stavčnofonetični sklop**: „s premorom oddeljeni del povedi“.

³³ Recimo to, da ima slovenščina jakostni in tonemski naglas, pa to, da danščina pa pozna sredni naglas (zadržanje tresenja glasilk pri izgovoru naglašene samoglasnika, Toporošič 1992: ESJ: 310: **stredni naglas**), nas ne sme napeljati, da bi enako razlikovalno vlogo pripisali ali našima dvema ali kar vsem trem naglasom.

za možnost dopuščanja variant (fonetičnih v govoru, tipografskih v pisavi), saj ta fizična *nekonstistentnost* različnih izraznih aktualizacij očitno še ne pomeni tudi nekonstistentnosti aktualiziranega pomena. V tem je *neenovitost* jezika, pojmovanega kot sistema, ki smo jo že večkrat omenili. *Kaj vse (v konkretnem aktualiziranem kontekstu) razločuje pomen, kaj od tega katero sestavino pomena, ostaja nedokončana zgodba, če empirično ne zamejimo gradiva (sistemskega konteksta), pa stopnje abstrakcije virtualnega znaka itd.*

Primer 2: Omejitev sistema alofonov in alografov pri transkribiranju

Pojavi besede, ki bi jo v knjižnem jeziku zapisali <jaz> (mene itd.): [jas], [jə̃s], [jə̃st], [jɜ̃s], [jɜ̃st] ... (Zemljarič Miklavčič/Stabej 2006: 8. prosojnica), torej /a/ z alofoni [a], [ə̃], [ɜ̃].

Zakaj recimo namenilnika *jest* (jesti itd.) [jest] ali pojma *jez* (jezu itd.) [jes] z [e] oz. /e/ ne moremo pridružiti v ta sistem pojavov, v katerem bi bil [e] opredeljen kot neosnovni alofon istega fonema /a/ kot [a], [ə̃], [ɜ̃] v govornih variantah pojma 'jaz'? Med omenjene pojave oz. bolikovne variacije zapise besede <jaz> ju ne moremo pridružiti prav zato, ker v morfemskem okolju teh dveh besed zapis <e> uresničuje samosvoj pomen – fonema /e/, ne pa fonema /a/.

Pri tem niso zaznamovane niti prozodične lastnosti: dolžina zlogotvornega samoglasnika, njegova (ne)naglašenost, vrsta naglasa. (Zasilen odgovor je seveda moč najti še drugje: v primerjavi z drugimi pojavitvami bi bilo verjetno ugotovljeno, da gre za bodisi „glavne“ bodisi „stranske“ variante glasu; ali pa bi menjali vidik in govorili o diahronem razvoju; ali pa sociolingvistično razlagali središnoslovensko dogovorjenost zapisa <jaz> za različne narečne variante.) In pri tem je relativno tudi, na kateri ravnini pomenskosti enačimo vse pojavitve: ali ravnina leksema (oz. iztočnice v slovarju, recimo v sistemu SSKJ: jaz), ali ravnina prvega/osnovnega podpomena takega leksema (recimo v sistemu SSKJ: jaz 1.), ali na posebej koncipirani (konotativno, slogovno, socialnozvrstno itd.) pomenski podravnini (recimo *ljubljski (mestni) govor* ali *slengonsko*).

2.2.2. Izrazna sredstva po jezikovnih ravlinah

Na koncu poglavja, ki obravnava semiotični vidik, še enkrat primerjalno razvrščamo pisna in govorna izrazna sredstva glede na ravnine idealnih osnovnih izraznih enot. V semiotično-strukturalnem-funkcijskem duhu jih v razdelimo na pet skupin (Tabela 2 spodaj), pri čemer so na levi osnovn(ejš)e enote, proti desni pa pomožn(ejša)a sredstva, ki jih je vedno težje opredeliti in razvrstiti tudi po jezikovnih ravlinah.

Kaj še lahko rečemo o izraznih sredstvih na podlagi te razvrstitve? Z vidika standardiziranja jezikovnih sredstev je jasno, da je lažje normirati tista na levi strani, ki so že po konceptu idealnejša, in tista v nižjih ravlinah, ki so po konceptu konkretnjša („bolj“ izrazno kot pomensko definirana). V sporazumevalnih situacijah, ki se odmikajo od standardne rabe (npr. spontani govor, sms-sporočila) pa bodo sredstva z desne strani tabele pogostejša kot v situacijah s standardno rabo.

Z vidika možnosti pretvorbe iz enega v drug izrazni sistem (npr. govora s transkribcijo) lahko na podlagi ugotovljenega v tem poglavju rečemo, da kompleksnost in nenatančnost **pretvorbe** raste v tabeli od leve proti desni. Od najnižje ravnine, kjer je izrazna plat jezikovnega sredstva pomembnejša od pomenske (samostojnega pomena na tej ravnini skoraj ni, gre le za njegovo *razlikovanje*) in kjer je pretvorba povsem konvencionalna oz. abstraktna, raste proti višjim ravlinam *strukturiranost pomena*: pomen čedalje bolj pridobiva na pomembnosti v primerjavi z izrazom, pretvorba se bolj podreja strukturiranju pomena, temu pa tudi strukturiranje izraza.

Tabela 2: Zmožnosti izraznih enot (sredstev) za razlikovanje in strukturiranje pomena

- Uvrstitev v raven ima za podlago primerjavo z osnovnimi enotami, ki gradijo višjo (pomensko) raven.
- Pri pisnih znamenjih je poudarek na ločilih; opisi rab ločil so mišljeni kot v SP²⁰⁰¹.
- Pri vseh izraznih enotah gre za namerno rabljene (tvorjene) enote v jezikovni funkciji.

sposobnost (linearnega) sestavljanja jezikovnega znaka na kompleksnejši jezikovni ravnini	oblike /forme	da (+ sintagmatska sposobnost)		deloma: sposobnost redundantnega razločevanja	ne (umestitev na raven ni gotova oz. ni enoznačna)
	virtualnega pomena	da (+ paradigmatična sposobnost)	ne		
jez. raven najmanjših gradnikov pomena: <i>fonem/grafem</i> ↓ <i>morfem</i>	govorni prenosnik	fon(em)	alofon, (ne)naglašenost	ton(skost), jakost; zlog; ³⁴ naglasno mesto	kolikost; glasovno barvanje
	pisni prenosnik	črka; opuščaj; nekatere neskladenjske rabe pike, ²⁹ vejice; ligatura	skladenjski vezaj, stična poševnica; neskladenjsko tripičje	(dodatno) diakritično znamenje; neskladenjska vezaj, oklepaj; deljaj; presledek	razlikovnost tipografske variante črke (font/pisava, velikost, barva, poudarjenost, velika/mala črka); nadpis/podpis
„besedna“ raven: <i>morfem</i> ↓ beseda (najmanjša samostojna enota s pomenom): ↓ <i>leksem</i> ↓ <i>frazem</i>	govorni prenosnik	„paradigmatična“ beseda	pojavnica (sistemsko razločevan glasovni niz); jakostni poudarek („besede“)	premor, mašilo; register; ^o	
	pisni prenosnik	števka, številka; „paradigmatična“ beseda; neskladenjska dvopičje in pomišljaj ³⁵ , matematični operatorji ipd. znanstveni simboli; krajšave in kratice;	grafična beseda (pojavnica), velika začetnica lastnega imena;		ponavlja; decimalno mesto
(nad)stavčna raven: <i>stavek ~ sintaksem</i>	govorni prenosnik	tonska /stavčnofonetična enota; izdišna enota, izjava; menjava govorca	tonski potek		nameren medmet: kašelj, smeh, zehanje ipd.
	pisni prenosnik	stavek, poved (npr. matematična formula), odstavek, poglavje	velika začetnica za začetek povedi; t. i. odstavčna/ alinejna številka, črka, pomišljaj, zaklepaj;	skladenjske rabe ločil (, : ; oklepaj); (ne)skladenjski narekovaj; enojna ali dvojna poševnica (nestična); končna ločila;	skladenjsko tripičje (izpusti); zamik, presledek pred odstavkom, poglavjem ipd.; vrstica/ kitica/ stolpci/ preglednica; neskladenjska klicaj, vprašaj; slog napisa/naslova, opombe, citata, primera;

³⁴ Zanimivo se je vprašati, ali zlog res nima nobene funkcije pri razločevanju pomena in gre samo za posebnost govornega prenosnika. SP²⁰⁰¹ omenja, da je zlog „lahko tudi samo pisen“, ko delimo besedo (§ 600, poglavje Deljenje).

³⁵ Dvopičje v pomenih ‘proti’, ‘deljeno z/s’ in pomišljaj ‘od .. do’ (predložni), ‘minus’ in ‘proti’ (nakazuje razmerje enot) delujeta kot piktograma z jasno ubeseditvijo.

2.2.3. Slovenski pravopis v funkciji znanstvenega zapisa govora

Pravopis kot jezikovni priročnik je namenjen kar najširši rabi jezika, katerega standardizira. Kot tak izhaja iz različnih praks rabe jezika in se vanje spet vrača; med njimi so tudi prakse zapisovanja govora – tradicije transkribiranja. Npr. intervjuji za tiskane oz. vizualne medije, zapisnikarstvo na sodiščih (npr. Novak 2006) in v množici drugih uradnih postopkov (zanimivi so govori na javnih sejah³⁶), zapisi govorov in predavanj, tudi stenografiranje; govor v literaturi, notranji monolog v literaturi z začetka 20. st. po zgledu psihiatričnih zapisov; podnapisi televizijskih oddaj in filmov. Različne znanstvene panoge, ki imajo opravka z govorom, so razvile svoje prakse zapisovanja: dialektologija, paremiologija, etnologija, fonetika/fonologija, gledališki govor (hrvaški primer je Starc 2006, slovenskega bi našli v raziskavah Katje Podbevšek), psihiatrija itd. S stališča slovenskega knjižnega jezika so za potrebe običajnega jezikovnega uporabnika te prakse morda nestandardne, a veljajo bolj ali manj na svojem področju.

Pravopis je namenjen predvsem zapisu standardne socialne zvrsti; v diplomu pa želim zajemati iz vseh socialnih zvrsti; torej bi bilo takrat, ko zajemam iz nestandardne socialne zvrsti, neustrezno, da uporabljam standardno kodiranje (pisni izraz) – v smislu pripisovanja pravilnosti (ustrezanja normi).³⁷ Vendar, če je pisni izraz standarden v smislu temeljnosti in povezovalnosti, potem mora biti sposoben kot širok prenosniški diasistem pokriti nekaj notranjih diferenciacij jezika (ne pa vseh). Pri diferenciacijah govora, ki jih pravopis ne more pokriti, se zdi, da imamo na voljo 4 možnosti zapisa:

Tabela 3: Transkribcije govora med možnostma ortografskost in fonetičnost

<i>ortografska transkribcija</i>	za transkribcijo ohranimo zgolj pravopis (zamejen nabor urejevalnih in poimenovalnih možnosti) in tako spregledamo nekaj govornih značilnosti ali jih celo potvorimo;
<i>modificirana ortografska transkribcija</i>	pravopis minimalno ali v večji meri dodelamo, da gotovo zajamemo določene govorne značilnosti (premik je bolj ali manj v smer fonetične pisave);
<i>fonetična transkribcija</i>	za zapis v celoti uporabimo fonetično pisavo;
<i>ortografsko popravljena transkribcija</i>	kombinacija zgornjih možnosti: pravopisni in nepravopisni zapis kombiniramo (vzporejamo), da damo za besedilo obe informaciji v dveh zapisih, kjer informacija ni razvidna iz ene vrste zapisa („druga“ informacija je način zapisa originalne informacije v standardni obliki).

Poimenovanje zadnjih treh vrst po Brinker/Sager 2001: 48.

Katera od možnosti je najbolj znanstvena? Ali res **fonetična transkribcija**,³⁸ ki se zdi najnatančnejša in zato najbolj informativno bogata? Funkcija fonetičnega zapisa je natančen zapis artikulacije glasov (nekako tako kot skiagram in sonagram, le zakodirano v kod znakov) v govoru. Zaradi te funkcije je pravzaprav predpogoj za glasoslovno jezikoslovno analizo (artikulacijsko področje fonetike). V osnovi historično izhaja iz enake fonetične pisave kot (slovenska) ortografska pisava (prim. opombo 27 na str. 20), le prirejena in precej razširjena je z znaki za alofone in prozodijo. Toda raziskovalcu, ki ga ne zanima obravnava glasoslovnih podatkov o jeziku, ni potrebno izbrati fonetične transkribcije le zaradi znanstvenosti.

³⁶ Transkribcije sej državnega zbora vključuje korpus Nova beseda: 1996–2004 711 sej ali 20 mio. besed, do 2007 pa že 27. mio. besed (*Nova beseda 2000* in kasnejše dopolnitve.), pa tudi korpus Fidaplus (2,04 mio.). Primer raziskave takega gradiva je npr. Krajnc 2004.

³⁷ Prim. opozarjanje na vprašljive primere uzakonjanja normativnosti v SP²⁰⁰¹, Vidovič Muha 2003: 119.

³⁸ Škofic 2006 govori o fonetični transkribciji kot o strokovni, znanstveni transkribciji. Oznaka „znanstven“ za fonetično transkribcijo ji pomeni predvsem njeno omejenost na ožjo sfero, strokovno rabo (str. 176), saj se ukvarja s prenosom narečnega besedila v poljudno, knjižno obliko (nas pa tukaj zanima znanstvenost transkribcije z vidika metodološkosti). Piše pa tudi o širšem pomenu znanstvenega zapisa okoliščin transkribcije – gl. poglavje I.3 v tej diplomu.

Dejstvo, zaradi katerega lahko oporekamo „a priori znanstvenosti“ fonetičnih transkripcij, je njihova nestandardiziranost. Pri tem kritika ne meri na pojav, da so s fonetično transkripcijo zapisana besedila pestra, neomejena v svoji fonetični pojavnosti. Gre za nekonsistentnost v stroki med konkretnimi razpravami pri uporabi možnega nabora znamenj in različne stopnje natančnosti in bogatosti zapisa fonetičnih in prozodičnih pojavov, včasih celo znotraj razprav. Npr. mešanje označevanja naglasov z ločevalnimi naglasnimi znamenji ali z enotnim pred zlog postavljenim opuščanjem po IPA (gl. prav primere v Škofic 2006), mešanje OLA in IPA naborov fonetičnih znamenj, včasih še celo z naborom SAMPA (gl. Tivadar/Jurjec 2003), pisanje z ločili ali brez njih, z velikimi začetnicami ali brez njih, s presledki ali brez ...

Na kratko povedano, hočem opozoriti na to, da obstaja tudi „modificirana fonetična transkripcija“. Jasnih meja med ortografsko, modificirano ortografsko in fonetično transkripcijo ni!³⁹ Ni čudno, da obstaja cela plejada modificiranih ortografskih (in fonetičnih) transkripcij (prim. npr. Toporišič/Gjurin 1981: 324–384 – Slovenska zvrstna besedila, poglavje Socialnozvrstna besedila). Potvorjevanje in izguba nekega deleža informacij o govoru, kar omenjam pri prvi možnosti transkripcije (Tabela 3 zgoraj), pa je v neki meri neizogibna pravzaprav pri vseh možnostih transkripcije.

Vsaka raziskava govora ima seveda pravico do svojega sistema/redakcije transkripcije, saj le-ta služi njenim specifičnim ciljem, gradivu, pa tudi usposobljenosti in energiji raziskovalca (podrobni transkripcijski sistemi kot npr. IPA so zelo zahtevni in je težko doseči konsistentnost pri njihovi uporabi). Kar se lahko naučimo glede znanstvenosti (se pravi: enoznačnosti, preverljivosti, konsistentnosti) transkripcije, je to, da je potrebno ob vsaki raziskavi govora, sistemu transkripcij dosledno uporabljati transkripcijski protokol, nabor znamenj in protokol pa eksplicirati, dokumentirati. To je tem lažje, čim bolj je sistem transkripcije naslonjen na eno standardno transkripcijo, pa naj bo to enostavnejši slovenski pravopis ali zahtevnejša IPA; z drugimi besedami – čim manj je osnovni tak temelj modificiran.

„Pravopis“ v širšem pomenu nam pomeni predpis in urejevalnih in poimenovalnih možnosti (standardnega) jezika, saj kot priročnik tradicionalno zajema pravila (~ slovnico) in slovar. Nanaša se na več jezikovnih ravnin, nikakor ne le na pomenskorazločevalno ravnino ali celo le na razmerje fonem-črka (kar je ožje ali laično pojmovanje izraza „pravopis“, včasih poimenovano tudi s pomočjo tujke „ortografija“). Slovenski pravopis je v temelju **fonetičen**, ker izhaja iz fonografskega pisnega sistema (črke kot znaki za foneme-znake),⁴⁰ na površju pa **oblikoglasen** (morfonološki), kar pomeni, da je njegova grafofonemskost drugotno, „na površju“ uporabljena za *razlikovanje morfemov in besednih oblik*. (Če bi bil fonetičen v vsej uporabljenosti tudi „na površju“, bi težil samo k razlikovanju fonov,⁴¹ kot npr. sistem Vuka Karadžića.) Ti dve opredelitvi veljata predvsem za črkopisno plat (ravnino) pravopisa; s pomočjo ločil pa pravopis razlikuje tudi skladenjske strukture (lahko bi skovali poimenovanje te funkcije: „grafosintaktičnost“), s pomočjo velikih/malih začetnic razlikuje pomenske strukture itd., o tem sem že razpravljajal zgoraj.

Ker se ta naloga povprašuje predvsem o besedah v govornem jeziku, ji pride na slovenski pravopis naslonjena transkripcija zaradi njegove oblikoglasnosti še kako prav. Vendar pozor: razlikovalni sistem glasovnih znakov in po njem **razlikovalni sistem oblik in besed v govoru** ne ustreza nujno v pravopisnem *slovarju* zajetemu sistemu morfemov in besed. Želimo zajeti

³⁹ Še toliko bolj to velja pri primerjavi slovenskega pravopisa in npr. angleškega v odnosu do npr. IPA: slovenski pravopis s pravorečjem je bolj fonetičen in manj historičen, bližje IPA kot angleški pravopis s pravorečjem.

⁴⁰ Kot sem že mnogokje omenil, oznaka fonetičnosti za slovenski jezik (pravopisa) pove tudi, da je v njem manjše razhajanje med aktualno standardno izreko in aktualno standardno pisavo – da je njegova pisava glede na živi jezik kljub historičnosti manj stara ali manj neživa kot pravopis angleščine, francoščine. Prim. opombo 13, str. 15.

⁴¹ Ortografske (običajne) fonetične pisave ne težijo k razlikovanju fonemov, ampak le fonov; niti ne moremo govoriti o razlikovanju alofonov. Definicija fonema izvira prav iz oblikoglasnosti, iz funkcije razlikovanja morfemov in kompleksnejših jezikovnih struktur. Razlika med terminoma fon in alofon odkriva razliko med deskriptivnim in funkcijskim vidikom glasovnih pojavov. Paradoks znanstvenih fonetičnih pisav je v tem, da je v fonetičnih znamenjih, s tem ko izvirajo iz katerekoli (običajne) fonetične pisave, princip fonemskosti že impliciran.

oblikoglasje aktualnega, konkretnega govora in ga združiti s sistemom v pravopisnem slovarju. Zatorej nam iz pravopisa univerzalneje kot slovar služijo *pravila*: in sicer pravila o grafofonemskih razmerjih in razlikovanjih (črke *in drugi znaki* nasproti fonemom), o primerjanju – sistemiziranju morfemov in besednih oblik (npr. načela glede pisanja skupaj oz. narazen, načela neupoštevanja prilagajanja po zvonečnosti itd.). Tako npr. za popolne redukcije fonemov uporabimo opuščaj, delne redukcije in prilagoditve po zvonečnosti pa zapišemo s standardno črko za osnovni alofon oz. fonem. Šele upoštevanje teh pravil naredi uporabno tudi leksikografsko primerjavo pisnih in govornih besedil – pri združevanju slovarjev, ki so strukturirani na podlagi morfemov, besed.

Primer 3: Oblikoglasnost in razlikovalni sistem oblik

Posamezne prevzete besede pogosto začnemo pisati fonetično (npr. lat. transkriptio, -onis > slov. transkripcija). Ko pa prevzamemo še sorodne besede z drugačno morfofonetično zgradbo (lat. scribere > transkribirati, skriboman), se ves razlikovalni sistem podomači, posloveni. Primer *transkripcija* se tako prilagodi pravilu zapisovanja zvonečnostnih parov (zapis nezvočnika po obliki pred samoglasnikom ali zvočnikom) – *transkripcija* (prim. SP²⁰⁰¹: § 686 in predhodni; ter § 224); ostajajoče izjeme so: deskriptiven, skripta.

Žal pa je pravopis tu in tam neenovit v pravilih – pri transkripciji, naslonjeni na pravopis, bo potrebno upoštevati tudi t. i. **pravopisne posebnosti**, če naj bo primerljiva s pisnimi besedili v besednih oblikah tam, kjer ne bi smelo biti razlik, pa so v zapisu – rečemo npr., da zaradi tradicije, historičnosti, v resnici pa zaradi raznih uporabniških vplivov na jezik, njegove multifunkcijskosti, interakcijskosti, živosti ...

Primer 4: Pravopisne posebnosti

Npr. „zapisovanje morfemov (*ležem lesti* proti *nizek nizka*, *z* – *s* nasproti *v*, *sčasoma* – *lišček*), zapisovanje pisanja skupaj narazen (*name* proti *na me*)“ (Toporišič 1992: ESJ: 204: **pravopisna posebnost**).

Upoštevanje fonetičnosti nad oblikoglasjem ima kljub upoštevanju pravopisne posebnosti pisanja *z* – *s* pri zapisovanju govora drugačne rezultate kot pri pisanju in razkriva nekatere razlike med govorom in pisanjem. Poglejmo izmišljen primer: pisno „s F1“, govorno [z '3f '3na]. V transkripciji StAnId (naš korpus) bi drugi primer zapisali <z *eF_ena>.

Več o transkripcijskih odločitvah in modificirani ortografski transkripciji (t. i. transkripcijski konvenciji) v gradivu diplome v poglavju III.11.

2.3. Izbor enot z besedne ravnine

V tej nalogi se želim pri slogovni interpretaciji osredotočiti na besedno raven, kar bi lahko z jezikovnoanalitičnega vidika bila čisto konvencionalna oz. svobodna izbira – vsaka jezikovna raven pove nekaj o jeziku, ne glede na prenosnik. Vendar, zakaj „so besede tisto, kar nekako sami od sebe najprej povežemo z jezikom“ (Jakopin 2002: 59)? Morda zaradi strukturalistične tradicije jezikoslovja. Zakaj pa je v strukturalizmu **beseda** v središču pozornosti, sistemiziranja; zakaj je pred očmi kot idealni jezikovni znak (gl. opombo 20, str. 17)? Natančnejši odgovor morda tiči v naravno- ali kognitivnojezikoslovni primerjavi jezikovnih znakov (sredstev), ali podobno morda v informacijski teoriji v Menzenrath-Altmanovem zakonu (gl. poglavje I.2.4).

Tukaj pa naj bo dovolj dejstvo, da je beseda **najmanjša (oblikovno) samostojna enota sporočila (povedi, besedila)**; lahko bi dodali še: **z lastnim (virtualnim) pomenom**. Taka ostane posamezna beseda v svoji izrazno-pomenski strukturiranosti, pa najsi jo prestavimo v drug kod ali prenosnik, iz enega sporočila v drugo; *temeljna dela njene izrazne in pomenske strukture se ne spremenita* (Primer 5 spodaj). Očitno je, da je tako pojmovanje besede je možno le za na poseben način strukturirane jezikovne znakovne kode⁴² – pravzaprav gre za teoretično *idealizacijo* izrazno-

⁴² Tipičen primer takega koda sta (alfabetski) pisni in (fonemski) govorni jezik; in vsi kodi, izpeljani iz njiju: Morsejeva abeceda, Braillova pisava ipd. Izrazna in pomenska strukturiranost besede pa se lahko spremenita ob prenosu v drug jezik zaradi njegovih drugačnih kulturnih (in izraznih) danosti, odraženih tudi v jezikovnem *sistemu*.

pomenske strukturiranosti, ki jo tukaj v nalogi imenujem **besednost/verbalnost**. Beseda je kot jezikovno sredstvo idealizirana podobno kot fonem, stavek itd. (prim. opombo 10 na str. 14).

Primer 5: Kohezija elementov besede

„Elementi besede kažejo večjo kohezijo na ravni besede kot v večjih enotah: korenski morfemi /angl. stems/ in obrazila ne morejo biti ločeni, razen z drugimi obrazili. Tudi zaporedje elementov besede se ne nagiba k variiranju.“ (Matthews 1997: 404: **word**).

Primere v podporo in mejne primere lahko najdemo tudi v govoru. Npr. z vidika besednega reda: v slovenščini je besedni red manj obvezen kot v nekaterih drugih jezikih (npr. nemščina, angleščina), pa vendar se v (tipičen primer) spontanem govoru marsikdaj do nerazumljivosti podre, skladienske strukture postanejo prelomljene in neizpolnjene – besede pa ostanejo v primerjavi z njimi cele, popolne. In mejni primeri: tudi v govoru imamo nedokončane besede; in sestavinske napake pri besedah, ki nastanejo zaradi motenj pri ubesedovanju (primer iz gradiva glej v nadaljevanju), v pisanju pa poznamo motnje zaradi disleksije in zatipkanja.

Primer iz besedila 1ned09Vg v korpusu StAnId:

007Žz [v] *čuj a veš *te pa () ... jaz 'mam en 'ga strica *Dankota *ko 'ma teto Vanko. in jaz rečem zmerom: kje *te sta Danko pa Vanka? ne! *Danka pa *Vanko! | PREMOLK | *toto zmerom zamenjujem al' ... | PREMOLK | ampak ko je še *pal taka p-ja ... *pol pa še *taki pomen opet- *opasno *vun 'zgleda ej. *Danka pa *Vanko ...*

2.3.1. Beseda in njen(i) sistem(i)

Pri razčlembi besedila – aktualizirane množice besed – se torej zgornja definicija besede krha in kliče natančnejše kriterije na pomoč, pač v odvisnosti od množice besed, ki jih vzame znanstvenosistemski kontekst v ozir, in od analitičnih izkušenj, ki jih ima na voljo. Točno to se namreč dogaja že pri banalni stvari, kot se zdi zapis govora: govorne nize na podlagi njihovega (virtualnega, od konkretne situacije neodvisnega) pomena izločiti iz govora kot besede in jim pripisati podobo grafično strukturirane in razločene besede. Delovni pogoj za določanje virtualnega pomena pa je pri tem *sistem (množica) drugih besed* v bodisi leksikonu (tezavru) jezika skupnosti bodisi spominu jezikovnega uporabnika – transkribtorja oz. zapisovalca; ta množica omogoča primerjanje in prepoznavanje besed.

Besede pa primerjamo tudi še potem, ko so že zapisane. Če vzamemo le govorno in pisno od prenosniških zvrsti jezika, ugotovimo, da (virtualno) enake besede najdemo v obeh; ne glede na sporazumevalne vloge in situacije. Torej lahko besede služijo za primerjavo notranjih diferencij jezika (prenosniških, situacijskih: socialnih, funkcijskih), pri čemer je *tertium comparationis* virtualna – pomenska dimenzija besede.

Oblika pa me zanima bolj kot pomen, ker v tej nalogi izhajam iz prenosniško zamejenega jezika in ker izrazna definiranost edina omogoča strojno obravnavo oz. manj vnašanja informacij na prvi stopnji računalniške obdelave (kaj nam slednje pove o izrazni dimenziji jezikovnega sredstva: bolj konkretna je od pomena). Izrazno-pomenske (z)možnosti besede strukturalistično razčlenim z dveh polov: sestavinsko, v linearnem okolju (sintagmatično); in v pomenskih razmerjih, ki prevevajo oba omenjena vidika (razlikovanje pomena, skladnja pomena – paradigmatsko). Na podlagi primerjanja aktualnih besed v govoru s temi platmi idealne besede nastanejo sistemi besed in kriteriji, ki besede uvrščajo v kategorije; gl. Slika 1 spodaj.

Na tej točki se je potrebno zavedati, *čemu kategoriziramo besede*. Običajno in ustaljeno je besednovrstno kategoriziranje ali zaradi medjezikovne primerjave (besedni dvo- in večjezični slovarji; primerjalne slovnice) ali zaradi razlikovanja samih besed v enem jeziku (normiranje paradigem; diahroni razvoj paradigem; enojezični besedni slovarji) ali zaradi natančnejšega razlaganja pomena nadbesednih enot (stavčnih členov, stavka, povedi, besedil, celo sporazumevalnih situacij). Skratka, gre za raziskovanje variiranja izrazno-pomenskih (z)možnosti

Več o kulturno in časovno pogojenih razlikah pri leksiki posameznika in skupnosti gl. npr. Suhadolnik/Janežič 1962: 45–46.

besed. Vsak taksonomski kriterij je tudi opisovalec izrazno-pomenske (z)možnosti uvrščene besede oz. skupine besed.

Slika 1: Strukturiranje izrazno-pomenskih (z)možnosti idealne besede



Zakaj me sploh zanima, kaj mi bo sploh povedala jezikoslovna kategorizacija besed na besedne vrste, na podrobnejše skladenjske kategorije, uvrstitev različnih besednih oblik pod leksem ali lemo? Ko transkribiram govor in pe ko ga kot gradivo nadalje obdelujem, besede govora kategoriziram. S kategorizacijo, ki zaradi prisojanja sistemske umestitve *vsaki aktualni* besedi zahteva veliko časa za delo, vnašam dodatne, vedno bolj abstraktne informacije, ki določajo nadaljnjo interpretacijo in morebitne nadaljnje stopnje strojne obdelave besedil.⁴³

V našem gradivu bomo *besede* jemali le kot *pojavnice* in kot („grafofonemske“) *različnice* – gl. naslednje podpoglavje. Dodatne kategorizacije besed, h katerim težijo sodobni korpusni jezikoslovci in ki so sicer zelo vabljive in mnogo obetajo, pa bomo prepustili drugim: besedna vrsta (angl. POS = **P**art **O**f **S**peech), oblikoskladenjska določitev (angl. MSD = **M**orpho**S**yntactic **D**escription), osnovna besedna oblika ali lema⁴⁴ (angl. lemma) ter njene odvisne besedne oblike (angl. word form), leksem in druge.

V prvi vrsti bo torej sad analiz *slika o obliko(slo)vnem variiranju besed (besednih oblik) v govoru v različnih pogledih: v primerjavi s pisnim jezikom, v različnih sporazumevalnih vlogah in situacijah, v govoru posameznika*. Jezikoslovna kategorizacija bo služila kot sredstvo metajezikovnega opisa tega obliko(slo)vnega variiranja objektnega jezika. Slog bo opisan na prej omenjenih ravneh s kvantitativnega vidika variiranja besednih oblik; katere so pogostejše, katere redkejšje, kakšni so njihovi deleži ...

⁴³ „Pri vsakem podatku, ki ga vnesemo, se moramo vprašati, v kakšno pomoč bo ta informacija računalniku. Ali je, na primer, podatek, da je neki pridevnik kakovostni in drugi vrstni, da je neki glagol naklonski in drugi glavni, koristen? Ali pri pregibnih jezikih, kot je slovenščina, koristi podatek o končnici /.../? In če gremo še korak nazaj, kakšne lastnosti pa sploh ima za računalnik koristen podatek? Takšna vprašanja so za jezikoslovca seveda izredno neljuba, saj slabo pozna delovanje stroja. Neljuba pa so tudi za inženirja, saj tehnologija še ni razvita do konca in težko je reči, kateri koraki bodo še potrebni na poti k zastavljenemu cilju.“ O dilemah pri gradnji oblikoslovnega in glasoslovnega slovarja SImflex, Verdonik idr. 2003.

⁴⁴ Neweklowsky v svoji raziskavi (1984: 384): „Pod *besedo* tukaj razumemo vse besedne oblike, ki pripadajo eni paradigmi.“ – prosti glagolski morfemi v njegovi raziskavi niso samostojne besede, v tej diplomi pa so.

2.3.1.1. Izrazno različne besede – pojavnica in različnica

Ko se izražamo o količini besed, govorimo o **pojavnica** (angl. type): večbesedni izraz, enobesednica, besedna zveza, ta stavek sestavlja 5 besed, članek je lahko dolg do 200 besed itd. Da jih lahko preštejemo, jih moramo zamejiti oz. izrazno razlikovati, čeprav na njihovo *pojavit* opozarja že njihova *umeščenost* v tok sporočanja. Njihova izrazna raz-ločenost v poteku besedila (običajno linearnem) je samo navidez posledica začetne in končne meje besede: v standardiziranem pisnem besedilu nam to pove presledek ali ločilo⁴⁵ oziroma ...

Na splošno bi morda lahko postavili grobo pravilo, da je pojavnica vse, kar se pojavi med dvema presledkoma in ločilom. Prvo izjemo tu predstavljajo kratice in okrajšave, ki jih pišemo s piko in jih navadno obravnavamo kot samostojno pojavnico s piko ali pikami vred (na primer *dr.*, *prof.*, *prim.*), tudi če vključujejo več besed (*itd.*, *tj.*, *t. i.*, *d. o. o.*).⁴⁶ (Gorjanc 2005: 63, citira iz disertacije Š.Vintar 2003(b): 41.)

Kaj pa v govoru, kjer premolki med besedami (ali druga fonetika v vlogi segmentacije) že v standardnem izgovoru niso vedno predvideni, kaj šele v nestandardnem? Krajšave sicer v govoru jasneje kot v pisnem jeziku razločimo kot besede, podobno še posamezno izgovorjene črke kratic. Najprej se naslonimo na izrazno sestavljenost besed, kasneje pa bo govora še o drugih vidikih, s katerimi si pomagamo pri razlikovanju besed (pregibnost, pomen ...).

Od drugih pojavníc posamezno pojavnico razločimo tako, da njen znakovni niz (osnovne in pomožne izrazne enote/sredstva) primerjamo z množico drugih nizov (besed). Kot posamezno pojmujeemo táko, ki ima lasten (edinstven) znakovni niz (hkrati tudi notranje strukturiran v morfemih), različen od drugih nizov, pa tudi ponovljiv v *enaki* obliki v drugih kontekstih, hkrati ima sintagmatsko sposobnost (ob njo lahko razvrščamo druge besede; prim. *načelo nerazdružljivosti in nezamenljivosti zaporedja delov dane besede*, SP²⁰⁰¹ § 485–487). Dokler štejemo vsako besedo v besedilu (vsako pojavit; torej nam tudi dve izrazno enaki besedi štejeta kot dve besedi-pojavnici), govorimo o pojavnica, ko pa štejemo v besedilu le *različne besedne oblike*, govorimo o **različnicah**⁴⁶ (angl. token; zdaj dve izrazno enaki besedi štejeta kot ena beseda-različnica).⁴⁷

Primer 6: Katere slovnične kategorije želimo razlikovati z grafonemskimi različnicami

Znan primer povedi v sledeči pisni obliki: <Gori na gori gori.> (Jakopin/Bizjak 1997: 516). Poved ima 4 pojavnice (Gori, na, gori, gori) in 2 različnici (gori, na). Koncipiranje različnice vsebuje že neko stopnjo abstrakcije (virtualnosti) v primerjavi s pojavnico, v tem primeru gre za abstrakcijo *kriterijev izraznosti*: zakaj <Gori> in <gori> (razlika v začetnici) nista 2 različnici? Stopnja abstrakcije je v tem primeru odvisna od tega, kaj želimo razlikovati, šteti: besede ne glede na vlogo zaznamovanja začetka povedi (v pisavi z veliko začetnico) ali pa glede na to vlogo, aktualizacijo.

Pri opazovanju izraznosti besed določa pojmovanje različnic natančnost kodiranja – ta vidik nas zanima pri transkribiranju, ko skrbimo, da bomo dve različnici enako zapisali. Primarnejši govorni jezik ima namreč veliko možnosti razlikovanja, v primerjavi s pisnim (npr. z ortografijo) govorimo o tem, da je redundanten (o tem že v poglavju I.2.2). Ko uporabljamo pisni

⁴⁵ Tako ločilo ali znak, ki ni v funkciji sestavljanja besede – „non-character“ oz. nečrkovni znak (za slednje prim. opombo 117, str. 89).

⁴⁶ Namesto strokovnega izraza različnica je mestoma primernejša natančnejša oznaka, npr. grafonemska različnica, za primere, ko pojavnice enačimo na podlagi njihovega zapisa z znamenji. Koncipiranje različnice pa se lahko nasloni tudi na druge kriterije: na kriterij oblikoskladenjske določitve (po njem so 3 pojavnice hiše rod. ed., hiše im. mn., hiše tož. mn. vzete kot različnice, ker imajo različne MSD; Neweklowsky 1988: 338), pa na nekoliko manj specifičen kriterij leme (ki primerov hiše ne bi razlikoval med seboj, bi pa primer <Gori na gori gori.> ločil na 4 različnice, saj ima vsaka beseda različno lemo: gori, na, gora, goret), pa še kak kriterij bi našli.

V zvezi s tem govori Neweklowsky (1988: 338) o ločevanju „leksikalne“ in „morfološke enakoizraznosti“, Jakopin/Bizjak pa o „večpomenskosti besednih oblik“ (1997: 516). Gl. še naslednji dve opombi.

⁴⁷ Z drugimi besedami: Pod količino različnic mislim na vse *izrazno neenake* pojavnice, ki se v neki količini besedila pojavijo.

jezik v funkciji transkribiranja, ne moremo nikdar preseči razlikovne zmožnosti govornih znakov, ampak se ji lahko le približamo ali jo, v idealnem primeru, izenačimo. Relevantno je v zvezi s tem razlikovati besede kot enakoizraznice (homonimi) glede na prenosnik: enakopisnice (homografe) in enakozvočnice (homofone; prim. Vidovič Muha 2000: 183).⁴⁸ Preprostejše kot je kodiranje, večja je možnost, da bomo pojavnice enačili kot različnice, saj prihaja z brisanjem razlik do informacijskega šuma. Po ortografski transkribiciji bo več enakopisnic (= manj različnic) kot po natančnejši fonetični transkribiciji, v le-tej pa bo še vseeno več enakopisnic (in manj različnic) kot enakozvočnic v samem govoru.

Primer 7: Odvisnost seznama različnic od natančnosti transkribicije

Poved v ortografskem zapisu iz zgornjega primera ima 3 enakopisnice (in 2 različnici), ob upoštevanju velike začetnice pa 2 enakopisnici – zadnji dve besedi (in 3 različnice). Če v zapisu uporabimo ločevalna znamenja v funkciji zaznamovanja jakosti in kakovosti samoglasnika (SP²⁰⁰¹: § 627; Gôri na gôri gori.), enakopisnic ni več, število različnic pa je izenačeno s številom pojavnice (4; Gôri ≠ gôri ≠ gori ≠ na). Enako kot v zadnjem primeru bi se s številom enakopisnic zgodilo, če označimo že samo jakostni naglas (npr. z ostrivcem kot enoznačnim znamenjem za naglas: Góri na góri gori.)

V nekaterih realnih situacijah je zanimivo še, kakšno stopnjo natančnosti kodiranja so uporabili tvorci in razumeli prejemniki sporazumevanja – razločevalni okvir prilagajajo udeleženci sporazumevanja vsaki aktualni sporazumevalni situaciji (t. i. diglosija, predvsem v menjavanju t. i. socialnih zvrsti jezika). Znotraj kakšnega razlikovalnega sistema deluje oblikov(a)nost neke (aktualizirane) besede v referenčni in razlikovalni funkciji? Glede na to, da se npr. fonetične različnice (ali neenakozvočnice) za Primer 2, str. 23 ([jas], [jøs], [jəst], [jɜs], [jɜst]), verjetno pojavljajo v različnih besedilih in torej *različnih sporazumevalnih situacijah*, se moramo vprašati, kdaj lahko govorimo o njih kot o variantah. In pri transkribiciji, če vzamemo ortografski črkovni nabor: bo zapis ena sam <jaz>, ali bodo uporabljeni še drugi: <jast>, <jez>, <jest> itd.?

Če povzamemo: zgornje dileme pridejo do izraza, ko želimo razločevalno pestrost besedja nekega govornega besedila ali (pod)korpusa oceniti po njegovem *seznamu različnic*. Brez določil zapisa (standardizirana transkribicija) in brez ozira na ta določila je ocena besedne oblikovnosti neutemeljena, neznanstvena.⁴⁹

V takem seznamu pa je problematična tudi umestitev besed, pri katerih smo v dilemi glede pisanja skupaj-narazen; npr. pisanje problematične besede narazen daje 1 pojavnico več v primerjavi s pisanjem skupaj (ali je več tudi različnic, pa je odvisno od ostalih besed v seznamu). Toda to (*nerazdružljivosti in nezamenljivosti zaporedja delov*) velja že za morfeme in celo za enoznakovne morfeme. V primerjavi z morfemi nam pomaga dodatna definicija besede kot morfemskega sklopa, funkcijskoskladenjsko zaokroženega – če je pregiben, je pregiben samo na koncu.

Problematične ostanejo nepregibne besede in t. i. prosti morfemi, od katerih je vsak običajno tudi že beseda. Prim. Suhadolnik/Janežič 1962 in Suhadolnik 1974, ki naj bi definirala besedo kot „grafično besedo“, tj. kot „črkovno zaporedje med dvema presledkoma oziroma med presledkom in skladenjskim ločilom“ (Neweklowsky 1984: 384, 385) – toda kako definirati proste morfeme in nepregibne besede v govornem prenosniku? Npr. nesklonljivi pridevniki, a v podrednem razmerju (kot odnosnica v besedni zvezi). Skupaj ali narazen: gremo na mini sprehod : mini golf, angora zajec : angora volna, na vikendu : vikend hišica, žiro račun, pojdi v avto : avtogaraža (SP²⁰⁰¹ § 487, 493–500, 504, 507, 518, 522, 529).

V govoru so besede zaokrožene zložno in tonsko: iz celega števila zlogov in kot naglasne enote (o naglasih glej SP²⁰⁰¹ §1023–1036). Ta vidika pomagata včasih tudi v pisnem prenosniku,

⁴⁸ Ne smemo pozabiti na struktural(istič)no podraven, na kateri tukaj govorimo o besedi: ne o enakoizraznosti leksemov („popolna enakoizraznost“; Vidovič Muha 2000: 183) oz. lem („delna enakoizraznost“, „mogoč/a/ samo znotraj slovarja“; isto mesto), ampak pojavnice („morfemska enakoizraznost“; ista, str. 182).

⁴⁹ O posebnostih analize besednega seznama (kot posebnega izbora nasproti celostni analizi besedila) in o posebnostih transbesedilne frekvence (pridobljene iz korpusa besedil) govori v uvodu svoje razprave Grzybek 2000: 141–142.

ko v nekaterih primerih grafične besede ločimo na podlagi razločevanja v govornem prenosniku, čeprav razločevanja v pisnem ni. Npr. [za:'to:] proti [za'to:]; tu je razločevanje s pomočjo *presledka* (sic – iz pisnega prenosnika!) šele sekundarno, tudi če ga pišemo v fonetičnem zapisu govornega niza [za:'to:] proti [za:'to:].

Nekatere rešitve pisanja skupaj oz. narazen so torej zgolj posledica pisne tradicije, prevladujoče prakse oz. konvencionalnosti odločitev (prim. recimo še primere različnih besednih enot v korpusu Nova beseda, sestavljenih iz črk, števil in drugih znamenj, Jakopin 2003: 147). Pri zapisovanju govora *pisne prakse* ni, je pa tradicija transkribiranja, ki jo največkrat sami soustvarjamo, ko določamo ali izbiramo transkripcijsko konvencijo. V naši so dilematični primeri pisanja skupaj oz. narazen označeni s podčrtajem na mestu možnega presledka (tako v besedilu kot besednem seznamu), podčrtaj pa služi sicer tudi cilju izpostavljanja večbesednih lastnih imen.

V pričujoči diplomii so bile zaradi obrazložitve načrtovanih analiz doslej pod drobnogled vzete enote dveh jezikovnih (osistemskih) ravnin: pomenskorazločevalne in besedne. Kasneje sledi še poglavje, ki se podrobneje obrača še na tretjo ravnino v tej nalogi – besedilno – in nekoliko na višje – diskurze (poglavje I.3); to pa ne pomeni, da vmesne ravnine povsem izpuščamo. V gradivu bomo kot enote šteli tudi povedi in izjave, a jih ne bomo podrobneje interpretirali. Kako jih definiramo, gl. v poglavju transkripcijske konvencije, ki določa njihov zapis – II.7.2.

2.4. Kvantitativni pretres jezikovnega gradiva s statistikami in pogostostmi

Zadnji pogled na manjše enote jezikovnega sistema, ki ga želimo predstaviti, ker iz njega v diplomii izhajamo, je kvantitativen (ostali so torej bili kvalitativni). Polja kvantitativne lingvistike obsegajo vse, kar je aktualno v današnjem jezikoslovju in celo v literarni teoriji, naj jih naštejemo le nekaj, ki so blizu poti moje raziskave. Vse naslednje oznake bi lahko imele spredaj še pridevnik „kvantitativna, -o itd.“: morfologija, slovnica, skladnja, pogostostni slovarji, besedne pogostosti, besedišče, stilistika (oz. *stilometrika*, za pojasnitev pojma gl. poglavje I.4.1), sociolingvistika, variiranje, računalniško jezikoslovje (gre za poglavja iz bibliografije kvantitativne lingvistike – BQL, Köhler 1995).

Kvantitativna jezikoslovna analiza povezuje tri izhodišča – tri postavke preračunavanj:

- gradivo (njegova velikost),
- enota, ki jo merimo, kvantificiramo, ter
- analitične kvantitativne vrednosti;

poleg tega vsega pa neobhodno še kvalitativni vidik: teoretična utemeljenost vseh treh postavk (izbor gradiva, definicija enote, interpretacije izračunov, grafov ...) in, v primeru uporabe računalnika, še strojna izpeljava raziskave za vse 3 elemente. Nasploh velja, da „/p/rešteti in izračunati /in lahko bi dodali – zinterpretirati/ se da marsikaj, in če gremo zelo v širino, se lahko rezultati proučevanemu delu po obsegu približajo ali ga celo presežejo“ (Jakopin 1995: 341). To velja že za enostavna štetja, kvantitativne metode za opis besedil oz. jezika pa imajo v mednarodnem jezikoslovju že dobro poldrugo stoletje tradicije (Tweedie 2005: 387),⁵⁰ in matematični modeli v naprednejših raziskavah dosegajo še dosti večje zapletenosti (prim. npr. izračune entropije, Jakopin 2002).

V tej diplomii smo od vseh možnih analitičnih kvantitativnih vrednosti osredotočeni na *statistične* (z vidika celote gradiva) in *pogostostne*⁵¹ (z vidika elementov gradiva) opise, torej na

⁵⁰ Citirani članek navaja kot eno prvih del kvantitativnega jezikoslovja raziskavo Angleža Augustusa de Morgana iz 1852, o kateri poroča T. C. Mendenhall, The characteristic curves of composition, v Science 1887, št. 11, 237–249.

⁵¹ Termin pogostost, pogostosten uporabljам kot sopomenko terminu frekvenca, frekvenčen. Varianti termina „pogostnost“ se poskušam prav tako izogibati, da bi ne prihajalo do nepotrebnih dvojnic; SSKJ med slovenskima dvojnica ne razlikuje.

najosnovnejši, najobičajnejši, najenostavnejši kvantitativni pretres jezikovnega gradiva. Le zaradi potrebe primerjav različno velikih besedilnih gradiv pogostosti preračunavamo v *delež*, ki jih imajo enote v gradivu, kar je spet zelo osnovna enota preračunavanj. Pogostosti razbiram za besedne oblike ali različnice, kot so pojasnjene v prejšnjem poglavju; razvrščam jih v pogostostne sezname, izračunavam razmerje različnic in pojavnic; dalje še seštevam kumulativno in predstavljam v grafu rasti besednega zaklada idr. Več o teh uporabljenih metodah preračunavanj v poglavju I.4.3, v nadaljevanju pa se posvetimo problematiki razmerja med velikostjo gradiva in kvantificiranimi enotami – elementi gradiva.

V korpusnem jezikoslovju (iz korpusov izvira največ podatkov za potrebe kvantitativnih raziskav, tako tudi v pričujoči diplomii) „novejše raziskave kažejo, da dejansko /večja/ količina daje tudi večjo kakovost; obvladovanje velike količine podatkov pa tehnološko ni več problem“ (Gorjanc 2005: 53).

Ustvarjalci korpusov za potrebe najširših slovnčnih raziskav se ženejo (ženemo) za vedno večjimi korpusi, toda zakaj? Zakaj pa fonetikom zadostujejo že krajše govorne zbirke, ki jih niti ne moremo imenovati še korpus? Razlog je v tem, da je za „razkrivajočo“ kvantitativno analizo potrebno izbrati dovolj majhen/velik vzorec gradiva glede na enoto, ki se jo analizira. Tako za raziskovanje fonov zadostuje nekaj deset ali sto besed, polnopomenske besede pridejo do izraza v korpusih od nekaj tisoč besed navzgor; ob odsotnosti besednovrstnih oznak delamo z besednimi oblikami, za katere so potrebni korpusi od nekaj deset tisoč do nekaj milijonov besed (za zelo pregiben jezik kot je slovenščina); besedni n-terčki od milijonskega gradiva navzgor, povedi šele v nekajstomilijonskih korpusih itd. (prim. Jakopin 2002: 58, 78).

Izhodiščni cilj te naloge je bil zbrati 100.000 besed velik govorni korpus, ga označiti z besednovrstnimi in oblikoslovnimi oznakami ter lematizirati; a zaradi zamudnosti transkribiranja in teoretičnih zank novega področja (malo tradicije govornih korpusov) je končni rezultat le polovica, in še to neoznačen korpus; za kvantitativno analizo besednih oblik bo velikost komaj zadostovala.

Poleg velikosti gradiva lahko izbiramo tudi velikost enote – jezikovnega sredstva. Ukvarjanje s kvantitativnim vidikom jezikovnih sredstev jezikovnega sistema niti ni tako redek prijem: v slovnici so marsikateri ugotovitve (tudi „pravila“) o običajnosti jezikovnih pojavov naslonjene na predpostavke velike pogostosti teh pojavov (prim. pogostost kot načelo naravnega jezikoslovja, opomba 83, str. 57), besedišče slovarjev se izbira po načelu pogostosti (prim. uvrstitev gradiva v SSKJ ali BSJ), stopice, rime idr. sredstva na različnih ravninah v literaturi so presojana kot gradniki literarnega sloga na podlagi njihove pogostosti v literarni stvaritvi itn.

V nadaljevanju dajem krajši pregled tradicije kvantitativnih jezikoslovnih raziskav slovenščine, katerih temeljne kvantificirane enote so bile besede. V pregledu niso zajete druge enote (fonemi, zlogi ipd.), kvantitativne literarne raziskave (katerih predmet so zgolj literarna besedila z vidika literarne interpretacije) in razprave s področja matematičnih ter tehničnih znanosti (te gredo običajno dlje od statistik, kar nas tukaj ne zanima; res pa ta manko kaže na revnost tega pregleda). Podobno niso zajete korpusne, bibliistične itn. kvantitativne študije o besedilih.

Tabela 4: Pregled kvantitativnih jezikoslovnih razprav o slovenščini

<i>enote v razpravi</i>	<i>naslov razprave, avtor, leto izdaj (za več gl. seznam literature)</i>
<i>besede (vrste, oblikoslovne kategorije, leme), „leksemi“</i>	Frekvenčna analiza besedišča v Tavčarjevi povesti Cvetje v Jeseni – Suhadolnik 1974
	Pogostnost in skladišna vloga sklonov v besedilu Šeligovega Triptiha Agate Schwarzkobler – Eri-Birk 1976
	Trubarjev Katekizem 1550: Konkordanca, indeks besed, pogostostni spiski – Neweklowsky 1984
	Pogostnost gramatičnih kategorij v slovenskih časopisnih besedilih – Neweklowsky 1984
	Zur Paradigmatik in Trubars Katechismus 1550 – Neweklowsky 1986
	Zur häufigkeit morphologischer Kategorien in slovenischen Prosatexten – Neweklowsky 1988
	BSJ: tabeli 11, 12 – 1999
Versuch einer computerunterstützten quantitativen Textanalyse am Beispiel des slovenischen Romans Namesto koga roža cveti – Pečnik 2002	

<p><i>razno: pisna znamenja, znakovni n-terčki, besede (oblike), besedni n-terčki, povedi; dolžine besed in povedi</i></p>	<p>Slovenski jezik, literatura, računalniki – Poniž 1974 nenatisnjeni pogostostni slovar iz sklopa priprave SSKJ, na SAZU – S. Suhadolnik idr. Nekaj števil iz Slovarja slovenskega knjižnega jezika – Jakopin 1995 Ali so rojstna imena krajša od drugih samostalnikov? – Jakopin 1996 BSJ: tabele 1–7 – 1999 Pogostnostna analiza besed iz elektronskega korpusa slovenskih besedil – Grzybek 2000 O nekaterih kvantitativnih kazalcih v slovenskem jeziku – Jakopin 2001 Entropija v slovenskih leposlovnih besedilih – Jakopin 2002 Nekaj zanimivosti iz besedilnega korpusa Nova beseda – Jakopin 2003</p>
--	---

Nasploh je v kvantitativni lingvistiki največ poudarka (več razprav, različni avtorji) na besedah, in sicer slovnično tako ali drugače označenih, kot minimum pa so v večini primerov podane statistike besednih oblik in pisnih znamenj. Iz slednjih bo možno zajemati primerjave za statistike besednih oblik v stilometričnih raziskavah korpusa StAnId. Od stilometričnih raziskav v slovenščini poznam le „Lahko računalnik prebira pesnike in ugotavlja njihov slog?“ Jureta Zupana (1994). Ta raziskava sicer sebe ne imenuje „stilometrična“ in je literarnega značaja, a pa prav tako kot moja presoja sloga avtorjev (Prešerna, Gregorčiča, Menarta) na podlagi pogostosti besednih vrst in njihovih različnic.

Za zaključek še razmislek o tem, kolikšna mora biti količina podatkov (velikost gradiva, korpusov), da bodo njene kvantitativne raziskave dale stanovitne informacije. Začnimo pri odnosu med enotami in celoto gradiva, kot ga opisuje znani **Menzerath-Altmanov zakon** – navajam ga v obliki, kot je razširjen za vse jezikovne ravnine: „Kolikor večja oz. kompleksnejša je jezikovna tvorba, toliko manjši oz. enostavnejši so njeni sestavni deli.“ (povzeto po Grzybek 2000: 145). Zakon lahko prevedemo v statistično obliko: Kolikor večje in kompleksnejše je jezikovno gradivo, toliko bolj v njem statistično prevladujejo enostavnejše, manjše in enoličnejše enote (jezikovna sredstva): ko delamo statistiko vseh različnih enot (različnic, takšnih ali drugačnih; angl. type), je pogostost enostavnejših enot bolj opazna, saj te enote izrazno (pregibno) manj variirajo same po sebi.⁵² To pa še ne pomeni, da so najpogostejše najmanjše enote enofonemske besede (v fonografskih jezikih) – te so kot jezikovna sredstva vendarle prerevne.

V korpusu StAnId prevladuje prosti govor, ki nastaja med vsemi zvrstmi besedil v najbolj kompleksnih okoliščinah: več govorcev ga oblikuje – dvosmernost sporazumevanja, časovno je bolj strnjen (manj razmisleka), krajše pasáže posameznega tvorca, enostavno kombiniranje z različnimi prenosniki (pisni, vidni, kretalni itd.) in mediji (telekomunikacijski, omejevanje vidnega stika itd.). Tudi sam prosti govor v kompleksnosti okoliščin variira (prim. podatke za korpus, poglavje IV.15), ampak lahko si predstavljamo, da bo v nenačrtovanem strnjenem večgovoru več najenostavnejših besed kot v počasnem intervjuju ob kavici, ali pripravljenem, celo branem monologu, ali knjižni literarni stvaritvi (ki ne vsebuje dialogov). Kompleksnost okoliščin sporazumevanja variira tako v besedilih govornega kot pisnega prenosnika. Pa vendar je za najbolj velike korpusne značilno, da so *za oba prenosnika* sestavljeni v največji meri iz besedil množičnih medijev (publicistike, književnosti, strokovnih člankov; RTV-oddaj, zborovanj, sej, predavanj; prim. Sinclair 1995, navaja ga tudi Gorjanc 2005: 49), za katere so prej značilne enostavne okoliščine sporazumevanja kot kompleksne, niso pa slednje izključene. Korpus prostega govora torej ima svojo vrednost, če iščemo raznolikost sporazumevanja, in v njem lahko pričakujemo med najpogostejšimi različnicami bolj pogoste najenostavnejše besede, s tem pa drugačne pogostostne sezname različnic kot v velikih korpusih.

Kaj še konkretno pomeni „kompleksna, večja“ jezikovna tvorba in kaj „majhni, enostavni, enolični“ deli v primeru pričujoče raziskave? *Kompleksno* govorno gradivo pomeni

⁵² Pravilo v obrnjeni veliki velja za opisne sisteme v znanosti in tako ne pove dosti novega: Tem natančnejši želi biti znanstveni opis jezika (ali drugega pojava), tem podrobnejše in enostavnejše kategorije oz. idealne enote uporablja; in tem kompleksnejši je ta opis zaradi pravil, ki celoto razčlenjujejo, strukturirajo in sistemizirajo te izbrane enote. Spet obrnjeno to pravilo upoštevam tudi kasneje pri definiranju sloga: v čim večji, kompleksnejši sistem primerjav vpletem besedilno gradivo, ki mu želim določiti slog, tem natančneje bom določil njegov slog. Prim. poglavje I.4.2 o strukturalnem modelu sloga in sledeča podpoglavja.

raznoverstnejše sporazumevalne dejavnike – med njimi več takih, ki najbolj vplivajo na notranjo diferenciacijo jezika: več govorcev, kombiniranje sporazumevalnih sredstev – medijev, prenosnikov, kodov. *Enostavni* deli pa so najpogostejši primerki, ki jih bomo dobili v statistiki jezikovnega govornega gradiva, kjer imamo za enote besedne oblike – različnice. *Bolj kompleksen* bo sestav besedilnega gradiva, ki ga bomo razbijali na pogostostni seznam različnic (korpusa ali izborov iz njega), bolj se bodo med najpogostejšimi primerki pojavljale besede, ki oblikovno najmanj variirajo: t. i. nepolnopomenske ali funkcionalne ali sinsemantične besedne vrste: slovničnorazmerne (veznik, predlog) in modifikacijske (členek, medmet). *Manj kompleksen* pa kot bo sestav besedilnega gradiva oziroma bolj natančno kot bo izbrano gradivo z vidika sporazumevalnih dejavnikov, bolj se bodo med najpogostejšimi besednimi oblikami pojavljale tudi takšne besedne vrste, ki oblikovno bolj variirajo (so manj enostavni deli jezikovnega gradiva): t. i. polnopomenske ali sinsemantične. Primerjava obeh sestavov besedilnega gradiva (bolj kompleksen ... / manj kompleksen ...) seveda *velja za isto količino gradiva*.

Na podlagi Menzerath-Altmanovega zakona lahko torej sklepamo, kaj je „/večja/ količina /,ki/ daje tudi večjo kakovost“ (Gorjanc 2005: 53; mesto smo navajali že zgoraj): ne le količina enot v gradivu, marveč tudi podatki o raznovrstnosti okoliščin, v katerih enote nastajajo, oziroma raznovrstnost gradiva. Za korpusne raziskave: Če je cilj raziskovalca dobiti jezikovno *raznoverstno* gradivo, mora sestavljati korpus, velik po raznovrstnosti sporazumevalnih okoliščin/dejavnikov; če je cilj jezikovno raznoverstnost kvantitativno relativno oz. (*premo*)sorazmerno ovrednotiti glede na kvantitete realnega sporazumevanja, mora iz realnega sporazumevanja izbrati sorazmeren korpus (ali iz korpusa sorazmeren podkorpus). Tako sodelujeta tudi kvantiteta in kvaliteta – sorazmerno in raznoverstno.

Cilj te naloge je ugotavljati slog, torej raznoverstnost jezikovnega gradiva. Za to potrebujemo dosti podatkov o okoliščinah/dejavnikih sporazumevanja.

3. SPORAZUMEVALNE OKOLIŠČINE BESEDILA

Kakršnokoli gradivo ima raziskava v neki znanosti, navada je, da se ob gradivu opiše tudi okoliščine njegovega nastanka. Pri opisovanju narečnega jezikovnega gradiva je utemeljitev zapisa okoliščin takšna:

Znanstveni zapis govornega dogodka (če gre za zapis umetnostnega besedila v narečju, govorimo o folklornem dogodku) ob natančni narečni transkripciji, ki navadno temelji na avdiovizualnem zajemu podatkov, zahteva tudi zapis načina in okoliščin, v katerih je besedilo nastalo – tak zapis namreč omogoča ne le analizo narečja na glasoslovni, oblikoslovni, skladenski in besedni ravnini, ampak tudi analizo govornega dogodka kot celote (s stališča slovstvene folkloristike, pragmatike itd.). (Škofic 2006: 175)

V nadaljevanju želim utemeljiti potrebo po zapisu sporazumevalnih okoliščin govornega gradiva te naloge še z drugega vidika kot v prejšnjem poglavju. Dalje še razlagam, kakšno je razmerje med metodologijama zapisovanja sporazumevalnih okoliščin in jezikovnega sporazumevalnega produkta.

Okoliščine sporazumevanja/besedil namreč jemljem kot **sporazumevalne dejavnike**. Antropogenim aktivnostim, v katerih prihaja do sporazumevanja, ni cilj sporazumevanje kot ozko „tehnični“ postopek gojenja medčloveškega odnosa s pomočjo jezikovnih znakov (besedila), ampak sporazumevanje kot pripomoček pri uresničevanju neke dejavnosti. Le naš vidik se usmerja samo na sporazumevanje. S tem, ko neko situacijo človeške dejavnosti imenujemo **sporazumevalna situacija**, želimo povedati, da nas zanima *samo funkcioniranje sporazumevanja* v njej; enako nas zanimajo okoliščine dejavnosti samo kot sporazumevalne okoliščine, torej tiste, ki imajo vlogo pri sporazumevanju (poudarjeno: dejavniki), in primerjanje njihovih vlog; in ko govorimo o **sporazumevalnem produktu**, nas zanima samo besedilo, ne pa tudi druge posledice dejavnosti, v katerih je bilo navzoče sporazumevanje – spremenjene socialne strukture, vedenje, rezultat dela, izdelki itd.

S tem, ko za gradivo naše naloge vzamemo neprekinjen posnetek govora v vsakdanu človeka, zajamemo širok spekter jezikovnih situacij, jezikovne raznolikosti. Pri *kvantitativni* jezikovni analizi takega gradiva je velika verjetnost, da bomo prišli do slovnice, podobne temeljni (prim. poglavje I.2.1.2);⁵³ ali pa do brezoblične, neuravnotežne gmote, ker je naš korpus vendarle majhen. A zanima nas jasna jezikovna (slovnična, besedna) slika specifičnih govornih situacij (diskurzov) – radi bi vzeli iz gradiva npr. samo spontani govor. Ne zanima pa nas samo ta v sodobnem jezikoslovju „popularna“ jezikovna zvrst, ampak celotno variiranje sloga govornega jezika glede na okoliščine.

Zakaj nas v to sili kvantitativna metoda analize jezika, ki smo jo izbrali (pogostostni sezname besed izbranega besedila ali (pod)korpusa besedil)? Z razlomom besedil na besede, majhne jezikovne enote, ki nosijo malo vsebine (podatka o mestu na po pogostostnem seznamu in pogostost pa sta precej abstraktna), se izgubi precej podatkov o širšem pomenu besedila, o sporazumevalni situaciji; ti podatki so sicer razvidni iz levega in desnega *sobesedila* besede (do ravnine stavkov, povedi ...; iz konkordanc besed). Zato smo se prisiljeni zateči k zunanjim podatkom (iz sekundarnega vira) o konkretni sporazumevalni situaciji – k podatkom o *zunanjem* *sobesedilu/kontekstu*. Če torej kvantitativne metode jezikovnega gradiva ne želimo zamenjati za kakšno kvalitativno, moramo biti kvalitativni vsaj v razločevanju sporazumevalnih situacij/okoliščin/dejavnikov. Kvantitativne analize pa zahtevajo tudi svoje značilnosti gradiva: predvsem zadostno kvantiteto (npr. analizo nekega diskurza, za katero bo premalo gradiva, bom pač moral izpustiti).

Model opisa sporazumevalnih okoliščin mora biti takšen, da natančno zajame prav tiste dejavnike, ki se pojavljajo in variirajo v govornem sporazumevanju; splet le-teh je na prvi pogled bolj raznolik kot v pisnem sporazumevanju (vsaj kar se tiče tiskanih medijev), ali pa vsaj drugačen. Tak model opisa bo razločeval mnogo dejavnikov, za vsakega pa nudil čim bolj enostavne, konkretne kategorije opisa (s tem se bomo bolj izognili tudi prejudiciranju predstav o vrstah besedil/sporazumevalnih situacij).⁵⁴ S podatki o sporazumevalnih okoliščinah bomo v modelu opisa ravnali podobno kot z jezikovnim gradivom v jezikovnem sistemu: jih strukturirali, tj. primerjali med seboj. Uporabnost tega modela pa se bo najbolj izkazala, če bomo s pomočjo izbora določenih in enovitih sporazumevalnih okoliščin izbrali tudi oprijemljivo jezikovno gradivo za kvantitativne analize.

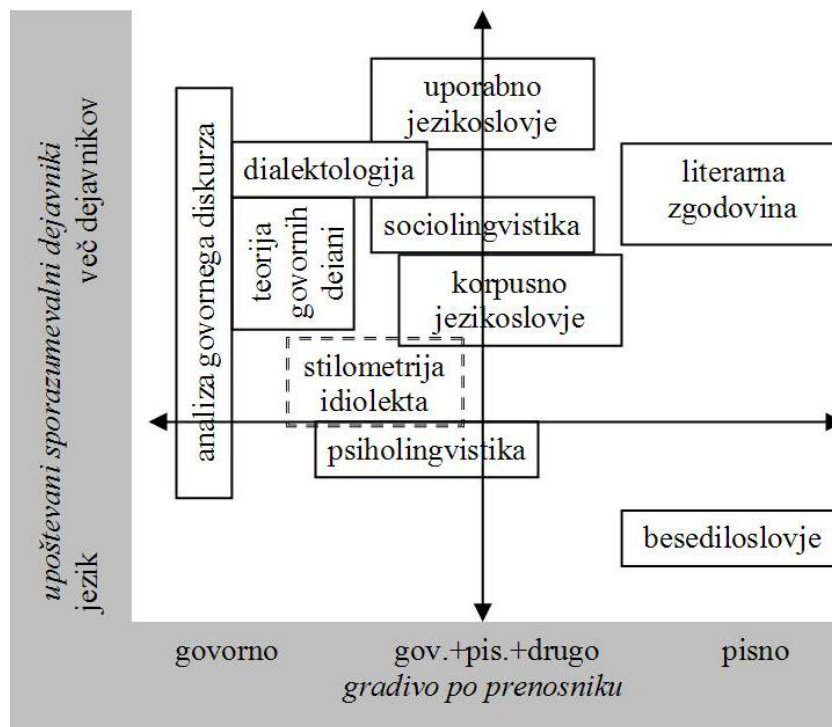
Ob jezikovnem gradivu se bomo nekoliko ukvarjali torej še s kompleksnejšimi pojavi z nekimi čisto nejezikovnimi področji (s tvorci in naslovniki (soudeleženci, njihovo dejavnostjo), s krajem, časom, prenosnikom, medijem, kodom ...), a še vedno kot jezikoslovci, in to brez poglobljene refleksije sporazumevalnih okoliščin (od kod le-te prihajajo in kakšne so same po sebi izven konkretnih sporazumevalnih situacij). S tema dvema vrstama gradiva hkrati se ukvarjajo še nekatere jezikoslovne vede; pozicijo te naloge v razmerju do njih (tradicionalno pojmovanih) hočem prikazati s pomočjo spodnjega grafa.

⁵³ Glede gradiva, iz katerega dobimo „temeljno“ slovnico: Enako se zgodi pri velikih (ogromnih) kvantitetah korpusov besedil, pa pri zajemu leposlovja, ki poustvarja veliko govornih situacij.

Glede tega, preplet česa (katerega gradiva) pripelje do temeljne slovnice kot rezultata jezikovnih analiz: S sistematskega vidika smo govorili o temeljni slovnici, ki nastane s prepletom izraznih sredstev na različnih jezikovnih ravninah (in še o drugih vidikih na *temelnosti* temeljne slovnice; gl. poglavje I.2.1.2). Toda ali ne gre pri temeljni slovnici v bistvu za preplet *socialnih situacij* (ne pa zgolj jezika kot njihovega produkta): za osnovne (pogostejše?) in za specifične (redkejšje?) situacije sporazumevanja?

⁵⁴ Za vsako zbirko podatkov nasploh velja, da je tem bolj fleksibilna, na čim bolj enostavne skupine so razporejeni njeni podatki. To načelo je veljalo tudi pri prenašanju sporazumevalnih okoliščin v računalniško zbirko podatkov v programu MS Access, gl. uvod v poglavje II.

Slika 2: Razmerje stilometrije idiolekta do drugih ved s funkcionalističnim pogledom na jezik



Naš pristop je korpusni, blizu celo gradnji referenčnih/uravnoveženih korpusov oz. takih, ki pokrivajo množico različnih sporazumevanj; pa vendar se od nje razlikuje po majhni količini in neuravnoveženosti gradiva (sploh pa ni pisnega gradiva, čeprav se pisnemu gradivu približujemo z ortografsko transkripcijo) in po bolj podrobni obdelavi okoliščin kot pri pisnih korpusih. Po drugi smo blizu analizi govornega diskurza po gradivu (govoru) in po zajemu govornih situacij (diskurzov); od nje pa se razlikujemo po metodi – kvalitativne podrobnosti (nad)skladenjskih enot nas ne zanimajo, pač pa le statistika besedja.

Do zanimive ugotovitve pridemo, če jezikovna sredstva in druge sporazumevalne dejavnike povežemo v isti sistem, v hierarhično razmerje, kot na zgornjem grafu: jezikovne enote od besedila navzdol kot manjše enote oz. nekompleksni dejavniki – kod, mediji, soudeleženci in socialne strukture pa od besedila navzgor kot večje enote oz. kompleksnejši sporazumevalni dejavniki. (Z drugimi besedami: Preskok z jezikovnosistemskega na sporazumevalnosituacijski vidik jezika je pravzaprav zgolj preskok na višjo ravnino jezikovnih-sporazumevalnih enot.) Ugotovimo, da je naše naslanjanja analize besedja na sporazumevalne okoliščine kot na okvir ali oporo dokaj *relativno*. To zato, ker sporazumevalne okoliščine niso apriorini temelj in imajo svoje ozadje, ki ga puščamo med raziskavo v nemar. Ravnina besed in ravnina spleta sporazumevalnih dejavnikov (ki določajo nek diskurz) sta po kompleksnosti pač toliko narazen, da to neabsolutnost dejavnikov spregledamo in vendarle še lahko marsikaj zanimivega povemo o jezikovnem gradivu. Moramo pa vzeti v zakup, da bo v primeru, da izberemo jezikovno gradivo zelo ozko, kot produkt zelo podobnih in enovitih sporazumevalnih okoliščin, jezikovno gradivo izkazovalo tudi enovitejši in specifičnejši diasistem (slovenskega) jezika. Vprašamo se lahko tudi, ali so razločnejše, natančnejše izbrane sporazumevalne situacije, njihov jezik – bolj *predvidljive*? Gotovo bolj od splošnejše nakazanih diskurzov – v raziskavi, metaopisu, ne pa realno, dokončno: prihodnje sporazumevalne situacije, njihova jezikovna sredstva ... so vedno odprti.

Več o tem, da izbor jezikovnega gradiva na podlagi izbora sporazumevalnih okoliščin deluje pri razbiranju sloga kot izbor sistema primerjav, v poglavju I.4.2.

3.1. Model sporazumevalnih dejavnikov

Preden naštejemo konkretne pojave, ki jih imenujemo sporazumevalni dejavniki, vzemimo v ozir še znane modele iz funkcionalističnega jezikoslovja, ki govorijo o sporazumevalnih dejavnikih oz. okoliščinah. Model Romana Jakobsona opisuje jezikovne funkcije konstitutivnih dejavnikov jezikovnega dejanja (sporazumevanja):

Slika 3: Jakobsonov model konstitutivnih dejavnikov sporazumevanja⁵⁵

<i>konstitutivni dejavnik</i>	pošiljalec / <u>tvorec</u>	stik ali kanal / <u>prenosnik</u>		<u>naslovnik</u> / (s)prejemnik
		sporočilo / <u>besedilo</u>	kontekst oz. cilj reference sporočila / tema ali predmet	
<i>možna jezikovna funkcija besedila glede na dejavnik (besedilna vloga)</i>	<i>ekspresivna / emocionalna (izrazna / čustvena / razodevalna)</i>	<i>poetična / estetska (lepotna)</i>	<i>referencialna (nanašalna / prikazovalna)</i>	<i>apelativna / konativna (vplivajska / pozivna)</i>
		<i>metajezikovna (jezikovna)</i>		
		<i>fatična / kontaktna (govornostikovna)</i>		

Sporočanjski model (ali veriga) Jožeta Toporišiča (SS²⁰⁰⁰: 696 sl.) je bolj usmerjen na zaporedje poteka sporazumevanja, kar nas tukaj ne zanima, refleksija konstitutivnih dejavnikov sporazumevanja pa ni več „v“ besedilu kot pri Jakobsonu, marveč v zavesti soudeležencev. Toporišič enako razčlenjuje dejavnike na 6 različnih, prav tako na dveh ravneh – same po sebi (ter v zaporedju modela nastopajoče) in v zavesti (obeh) ssporazumevalcev. Razlika med šestimi dejavniki Jakobsona in Toporišiča je malenkostna: eden od dejavnikov pri Toporišiču se ne imenuje tema, ampak stvarnost; toda kaj pa je zavest in izražanje o stvarnosti drugega kot tema izražanja?

Naš model v nalogi je sinteza⁵⁶ doslej omenjenih dejavnikov; je statičen (ne ozira se na proceduralnost, razen kolikor pač med kakšnimi dejavniki razločamo tudi na podlagi neokratnega delovanja v sporazumevanju); je sintetičen: zajeti skuša vse, kar k sporazumevanju pripomore, v njem nastopa in ga oblikuje; je funkcionalističen in strukturalističen, ker presoja in primerja različne okoliščine glede na njihovo vlogo v sporazumevanju kot **sporazumevalne dejavnike**.

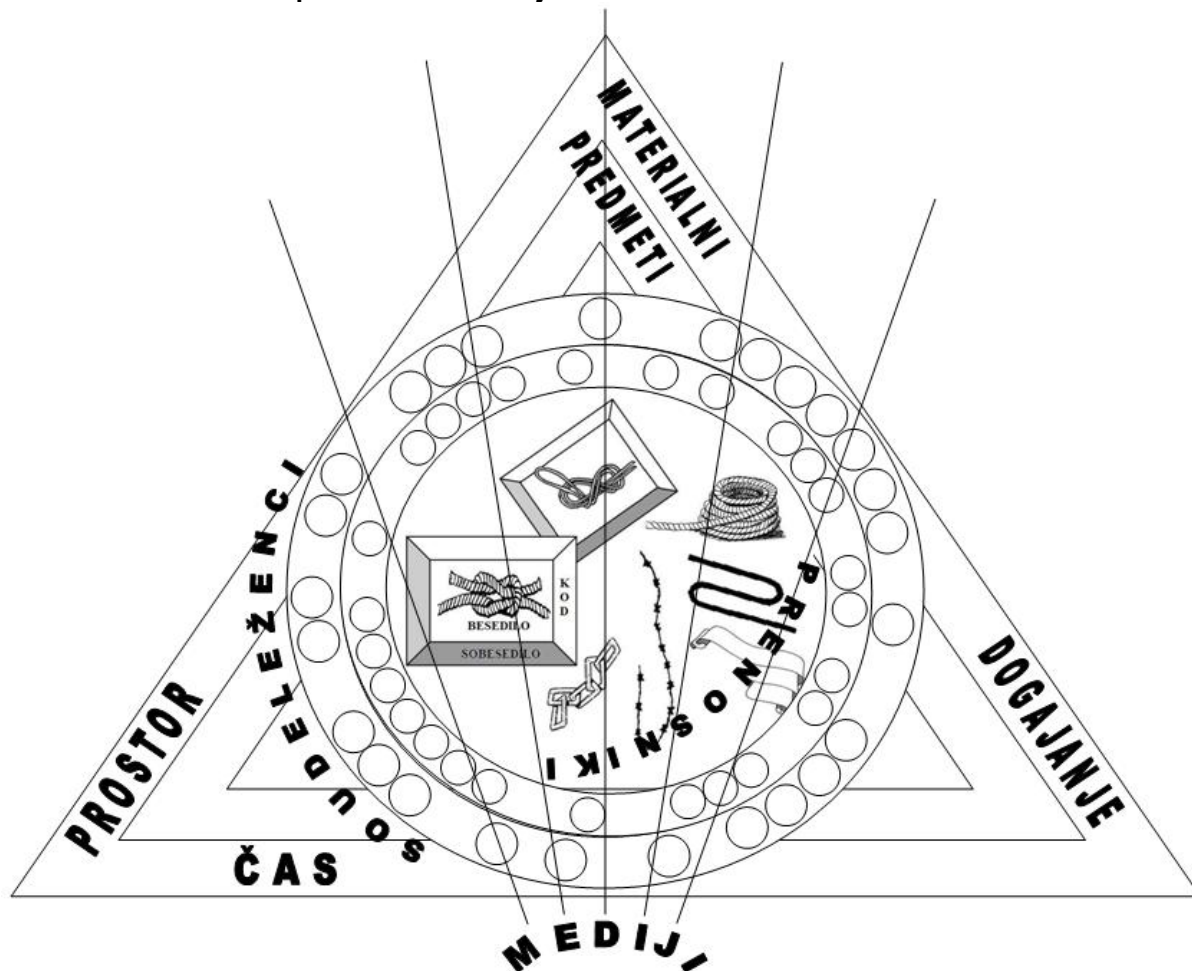
Ta model je zgolj opisni, raziskovalni pripomoček (na poti h konkretnim analitičnim prijemom, metaoznaki besedil), poimenovanje dejavnikov in skica sta le eden od konceptualnih pristopov. Model želi dajati celovit pregled nad sporazumevalnimi dejavniki in jih tudi deliti v 4 skupine. Le-te so opisane v naslednjih treh podpoglavjih in v naslednjem poglavju, začenši pri od besedila najbolj oddaljenih dejavnikih:

- 1)) neodvisne okoliščine sporazumevanja: prostorski, časovni in dogajalni okvir sporazumevalne situacije ter druga materialna, neživa stvarnost;
- 2)) soudeleženci sporazumevanja: socialni okvirji in personalne značilnosti, oboje kot stalno in kot aktualno ob sporazumevalni situaciji;
- 3)) sporazumevalna sredstva: mediji, prenosniki, kodi: različni možni in izbrani v sporazumevalni situaciji;
- 4)) besedilo oz. sporazumevalni produkt: zaokroženo besedilo in njegova sobesedila.

⁵⁵ Jakobsonov model je povzet po Jakobson 1989(1989): Lingvistika in poetika. *Lingvistični in drugi spisi*. Ljubljana: ŠKUC, ZIFF (Studia humanitatis, letnik 4). 147–190.: 153–159, podčrtani izrazi pa so Toporišičevi (za besedilne vloge vzeto iz SS²⁰⁰⁰: 724–725).

⁵⁶ V našem modelu je dodan poseben koncept medija/posrednika, ki je v Jakobsonovem modelu skrit v kanalu/prenosniku. S tem dodatkom sta v našem modelu tudi prenosnik in kod bolj specifično določena: naš „medij“ prevzame vlogo Jakobsonovega „prenosnika“, Jakobsonov „kod“ pa razdelimo po strukturalističnem pojmovanju jezikovnega znaka – naš „prenosnik“ nam pomeni nabor možnosti zaznamujočega, naš „kod“ pa nabor možnosti zaznamovanega.

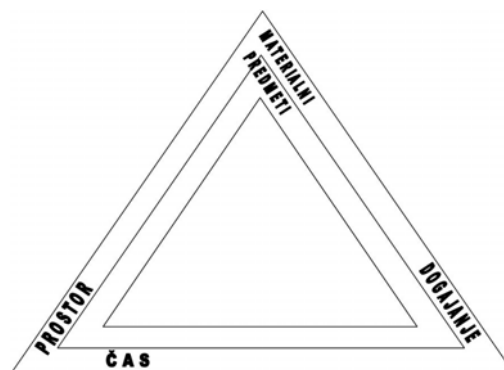
Slika 4: Model sporazumevalnih dejavnikov



V nadaljevanju so dejavniki zgolj sumarno teoretično opisani, za vsakega pa sta podani v okvirčku tudi šifra in oznaka podatka – metaoznake besedila, s katerim želim „zgrabiti“ dejavnik v korpusu. Oznake s številko so oznake besedil kot celot (npr. 18 prostor), oznake brez številke (npr. prostor) pa so tekoče oznake sprememb okoliščin med besedilom, v partituri transkripcije. Različne okoliščine so lahko opisane pod podatkom 0 ostalo in oznako razno. Podrobneje je zapisovanje teh podatkov opisano v poglavju II.8.

3.1.1. Neodvisne okoliščine sporazumevanja

Na sliki modela sporazumevalnih dejavnikov smo neodvisne okoliščine sporazumevanja prikazali kot trikotnik, ki sega najdlje v stran od besedila. Gre za tiste okoliščine, za katere smo pravzaprav že lahko v dvomu, ali še sploh imajo vlogo pri sporazumevanju. Zdi se, da vplivajo na besedilo/sporazumevanje, a niso neposredno odvisne od soudeležencev sporazumevanja. Navajeni smo jih omenjati in opisovati, da si sporazumevanje lažje predstavljamo in ga primerjamo z drugimi situacijami. Tu in tam kak del teh okoliščin postane predmet reference besedila ali (materialni predmeti) celo postane sporazumevalno sredstvo.

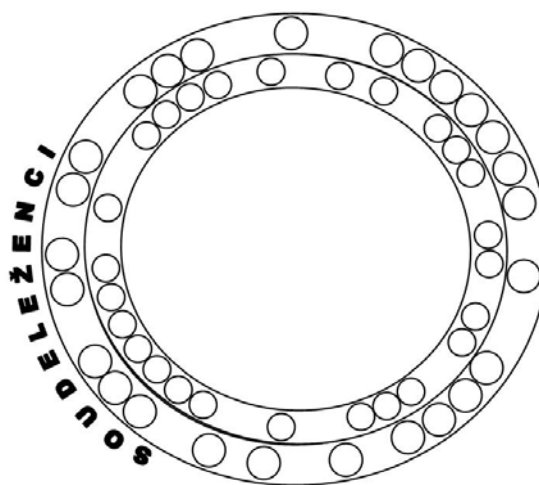


<i>DEJAVNIK</i>	<i>širše</i>	<i>ožje</i>
PROSTOR je omejevalec gibanja in prenosnika, njegove materialne narave. MATERIALNI PREDMETI so manjši deli tega prostora, ki imajo bolj specifično vlogo v človeškem življenju (dogajanju, spodaj) in so zato večkrat predmet referenc besedila.	Označimo geografsko razsežnost, odvisno od širine jezikovne raziskave: država, pokrajina; naša najširša oznaka prostora je [17] kraj (naselje).	Ožja oznaka prostora nam daje sliko o potencialnih možnostih razvoja sporazumevalne situacije: gibanje soudeležencev, omejitve medija, prenosnika: [18] prostor – kot soba, lahko celo pozicija v prostoru ali opis gibanja med sporazumevanjem – [prostor].
Oznaka ČAS ni nič drugega kot oznaka tekočega obsega sporazumevalnega dogajanja glede na drugo običajno dogajanje v ritmu človeka, narave.	Najširša oznaka je umestitev v družbeni/kulturni konsenz, torej brez ozira na ritem življenja soudeležencev sporazumevanja: [06] datum snemanja (leto, mesec, dan).	Ožja oznaka umešča v dnevni ritem življenja – dnevni čas: [15] čas začetka, [16] čas konca – ali daje podatek o časovnem obsegu: [14] trajanje.
DOGAJANJE IZVEN SPORAZUM. SITUACIJE je običajno obsežnejše dogajanje, ki sporazumevalno situacijo presega tako časovno kot po številu soudeležencev. Posebej je treba omeniti dogajanje, za katerega je predvideno, da ga uporabnik gradiiva oz. analitik lahko zgreši, pa je važno za razumevanje gradiiva.	Na to mesto sodijo obrazložitve o namenih diplome, o odločitvah glede obdelave in sami proceduri obdelave; bolj splošne utemeljitve in konkretni podatki: [A] ID stopnje korpusa, [B] ustvarjalec (korpusa), [C] transkripcijska konvencija itd., [02] datoteka, [03] posnetek, [07] začetek prepisovanja, [08] zadnje urejanje in podatki o datotekah v računalniški strukturi; [19] kvaliteta (posnetka).	Nekatera dogajanja so zapisana najprej v dnevnik snemanja in nato naprej – pod splošno oznako [00] ostalo za posamezno besedilo (npr. o zlepljanju posnetkov v eno besedilo) ali [razno]; za tekoče aktualnosti –, nekatera dogajanja pa so zaznavna tudi na zvočnem posnetku le kot šumi, hrup, motilci prenosnika in medijev in opisana ob oznaki [šum].

Komentar: Kaj je „neodvisno“? Življenje je povezano, tudi te „neodvisne“ okoliščine so delo človeških rok: skupnosti ustvarjajo prostore za institucionalne dejavnosti, stavbe, naselja, kulturno krajino; tudi merjenje in pojmovanje časa je zamišljeno po merah človekovega duha in potreb skupnosti; enako trajne aktivnosti (pravni red, tradicije ipd.) v človeški zgodovini. Gre le za abstraktnější, od konkretnega (od-ločenega) sporazumevanja bolj oddaljen, neaktualen sklop okoliščin; bližji sklop pa je predstavljen v naslednjem podpoglavju. Ločitev med tema sklopoma: preprosto izbor konkretne sporazumevalne situacije z opisom njenih meja – začetka in konca (ne v smislu linearnega časa, ampak v morju človeške kulture, dejavnosti, zgodovine).

3.1.2. Vpliv soudeležencev na sporazumevanje

Na sliki modela sporazumevalnih dejavnikov smo soudeležence sporazumevanja prikazali kot posameznike in skupine na kolobarju, ki obdaja sporazumevalna sredstva in besedilo. Ta celostni zajem okrog besedila prikazuje pomembno vlogo, ki jo imajo soudeleženci: vse sporazumevanje, tvorba in sprejemanje, gre skozi njihovo dojetje stvarnosti, celo v prejšnjem poglavju omenjene „neodvisne“ okoliščine sporazumevanja. Pa vendarle imamo soudeležence za dejavnik



sporazumevanja le v tistih primerih in na takšen način, ko in kot njihov vpliv lahko spoznamo kot kultiviranost, antropogenost, npr.:

razlike med besedili so pogojene predvsem s posameznikovo jezikovno kultiviranostjo, manj pa z okoliščinami pojavljanja besedila (formalnost : neformalnost govornega položaja, javnost ter prenosnik). (Kranjc 2004: 395)

<i>DEJAVNIK</i>	<i>razmeroma stalno</i>	<i>aktualno v situaciji</i>
POSAMEZNIK kot bodisi tvorec bodisi sprejemnik je glavna okoliščina sporazumevanja; le preko njega ostale okoliščine pridejo do izraza v besedilu in vplivajo nanj. Kljub množici posebnih oznak ga v korpusu najpodrobneje opisuje vsebina njegovega govora.	Stalna je osebna identiteta govorca, od katere označujemo: [I] ID govorca, [II–IV] ime/priimek/vzdevek, [VI] poznanost (ustvarjalcu/uporabniku) in [X] razmerje do 001Mz, [V] spol, [VII] starost, [VIII] govorno poreklo (narečno/socialnozvrstno, [D] jezik), [IX] izobrazba, [0] ostalo. Druge možne oznake: poklic, družbeni razred, anketiranje njegovih prepričanj ...	Tudi posameznik se iz besedila v besedilo spreminja v skladu z drugimi okoliščinami; vzamemo statistiko njegovega govora v besedilu: [10–12OB] [izjave/povedi/besede, aktualno socialno vlogo [XOB] vloga] in relevantne posebnosti v posameznem besedilu [00OB] ostalo; lahko bi primerjali še npr. njegovo osebno psihofizično počutje (po poligrafu).
SOCIALNE SKUPINE sicer slonijo na posameznikih, a niso odvisni le od posamezne osebe, zato jih omenjamo posebej. Oznake so precej abstraktne, za podrobnejšo analizo bi bilo potrebni sociogrami, dosti bi povedali tudi kvalitativni pretresi jezikovnega gradiva.	Stalnih socialnih pojavov in njihovih značilnosti posebej ne opisujemo (npr. institucij, literarnih slogov oz. obdobji; razlik v socialnem statusu, ki se odražajo v vikanju, naslavljanju ali v strukturi besedil), razbiramo le [33] govorni položaj] (samo razlikovanje javni/zasebni, možno bi bilo še formalni/neformalni ali uradni/neuradni). Nekaterih metaoznak sporazumevalnih situacij in besedil (z omejenimi nabori vrednosti!) pa lahko hkrati uvrstimo tudi kot oznake socialnih pojavov: [31] socialna zvrst (jezika), [32] namen sporazumevanja, [34] funkcijska zvrst (jezika).	Trenutno socialno situacijo sporazumevanje dobro odlikava in jo opisuje več oznak: [21] tema (ni je stvari, o kateri ne bi mogli govoriti), [22] dejavnost in [dejavnost: [23–24] število govorcev/soudeležencev, [25] zavest snemanja, [35] smer sporazumevanja. O smeri sporazumevanja gl. spodaj.

Od soudeležencev (pa še od medija) je močno odvisna **smer sporazumevanja**, ki jo lahko vzamemo tudi kot samostojen sporazumevalni dejavnik. Prav med govorom, sploh pri mediju osebnega stika njeno spreminjanje najlažje opazujemo. Informacije (sporočila) lahko *znotraj besedila* tečejo:

- le od tvorca proti naslovniku, ki sporočilo sprejme⁵⁷ (pa čeprav je tvorec več – govorijo drug mimo drugega) – enosmerno sporazumevanje;
- lahko pa se nagovorjeni naslovnik odzove in obrne smer nazaj k bivšemu tvorcu s povratnim sporočilom – ker tok sporazumevanja ni prekinjen in so ostale okoliščine še iste, je to še vedno del enega samega besedila – dvosmernega;
- vedno pa je sprejemalec sporočila vsaj tvorec sam (če poleg njega ni naslovnika, smeri pravzaprav ni).

Z vidika smeri sporazumevanja so v gradivu še kompleksnejši primeri, ki pa so ponavadi le del besedila, zato jih podrobneje ne reflektiramo: primeri, ko je neko besedil sekundarno, terciarno itd. rabljeno, npr. dialog v poročilih, ki ga poslušamo doma pred TV, objektivi jezik v

⁵⁷ Šele sporočanje + sprejemanje = sporazumevanje. Na sprejemanje kot sporazumevalno aktivnost in dejavnik sporazumevanja pogosto pozablamo; a lahko sporazumevalne okoliščine oz. besedila kategoriziramo tudi s stališča poslušalca (Vogel 2004: 458). Pri opisu okoliščin za naše potrebe ne moremo presoditi, kdo je *naslovnik* sporazumevanja (oseba, na katero tvorec misli pri ubesedovanju), niti, kdo je prejemnik (ki sporočilo sprejme in se miselno odzove nanj) – ne poglobljamo se za ti presoji v besedilo; ocenimo le število *udeležencev* sporazumevanja – oseb, ki jih jezikovna sredstva preko medija dosežejo ne glede na to, to hkrati naslovniki in prejemniki. (Za vir obrazložitev pojmov naslovnik, prejemnik gl. Vogel 2004: 453.)

metajezikovnih razpravah; nasploh citati ... Za potrebo ločevanja prostega govora od branih besedil se je zdelo smiselno vpeljati tudi oznako 30) prostogovornost, ki s pomočjo lestvice vrednosti opisuje, koliko je v besedilu navajanja.

3.1.2.1. Idiolekt – koncept diasistema jezika, vezan na posameznika

Način izražanja posameznika, opisanega v zgornji preglednici, imenujemo **idiolekt**. Diasistem jezika, ki ga razbiramo iz posameznikovega sporočanja oz. navežemo na okoliščino – posameznik, je najdoločnejša stopnja med diasistemom posameznega besedila in diasistemom kakega sociolekta (o le-tem gl. nižje spodaj).

„Splošno gledano, tipanja, ki spremljajo koncept idiolekta, le prevajajo potrebo po vmesni entiteti med govorom /rabo/ in jezikom /sistemom, shemo/, ali, če hočete, potrebo po konceptu govora, ki je sicer že institucionaliziran, a ga še ni mogoče radikalno formalizirati, kakor je mogoče jezik.“ (Barthes 1990: 148; pojasnila v poševnicah dodal R. B.)

Pred tem Barthes navaja te definicije oz. „tipanja“ idiolekta: „Idiolekt je jezik, kolikor ga govori en sam posameznik (Martinet) ali celotna igra navad enega samega posameznika v nekem trenutku (Ebeling).“ In omenja Jakobsona, ki da je spodbijal pomembnost koncepta idiolekta, češ govornica je zmerom družbena, tudi na individualni ravni, ko komuniciramo (Barthes 1990: 147). Gotovo je jezik oz. besedilo, ločeno od svojega tvorca, družbena tvorba, a to že ne pomeni, da posameznikove diferenciacije te tvorbe ne bi smeli razlikovati od tvorb drugih posameznikov. Pravzaprav govorimo o slogih.

Na začetku XIX. stoletja je W. von Humboldt uveljavil znamenito distinkcijo med *érgon* in *énérgεια*: jezik, ki je pasivni instrument sredine, in jezik, ki je ustvarjalni podvig posameznika. Jezik je tvorba ustvarjalca. Le-ta živi v sredini, ki ga sprejema. Zato podlega zakonom psihologije in sociologije, na katerih se snuje vsakokratna podoba življenja. Jezik je torej neposredno odvisen od posameznika, od njegovega načina življenja in od okoliščin tega življenja; odvisen je od človekovega temperamenta, kulture, starosti in od spola. Takšen in tak jezik je torej bistveno dejstvo stila. (Pogačnik 1967: 178)

Za način izražanja, ki veže osebe istega socialne skupine (pa tudi institucije, gibanja ...; o socialnih skupinah gl. zgornja preglednica), bi lahko tukaj, kjer se osredotočamo predvsem na idiolekt, uporabili izraz **sociolekt** kot edini nasprotni izraz idiolektu (čeprav imamo še izraze register, žanr, diskurz, socialna in funkcijska zvrst ...).

Idiolekt imamo za stalne jezikovne specifičnosti osebe, ki odkrivajo njene osebne lastnosti in v glavnem ne variirajo glede na okoliščine (sociolekt pa to relativno bolj). Pri virih interpretacije idiolekta pa moramo paziti: posameznikove stalne osebne lastnosti, ki bi jih morda razpoznali iz slogovnih analiz, so razpoznavne tudi iz tipičnih nejezikovnih lastnosti posameznika in njegovega okolja, opisanih kot okoliščine sporazumevanje, še bolj enostavno kot naše analize. Metodo razpoznavanja govorcev iz paraverbalnih značilnosti govora pa omenjam na koncu uvoda v poglavje I.5.1.

Soudeleženci kot sporazumevalni dejavnik imajo veliko lastnosti, tukaj pa označujemo vendarle samo tiste, ki se nanašajo na sporazumevanje. Na tem področju pa je povezanost z dejavniki iz naslednjega podpoglavja (sporazumevalnimi sredstvi) zelo močna:

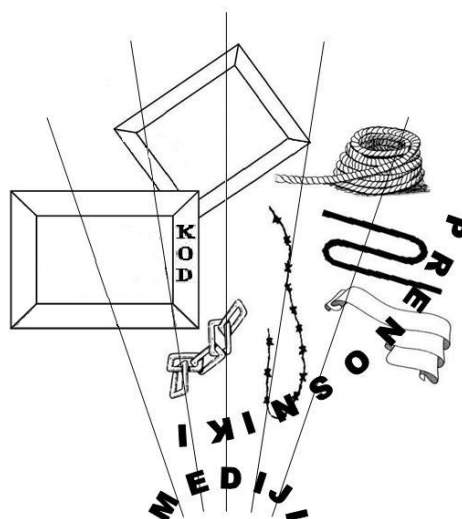
- izbor besedila (citata, intertekstualnost) je izbor njegovega avtorja in ozadja;
- izbor koda (nacionalnega jezika ali socialne zvrsti) je izbor njegove skupnosti, odprtosti sporazumevanja (npr. manjše ali večje zahteve po t. i. funkcionalni pismenosti), razmerij v skupnosti itd.;

- izbor medija (posrednika) je izbor soudeležencev in njihove možnosti sodelovanja v sporazumevanju;
- izbor prenosnika (kanala, prenosniške zvrsti jezika) je izbor oblikovanja jezikovnih sredstev in strukturiranja pomena;⁵⁸ gre tudi za kombiniranje prenosnikov in „večdimenzionalnost“ sporazumevanja.

Ločitev med to in naslednjo skupino sporazumevalnih dejavnikov je preprosto konkreten izbor sporazumevalnih sredstev v sporazumevalni situaciji.

3.1.3. Sporazumevalna sredstva

Na sliki modela sporazumevalnih dejavnikov smo sporazumevalna sredstva prikazali tri načine: kot črte, ki presegajo marsikatero okoliščino sporazumevanja (mediji), kot material vezi oz. besedila (žica ali najlonska vrv ali vrv iz preje ali usnjen trak ali veriga ...; prenosniki), kot okvir besedila (kodi kot nabori jezikovnih sredstev).



<i>DEJAVNIK</i>	<i>druge možnosti, lahko sodelujoče kot kombinacija</i>	<i>prevladujoča izbira</i>
MEDIJ (POSREDNIK, vmesnik) je najabstraktnješe od naših sporazumevalnih sredstev; pomeni nam zbirni izraz za predmet/prostor/odnos/napravo/dejavnost, ki omogoči, da sporočilo v svoji izrazni obliki (prenosniku) pride od enega do drugega soudeleženca; ki posreduje izraz/prenosnik jezikovnih znakov, zagotavlja stik. Lahko bi rekli, da poimenujemo nekaj neoprijemljivega; opazimo ga šele takrat, ko ga ni: zvok govora ne pride do nas, zapisa sporočila ali kretnje ne vidimo (po zvočnem/svetlobnem/električnem/otipnem itd. signalu).	Npr. druge možnosti: pisni/vizualni stik (dopisovanje, nagrobnik, slika, teletekst), zvočni stik (radio, telefon, mobilni, zvočnik ali osebni stik brez vidnega in telesnega), vizualno-zvočni neosebni (TV, videokonferenca, po internetu) idr. Tekoče spremembe in kombinacije medijev so opisane tudi pod oznako <u>navzočnost</u> .	Oznaki <u>[20]</u> medij]. Npr. izbrani medij je osebni stik (zvočno-vidno-telesni, običajno le prva dva načina, a tega ne razločujemo).
PRENOSNIK (KANAL) je najkonkretnješe od naših sporazumevalnih sredstev; pomeni nam materialno možnost izražanja, v kateri ima kod/jezik nabor možnih oblik/izrazov jezikovnih znakov. ⁵⁹	Govor, ki je posledica ubeseditve drugih prenosnikov v govornem, ima oznako <u>pren</u> ; običajno gre za branje, posebej pa je označeno še recitiranje, petje ...	Na posnetku je razviden je le govorni, v transkribciji „čistega“ besedila pa le govorni (trak [v] v partituri transkribcije). Paraverbalne govorne zvoke pojasnjujeta oznaki <u>nv</u> in <u>iz</u> , ostale nezvočne pa <u>signal</u> in <u>šum</u> .

⁵⁸ Posebnost govornega prenosnika ni toliko v njegovi fizični naravi, ločeni od človeka (kot pri pisnem prenosniku), pač pa v tesnejši psihofizični povezanosti s človeško sposobnostjo sporazumevanja. Ta sposobnost je odražena tako v fizičnem tvorjenju/sprejemanju jezikovnih sredstev kot v njihovem (spet fizičnem) oblikovanju oz. strukturiranju (skladanju, povezovanju).

⁵⁹ V primerjavi z medijem lahko rečemo, da je prenosnik specifična oblikovanost materiala za potrebe jezikovnega znaka, medij pa fizikalno delovanje istega materiala – medij deluje tudi, če sporočila ni in iz delovanja ne razločimo oblikovanosti. (Termina prenosnik in posrednik sta si nekoliko prepodobna za jasno razlikovanje; za primerjavo prenosnika in medija gl. tudi opombi 5, str. 12, in 56, str. 38.)

<i>DEJAVNIK</i>	<i>druge možnosti, lahko sodelujoče kot kombinacija</i>	<i>prevladujoča izbira</i>
KOD (NACIONALNA / SOCIALNA ZVRST JEZIKA) nam pomeni vednost/znanje, ki uporabniku jezika omogoči, da prenosniško/materialno oblikovanost besedila prevaja v vsebino/pomen. V idealni podobi je toliko temeljna, da je prenosljiva z enega uporabnika na drugega.	Krajše pasaže tujih jezikov v gradivu imajo oznako [tuje], daljše pasaže sploh niso transkribirane, a je o tem opomba. Tekoče spremembe socialnih (ali drugače definiranih, npr. časovnih) zvrsti jezika niso označene.	Kot privzet nacionalni jezik se v gradivu smatra slovenščino, prevladujočo socialna zvrst posameznega besedila pa pojasnjuje vrednost podatka [31) socialna zvrst] (jezika).

Komentar: Ko to skupino sporazumevalnih okoliščin primerjamo z „neodvisnimi“ in „soudeleženci“, se lahko vprašamo: Kje so razlike sporazumevalnih sredstev? Mediji včasih prebijejo vse (osebni stik in tisk – tudi zapis-tisk zvoka – nista omejena časovno, sta pa prostorsko – preko prostora še število udeležencev –; telekomunikacije obratno), pri tem so bistveno odvisni od materialne narave prenosnika. In po drugi strani: ko menjujemo prostor, se deli osebnega stika (govorno-slušni stik, vidni stik, telesni stik) hitro lahko spremenijo. Prenosnik določa strukturiranje verbalnih in paraverbalnih enot jezikovnih znakov, besedila. Toda pri nekaterih prenosnikih (npr. kretnje kretalnega jezika) razlike med kodi (ki so na skupnosti uporabnikov omejeni nabori jezikovnih znakov) skoraj izginejo (čeprav: poznamo kulturne razlike med pomeni telesnih govoric).

Vendarle se „sporazumevalnim sredstvom“ pozna, da so od vseh sporazumevalnih okoliščin najbližje besedilu: večina izbranih za posamezno besedilo ima še modifikacijo znotraj samega besedila. Predvsem zaradi sporazumevalnih dejavnikov je bilo potrebno v partituro transkripcije dodati traka [pv] (paraverbalno) in [Ozadje], na katera je predvideno zapisovanje tekočih sprememb sporazumevalnih sredstev.

Vse doslej našteje skupine sporazumevalnih dejavnikov smo lahko same med seboj razdelili znotraj odnosa konkretno (izbrano, aktualno, sodelujoče,) vs. abstraktno (neizbrano, nerelevantno, neaktualno, kar ni vsebina sporazumevanja, kar ni sodelujoče); to so na sliki modela skušali izraziti dvojni liki – dva trikotnika, dva kolobarja, po več primerov sporazumevalnih sredstev. Tudi ta delitev je, podobno kot delitev na skupine, le umetna, sistematska zaradi lažjega opisa in odkrivanja razlik med dejavniki; ti so pri razumevanju vendarle močno prepleteni med seboj.

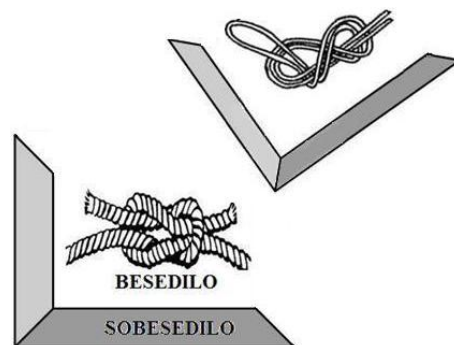
Tabela 5: Ohlapna in aktualna plat dejavnikov sporazumevanja

<i>skupina dejavnikov sporazumevanja</i>	<i>ohlapno ozadje sporazumevalne situacije</i>	<i>aktualizirano v sporazumevalni situaciji</i>
<i>neodvisne okoliščine sporazumevanja</i>	širši, obsežnejši vidik okoliščine	ožji, podrobnejši podatek o okoliščini
<i>vpliv soudeležencev na sporazumevanje</i>	razmeroma stalna identiteta in socialne strukture sodelujočih	aktualno dogajanje, razmerja in vloge v situaciji
<i>sporazumevalna sredstva</i>	druge možnosti, lahko sodelujoče kot manj pogosta kombinacija	prevladujoča izbira
<i>sporazumevalni produkt – vezana jezikovna sredstva/znaki</i>	diskurz, sobesedilo, sklopi besedil	posamezno besedilo, njegovi deli in parabesedno

Podobno notranjo dvojnost vlečemo tudi naprej, na raven sporazumevalnega produkta – besedila.

3.2. Vrste besedil in meje med besedili

Za nekatere prejšnje okoliščine obstaja tudi oznaka *zunanj* (*sporazumevalni*) *kontekst* (za razliko od *notranjega* – sobesedilnega). S tem, ko govorimo o zunanjem kontekstu, že opazujemo, primerjamo več besedil, več sporazumevalnih situacij (v realnosti jih je neskončno mnogo). Ali lahko govorimo o besedilih kot primerljivih „enotah“? Raje ne, saj je naslednja kompleksnejša ravnina, v odnosu do katere bi lahko besedila lahko bila gradbene enote ali sestavni deli, nejasna, preobširna, preraznolika. Lahko sicer besedila razmestimo v različne sporazumevalne „toge“ – diskurze, pa žanre, sloge, obdobja, zvrsti ... (torej zbirno). Te „skupine“ besedil opisujemo pravzaprav s sporazumevalnimi dejavniki, ki smo jih našeli v predhodnih poglavjih; to so hkrati tudi kriteriji, po katerih predalčkamo konkretna besedila na vrste, v nikdar zaobsežen sistem. *Vrste besedil so vrste sporazumevalnih dejavnosti, njihove meje pa meje sporazumevalnih dejavnikov.*



Vse gradivo diplome (transkribcije in njim vzporedni posnetki za neprekinjeno obdobje dneva) je torej razdeljeno na „besedila“, vsako pa ima svojo oznako (01 ID besedila) in en sam opis vrste (28 oblika sporazumevanja), ki ni posebno strukturiran; več povedo oznake besedila za njihove sporazumevalne dejavnike, ki so našete zgoraj.

Tako torej konceptualno (lahko bi rekli tudi: vsebinsko) razmejimo besedila, nato jih je potrebno še tudi praktično; nekatere oznake dejavnikov tako dobimo šele po praktični razmejitvi – statistike ipd.). Najobičajneje lahko zamejimo besedila podobno kot manjša jezikovna sredstva, namreč z določanjem njihovega začetka in konca njihovega izraza (oblike/forme) v (linearnem) toku sporazumevanja. Narava zamejitve je – ali enaka naravi prenosnika: pisna besedila zamejujemo vizualno oz. prostorsko (s prazno površino, razmikom), govor, zvočne znake in kretnje pa časovno (s premolkom, tišino, nepremičnostjo; z obdobjem trajanja); – ali pa zamejujemo količino gradbenih jezikovnih enot: toliko pisnih/govornih znamenj, kretenj ... Tako besedilo z vidika nižjih jezikovnih ravnin označijo tudi podatki o njegovih sestavnih delih: (09) n-govor (poenostavljeno število govorcev), (10-12) izjave/povedi/besede (število enot), (13r) gostota (časovna strnjjenost besed pri ocenitvi ob rezanju) oz. (13) besede/s (gostota besed na sekundo vzporednega posnetka), (29) vgnezdeno (večji deli besedil, ki bi jih v drugih okoliščinah tudi lahko imenovali besedila).

3.2.1. Meje med besedili v tej nalogi

Kontinuiteta govora ima pri rezanju posnetkov prednost pred kontinuiteto tišine: še tako *kratek sklop govora* (npr. par sekund), zamejen s tišino, je lahko označen kot besedilo. Kjer razločitve kontinuitete govora s pomočjo tišine ni ali pa ni jasno, kje v tišini jo narediti (neprekinjen *dolg sklop govora*), meje besedil niso (točno) meje govora, ampak meje govor spremljajočih dejavnikov, torej krajših od govora.

Nasprotno so *krajši sklopi tišine* (nekje dolžine do dveh minut) na mejah med dalj časa trajajočimi dejavnostmi enostavno prisojeni predhodnemu ali naslednjemu besedilu oz. dejavnosti; enostavno se jih ne splača beležiti v taki obliki – kot že rečeno, do dolžine dveh minut. S „tišino“ mislimo obdobje brez govora, so pa iz posnetkov te tišine slišni drugi zvoki, ki raziskovalcu pomagajo razumeti dogajanje.

Čprav bi tudi en sam *daljši sklop tišine*, v katerem se zamenjajo odločilne sporazumevalne okoliščine (glej spodnji kriteriji za meje govora), bilo možno razdeliti na več „samostojnih“ „besedil“ oz. transkripcij, tega ne počnemo. Dlje kot dve minuti trajajočo tišino (ali vmesno obdobje, ko je bilo snemanje izpuščeno) puščamo kot eno samo „besedilo“. Oznaka „besedilo“

nam v takšnih primerih pomeni torej samo enoto gradiva, ne pa tudi množice jezikovnih znakov (vsaj ene besede ...).

Kriteriji za meje besedil so vpisani za vsako posamezno besedilo pod oznakama [26] povod [začetka] in [27] povod zaključka; najprej en sam, najvišji kriterij iz spodnje lestvice, nato vsebinska obrazložitev okoliščine.

Tabela 6: Kriteriji za ločevanje govora in spremljajočih okoliščin na enote gradiva – „besedila“

<i>kriterij</i>	<i>siceršnji podatki o sporazumevalnem dejavniku</i>	<i>obrazložitev rabe</i>
Kriteriji za sosledja neprekinjenih sklopov govora in od 2 minuti krajših sklopov tišine:		
prostor:	[18] prostor	Se nanaša na gibanje nosilca mikrofona, potrebno pa je pojasniti tudi, koga nosilec mikrofona zapusti/sreča. S spremembo prostora se pogosto vzpostavijo tudi drugačni spodnji kriteriji, pa jih v takih primerih posebej ne navajamo.
navzočnost:	[18] prostor, [20] medij, [23] število govorcev, [24] število soudeležencev	Se nanaša na gibanje, ki spremeni prisotnost/odsotnost sogovorcev/soudeležencev govora, pa tudi, kadar so uporabljeni mediji, da se vzpostavi kontinuiteta govora (npr. napoved po zvočniku – njen začetek in konec; začetek pogovora po telefonu).
dejavnost:	[22] dejavnost	Začetek/konec dejavnosti, ki ob sebi zaznamuje sporazumevanje oz. ga spremeni glede na sporazumevanje pred dejavnostjo: npr. obed, učenje, maša ...
Kriteriji za sosledja kratkih sklopov govora in od njih daljših sklopov tišine (posamezen sklop tišine pa ob tem več kot dve minuti):		
kontinuiteta:	fonografski zapis govora v transkribciji, časovno umeščen v partituri zapisa	Opis zgoraj nad tabelo.
razločnost:	[19] kvaliteta (posnetka) in [šum], trak [Ozadje]	Razločnost posnetka se tako poslabša, da ni zapisa govora in iz transkribcije razbiramo le tišino.
vir:	[14] trajanje, [15–16] čas začetka/konca	Začetek/konec zapisa (posnetka in transkribcije) je odvisen od začetka/konca gradiva, ki je na voljo.

3.2.2. Sobesedilo

Ko že poznamo meje besedil, ne moremo mimo zvez, ki med posameznimi besedili vendarle obstajajo. Besedila v teh zvezah imenujemo *sobesedilo* ali s širšega vidika *notranji* (sporazumevalni) *kontekst*; za razliko od *zunanjega*, omenjenega zgoraj, ta izraz poimenuje druge sporazumevalne produkte, njihov vsebinski in oblikovni pol jezikovnega sredstva. Sobesedilo je v podobnem odnosu do besedila kot so med seboj širše (kontekstualne) in ožje (aktualne) variante sporazumevalnih okoliščin – prim. Tabela 5, str. 44.

Podatek o zunanjem kontekstu nekega besedila najdemo v gradivu prikazan kot podatek o umestitvi v tok sporazumevalnih situacij – podatki o neposrednem prejšnjem in sledečem besedilu. O tem govorita zaporedno šifriranje besedil v [01] ID besedila (npr. 6pet07Er, 6pet08Pn, 6pet09Dm) in podatek o materinski transkribciji, ki še ni razdeljena na enote „besedila“ po kriterijih za meje besedil [01Tr] ID transkribcije. Sporazumevanja v korpusu je namreč v začetni fazi obdelave transkribirano po danih posnetkih in šele nato razdeljeni po kriterijih za meje besedil: prvo stopnjo gradiva imenujemo *StAnId.3Transkribcije* in enote gradiva „transkribcije“, drugo stopnjo pa *StAnId.4Besedila* in enote gradiva „besedila“.

Notranji kontekst oz. vsebinske sobesedilne povezave niso izrecno označene, na ravnini posameznih leksemov bi za to potrebovali pomensko določitev, slovar, na višji ravni sinopsise ali izvlečke posameznih besedil. Prava so-besedila (na besedilni ravni) lahko iščemo tako med besedili zunanjega konteksta kot med besedili s podobno vsebinsko oznako – [21] tema, [32] namen [sporazumevanja], [34] funkcijska zvrst.

Ko v korpusu delamo raziskave delov korpusa, sestavljamo podkorpuse; kriterij za povezovanje besedil v korpuse je lahko zunanji ali notranji kontekst ali drug podatek, za izbor na

podlagi katerega zadostuje že samo to, da izbrana besedila sploh so v tem korpusu (npr. statistika velikosti besedila). Tudi podkorpusni izbori pa imajo nekatere svoje metapodatke. Za govorce so to statistike njihovih izrečenih enot (10O) izjave, (11O) povedi, (12O) besede) ter podatek, v koliko besedilih v korpusu nastopajo (XIO) prisoten v besedilih). Za posamezna besedila pa je to podatek, kateri govorce nastopajo v njih (XI) prisotni govorce).

Besedilo z vidika sporazumevalnih dejavnikov nikoli ni enovito (homogeno) – s tega vidika je vendarle abstrakten, idealen analitični predmet (enota). Metaoznake zanj (oznake posameznih sporazumevalnih dejavnikov) veljajo v nekaterih primerih le za večino besedila, se pa vmes tudi spreminjajo. Bistvene spremembe so zabeležene v partituri na pomožnih trakovih [pv] in [Ozadje] – gl. poglavji II.7.3 in II.8.1.

4. SLOG RABE JEZIKA

Slog ali stil⁶⁰ je pojem, široko rabljen v najrazličnejših diskurzih, tematikah, ne le v ukvarjanju z jezikom. Poglejmo nekaj gradiva iz SSKJ (prekrivno gesli *slog* in *stil*):

Primer 8: Kaj pravi SSKJ o slogu oz. stilu

<i>podpomen</i>	<i>primeri, obrazložitev primerov, kvalifikatorsko pojasnilo</i>
1. 'kar je določeno z izborom in uporabo izraznih, oblikovnih sredstev v posameznem delu ali v več delih'	oblikovati nov stil; določiti, proučevati slog; ima jedrnat, slikovit stil – 'se jedrnato, slikovito izraža' analizirati, opisati stil avtorja (Cankarjev, Jakopičev stil; Jakopičev, Prešernov slog), dobe lingv. 'tipični izbor jezikovnih prvin v besedilu': časopisni stil 'časopisna zvrst jezika'; lit. baladni stil; umetnostni slogi/stili in smeri: arhitekturni (hiše v alpskem, panonskem slogu), glasbeni slogi; slogi v književnosti (biblijski slog; osebni slog pisanja); um. baročni, bidermajerski stil
2. 'uporaba zlasti sintaktičnih jezikovnih sredstev in izrazov glede na normo'	izboljševati, knjiž. brusiti stil; imeti slab stil; napake v stilu izboljševati, popravljati slog; slog šolskih nalog
3. s prilastkom 'skupek značilnih elementov pri opravljanju kake dejavnosti'	imeti poseben/svoj delovni stil/slog; nov režijski in igralski stil; zanimiv slog dirigiranja; slog vožnje; stil oblačenja; življenjski stil 'stalne značilnosti človekovega ravnanja, vedenja'
4. 'individualen način izvajanja športne discipline'	stil smučarskih skokov šport. ustaljen, predpisan način izvajanja športne discipline: plavalni slogi, rokoborba v grško-rimskem slogu
terminološko gnezdo	ptt odprti, tajni slog; tisk. črkovni slog; um. časovni, dorski, gotski, krajevni, osebni slog

Skušajmo povzeti, kateri vidiki so skupni vsem pojmovanjem sloga:

- 1)) (značilni, tipični) izbor elementov, določenost oz. opisanost z njim preko atributa – prilastka;
- 2)) stalnost teh značilnosti, ali njihovo delno spreminjanje;
- 3)) pripisanost posameznemu (*individualnemu*) delu/osebi ali pa več delom/osebam oz. dejavnosti.

V tej nalogi nas zanima le slog na jezikovnem področju, torej **jezikovni slog** – tipični izbor jezikovnih prvin v besedilu (gl. zgoraj, primeri za 1. podpomen). V prihodnjih poglavjih bomo na kratko pogledali, kakšna je tradicija pristopov k zgoraj naštetim vidikom sloga (izbor značilnih elementov ter njihova stalnost in pripisanost nekemu okviru), nato pa na tej podlagi pojasnili, kako se lotevamo jezikovnega sloga v tej nalogi.

⁶⁰ Slednja prevzeta beseda je morfološko bolj prisotna v poimenovanju ved, ki se ukvarjajo s slogom: stilistika, stilometrija; v nalogi pa se bomo skušali izogibati dvojnicam in uporabljati le domačnejšo besedo *slog*. V besedilu sledečih poglavij se oznaka slog nanaša na *jezikovni slog*.

Navaden postopek opisovanja (analize, primerjave, vrednotenja) sloga je tak: branje besedila in ozaveščanje o intuiciji z opisovanjem jezikovnega sloga: je jedrnat, slikovit, statičen, subtilen ipd. Tako metodo, običajno značilno za *literarno kritiko*, skušamo narediti bolj objektivno z jezikoslovnim pristopom, to je objektiviziranje občutkov o posebnostih besedila z izražanjem o konkretnih slovničnih in slovarskih elementih jezika (npr. zvočnost, besedišče, fraze). Uporabljanje jezikoslovnega metajezika in kategorij za slog nam lahko pomeni zagotavljanje dovolj objektivnih meril, da bodo lahko našo oznako sloga preverili tudi drugi (Freeborn 1996: xxi). Pa vendar tudi jezikoslovne kategorije niso absolutne; ob variabilnosti govornih situacij se pokaže nezadostnost tradicionalnega strukturalističnega modela in paleta možnih pristopov s skupnim imenovalcem analiza govornega diskurza.

Tudi pri metodah na podlagi računalniškega korpusnega in kvantitativnega pristopa in se zdi, da so te metode same po sebi *objektivne*, takšni torej tudi rezultati; kvantitativna merila so ja dana vnaprej, tudi načela transkripcije in ureditve korpusa, razčlemba naredi računalnik, vse skupaj se zdi mnogo preabstraktno, da bi bilo lahko subjektivizirano s strani človeške zavesti raziskovalca, analitika. Pa vendar ta „mit objektivnosti“ tudi za računalniško jezikoslovje ne drži. Kot smo pokazali v prejšnjih poglavjih, so vsi postopki ustvarjanja korpusa in pripravljanja analiz posledica odločitev raziskovalca, v kolikor seveda ne opusti kakšne odločitve zaradi nenatančnosti, pomanjkanja časa ali privzame sadove drugih raziskovalcev (standardni računalniški jezik, program za transkribiranje, oblikoskladenjski določevalnik, model pogostostne liste različnic ...).

Toda, kaj pomeni prepoznavanje objektivnih značilnosti sloga kot takega? Brez poznavanja besedila nikakor ne gre. Nato na podlagi občutka postavimo hipoteze o slogu in jih preverimo z ustreznimi metodami (računalniški korpus in kvantitativne analize), svoja merila (jezikoslovne kategorije) pa ovrednotimo uvodoma in pri interpretaciji.

4.1. Tradicionalni stilistični pogledi

Zgoraj opisana metodološka smer ravnanja s slogom (značilna sicer za 20. stoletje; od intuicije o slogu do sistematičnega razlaganja z jezikoslovnimi kategorijami) je vzporedna tradicionalnemu razvoju ved, ki uporabljajo slog. Prim. obrazložitev osnovne vede o slogu, **stilistike**:

Preučevanje sloga v jeziku;

tradicionalno: preučevanje variacij v rabi v literarnih in drugih besedilih;

danes, bolj splošno: preučevanje katerih koli sistematičnih variacij, bodisi v pisanju ali govoru, ki se nanašajo prej na tip diskurza ali njegov kontekst kot na razlike dialekta /geografsko vezanega govora/. Npr. slog, pripisovan javnim predavanjem, se razlikuje od sloga priložnostnih pogovorov med prijatelji; slog molitev v cerkvi vključuje intonacijo ipd. pri branju; itd.

Ob tem je raziskovalno področje *literarne stilistike* (kot preučevanja relevantnih razlik, drugačnih, kot so rabljene v dialektu ali jeziku) primerjava posameznih piscev, obdobj, žanrov.“(Matthews 1997: 357: *stylistics*.)

V 20. st. je stilistika nadgradila študije o *elocutio* v retoriki. „V največji meri je blizu literarni kritiki, /.../ njen najobičajnejši material je literaren, torej je osredotočena na besedilo. /.../ Njen namen ni zgolj opisati formalne poteze besedil same po sebi, ampak pokazati njihovo funkcionalno pomenljivost za interpretacijo besedila; včasih tudi z nanašanjem na odnos med literarnimi učinki in jezikovnimi povodi.“ (Wales 1994: 437–438: **stylistics**). V drugi polovici 20. st. so na stilistične prijeme v želji po izoginitvi nedoločnosti in manipulaciji impresionističnih sodb vplivali sodobni jezikoslovni (generativna slovnica, analiza diskurza, pragmatika) in literarnovedni modeli. Razločevanje stilistike na *literarno* in *jezikoslovno* ni jasno, saj razločuje vidika, ki pa sta

hkrati uporabljana: literarnost pomeni nagibanje k ukvarjanju z literarnimi besedili, jezikoslovnost pa jemanje modelov iz jezikoslovja.⁶¹

Vendar termin stilistika (ali *splošna stilistika*, angl. general stylistics) pokrije tudi analize neliterarnih variacij jezika (v angl. domač termin zanje je *register*). Na tem širokem področju prihaja blizu sociolingvistiki, predmet sociostilistika se ukvarja z jeziki piscev kot socialnih skupin (šol).⁶²

Stilometrija (tudi stilometrika) je poddisciplina stilistike, ki

uporablja statistične analize z namenom *določevanja avtorstva* besedil; zato zadeva pojmovanje sloga predvsem kot idiolekt. /...; npr. na primerih psevdo-Shakespearja in svetopisemskih besedil/. Običajno preučevane jezikovne značilnosti v stilometriji so besedna in stavčna dolžina, konektorji, kolokacije. Le-te niso nujno /očitne/ značilnosti iz ospredja /angl. foreground/, ampak takšne, da se zanje predvideva, da jih avtor uporablja relativno nezavedno in zato dokaj stanovitno skozi vse svoje delovanje. Postopki so v osnovi primerjalni: primerjanje izbora spremenljivk v spornem besedilu s tistimi v avtentičnem besedilu. V nasprotnem primeru je težko dognati normo za primerjavo. (Wales 1994: 439: **stylometry**)

Pravzaprav je naddisciplina stilometriji *računalniška stilistika*, ki ni nič drugega kot polje računalniškega in korpusnega jezikoslovja, ki se pač ukvarja s stilom (pogosto kot idiolektom). Dejavnosti, ki uporabljata podobne statistične pristope, sta še pripisovanje avtorstva, forenzična lingvistika itd.⁶³

V katere od naštetih prijemov sodi pričujoča raziskava? Stilistična dejavnost v njej je usmerjena k neliterarnim besedilom, predvidevanja in ugotovitve o slogu se nanašajo na variacije v govoru (prenosniško zamejeno, ampak v konkretnem korpusu je malo literarnih pasusov). Torej ne moremo govoriti o *literarni* kritiki ali stilistiki. Po statističnih analizah, preučevanih jezikovnih značilnostih (besede) in pojmovanju sloga kot idiolekt sodi v bistvu na področje stilometrije in računalniške stilistike. Ne gre pa za pripisovanje avtorstva, temveč zgolj za ocenjevanje tega, kaj lahko označimo kot značilno za posameznika – avtorstvo je znano, primerjava posameznikov med seboj pa je le eden od vidikov variacij govora na podlagi zunajjezikovnih dejavnikov.

4.1.1. Pomen sloga v literarni vedi

Stilistike se v zadnjem času ne pripisuje ne literarni vedi ne („golemu“) jezikoslovju (prim. opombo 71, str. 52). Glede na to, da je objekt te diplome neliterarno gradivo in da se ga razlaga s pomočjo jezikovnih struktur, se morda zdi vpletanje literarne vede nepotrebno. Kar se pri fenomenu sloga ne da razložiti s pomočjo jezikovnih (slo)strukturnih ali pragmatičnih enot, značilnosti, ravnin itd., je vprašanje, kaj naredi slog (oz. inherentne značilnosti sloga) za razločevalnega, posebnega, značilnega, edinstvenega ... Očitno gre za nekaj *zunajjezikovnega*, ki *pride do izraza*⁶⁴ v jezikovnem slogu, in v tem je vprašanje sloga blizu vprašanju literarnosti. Ti dve vprašanji združuje opis **literarnega (književnega) sloga**. Le-tega lahko pogojuje:

⁶¹ Toporišič 1962: 40: „bistvene razlike med obema stilistikama ni: stilistika je ena sama. Razlika je samo v predmetu: t. i. jezikoslovno zanima stil vseh sporočil /.../ Predmet literarne stilistike bi bila v nasprotju s tem le umetniška besedna sporočila“.

⁶² Povzeto po Walesovi, isto mesto. Omenja še druga sorodne vede, navajam jih kar v angl., kjer so termini očitno doma: critical linguistics, expressive stylistics, formalist in functional stylistics, linguistic criticism, new stylistics, radical stylistics in besediloslovje.

⁶³ Članki, ki ponujajo pregled kvantitativnih metod za te namene, so npr. Holmes 1994, Tweedie 2005 in Tuldava 2005.

⁶⁴ V zvezi s tem najdemo „dva temeljna pomena termina slog /.../ prvič dosledno formulirana pri Platonu in Aristotelu: za Platona je slog specifična lastnost, kvaliteta, ki jo neko delo ima in je drugo nima, za Aristotela pa je generični pojem, ki označuje različne lastnosti dela in ga ima vsako delo. Platonova ideja sloga je izšla iz gr. pojma logosa, v katerem se zedinjata misel (*ratio*) in govor (*oratio*). Slog nastane, ko se poetska vsebina odene v sebi primerno formo. Če te nima, potem delo nima sloga. Samo v absolutni dovršenosti misli in oblike nastane

izpovedni in estetski namen ter ustvarjalna moč besednega oblikovalca (individualni slog; duhovno-čustvena posebnost; način obnašanja, mišljenja, govorjenja posameznika; njegova percepcija); njegov družbeni položaj in izobrazba, poklicna, pokrajinska, rodovna in narodna pripadnost (nacionalni slog); okus časa (zgodovinski slog); literarni zgledi, zvrstna pravila; snov ubesedenja in še marsikaj (povzeto po Kmecl 1995: 42–43);

poleg tega lahko še esenca (nedojemljiva z logično analizo, pač pa z intuicijo), izbor izraza (ob enem možnem referentu oz. denotatu), razmerje do jezikovne norme (njena deviacija), učinek besedila (na bralca) ... (Dodano iz Dukat 1985: 759–764): **stil**.)

Definicije in kriteriji vrednotenja sloga se lovijo okrog tega, na kateri pogojevalec sporazumevanja in jezikovnega izraza se nasloniti, in tako tudi o tem, ali slog sploh obstaja (namreč v virtualnosti jezika kot univerzalija, nevezana na aktualizirano). Najbolj razširjene teorije sloga se osredinjajo na individualnost: dela, govornika/pisca (**idiolekt**), skupine (sociolekt), žanra, obdobja. Toda gre za individualnost, lastnost v jeziku: kako je torej možno sporazumevanje, doumetje takega jezika oz. sporočila?

To vprašanje smo srečali že pri semiotiki (poglavje I.2.2): v konkretnem besedilu je jezikovni znak aktualiziran (rekli bi tudi: individualiziran, ker vzpostavi individualno referenčno razmerje s stvarnostjo), njegov (virtualni) pomen pa obstaja le v zavesti tvorca in sprejemnika. Podobno slog ne obstaja brez izraza (kompleksnejšega, generičnega skupka izraznih lastnosti), njegov pomen pa zmoremo jezikovni uporabniki dojeti v virtualnosti. Na kompleksnejših jezikovnih ravninah (kompleksnejših od ravnine običajnega jezikovnega znaka – besede) se s tem ukvarjajo še druge jezikoslovne vede. Pravzaprav gre za zunajjezikovne dejavnike, ki jih poskušajo z jezikom oz. sporazumevanjem povezati tudi kognitivno jezikoslovje, psiholingvistika in sociolingvistika (le-ta npr. namesto obstoja sloga problematizira obstoj idiolekt nasproti sociolektu). Prim. poglavje I.4.1.2; v njem razlagam pogled na slog kot idiolekt, ki smo ga tukaj že razložili kot vezanost izraznih slogovnih značilnosti na individualnost, še z vidika ravnine in vsebine tega izraza.

Bolj preprosto kot vprašanje, kaj je literarni slog, je vprašanje, kaj nam kaka slogovna značilnost pove. Seveda je to odvisno tudi od analitičnega izhodišča – vprašanja, kaj želimo izvedeti iz literarnega primerka. Morda se zdi samoumevno, da nas zanima tisto, kar je edinstveno v literarnem delu in da je to slog. Analitično izhodišče žal pri vseh analitikih in stilističnih teoretikih ni ozaveščeno. Eni se zavedajo, da si je o jezikovnem delu možno zastaviti na tisoče vprašanj, drugi pa se začno spraševati le o tisti slogovni značilnosti, ki *nagovori njihovo osebno slogovno dojemljivost*, intuicijo oz. zanimanje – le-to pa ni absolutno niti med ljudmi različnih dob niti med sodobniki niti v ontogenetskem razvoju posameznika niti v njegovih vsakdanjih življenjskih okoliščinah. Tudi je težko tako receptivno zanimanje bralcev (ciljne publike) natančno predvideti. Potem ni čudno, da je toliko različnih definicij sloga.

Recimo, da nas zanima psihološki oris osebnosti literata (vpliv psihologije s preloma 19. v 20. st.) – slog in oblike izražanja nam lahko služijo za odkrivanje duševnih pogojev, ki delujejo pri nastajanju literature.⁶⁵ Lahko nas zanima posredna karakterizacija literarnega lika z določenimi jezikovnimi sredstvi (sic! – slogom); itd.

4.1.1.1. Govor v književnosti in publicistiki

Na tem mestu dodajam še razmislek, ki se nekoliko odmika od predstavljanja vidikov sloga.

Realni naravni govor, zbran v gradivu te raziskave, imamo možnost primerjati z govorom (pogovori, izpovedi in govori oz. monologi, notranji monologi) v umetnostni književnosti oz.

neponovljivo čudo književnega dela – celote, v kateri sta lepota in resnica samo formalni oziroma vsebinski vidik.“ (Dukat 1985: 760.)

⁶⁵ Ocvirk 1978: 30–56: poglavje III. Literarnoteoretični pogledi na pesniško osebnost, 2. Psihološki oris osebnosti.

literaturi. K temu nas spodbuja dejstvo, da primere tega govora iz tega vira najdemo tako v *tradicionalnih* slovnica, slovarjih in (računalniških) korpusih, ki jezikoslovcu služijo kot standardna metajezikoslovna opora.⁶⁶ Biti pa moramo pozorni na fiktivnost, mimetičnost v literaturi; na podrejenost govora in njegovega prikaza ustvarjalnim intencam avtorja.

Pogovor v romanu ali kratki zgodbi odslikava avtentični pogovor, a ga nikdar ne kopira. Saj ne uporabimo tipičnih netekočnosti, razen za občasno oklevanje ali nedokončane povedi. In, z literarnega vidika, pogovarjanje v romanu ima mnogo večji pomen kot pogovor iz resničnega življenja v zvočni obliki. Pospeši dogajanje in obogati naše védenje o /literarnih/ značajih. (Freeborn 1996: 218.)⁶⁷

Govora pa tako ubesedovalno pretvorjenega ne najdemo le v književnosti, ampak v vsakem njegovem posrednem poročanju, tudi v (funkcijskozvrstna delitev po Toporišiču) praktičnosporazumevalnih, strokovnih in publicističnih besedilih.

Ta pretvornik /iz govora v pisno sporočilo in obratno/ nas v okviru *stilistike poročevalstva* zanima le kot tista faza (ustnega ali pisnega) sporočanja, v kateri je v vidno ali slušno tvarnost »odeto« sporočilo podvrženo šumu (ki zmanjšuje obvestilnost (informacijo), večja entropijo). Pretvornik namreč ni gladko, v vseh okoliščinah sporočanja sposoben npr. pisnemu prenosniku prilagojena sporočila brez šuma prepuščati in hkrati pretvoriti v slušno sporočilo brez dodatnih procesov. Važno je torej, da sporočilo pred vstopom v pretvornik »prepariramo« tako, da bo šum pri pretvorbi čim manjši ali odpravljen. (Korošec 1986: 394.)

4.1.2. O slogu v jezikoslovju

„Ker so slogovne značilnosti v temelju značilnosti jezika, je slog v nekem smislu sinonim za jezik: enakovredno lahko govorimo o *jeziku* /⁶⁸; v slov. tudi o *govorici*/ nekega besedila. Kar je vendarle vsebovano, je to, da je /ta/ jezik na nek način razločevalen, značilen“ (Wales 1994: 436: **style**) ali svoj(ski) za besedilni primerek, ki mu slog pripisujemo.

Za ločitev stilističnih raziskovalnih polj jezikoslovja in literarne vede pa ne zadostuje razlikovanje z govorjenjem ali o slogu ali o jeziku/govorici. Vidiki in prijemi obeh polij se močno prepletajo, nakažemo pa lahko *tradicionalno* izhodišče jezikovne stilistike v slovenskem prostoru.

Izhodišče stilistike opišemo glede na proces nastajanja sloga: tvorec iz sistema („mehanizma“, nerealiziranosti) vzame abstraktno jezikovno sredstvo ter ga po svoje oblikuje, veže oz. aktualizira (v „organizem“, realiziranost).

1)) Bolj na začetek tega procesa sodi *genetična stilistika*: ukvarja se z **vzroki** sloga, zanima jo razmerje med posameznim ustvarjalcem (posameznik ali kolektiv) in njegovim izrazom, tipična je v literarni kritiki. O tem izhodišču sem omenil zgoraj, da je najbolj razširjeno za teorije sloga v literarni vedi (vezanost sloga na individualnost).

2)) S konca tega procesa je *deskriptivna stilistika*: ukvarja se z **učinki** sloga (odvisnimi od pomenov oz. funkcionalnosti struktur v jeziku), zanima jo razmerje med konkretno jezikovno obliko in mislijo. (Povzeto po Pogačnik 1967: 179–180.) Namesto o učinku lahko govorimo tudi o stilistični funkciji.

⁶⁶ Namreč v slovenskih razmerah: slovnice J. Toporišiča in še bolj predhodne, SSKJ in SP²⁰⁰¹, korpusa FIDA in Nova beseda, pa še kaj. Razlogi za navajanje govora iz književnosti za primere in dokumentiranost tega navajanja ob primerih so v teh priložnostih različni; še najbolj gre korpusom, ki pa po svoji naravi dajejo manj metajezikoslovnih informacij. Pa vendar lahko ocenimo, da si je natančen *situacijski kontekst* govora v posameznih primerih težko predstavljati in tako ovrednotiti njegovo upoštevanost, ki je bistvena pri metajezikoslovni opredelitvi aktualiziranega jezikovnega pojava. Za probleme transbesedilnih besednih seznamov prim. opombo 49 na str. 31.

⁶⁷ Avtor se osredotoča predvsem na dialoški govor iz vsakdanjega življenja (t. i. spontani govor), kar pa je v tej raziskavi le ena od notranjih diferenciacij prenosniško pojmovanega govora.

⁶⁸ Slog je lahko sinonim le za jezik kot *parole* oz. za rabo jezika, nikakor pa ne za jezik kot sistem jezikovnih sredstev (prim. Korošec 1998: 9; tam avtor kritizira Toporišiča za nerazločevanje v strokovnih izrazih *funkcijska zvrst jezika* in *funkcijski stil* v ESJ).

Za drugo izhodišče menim, da je blizu jezikoslovnemu ukvarjanju s slogom, tuje pa ni tudi literarni stilistiki.

4.1.2.1. Jezikovne ravnine stilemov

Nasploh lahko slog prepoznavamo na različnih jezikovnih ravninah, čeprav pri konkretnih besedilih v „tipičnem izboru jezikovnih prvin“ niso nujno prisotne vseh ravnine, ampak le nekatere. Konkreten splet jezikovnih značilnosti, na katere zelo določno navežemo izraz in vsebino sloga, imenujemo **stilizem** ali **stilem**. V slovenščini kakšnega stilističnega slovarja⁶⁹ ali slovnice, ki bi opisala stilistična sredstva na različnih ravninah, nimamo; na tej ali oni ravnini⁷⁰ omeni kaj Toporišičeva SS²⁰⁰⁰ – iz nje primer ene od višjih ravnin:

Primer 9: Slog nad ravnino skladnje – pragmatika

„Posplošeno lahko rečemo, da vsako (celostno) globalno govorno dejanje določa stil besedovanja, tj. narekuje vrsto slovničnih zgradb, ki so posledica izbirnih postopkov iz pomensko enakovrednih možnosti. Da gre tudi za pragmatične, ne samo stilistične razlike, nam kažejo dvojice kot »Podaj mi no sol!«, »Prosim, podaj mi sol.« ali »Bi mi, prosim, podal sol?« (vljudnostne izbire).“ (Toporišič SS²⁰⁰⁰: 741–742, poglavje Sporočanje: Pragmatika: (Vele)govorna dejanja in besedovanje.)

Izrazno že literarna teorija **deli slog po jezikovnih ravninah**: najnižji je zvočni slog (literarni: evfonija) v govoru ali likovni v pisanju, potem besedni slog (literarni: tropi) in stavčni slog (literarni: figure in razširjene figure). (Kmecl 1995: 46 in 145.) Ta delitev je kar vzporedna običajni jezikovnoravninski (fonemska, morfemska, leksemska, sintaktična raven). Vendar se moderno jezikoslovje ozira tudi po višjih oz. kompleksnejših jezikovnih enotah (ravninah?), katerim prav tako lahko pripišemo slog: besedilo (npr. nagovor, tudi pogovor), nadbesedilo (npr. knjižna zbirka, časopis, dramska trilogija, TV-poročila) oz. diskurz (npr. trajne debate), jezikovni produkt posameznika (opus) ali skupnosti (npr. tezaver, nacionalni korpus), nekega obdobja, kulture ... Te kategorije skuša zajeti tudi literarna teorija, le da ne govori več samo o literarnem slogu, temveč o literarnih vrstah, žanrih, zvrsteh, obdobjih, zgodovinskih slogih ...⁷¹

4.1.2.2. Deskriptivna stilistika: slog kot učinek nekaterih izraznih sredstev

/S/til je vsebovan v izboru iz jezikovnih sredstev, je funkcija izbranih jezikovnih sredstev, najsi že gre za glas, naglas, besedo in njeno obliko, sintaktično zvezo ali mesto v govorni verigi (torej za besedni red); in če gre za tiskano sporočilo, je stilotvoren tudi izbor pravopisnih in pisnih navad in značilnosti. (Toporišič 1962: 38)

Slovenska jezikovna stilistika pa se je predvsem ukvarjala s preprostimi izraznimi sredstvi –besedami (gl. naslednja podpoglavja) in govornimi o **nevtralnosti** (nezaznamovanosti, objektivnosti, slogovni neopaznosti) in **slogovni zaznamovanosti** (opaznosti, konotiranosti) izraznih sredstev.

⁶⁹ Omenja ga Toporišič 1992: ESJ: 311: **stilistični slovar**: „Slovar stilno zaznamovanih sredstev, zlasti besednih, oblikoslovnih, skladenjskih.“

⁷⁰ Nekaj primerov: o stilno nevtralnem besednem redu (668) in o stilnem izražanju gotovosti v pomenski podstavi stavka (525) s področja skladnje, o stilni vrednosti izpeljav samostalnika iz glagola (169) iz besedotvorja, o stilni vrednosti pri prehodnosti glagola (355) v oblikoslovju in o stilni vrednosti besed (123–133) iz besedoslovja. Vse iz Toporišič SS²⁰⁰⁰.

⁷¹ Kmecl (1995: 144) na koncu poglavja o literarnem slogu: „Torej se literarna teorija na robu stilistike oziroma poetike jezikovnega izražanja že tesno približa drugemu obsežnemu, znanemu, le da povsem književnovrstnemu poglavju – stilistika je le bolj ali manj vmesno področje med jezikovnimi in književnimi vedami – nauk o književnih zvrsteh in vrstah.“

Bistvo stila je tako v našo pozornost zbujujočih elementih (imenujmo jih stileme, Ocvirk in drugi za njim jih imenujejo stilizme) kakor v odnosu le-teh do elementov, ki naše pozornosti posebej ne pritegujejo. Poslednji, tj. prvine, ki naše pozornosti ne pritegujejo, so neogibni protipol stilemom: pomenijo nevtralni, običajni izbor jezikovnih sredstev, stilemi pa v nasprotju z njimi – bolje bi bilo celo reči: zaradi nasprotja z njimi – aktivni, manj pogostni, včasih celo izjemni izbor jezikovnih sredstev. Da se izrazim primerno marksistični terminologiji: eni in drugi skupaj tvorijo dialektično celoto, saj so eni z drugimi določeni in določljivi, kakor uči relativnostna teorija. /.../ če dani estetski ali kakršenkoli že tekst obdelam natančno glede stilemov – kateri so, kakšni, kako pogosto nastopajo, kakšen je njihov medsebojni odnos, morda skupni imenovalc – je s tem tudi povedano, kakšen je stil danega teksta. (Toporišič 1962: 38)

Ukvarja se torej z „doživljajno vrednostjo vseh jezikovnih sredstev“, ne le izbranih slogovnih in ne le vezanih na posameznika. V tem je zanjo značilno še dvojje: da je blizu zgoraj opisani deskriptivni stilistiki – doživljanje izraznih sredstev kot njihov *učinek* – in da skuša ubežati subjektivnemu razlaganju sloga s pomočjo *jezikovnega čuta* in *okusa*. Objektivnost razlaganja sloga naj bi bila torej v tem, da se raziskuje „tiste primere, kjer je kakšno stvar mogoče povedati na dva ali več načinov, pri tem pa je bistvo povedanega isto“. In merilo o tem, katere od teh možnosti so nevtralne, katere pa zaznamovane? Večja *vsakdanjost*, *običajnost* je razlog, da nevtralnih sredstev (izraznih lastnosti) ne dojemamo kot slogovno učinkujočih.⁷²

Primerjanje različnih izraznih sredstev, ki da povedo *isto*, je virtualno, jezikovnosistemsko. V konkretnosti, aktualiziranosti pa ima vsako izrazno sredstvo svoj *izraz* (celokupnost aktualiziranega pomena), ne učinkujejo pa vse sestavine izraza slogovno, ampak le nekatere, „izbrane“. (Če namesto o izrazu/pomenu govorimo o stilu, vnesemo nekaj zmede, npr. „stil je vsakršen izbor iz jezikovnih sredstev“.) Stilistika zato pri presojanju učinka izraznih sredstev v konkretnem sporazumevanju le-ta naveže na zunajjezikovno realnost, pri tem pa prestopi meje jezikovnosistemskosti in izhodišče zgoraj koncipirane deskriptivne stilistike. Najprej se predpostavi, da jezikovno sredstvo dobro izrazno učinkuje, ko neka predmetnost najde svojo „najpopolnejšo ubeseditev“, ko je doseženo „idealno razmerje med znakom (tj. besedo) in označenim (tj. tistim, na kar se znak, beseda nanaša).“⁷³ Potem se razloži, da v bistvu slogovno zaznamovano sredstvo kot tako učinkuje samo zato, ker ubeseduje „manj navadno, posebno ali nenavadno *dojeto* resničnost. Tam /pa/, kjer je ta resničnost ali predmetnost čisto *vsakdanja*, pričakujemo, da bo taka tudi njena ubeseditev“ (ta in navedki v zgornjih dveh odstavkih so iz Toporišič 1965: 157–162; poudarke napravil R. B.).

4.1.2.3. Vsebina sloga: konotativni pomen

Preprosto rečeno, ugotovitev iz prejšnjega odstavka ugotavlja, da ima slog (tj. slogovne lastnosti ali vrednosti jezikovnih sredstev, ena ali več teh slogovnih vrednosti) tudi svoj pomen, *svojo* referenco v zunajjezikovni resničnosti. (Pri slogu torej ne gre le za oblikovno plat jezikovnega znaka, za zaznamujoče, kot bi lahko sklepali iz prve razlage sloga v Primer 8, str. 47: „tipični izbor jezikovnih prvin v besedilu“.) Pri *leksemih* imenujemo pomen njihovih slogovnih prvin **konotativni pomen**. Za jezikoslovce pa še vedno ostaja problematična pojasnitev tega,

⁷² „Izmed več izraznih možnosti je nato stilistika tisto, ki je najvsakdanjejša, imenovala nevtralno. Mi bi lahko rekli, da je taka izrazna možnost stilsko neopazna ali nezaznamovana – manj običajna možnost (ali možnosti) pa je zaznamovana ali opazna.“ Toporišič 1965: 159.

⁷³ Idealno referenčno razmerje Toporišič (1965: 161) dalje razlaga kot zadosten razlog dobrega učinkovanja zato, ker da omogoča sprejemniku tega sporočila, da ga dobro razume (da „v sebi najnatančneje obudi predstavo predmetnosti, ki je govorečega nagnila k ubeseditvi“). Dodaja še, da ta sprejemnik uporablja *isti jezik* kot tvorec. Kaže, da je težko opredeliti in členiti, kaj vse vpliva na to, ali/da tvorec in sprejemnik enako razumeta sporočilo: ali natančnost referenčnega razmerja (samo do denotata, ali tudi do konotiranega?), ali isti jezik; ali druge okoliščine, ki vplivajo na (spo)razumevanje, lahko bi govorili npr. o elementih sporazumevalnega oz. sporočanjekega modela ipd. Še, kaj je denotat: realni, zunajjezikovni predmet (objekt), na katerega se nanaša jezikovni znak. V okviru jezikovnega sistema je od posameznega odtrgan (Vidovič Muha 2000: 45).

katera oblikovna prvina jezikovnega sredstva zaznamuje ta pomen in na kaj točno se ta pomen nanaša (referira).

Vidovič Muha (2000: 97–98) v okviru leksema trdi, da je konotativni pomen „vezan na celotni leksem ali pa na katerega izmed njegovih pomenov.“⁷⁴ V okviru leksikalne sopomenskosti (2000: 158–159) ugotavlja podobno, da je „smiselnost poimenovalne (izrazne) različnosti istega denotata /.../ utemeljena s stilno vlogo sopomenk“. Ko dalje razčlenjuje izrazno utemeljenost slogovnih variant sopomenk, jo najde v besedotvornih (obrazilnih) morfemih: „morfemi so temeljni nosilci jezikovnosistemske stilistike“.⁷⁵ Lahko torej rečemo, da so nekateri obrazilni morfemi prav tako obremenjeni tudi s konotacijo, le-ta pa ni njihov osnovni pomen (to je še vedno denotat).

Konotativni pomen govori o razmerju do denotata; o spoznavnem vidiku denotata. Tvorec kot spoznavni subjekt (Vidovič Muha 2000: 97) želi izraziti tudi svoje spoznavno razmerje, ker to nekaj pove o dojetanju denotata. Lahko bi rekli: konotacija je dodatna, redundantna informacija, ki zagotavlja celostno koherenco, „najpopolnejšo ubeseditev“ oz. idealno vzpostavitev referenčnega razmerja tudi pri naslovniku sporočila (glej zgoraj ob opombi 73). Običajno si konotacijo predstavljamo kot tvorčevo razumsko ali čustveno razmerje do denotata. Konotacija (in z njo slog) pa lahko *govori* tudi o drugih **sporazumevalnih okoliščinah, ki vplivajo na spoznavanje denotata**, o kontekstu v najširšem pomenu: o času, kraju, stroki,⁷⁶ mediju, soudeležencih sporazumevanja, obdobju, jeziku, kulturi itd., v katerih se denotat pojavlja. Vendarle ostaja konotacija inherentna sestavina jezikovnega sredstva, učinkujoča tudi v virtualnosti, jezikovnosistemske; njen učinek ni odvisen od aktualiziranih sporazumevalnih okoliščin; sporazumevalne okoliščine, o katerih govori, so navezane na denotat in so skupaj z njim lahko virtualne.

Primer 10: Okoliščine dojetanja denotata, zajete v konotativnem pomenu

Razlaga zgleđa, ki ga omenjam še v 5. skupini, Primer 11 spodaj: *bôs(s)*: V slovenščini (pomenski kalk in izrazna (fonetična) prevzetost pa sta iz ameriškega okolja) je slogovni učinek besede *bôs(s)* v tem, da se udeleženec sporazumevanja pri spoznavanju denotata ‘šef, vodja’ vživi v ameriško okolje in obudi predstav šefa na prav takšen, s hoteno konotacijo predviden način. Predstava pa je hkrati drugačna od predstav šefa v drugih okoljih, ki so zajete v druge izraze: doma (vodja, predstojnik; šef, šefica) ali kje drugje (morda kapo(dibanda); komandir; lider; firer ipd.). V angleščini pa ima ta beseda verjetno tudi več učinkov, od nevtralnega do slogovnih in pragmatičnih.

Za zgled vzemimo **SSKJ** in njegovo pomensko opredelitev izraznih sredstev, ki so v njem (leksemov: besed, tipičnih zvez, stalnih zvez, fraz(emov), terminov, posameznih pomenov ipd.).⁷⁷

⁷⁴ Zraven sodi še pojasnilo iz istega vira, da konotativni pomen ni značilen za vsak leksem in da je takšnih leksemov v jeziku manjšina. Glede na temeljni, denotativni pomen leksema, je postranski, dodatno pomensko obremenjujoč.

⁷⁵ Ugotovitev sloni na zgledih: pis-ec, pis-ar, pis-un (obrazilni morfemi –ec, -ar, -un kot sopomenski imajo denotat ‘tisti, ki (kdor)’); otroč-ič, otroč-ek (obrazilna morfema –ič, -ek kot sopomenska imata denotat ‘majhen/čustveno pozitivno’). Vidovič Muha 2000: 159. Opomba: avtorico bolj kot stilni vidik zanima sopomenskost morfemov (v okviru pomenskih medleksemskih razmerij).

⁷⁶ Trditev v zgornji povedi ima presenetljivo posledico: termini so pojmovani kot zaznamovani, slogovno nevtralnimi, torej imajo konotativni pomen! Vidovič Muha (2000: 46) trdi drugače („samo denotativni pomen je lasten vsemu strokovnemu izrazju“), ker veže konotacijo predvsem na čustvenost spoznavnega subjekta (isto mesto) in ker enoznačno odslikavo predmetnosti pri terminih veže samo na denotativni pomen. Menim, da je ta vezava preozka, čeprav je težko postaviti konceptualno mejo med konotativnim in denotativnim pomenom. Prim. SSKJ: kvalifikator, ki označuje rabo pomena (torej denotata) izraznega sredstva v neki stroki, vedno stoji *pred* razlago (= vsebina pojma, torej denotat; gl. nadaljnji tekst); funkcionira kot *spoznavno* pojasnilo o konotaciji in s svojo prisotnostjo vrednoti izrazno sredstvo tako, da ga ne pripiše med nevtralna izrazna sredstva.

⁷⁷ Avtorji govorijo o „osnovnih izraznih jezikovnih sredstvih (od osnovnega besedišča do navadnih stavčnih konstrukcij, od pisave do izgovora in naglasa)“. SSKJ je potrebno gledati tudi z vidika drugih njegovih izhodišč, ne kot apriorni standard slovenskega jezika: je enojezični, splošni slovar, t. i. srednjega tipa obsega, kolektivno delo, iz 2. polovice 20. st. z oziranjem v 19. st., informativno-normativen, slovar *knjižnega* jezika (knjiga kot medij in tisk kot prenosnik; knjižni jezik kot standardni, socialni, funkcijski delni sistem jezika, „bleda podoba resničnega, živega jezika“, „deindividualizirani jezik“; prevlada leposlovja) in še kaj. Citirano je iz Klopčič/Suhadolnik 1970): 137.

Razlaga prinaša le vsebino pojma, posebnosti dodatno pojasnjujejo *kvalifikatorji*, oboje bodisi utemeljuje bodisi dopolnjuje *ilustrativno gradivo*, razmerja med besedami pa so nakazana z *umestitvijo* v strukturo slovarja. Slogovne označitve (in s tem konotacijo) iščemo med **kvalifikatorji** različnih vrst, ki vsaka z drugačne strani opredeljuje slovarske enote. Ne zanima nas toliko tista skupina kvalifikatorjev, ki je enakomerno razporejena po vsem besedišču (slovnični in pomenski), ampak druga skupina, ki besedišče *vrednoti v nasprotju z nevtralnim* večinskim besediščem (ki teh drugih kvalifikatorjev sploh nima): stilno-plastni, ekspresivni, časovno-frekvenčni, normativni, terminološki kvalifikatorji (za delitev na ti dve skupini prim. Hajnšek Holz/Novak 1970: 140).

Naslednji primer je nekoliko drugačna delitev *kvalifikatorskih pojasnil*, ki jih v slovarju ni bilo mogoče kategorizirati kot gornjih vrst kvalifikatorjev.

Primer 11: Označevanje sloga s kvalifikatorskimi pojasnili

<i>skupina kvalifikatorskih pojasnil</i>	<i>primer iz SSKJ: kvalifikatorsko pojasnilo</i>
1)) opozarjajo na vezanost besede na določeno besedno zvezo	
2)) opozarjajo na odvisnost posameznega pomena od določene besedne, stavčne ali oblikoslovne kategorije	
3)) obravnavajo prehajanje besed iz ene besedne kategorije v drugo	
4)) opozarjajo, za kakšne namene se lahko beseda ali besedna zveza v določenem pomenu uporablja	dober : kot voščilo pri jedi dober tek
5)) kažejo na omejenost izraza na določeno okolje	bòs / bòss : v ameriškem okolju
6)) krajevno in časovno opredeljujejo pojme	čajnica : v vzhodnih deželah
7)) opozarjajo na vezanost pojma na kak nazor, miselnost ali pojmovanje (Kozlevčar 1970: 139)	desetnica : etn., po ljudskem verovanju absurd : filoz., po Camusu

V zgornjem primeru so prve tri skupine primerljive s slovničnimi in pomenskimi kvalifikatorji, zadnje 4 skupine pa s kvalifikatorji, ki „vrednotijo“ – govorijo o konotaciji (slogu), ki jo ima beseda *sama po sebi*, ali o *sporazumevalnih okoliščinah*, v katerih šele beseda dobi slogovni učinek (predvsem 4. skupina).

4.1.2.4. Sistemska in strukturalna umestitev razlikovanja sloga

Prav ob primeru 4. skupine zgoraj se moramo pozorno vprašati, na kateri **jezikovni ravni** se dogaja **razlikovanje učinka**. Sloga namreč ne vežemo samo na jezikovno raven slogovnih izraznih značilnosti, na katere vežemo poseben učinek (oz. posebni dodatni pomen, priokus, konotacijo v širšem smislu). Slog razlikujemo na višji jezikovni ravni od prejšnje: na tej ravni **„enačimo“ oz. primerjamo jezikovne enote**, npr. leksem: besedo (*izrazno*) in denotat (pomensko) kot *tertium comparationis*.

Učinek iz 4. skupine (voščilo) ni vezan samo na besedo „dober“, ampak izhaja iz razmerja med besedno zvezo kot besedilom in njegovimi zunajbesedilnimi (zunajjezikovnimi) okoliščinami (osebe pri prehranjevanju). Raven, na kateri enačimo slogovne variante in razlikujemo njihove učinke, ni raven leksemov, ampak na raven sporazumevalnih situacij (besedila kot dejanja).⁷⁸

⁷⁸ Posebna jezikoslovna veda, pragmatika, ki se ukvarja s prav to ravno sporazumevanja in vsemi učinki na tej ravni (besedil kot dejanj, npr. ekspresivnim pri voščilu Dober tek!), bi te učinke imenovala **pragmatični učinek**. Prim. koncept performativa in Primer 9, str. 52.

Termin pragmatični učinek se nanaša na vse variante učinka na tej, *sporazumevalni* ravni. Ker so enote sporazumevalne situacije, sestavni deli le-teh pa tudi aktualne zunajjezikovne okoliščine in njihovi vplivi na funkcioniranje jezika, ne moremo več govoriti o *jezikovni* ravni. Na jezikovni ravni enote, primerljive sporazumevalnim situacijam, imenujemo jezikovna sredstva/enote, njihov „pragmatični učinek“ pa imenujemo pomen (celokupnost pomena, ki izhaja iz izraznih sredstev).

Lažje predstavljivo je razčlenjevati učinke sloga na manj kompleksnih, tradicionalnih jezikovnih ravnanjih, od morfemske do besedilne.

Pogoj za primerjanje izraznih značilnosti določenih enot (nevtralnih ali zaznamovanih) je **sistem primerjav** na njihovi ravnini. Dovolj sta že dve enoti (npr. dve sopomenki, dve povedi, dve besedili, dve sporazumevalni situaciji), množica pa je lahko mnogo obsežnejša, ali bolj heterogena (npr. ne le slovsko izpisane besede, ampak konkretno besedilo, ali več besedil avtorja, ali raznih avtorjev, korpus, večjezični, diahroni korpus ...).

Primer 12: Enote, ki jim vrednotimo zaznamovanost: njihove jezikovne ravnine in sistemi primerjav

Poskusimo si predstavljati fikcijo – opis obeda v romanu, pri katerem si osebe voščijo „Dober tek!“ – ali pa navedba te fraze prav v tej pričujoči povedi. Ali zaradi fiktivnosti, virtualnosti okoliščin še lahko govorimo o pragmatičnem učinku *vsakdanjega voščila* iz realne situacije? Ali pa gre za slogovni učinek realističnega opisa (enote v literarnem besedilu, romanu) oz. za slogovni učinek primera in argumenta (enote v strokovnem besedilu – v tej diplomiji)?

T. i. stilno-plastni kvalifikatorji v SSKJ (bibl., knjiž., pesn., pisar., publ., vznes.; nar., otr., pog., nižje pog., žarg.; SSKJ, § 131–141) označujejo leksiko, ki jo „slovar, ki je nosilec leksike knjižnega jezika, /.../ prinaša samo v taki meri, kolikor se kot stilni element knjižnem jeziku lahko uporabljajo.“ (Vidovič Muha 1971/72(: 180). *Samo v konkretnem gradivu knjižnega jezika* so sprejemljivi „kot funkcionalni stilizem“ (ista, 178) in samo v tem konceptu slovarja imajo takšen kvalifikator. V Svetem pismu (v religioznem diskurzu), v publicistiki, v pisarniškem diskurzu, med otroci, v vsakdanjih pogovorih ta leksika ne bi imela posebnega slogovnega učinka in tako ne bi zaslužila oznak bibl., publ., pisar., otr., pog. itd.⁷⁹ So pa ti sporazumevalni diskurzi dosti bolj omejeni kot vseslovenski diskurz (zelo širok sistem primerjav), ki ga zajema SSKJ.

Ob *obsežnosti* sistema primerjav se pokaže **relativnost sloga**:⁸⁰ če neka jezikovna enota nima primerjave, kot slogovne oz. tipične zanjo pojmuje vse njene izrazne značilnosti, izbora ni (lahko rečemo tudi: nevtralna je); ista enota bo v primeru devetindevetdesetih primerjav imela manj slogovnih značilnosti (tipičnih, izbranih, svojiskih) kot v primeru ene primerjave.

Slog je relativno odvisen tudi *izrazne kompleksnosti jezikovne enote*: enote (tj. vsa jezikovna sredstva, uvrščena na isto jezikovno raven) so različno kompleksno sestavljene, ene bolj druge manj natančno in določno izražajo, s tem pa jim analitik (ali naslovnik) lahko *pripíše* več ali manj slogovnih značilnosti. Pozor: gre za **pripisanost** (konotativnega, tj. **slogovnega pomena**) v razlikovanju izraznih značilnosti; stopnja dejanske, inherentne izrazne obremenjenosti s slogom ni premosorazmerna⁸¹ s stopnjo pomenske obremenjenosti, vsaj ne vedno, saj jezikovni sistem ni toliko enovit. Prim. npr. tvorjenko 5. stopnje in tvorjenko 1. stopnje, 4-stavčno in 1-stavčno poved, petdejanko in enodejanko (dramo) itd. Pri bolj kompleksnih enotah (npr. daljše ali bolj strukturirane besede, npr. osebna glagolska oblika 5. stopnje tvorjenosti deležnika in v predpretekliku) lažje razlikujemo slog (pripišemo slogovnemu pomenu izbrane oblikovne značilnosti) kot pri enostavnejših (krajše ali manj strukturirane besede, npr. t. i. prosti morfemi).

Termin slog pa je uporabljan širše, za vse ravni (jezikovne in sporazumevalno), in sicer zgolj kot izbor posebnih značilnosti učinkov (pragmatičnih ali izrazno-pomenskih).

⁷⁹ V citirani razpravi je skušala Ada Vidovič Muha stilistično podrobneje razčleniti zvrsti jezika in izvore njihove leksike. Paradoks je v raztegotovanju *knjižnega jezika* (kot ga določa koncept SSKJ) na najširše situacije *sporočanja* v slovenščini; jezikovnosistemska SSKJ nasproti aktualnemu sporočanju: „subjektivna“ in „objektivna“ sporočila sta dve osnovni pojavnici slovenskega jezika iz naslova razprave. Morda bo boljše izhodišče za takšne razprave kak interaktivni slovar na podlagi referenčnega (tj. iz še več zvrsti jezika z dokumentiranim gradivom) korpusa v prihodnosti.

⁸⁰ Kar tukaj razlagamo kot relativnost sloga, je povezano z neobveznostjo konotativnega pomena leksema pri Vidovič Muha 2000: 29: po njej je konotativni pomen zgolj „možen“ nasproti obveznemu denotativnemu in kategorialnemu pomenu. V opombi 25 na isti strani pa govori tudi o primeru kletvice, kjer bi bilo lahko mogoče govoriti „o prehodu oz. prevrednotenju konotata v denotat“.

⁸¹ O povezavi obrobni, stilistično zaznamovanih besed z izrazno in pomensko členjenostjo pišeta že Suhadolnik in Janežičeva (1962, 48): „Statistiki so izračunali: čim dalj od središča, oz. čim nižje na frekvenčni lestvici stoji beseda, manj pomenov ima in iz več fonemov je narejena.“

4.1.2.5. „Obstranskost“ sloga

Nazadnje se dotaknimo zgodovinskih, metajezikoslovnih okoliščin, v katerih slovensko jezikoslovje ločuje besede na nevtralne in slogovno zaznamovane. Za SSKJ je rečeno, da je večina oz. 60 % njegovih izraznih sredstev nevtralnih (Vidovič Muha 1971/72: 179). O nevtralnih sredstvih je še rečeno, da tvorijo „osnovo jezikovne realizacije“, da jih „lahko brez pomislekov uporabimo v vsakem jezikovnem sporočilu“ (isto mesto). SSKJ (§ 42) pravi, da so „bolj znane in široko rabljene besede iz nevtralne plasti razložene razmeroma krajše, besede iz ožje rabe pa imajo navadno daljšo, podrobnejšo razlago“.

Ko jezikoslovec določa vrednost slogovne zaznamovanosti oz. nevtralnosti, skuša sodbo opreti na predstave⁸² o objektivni rabi (pogostost in funkcionalnost v sporazumevalnih situacijah) in stopnjo abstraktnosti v pomenski členitvi ujezikovljene realnosti. Važna je tudi konkretna slovenska metajezikoslovna situacija, kjer je merilo zaznamovanosti naslonjeno na samosvoj koncept *slovenskega knjižnega jezika* in celo *konkretnega projekta*, SSKJ: na oceno, kaj je *temeljno* med slovenskim izraznimi sredstvi⁸³; prim. tudi poglavje I.2.1.2). Poenostavljeno pa je za slogovno zaznamovana jezikovna sredstva mogoče reči, da so med vsemi jezikovnimi sredstvi nasploh **obstranska**: bolj redka oz. manj običajna (prim. komentar k sledeči tabeli), v odnosu do temeljnega sporočila – npr. denotata (za raven besede) – pa ob robu referiranja do ujezikovljene realnosti; zato pa zelo natančno členijo realnost oz. celo vplivajo na spoznavanje te realnosti.

Tabela 7: Ocena razmerij med besedišči: SSKJ vs. BSJ in nevtralno vs. zaznamovano

	SSKJ			BSJ	SKUPAJ: BSJ + SSKJ		
	SSKJ	nevtralno (min. 60 %)	ne-nevtralno (max. 40%)		SKUPAJ	slogovno zaznamovano	79 %
<i>iztočnice</i>	93.154	55.892	37.262	178.457	271.611	215.719	
<i>(pod)gesla (zaokroženo na tisoč)</i>	107.000	64.200	42.800	183.000	290.000	225.800	

Komentar: SSKJ in BSJ kot slovarja izhajajoč iz istega leksikografskega izhodišča zajemata *nabor poimenovalnih možnosti* slovenskega jezika. Podatki o njunem besedišču torej dajejo predvsem sliko o raznolikosti leksemov v njihovi poimenovalni funkciji; ti podatki so eksaktni in so v sivih poljih. Ostalo (v belih poljih) so ocene, ki povezujejo dve predpostavki: da so bile v SSKJ uvrščene najpogosteje in v BSJ redkeje uporabljane poimenovalne možnosti (Sklepna beseda v 5. knjigi in enotni izdaji SSKJ) in da je 60 % poimenovalnih možnosti v SSKJ nevtralnih (ocena že zgoraj po Vidovič Muha 1971/72: 179). Znano je tudi razmerje med poimenovalnimi možnostmi BSJ in SSKJ: 2/3 proti 1/3 zbranega gradiva (na 6 milijonih izpisanih listkov).

Temu pridružujemo sklep, da besedišča iz BSJ ne moremo šteti med nevtralne poimenovalne možnosti, ker niso uvrščene v SSKJ (vprašanje je, ali samo na podlagi njihove redkosti⁸⁴). Od vseh poimenovalnih možnosti, ki jih je zajel projekt SSKJ+BSJ, je torej večina slogovno zaznamovanih, manjšina pa nevtralnih: med vsemi iztočnicami je prvih

⁸² Ne gre zgolj za jezikovni čut jezikoslovca; delo je utemeljeno z leksikografsko teorijo in prakso; npr. na zastavitvi izpisovanja, splošni, frekvenčni in drugih kartotekah SSKJ, uvrščanju v slovar itd.

⁸³ Skušam povezati nekaj sorodnih pogledov: temeljno ~ nevtralno ~ nezaznamovano. Na opisovanju nasprotja zaznamovano nezaznamovano pri jezikovnih prvinah sloni teorija zaznamovanosti, kot sta jo razvijala N. Trubeckoj in R. Jakobson. Iz te izhaja večstopenjska teorija naravnosti (v naravnem jezikoslovju), ki kot bolj naravno opredeljuje tisto, kar je za možgane lažje (Dobrovoljc 2005 : 11). Koncipiranje (stilno) nezaznamovanih sredstev se pogosto skuša nasloniti na merila, kot so (strani iz Toporišič 1965:) vsakdanost (159), najpopolnejša ubeseditve (161), velika pogostost (162, 164). Ta merila spominjajo na merila za utemeljevanje lestvic naravnosti: načelo najmanjšega napora, včlenjenosti v stavek, pogostosti in druga. (Dobrovoljc 2005: 27–49).

⁸⁴ Jurančič/Legiša 1970): 137: „Sprejeli smo besede, ki jih knjižna raba, predvsem novejša, zanesljivo potrjuje. Pri tem smo se ozirali na listkovno gradivo in frekvenco. Ravnali smo se po načelu, da je treba zajeti besedni zaklad v vseh plasteh današnjega knjižnega jezika, kakor živi v bolj ali manj široki rabi.“

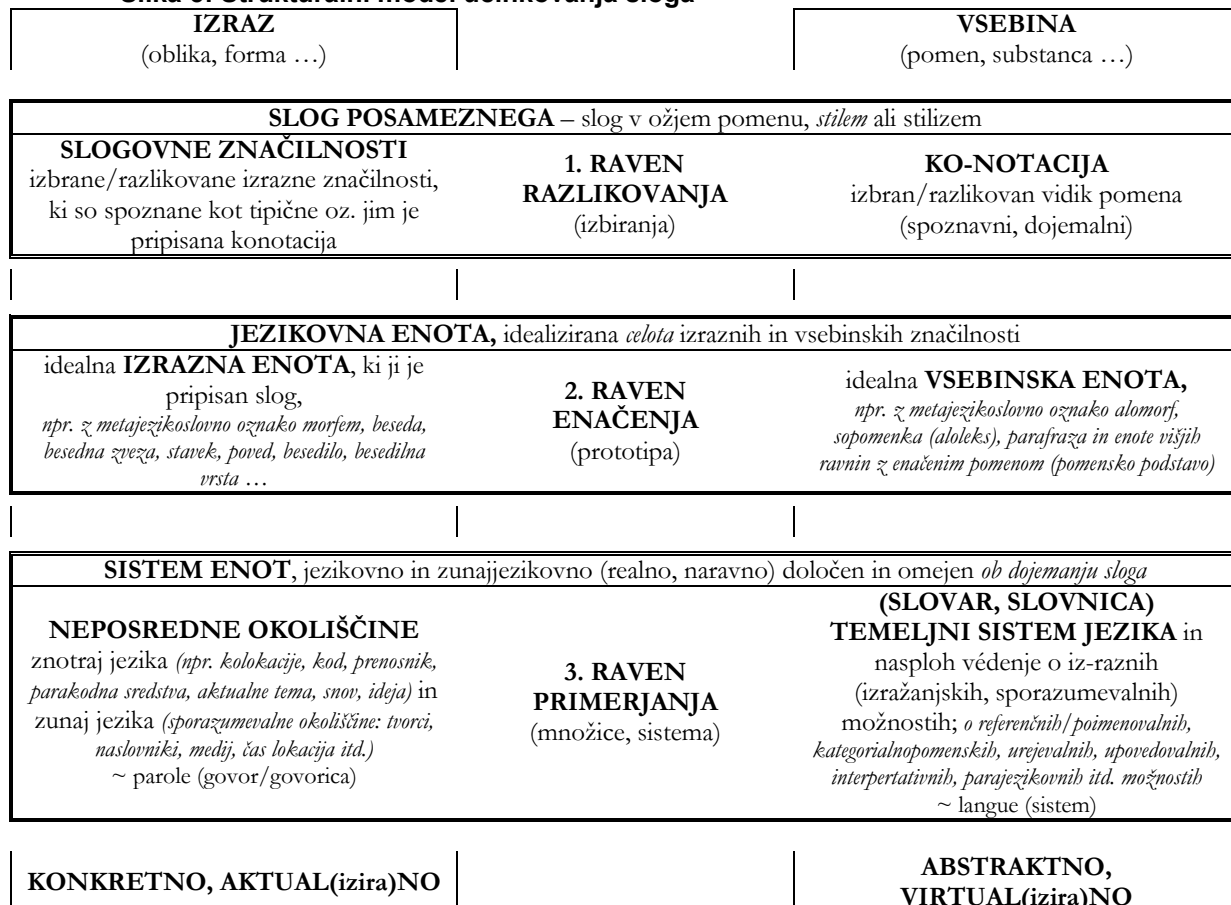
178.457+37.262= 215.719, drugih pa 55. 892 ali štirikrat manjša količina. Če govorimo o tem, da so slogovno zaznamovana jezikovna sredstva bolj redka, se torej to nanaša predvsem na nizko pogostost njihove rabe, ne pa na njihov skupinski delež v besednem zakladu slovenskega jezika.

V tem poglavju opisani problemi sloga so bili trd oreh za literarna izhodišča stilistike: konotativna oz. „spoznavna moč“ stilemov, „ravnina“ in „sistem“ njihovih primerjav, „relativnost“ in „pripisanost“ ter „obstranskost“ med vsemi izraznimi/jezikovnimi sredstvi. Terminologija v drugih teorijah stilistike tako govori o *jezikovnem čutu* in *intuitivno* spoznavni esenci sloga ali literarnoreceptivno o bralčevi duševnosti; namesto o sistemu razpoznavanja sloga se govori literarnozgodovinsko o avtorju (*idiolektu*) in njegovi duševnosti; namesto o obstranskosti o *izbiri* tipičnih lastnosti (ene variante – denotata); govora je ponekod samo o oblikovni plati sloga ali pa se v obstoj sloga celo dvomi.

4.2. Poskus strukturiranja sloga z modelom

Raziskovanje sloga v tej diplomii ne sloni na intuiciji oz. občutku, kakšen slog (ali kakšni slogi) je prisoten v besedilnem gradivu; nimam neke hipoteze o slogu. Izhodišče je vprašanje, *kaj in kako lahko nekaj povem o slogu*. Zato skušam raziskovanje sloga v govoru „vkleščiti“ v formalnogični model, ki razlaga, kako deluje slog v jeziku konkretno od vzroka do učinka. Pri tem se v celoti naslanjam na jezikoslovni vidik sloga, saj mi ta omogoča, da skupaj povežem različne uporabne plati korpusno urejenega gradiva in metod kvantitativnega oz. računalniškega jezikoslovja: različne ravnine segmentacije čistega besedila (oblikovna in pomenska variantnost besed; besedila), zunajjezikovni podatki (okolščine sporazumevanja), statistično razbitje in obdelava teh enot v računalniški obliki.

Slika 5: Strukturalni model učinkovanja sloga



Formalnologični model učinkovanja sloga, ki sem si ga drznil sestaviti, izhaja iz strukturalističnih semioloških načel iz 1. polovice 20. stoletja. *Slog kot jezikovno sredstvo (znak)* je v temelju **dihotomno razdeljen** na izraz (oz. konkretno) in vsebino (oz. abstraktno) – gl. Slika 5 zgoraj, horizontalna delitev na levo in desno plat.⁸⁵ Vertikalna delitev na sliki pa temelji na **načelu dvojne artikulacije**, ki je tukaj razširjena na 3 ravni, nekakšne spoznavne (formalnologične) dimenzije sloga. Na kratko rečeno gre najprej za **1** raven jezikovnih enot, pri katerih slog raziskujemo (srednja raven), potem **2** raven tipičnih slogovnih značilnosti (vrhnja, jezikovno najpreprostejša raven; razločevanje) in nazadnje za **3** raven sistemske primerjave (najgloblja, jezikovno najkompleksnejša raven, ki tudi presega sam jezik kot naravni pojav).

Posamezni pojmi, ki se pojavljajo v delih modela (izraz, vsebina, izbiranje, konotacija, razlikovanje, enačenje, sistem primerjanja itd.) so bili predstavljeni že v prejšnjem poglavju s podpoglavji – I.4.1.2, razdelitev na ravni pa predvsem v I.4.1.2.4.

Te 3 ravni nimajo zveze z strukturalnimi jezikoslovnimi *ravninami* jezikovnih enot oz. je kot ena od jezikovnih ravnin lahko opredeljena le 2. raven. Preostali dve ravnini v modelu, 1. in 3., sta bolj relativni in manj oprti na koncepte jezikoslovnih ravnin. Zajeti skušata bodisi *vse* okoliščine, ki so pri sporazumevalnem učinkovanju sloga *relevantne* (3. raven), bodisi *vse* relevantne izrazilne podrobnosti (1. raven) pri konkretnem stilemu. Pri tem 3. raven oz. „vse relevantne okoliščine“ pomeni hierarhično umestitev 2. ravni v odnosu do kompleksnejših jezikovnih (in sporazumevalnih) struktur, 1. raven oz. „vse relevantne podrobnosti“ pa hierarhično nadrejeno umestitev 2. ravni v odnosu do manj kompleksnih jezikovnih struktur.

Ostaja dilema, ali je 2. raven v modelu sploh potrebna; ali ni o slogu *vse povedano že* s 1. in 3. ravnjo, ki predstavljata dvojnost v načelu dvojne artikulacije jezikovnega znaka. Se pa pri razlaganju učinkovanja sloga zdita 1. in 3. raven vsaksebi preveč narazen (npr. izbrane izrazne značilnosti proti védenju o sistemu izraznih možnosti). S tem, ko na 2. ravni govorim o enoti (enotnost ~ konceptualna nedeljivost, neartikuliranost) oz. prototipu,⁸⁶ dobim most med obema artikulacijama sloga s 1. in 3. ravnjo. O tem prim. še razmišljanje v podpoglavju I.4.3.3.

Z definiranjem 3. ravni primerjave se raziskovalec odloči ali za objektivni (empiričen; če izbere točno določen sistem primerjav) ali subjektivni pristop (vzvišen, „tradicionalno humanističen“; če ne izbere sistema primerjav ali ga določi le ohlapno, ne za vse enote, ki jih primerja). Namreč, pri preprostih, majhnih primerjavah (znotraj preprostih sistemov primerjav, npr. 2–9 primerjav) odkrijemo le malo lastnosti, malo specifičnih stilemov (izraza in vsebine) pri predmetu slogovne analize. V ohlapno določenem sistemu primerjav pa raziskovalec napaberkuje več stilemov, lastnih predmetu analize, in to z več jezikovnih ravnin; ne zna pa pojasniti, s čim je predmet primerjal in zakaj nekatera njegova izrazila učinkujejo slogovno, druga pa ne. Pri tem subjektivnem pristopu je po mojem modelu razlaga ta: predmet je primerjan znotraj mnogo

⁸⁵ V bistvu gre za dva osnovna pola (gl. Slika 5, levi in desni stolpec), ki se ponavljata v vseh ravneh modela: izraz in vsebina; forma in substanca; konkretno in abstraktno; aktualno in virtualno; govorica (parole) in jezikovni sistem (kot langue). Ne gre za dve fizični plati (jezikovnega znaka katerekoli jezikovne ravnine), ampak dva antropološko pogojena koncepta. Človek, uporabnik znaka, je tisti, ki ob *očitni konkretnosti izrazne plati znaka* na prav to isto konkretnost naveže še neko abstraktno *dogajanje*. Prim. Barthes 1990: „zelo radi namreč znak razumemo le kot označevalec“ (160) in „Pomen lahko razumemo kot proces; je dejanje, ki povezuje označevalec z označencem, dejanje, katerega proizvod je znak.“ (167).

⁸⁶ **Prototip** je subjektivno pogojena predstava nekega koncepta, pojava; v primerjavi z njo so neprototipični pojavi (ki pa jih zajamejo podrobnejši znanstveni kriteriji), obrobní. Primer iz psihologije, od koder pojem prototipičnost prihaja: Prototip *ptice* so nestrokovnjaku vrabec, sinica, orel s svojimi lastnostmi (dve nogi, kljun, peruti, zna leteti), znanstvena klasifikacija pa zajame tudi mejne ali „obrobne primere“ – noj, pingvin (Dobrovoljc 2005: 40). Enako je mogoče razlagati tudi razumevanja jezikovnih kategorij in nestrokovne rabe, s katerimi se srečujemo tudi v tej diplomí: vsakdanji pogovori kot prototip (po)govornega jezika; časopisi in RTV kot tipični mediji/posredniki (neprototipično pa je za medij označevati osebni stik); beseda in njena ravnina kot tipični jezikovni znak; nerazlikovanje besede in pojavnice, besedne oblike, leme, leksema itd. Tudi **pri slogu** lahko domnevamo, da uporabljamo prototipe bodisi slogovnih značilnosti (literarni npr. rime, tropi, figure) bodisi sistemov primerjanja (npr. SSKJ). V *zgornjem modelu* pa je prototipičnost 2. ravni v enačenju t. i. „slogovnih variant“ med seboj (ne v tem, da je prototip t. i. nevtralna varianta, ampak v prototipu jezikovne ravnine, ki ji pripadajo vse variante).

večjega sistem primerjav (npr. nekaj sto primerjav), ki je prisoten v raziskovalčevi (pod)zavesti in strukturiran po njegovih subjektivnih literarnih normah, idealih, prototipih, izkušnjah, vedenjih (lahko bi uporabili izraz iz jezikoslovja: po raziskovalčevi slogovni „zmožnosti“).⁸⁷

Poglejmo primer. Izhodišče za opazovanje modela je konkreten izsek iz besedila (gl. Primer 13 spodaj, vrstica 3.1, sivo). (Gradivo je deloma izmišljeno deloma povzeto po virih; važen za posamezen zgled je odnos med njegovimi ravnmi in poloma.) Konkretno jezikovno enoto iz 3.1 primerjamo znotraj 3.2 (sistem v 3.2 si umislimo, podobno kot za 3.1 se vanj vživimo in s tem sledimo tvorcu, ki se je izrazil; prim. vrstico „vir“). Ob izboru konkretne jezikovne enote iz okoliščin hkrati dobimo, kar je v vrstici 2.1. Ob dejstvu, da enote iz 3.2 v kontekstu (3.1) enačimo kot alternative, se izloči še 2.2.

Čeprav v bistvu umevanje vseh treh ravni (pri tvorcu in naslovniku) poteka hkrati, je raven 1 tisto, kar konkretno ostane, poseben slogovni učinek v linearnem teku sporočanja, poenostavitev vseh primerjav, ki jih izsledimo v podrobnejši analizi. Analiza – razčlenjevanje, razlikovanje primerov namreč izloči najmanjši nedeljivi pomen (1.2) tega okusa in izrazno značilnost (1.1), ki ji je ta okus možno pripisati.

Primer 13: Besedni primeri za strukturalni model učinkovanja sloga

raven > po ⁸⁵ učinkovanja	zgleđ 1	zgleđ 2	zgleđ 3	zgleđ 4
1	1.1 izraz stilema	(navezava na koren besede oz. kar na celotno paradigmo)	daljšanje osnove s -t- (navezava na celotno paradigmo)	izgovorjeno z višjim registrom, pojoče in zateglo
	1.2 konotacija	(nevtrarno)	neknjiž., pravopisno prepovedano	šalj., iron.
2	2.1 izraz	očeta	očita	oče
	2.2 vsebina	‘moškega v odnosu do svojega otroka’	‘moškega v odnosu do svojega otroka’	‘moški v odnosu do svojega otroka’
3	3.1 konkretni sistem	Zgled v SSKJ brez kvalifikatorja: „ima še očeta in mater“.	Otrok v šolskem spisu: „Vsi otroci nimajo očita.“	Družinski pogovor: „A oni Geza je jegov foter, mislim oče?!“
	3.2 abstraktni sistem	očeta, ateja, očija, očka	očita, očija, očeta	oče, ata, foter, stari
vir	iz SSKJ oče, 1. (pod)pomen	SP ²⁰⁰¹ : § 785 in geslo ôči (str. 1004)	v lokalnem družinskem okolju (SSKJ ata v družinskem okolju)	geografsko in socialno zamejeno (SSKJ tatek v meščanskem okolju)

Primer 13 prikazuje predvsem variacije 1. in 3. ravni modela od zgleđa do zgleđa, 2. raven pa ostaja več ali manj ista (gre za primere pripisovanja sloga besedni jezikovni ravnini). Konkretni poli učinkovanja (n.1) pri tem že po svoji naravi bolj očitno variirajo kot abstraktni. V zgleđu 1 leksem oče ni zaznamovan, bodisi zaradi konteksta (3.1, protipomenka mati) bodisi zaradi vloge nevtralne razlagalne besede v SSKJ. V drugem sistemu oz. okolju (3.2) ima lahko isti leksem (slogovno) zaznamovanost – zgleđ 3, še vedno pa te zaznamovanosti ni mogoče pripisati posebni izrazni značilnosti, kot v zgleđih 2 in 4. Če primerjamo ta dva zgleđa, ugotovimo, da vsebina slogovne zaznamovanosti (1.2) ni vezana nujno na jezikovne priročnike (sistem primerjav v zgleđu 4 je zelo relativen, gl. 3.2 oz. opombo 88) in da je stilem lahko izražen (1.1) tudi s pomočjo

⁸⁷ Zanimivo, na podobni trditvi o subjektivnem pristopu k slogovni analizi sloni neka druga slovenska stilometrična (torej empirična!) analiza: da literarni kritik potrebuje dalj časa in več branja, da izloči slog, značilnosti besedila (Zupan 1994). Analiza po praksi umetnih inteligenc uporabi računalniško „nevronska mrežo“ za tak način razpoznavanja sloga.

⁸⁸ Poimenovalne možnosti so napaberkovane po SSKJ. Slovar sopomenk besede oče pa recimo ponuja tudi urejevalnik besedil Microsoft® Office Word 2003 v slovenski priredbi: oče, ata, oči, očka, pobudnik, utemeljitelj, prednik.

parajezikovnih sredstev (spet zgled 4). Pri zgledu 2 lahko v slogovnost podvomimo: lahko si mislimo, da otrok uporablja izraz „ôčita“ povsem nevtrarno, slogovno pa isti izraz razume kakšna starejša oseba z bogatejšo jezikovno vednostjo ali z drugimi družinskimi navadami oz. učiteljica, ki popravlja šolski spis in s tem izobrazuje, vzgaja.

4.3. Izbrane stilometrične metode

S pomočjo predstavljenega modela učinkovanja sloga bom v nadaljevanju razložil stilometrične primerjave besedilnega gradiva na podlagi dveh preračunavanj besednih oblik (različnic in pojavnic), ki smo jih napovedali že v poglavju I.2.4:

- 1)) pogostost različnice in zaporedno mesto različnice v pogostostnem seznamu različnic
- 2)) ter razmerje različnic in pojavnic.

In še opomba: Ko v nadaljevanju govorim o *besedilnem gradivu*, želim s tem izrazom pokriti različno velike količine jezika:

- posamezno besedilo iz korpusa (razrezanega po kriterijih za besedila, mišljena je polna besedilna enota),
- (pod)korpus besedil – več besedil iz korpusa, zbranih skupaj v enoto po določenih združevalnih kriterijih (običajno gre za sporazumevalne okoliščine),
- vloga govorca – kar posamezen govorec spregovori znotraj besedila ali korpusa: ne gre za besedila kot celote, marveč iz posameznega besedila ali podkorpusa besedil „iztrgane“ izjave, ki so jasno pripisane določenemu govorniku (znotraj posameznega besedila pa so iztrgane vse takšne izjave govornika).

4.3.1. **Umestitev pogostosti različnice v slogovni model**

Slogovno variiranje, znotraj katerega želimo določiti slog različnih besedilnih enot v tej nalogi, so pogostostne značilnosti besedišča – različnic; to je razlaga stilometrije v tej diplomii. Vsako besedilno gradivo, za katerega ugotavljamo slog, razbijemo na pogostostni seznam različnic.

Slika 6: Pogostostni seznam različnic v slogovnem modelu

IZRAZ			VSEBINA
SLOG POSAMEZNEGA, STILEM = razlike v pogostostih istih različnic			
SLOGOVNE ZNAČILNOSTI = razlika v pogostosti (deležu) različnice in njenem zaporednem mestu v pogostostnem seznamu med primerjavama	1. RAVEN RAZLIKOVANJA (izbiranja)	KO-NOTACIJA = večja/manjša izražajnska važnost različnice med drugimi v gradivu	
JEZIKOVNA ENOTA = različnica v pogostostnem seznamu posameznega besedilnega gradiva			
IZRAZNA ENOTA = pogostost različnice v besedilu (kot delež)	2. RAVEN ENAČENJA (prototipa)	VSEBINSKA ENOTA = izražajnsko razmerje različnice do celega seznama	
SISTEM ENOT = ena oz. več gradiv, ki iz katerih vzamemo sezname za primerjavo			
NEPOSREDNE OKOLIŠČINE = posamezno besedilno gradivo, njegov seznam različnic in sporazumevalne okoliščine besedila	3. RAVEN PRIMERJANJA (množice, sistema)	SISTEM JEZIKA = besedilna gradiva v primerjavi in njihovi dejavniki; iste okoliščine, ki gradiva ponujajo v primerjavo	
KONKRETNO			ABSTRAKTNO

Zgornja slika daje pregled, kaj nam v stilistični primerjavi po našem modelu pomenijo različni elementi primerjave. *Enota primerjave*, ki jo želimo vzeti pod drobnogled, zapisana na 2. ravni, je različnica gradiva skupaj s svojo pogostostjo in mestom v seznamu različnic. Naslednji dve podpoglavji razlagata 1. raven, raven razlikovanja, za to enoto primerjave. Poglavje I.4.3.2 nato pojasnjuje druge enote stilometričnih primerjav, poglavje I.4.3.3 pa razlaga 3. raven našega modela za vse enote primerjav hkrati; sistemi primerjav niso nič drugega kot pari besedilnih gradiv (npr. besedilo in besedilo ali podkorpus in podkorpus), ki jih primerjamo.

4.3.1.1. Pogostost različnice kot kvantitativno izražanje sloga

Osnovno merilo analize sloga je kvantitativno merilo **pogostosti**. Zakaj pogostost (besed)? Predvsem je preprost pripomoček, ki abstraktno besedišče nekega besedila **aktualizira**. (Razbijanje nekega linearnega⁸⁹ besedila na besedni seznam (slovar, leksikon) je pravzaprav že metajezikoslovno, formalni jezikovni sistem ustvarjajoče dejanje; poleg jezikovnih informacij, ki so shranjene v samih besedah, je kriterij uvrstitve v seznam edina informacija, ki ostane iz celote nerazbitega besedila.⁹⁰) Če ob tem velja, da je seznam po pogostosti tudi urejen (ne recimo po abecedi različnic kot pri običajnem slovarju), je uporaben kvantitativni podatek tudi mesto v seznamu, razmerje do pogostosti ostalih enot v seznamu, za primerjavo pa lahko vzamemo le del seznama, npr. nekaj najpogostejših besed. Pri analizi sloga je tako uporaben vsak podatek, ki preučevano slogovno značilnost veže na aktualno besedilo (slogovni učinek je bistveno pogojen z aktualiziranostjo v besedilu); tako pogostost kot razmerja med enotami pogostostnega seznama besed temu ustrezajo.

Osnovni podatek o pogostosti predstavlja naravno število (absolutna vrednost), za primerjavo različno velikih besedilnih gradiv pa je bolje, da to vrednost pretvorimo v delež: delimo ga s številom vseh pojavnih v gradivu. Tako preračunane vrednosti ne spremenijo zaporedja v seznamu.

Termin pogostost nikakor ne pomeni nujno le tega, da so enote *zelo* pogoste. Nasprotno, tudi enkratnice imajo svojo pogostost (z vrednostjo 1). Vsekakor pa je važno vprašanje, kje na **lestvici pogostosti** je pričakovati slogovno učinkujoče in razpoznavne enote.

Ena od osnovnih značilnosti sloga je njegova stalnost (prim. uvod o slogu, poglavje I.4). V jezikovnem slogu torej ponavljanje, „rekurenca“ (vračanje, ponovno pritekanje) jezikovnih sredstev. Iz tega kriterija bi torej sledilo, da so najpogostejša jezikovna sredstva najbolj značilna.⁹¹ Druga značilnost sloga je izbranost nekaterih njegovih elementov; teorija o stilni nezaznamovanosti (in teorija naravnosti) pa nam govori, da so ravno najpogostejši elementi naravni za jezik, nevtralni, nezaznamovani (prim. opomba 83 na str. 57; tam smo obdelali to navidez paradoksalno razmerje pogostosti in „obrobnosti“ sloga). Tretja značilnost sloga, pripisanost posameznemu (besedilu, besedilom posameznika itd.), pomaga pojasniti, od kod vzamemo sezname enot. Imamo dva ali več pogostostnih seznamov, prvi je izdelan iz posameznega gradiva, ki mu določamo slog, vsak naslednji seznam pa je primerljiv sistem.

⁸⁹ Kot besedila v širšem pomenu so lahko šteti tudi sami sezname besed, tabelarično, grafično ali zvočno drugače organizirane besede in druge jezikovne enote.

⁹⁰ Podatek o pogostosti enote v seznamu (besede kot iztočnice, oblike, leme itd.) lahko pripišemo različnim seznamom in v vsakem primeru dopolnimo informacijo, s katero se („sistemski“) seznam približa aktualnemu besedilu, iz katerega izhaja.

⁹¹ Nasprotno mnenje pa je, da so prav najpogostejše besedne enote *neodvisne od stila*, ker so zastopane tudi v raziskavah drugih besedilnih zvrsti – Neweklowsky (1984: 399) tako primerja rezultate iz publicistike z rezultati iz leposlovja, ki jih vzame iz Suhadolnik/Janežič 1962. Tudi pri primeru iz SSKJ smo središčno, najpogostejše besedišče imeli za nevtralno.

Iz sinteze vseh treh kriterijev (moč ponavljanja, zaznamovanost z redkostjo, sistemi primerjav) lahko predvidimo **6 predpostavk o slogovni zaznamovanosti različnic s pogostostjo**:

- 1)) zaznamovane so enote tistih pogostosti, ki odstopajo ob povprečne pogostosti iste enote;
- 2)) bolj odstopajo od povprečja, bolj so zaznamovane;
- 3)) če odstopajo od povprečja v smeri večje pogostosti, so bolj zaznamovane kot pri odstopu v smer manjše pogostosti;
- 4)) pogostejša v seznamu je enota, bolj zaznamovano je njeno odstopanje;
- 5)) enota, ki je v seznamu zelo pogosta, v primerjanih seznamih pa je sploh ni, je zelo zaznamovana;
- 6)) bolj širok kot je sistem primerjav (velikost posameznega gradiva in količina primerjanih gradiv), bolj zaznamovano je odstopanje enote od povprečnega deleža.

Pogostosti torej dodatno primerjamo kvantitativno še tako: za različnice, ki se nahajajo v obeh gradivih, izračunamo povprečen delež in odstop od tega povprečja, ostale različnice pa označimo z dvočrkjem „BP“ (brez primerjave; če se različnica pojavlja samo v enem gradivu) ali „pp“ (premalo primerjav; če se različnica pojavlja v več kot enem gradivu, a ne v vseh primerjanih, kar onemogoča izračun povprečja). Nato za primerjavo vzamemo le 20 najpogostejših besed (ki naj bi po ocenah pokrile vsaj četrtino vseh pojavnici v gradivu) iz posameznega gradiva in na vsakem seznamu označimo 5 različnic, ki najbolj izstopajo po izračunu in tudi vse „BP“; preostala neoznačena manjša izstopanja pa so razvidna iz izračunanih podatkov.⁹²

Primer 14: Pogostostni seznam različnic za potrebe stilometrične analize

tip primerjave	Primerjava n: opis, npr. govorec proti sobesedilu ... (opis podtipa primerjave, npr. različne okoliščine) /koliko je različnic skupaj v vseh primerjanih seznamih po 20, npr: 23 različnic v spodnji primerjavi					
	besedila: 1ned: 02, 04–06 ... 6pet: 01–03, 05 ...					
	1. gradivo: sobesedilo (govor brez 001Mz)			2. gradivo: samo govor 001Mz		
	R= 6.265	P= 32.822	R/P= 0,191	R= 3.173	P= 13.485	R/P= 0,235
zap. št.	različnica	delež [%]	razlika	različnica	delež [%]	razlika
1	je	3,66	0,38	pa	3,71	0,05
2	pa	3,6	-0,06	ne	3,21	0,30
3	ne	2,61	-0,30	ja	2,97	0,35
4	ja	2,27	-0,35	je	2,91	-0,38
5	da	1,96	0,21	*e	1,7	0,46
6		

Seznama različnic in vse važne izračunane podatke prikažemo zaradi boljše preglednosti vzporedno v tabeli. V glavi tabele pa so še podatki o gradivu, ki ga primerjamo, in v zvezi z različnicami še podatek o številu vseh različnic v primerjavi (minimalno 20, več pa ko jih je, bolj raznoliki so seznam). Gradiva so od leve proti desni razporejena

od največjega proti najmanjšemu (gl. tudi vrednost P kot pojavnice).

⁹² Zakaj 20 najpogostejših besed? Potrebno je bilo izbrati neko univerzalno količino. Bolj kot da bi izbirali „okrogle“ številke (10, 20, 30, ... 100) se je zdela univerzalna količina seznam najpogostejših besed, ki pokrije nek univerzalen delež (polovico, tretjino, četrtino, petino besedišča – pojavnici), pa ni ne predolg za grafične predstavitev na A4 listu ne preskromen, da bi ne razkril razlik v pogostostih najpogostejših različnic za našo velikost besedilnega gradiva. Polovico pojavnici pokrije v govornem korpusu okrog 120 najpogostejših besed, tretjino okrog 35 besed, četrtino 18 itd.; seznam najprimernejše dolžine se zdi torej 20 besed. V pisnih korpusih ter korpusih bistveno drugačnih velikosti so razmerja seveda drugačna.

Zakaj 5 najbolj izstopajočih različnic? Ta vrednost pa je izbrana kvečjemu po psihološki značilnosti človeškega dolgoročnega oz. srednjeročnega spomina, da si je sposoben zapomniti nekje do 7 enot v seznamu; število pa je po zaznamovanosti z našim desetiškim številskim sistemom zaokroženo na 5.

4.3.1.2. Slogovni pomen različnice in njene pogostosti

Kako lahko izrazno pojmuje besedo in kaj nam o posamezni besedi pove njeno metajezikoslovno kategoriziranje, smo razdelali že v poglavju I.2.3. Usmerjeni smo torej v besede kot sestavne dele kompleksnejših jezikovnih enot, pa tudi pri besedah slogovno ne preučujemo vsega. Ob strani pustimo njihovo etimologijo, družljivostne in vezljivostne zmožnosti, fonetično podobo, tvorjenost oz. morfemsko sestavo, podrobnejšo umestitev pomena in še kaj. S transkripcijo zabeležimo fonemsko, pregibno in s tradicijo zapisa utemeljeno razlikovalnost besede – različnice. Različnica v kontekstu (gl. Priloga 5) sicer marsikaj pove, prepoznamo jo lahko kot leksem z aktualnim pomenom.

Različnica v pogostostnem seznamu pa je oropana konteksta in pove zelo malo. Pomen, ki ga lahko bralec razbere, je reduciran s seznamom:

- vsebina (pomen) različnice v seznamu je unija ali vsota vseh verjetnih denotativnih in kategorialnih pomenov, ki jih lahko pripišemo temu sklopu fonemov (enakoizraznici – transkripcijski enakopisnici);
- vsota pomenov je brezoblična, razmerja med verjetnostmi možnih pomenov (oz. konkretnimi v besedilu) niso razvidna (razen v primeru, da je pogostost 1 in ima oblika le eno možno kombinacijo denotativnega in kategorialnega pomena – to možnost izkoriščajo tudi POS-taggerji);
- dodana pogostost sicer izraža neko razmerje do besedilnega konteksta, kot smo nakazali v prejšnjem poglavju, a je tudi to razmerje zelo abstraktno.

Tako razbran slog ima malo vsebine – ne pove kaj dosti (celotne) predstave o posameznem (o sporazumevalni situaciji, besedilu, posamezniku), zaradi revnega sistema same različnice (kot besedne enote, in v seznamu) in zaradi revnega sistema primerjav (samo 2 besedili ali nekaj več primerjanih). *Pogostost posamezne različnice da le predstavo o izražanjški pestrosti zaklada besednih oblik besedila, definirani z grafofonemskimi različnicami.* Ker je slovenščina zelo pregiben jezik, je samih grafofonemskih razlik med pojavnicami sicer precej, a vseeno ne povedo dosti, dokler jih ne povežemo z dodatnimi oznakami (oblikoskladenjske določitev, podpomeni itd.). Če pa vzamemo skupaj 20 najpogostejših različnic, pa dobimo že malo bolj konkretno predstavo o četrtini zaklada besednih oblik besedila.

Različnice povedo strašno malo – bi lahko izbrali kakšne druge enote ali preračunavanja ob tako označenem gradivu, kot ga imamo?

Zanimive bi bile nepregibne (oz. t. i. slovničnorazmerne ali sinsemantične) besedne vrste – med najpogostejšimi različnicami je večina takih.⁹³ Obstajajo stilometrične raziskave, ki se naslanjajo na te besede (prim. Holmes 1994: 90–91). A kaj, ko te besedne vrste pogosto konvergirajo med seboj in z zaimki (potrebna bi bila besednovrstna določitev), medmeti pa v govoru nimajo niti stalnih oblik (za analizo teh bi bila potrebna poleg besednovrstne določitve še lematizacija z definiranjem idealne oblike medmeta). Brez teh dodatnih oznak bi lahko vzeli nepregibne besedne vrste le ne glede na verjetnost, da konvergirajo. Druga hitra rešitev bi bila, da bi samo za te besedne vrste preverili možne besednovrstne določitve v strojnem določevalniku. Pa kaj, kot ne vemo, koliko ima ta v svojem učnem gradivu primerov govornega gradiva – verjetno ne bi razkril vseh možnih pomenov besednih oblik. Tako s pomočjo teh dveh bližnjic pač ne moremo *zagotovo* in natančno interpretirati boljše vsebine iz stilometrične analize najpogostejših besednih oblik. To ostane odprto vprašanje, dokler nekdo ročno ne začne določevati besednih vrst in lem govora.

Pogost prijem (stilometričnih) raziskav netagiranega gradiva korpusov je tudi preračunavanje dolžin besed v zlogih ali znakih. Ker pa imamo tu opravka z govornim gradivom,

⁹³ Neweklowsky (1984: 387) ugotavlja, da niti iz 20 najpogostejših besednih oblik (21,65 % korpusa) niti 20 najpogostejših besed (lem z odvisnimi pregibnimi oblikami; 26,43 % korpusa) ni mogoče sestaviti stavka – njegov korpus iz člankov časopisa Delo, na podlagi katerega dobi pogostostni seznam in pride do tega sklepa, je velik 10.654 pojavnic, 5143 (grafofonemskih) različnic, 2886 različnih lem.

za katerega je značilno krajsanje (redukcije, zlitja itd.) ali daljšanje (vrinjeni, protetični itd. glasovi) fonske sestave besed in v primerjavi spremembe števila vokalov tudi zlogov, bi bilo to preračunavanje na podlagi ortografske transkripcije za lase privlečeno – veljale bi samo primerjave besedil, prepisanih v tej isti transkripcijski konvenciji.

4.3.1.3. Ovinek k pogostostim leksemov

Besede same vendarle tudi dosti povedo (predvsem besede kot leksemi), a moramo za to besedam v gradivu *posamič dodati še nekaj več podatkovne vrednosti* – oznako: lemo, ime besedne vrste, oblikoslovne kategorije, konkordanco (primeri rabe), definicijo pomena, sopomenke, geselski članek leksema ...

Z vidika pogostosti smo govorili o „besednih enotah“ pogostostnih seznamov: gre nasploh za tako ali drugače razlikovane enote besedne ravnine (pojavnice, različnice, leme, leksemi, besedne vrste ipd.), čeprav se v tej diplomii lotevam le pojavnice in različnic, ker gradivo ni dlje označeno. S samega kvantitativnega vidika vrsta enote (besede) ni važna, pač pa je z njo povezano pričakovanje, kje na lestvici pogostosti bo največ odstopanj od povprečja.

6 predpostavk o slogovni zaznamovanosti s kvantiteto iz poglavja 0 torej velja le za različnice, za (grafofonemsko) *obliko* besed; pa še to le znotraj posameznega nacionalnega jezika, saj jeziki po svoji fleksijskosti bistveno razlikujejo. Za drugače razlikovane enote besedne ravnine bi dodali še 7. predpostavko:

7)) kompleksnejša kot je enota (pomensko; skladenjskofunkcijsko polnopomenske bolj od slovničnorazmernih besed; tudi izrazno daljše besede bolj od krajših), bolj je zaznamovana.

Suhadolnik in Janežičeva v naslednjem citatu recimo govorita o leksemih; znotraj njih podrobneje razdelata jedrno besedišče, znotraj tega pa polnopomenske besede. Citat je sicer dolg, navajam ga kot važno razpravo, ki bi jo zaradi starosti objave lahko hitro spregledali, in ker naj sklepanje o pomenu pogostosti leksemov v njej služi kot primerjava mojemu sklepanju o pogostostih (grafofonemskih) različnic.

Živa leksika /preostali skupini sta t. i. nova in mrtva leksika; .../ se razlikuje tudi po pogostnosti v rabi, po frekvenci. Z njo so najtesneje povezana vprašanje vloge in pomenov posameznih besed. /.../ Besedne vrste so v slovnici obdelane kot med seboj enakovredne kategorije, vendar jih je treba za leksikalno in stilistično rabo porazdeliti v tri skupine: *polne besede* so glagoli, samostalniki, pridevniki in prislovi. /.../ K *pomožnim besedam* spadajo predlogi, vezniki in členki. /.../ V kategorijo *indiferentnih besed* pa, ki se vežejo s polnimi in pomožnimi besedami, spadajo: števnik, zaimki in medmeti. /pomožne in indiferentne besede v nadaljnjih primerjavah imenuje skupaj pomožne/

Najpogosteje uporabljamo v govoru in pisavi (v govoru seveda več) pomensko prazne, t. i. pomožne besede, npr. *v, na, in, da, nič, se on*; tem slede po pogostnosti pomožni glagoli in polne besede z zelo širokim pomenskim obsegom, npr. *biti, iti, imeti, oče, človek, mati, rad, velik*. /.../

Natančnejšega števila besed /.../ osrednji skupini nevtralne leksike ni mogoče določiti, vendar je dokazano, da vsak človek lahko shaja s pet tisoč leksemi. /.../ Človek izgovori na dan od deset do osemnajst tisoč besed in pri tem uporabi kvečjemu tisoč različnih slovarskih enot, to je leksemov. /.../ Orientacijski odnos med besedami posameznega teksta in njegovimi leksemi je 7 ali 8 : 1, oz. bolj preprosto povedano: tekst s 40.000 besedami ima kakšnih 5.000 leksemov.

Teh 5.000 leksemov razpada v dve močni skupini:

1. v 100 pomožnih leksemov, to je predlogov, veznikov, števnikov, zaimkov in členkov ter

2. v 4.900 polnih leksemov, to je glagolov, samostalnikov, pridevnikov in prislovov.

/.../ Da bo odnos med podskupinami polnih leksemov še očitnejši, vzemimo za izhodišče namesto tipičnih 5.000 leksemov najvišje možno merilo, to je 25.000 leksemov, ker so strašno redki pisatelji, katerih besedišče seže čez to številko, mnogo pa jih je globoko pod njo (Dante 5.800, Shakespeare 15.000, Baudelaire 25.500, Claudel 50.000). Med temi 25.000 polnimi leksemi bo:

50 t. i. *vodilnih leksemov*, ki pokrijejo 19 % celotnega števila polnih besed /so v vsakem stavku; .../

4.000 *osnovnih leksemov*. Ti pokrijejo 80 % polnih besed in tvorijo jedro, osnovo človekovega govora.

Preostalih 20.000 leksemov pokrije samo 1 % polnih besed. To so leksemi z majhno frekvenco, so pa pomensko jasno opredeljeni in jih zato imenujemo *značilni leksemi*. V njihovi izbiri in rabi se kažejo razlike med *posamezniki*.

/Nadalje omenja primera, da imata 2 strokovnjaka vsak na svojem področju še vedno/ nad 60 % skupnih frekvenčno najmočnejših besed. Kadar pa se lotita iste teme /.../, ne bo med njima ne v izboru ne v frekvenci skoraj nobenega razločka.

Po vsem tem bi bil študij leksike posameznih avtorjev in vzporedno s tem preiskovanje njihovega stila s statistično analizo nesmiseln. A stvar ni taka. Pri umetniku prevladuje, kot smo že poudarili, individualna izbira leksike, in zato se bolj ali manj oddaljuje od norme. Neredko se zgodi, da so avtorju katere besede posebno ljube, in te vnaša tako pogosto v tekst, da se vzpno v *frekvenčni lestvici visoko nad povprečje*. Ker so te besede ključ za razumevanje avtorjeve osebnosti, jih imenujemo *ključni leksemi*.../

Odnos med besedami, ki jih človek pozna, in med tistimi, ki jih navadno uporablja, pa je približno 11 : 1.“ (Suhadolnik/Janežič 1962: 48–49. Poudarke napravil R. B.)

4.3.2. **Kvocient različnic in pojavníc (R/P) v slogovnem modelu**

Ob prejšnjih dveh kvantitativnih metodah, ki izhajata iz pogostosti posamezne različnice besedilnega gradiva (gl. Slika 6, str. 61), v tej nalogi uporabljam še eno kvantitativno metodo, ki predstavi vse različnice z njihovimi pogostostmi hkrati.

JEZIKOVNA ENOTA = pogostosti različnic celega besedila skupaj		
IZRAZNA ENOTA = razmerje različnic in pojavníc	2. RAVEN ENAČENJA (prototipa)	VSEBINSKA ENOTA = izražanjška pestrost besedila kot celote

Prva predstavitev vseh različnic besedilnega gradiva z njihovimi pogostostmi hkrati je **razmerje različnic in pojavníc** (angl. type-token ratio), aritmetična vrednost iz števila različnic, deljenega s številom pojavníc v besedilnem gradivu. V naši diplomii ga bomo označili „R/P“. (V angleški literaturi $R = V/N$: R kot „ratio“, V število različnic oz. „types“, N število pojavníc oz. „tokens“.)

Primer izračuna za 1. gradivo, gl. Primer 14 zgoraj: $R = 6.265$, $P = 32.822 \rightarrow R/P = 0,191$.

Pri tej metodi slogovne razlike med primerjanimi gradivi pomenijo relativno večjo ali manjšo pestrost zaklada besednih oblik, je pa pri drugem razlika med gradivi manj odvisna od velikosti gradiva (v pojavnícah).

Pogoste kritike v literaturi letijo na omejenost tega izračuna glede na značilnost jezika: z večanjem števila pojavníc se število različnic ne večja premosorazmerno, ampak se bliža neki končni vrednosti (Holmes 1994: 92).⁹⁴ Relativnost tega razmerja bom skušal upoštevati tudi pri njegovem stilističnem interpretiranju.

4.3.3. **Sistemi primerjav sloga**

Nazadnje smo dolžni pojasniti le še 3. raven slogovnega modela, in sicer njen abstraktni pol, saj je njen konkretni pol nič drugega kot besedilno gradivo, iz katerega delamo seznam različnic. Abstraktni pol pa pomeni vsa druga besedila (in iz njih razlomljene pogostostne sezname različnic), ki jih postavimo v primerjavo – sistem primerjav: običajno dve primerjavi, lahko pa več. Pri izbranih kvantitativnih metodah in korpusnem pristopu ta sistem primerjav pravzaprav niti ni abstrakten – gre za konkretna besedilna gradiva.

Besedilna gradiva za primerjave izbiramo v različnih količinah (gl. spodnji seznam: zgoraj so večji, spodaj količinsko manjši sistemi primerjav), vseeno pa želim izbrati le takšna gradiva, ki so čim večja (nekaj tisoč pojavníc za posamezno besedilno gradivo v primerjavi) in ki imajo enakovredno količino tudi v primerjanem besedilnem gradivu v primerjavi.

⁹⁴ V kvantitativni lingvistiki je zelo znana študija G. Kjetsaa iz l. 1979, v kateri so bile opravljene primerjave izračunov R/P različnih odlomkov romana Tihi Don, vseh dolgih isto količino – 500 pojavníc. (Holmes 1994: 92).

5 SISTEMOV PRIMERJAV:

- 1)) **korpus** (prenosniška skupina):
 - # 1 podkorpus vsega prostega govora s pisnimi korpusi;
- 2)) **okoliščine** sporazumevanja med seboj:
 - # 2 različne okoliščine (podkorpusi ali vsaj besedila);
 - # 3 posamezna besedila istih okoliščin;
- 3)) **govorci** med seboj:
 - # 4 v istih okoliščinah (podkorpus ali vsaj besedilo);
- 4)) vloge izbranega govorca proti sobesedilu (**sobesedilo**):
 - # 5 podkorpus vseh besedil, kjer govorec nastopa (različne okoliščine);
- 5)) **vloge** izbranega govorca med seboj:
 - # 6 izbrani govorec v različnih okoliščinah (podkorpusi ali besedila);
 - # 7 izbrani govorec v istih okoliščinah (več besedil).

Že na prvi pogled je vidno, da ne primerjamo samo idiolekta (zadnji trije sistemi), ampak celo sociolekte, definirane z vidika različnih okoliščin (prva dva sistema). Iz tako zastavljenih analiz želim ugotoviti, kako se naš slogovni model izkaže za različne količine besedil in za različne okoliščine sporočanja, in tudi, kaj so sploh stalnosti v govoru. Da se ne izgubimo med množico primerjav, ima vsak sistem svojo ključno besedo (odebeljeno) in – pod njo bo možno najti tudi primerjave v poglavju analiz (IV). V nadaljevanju pa nekaj komentarjev in predpostavk o posameznih sistemih primerjav.

AD korpus:

Glede razlike pisna – govorna besedila je možno predvideti naslednje: pri pisnih je več refleksije kot pri (spontanih) govornih; slednja kažejo gotovo večjo kaotičnost, preprostejšo skladnjo in, kar bomo lahko preverili, preprostejše besedišče.

AD okoliščine:

Za poimenovanje mojih seznamov besedja besedilnih gradiv, definiranih s pomočjo okoliščin, i in ker se nanašajo na več govorcev, uporabljam izraz *sociolekt* (zakaj, gl. poglavje I.3.1.2.1).

AD govorci, vloge, sobesedilo:

Besedilnim gradivom teh primerjav bi sicer lahko odrekli jezikovnost (uporabnost oz. pomenljivost za druge jezikovne uporabnike), češ da so samo del idiolekta in ne tudi kateregakoli sociolekta, pa ta vidik zanemarimo. Predvidevamo, da je količina dejanskega idiolekta posameznika (razumljivega le njemu samemu), zanemarljivo majhna ali nična, sploh v nemonoloških situacijah.⁹⁵

Možno bo pa preveriti tudi trditev, da „č/lovek izgovori na dan od deset do osemnajst tisoč besed“ (Suhadolnik/Janežič 1962: 49).

AD sobesedilo:

V tem sistemu primerjav je besedilno gradivo v ospredju govor izbranega govorca, primerjalno gradivo pa ves preostali govor – v istem besedilu ali besedilih izbranega podkorpusa, v katerih govorec nastopa.

⁹⁵ Običajno jezikovne prvine nasploh seveda uporabljamo vsi, kar nam omogoča komunikacijo; izjeme so le take prvine, ki jih iznajde nek avtor sam; pa še te so na nek način primerljive z ustaljenimi, kar spet omogoči njihovo razumevanje.

Zanimivo pa je še drugo vprašanje: Ali o nekem posamezniku nekaj pove le jezikovna tvorba „iz njegovih ust“? Očitno ne (prim. npr. 2. in 3. osebo pri osebnih glagolskih oblikah in zaimkih, lastnoimenske kategorije, na nadbesedilni ravni pripoved o nekem), toda to ni več izrazni slog opisovanega posameznika, ampak tistega, ki se o njem izraža; literarno pravimo, da gre za neposredno karakterizacijo.

5. KORPUSNA ORGANIZACIJA IN RAČUNALNIŠKA OBDELAVA GRADIVA

Ob pripravljanju gradiva diplome za jezikoslovno analizo na računalniku se nisem izognil skušnjavi „trendovsko“ ga poimenovati – korpus. S tem sem tudi svoje delo uvrstil v skoraj že pol stoletja star pristop korpusnega in računalniškega jezikoslovja, ki zmore iz računalniških zbirk besedil potegniti komaj kaj bolj utemeljena spoznanja kot jezikoslovci in raziskovalci literature, ki se lotevajo besedilnih zbirk izven računalnika, čeprav so pričakovanja uporabnikov omenjenega pristopa mnogo večja. Metode analiz in uporabe korpusov pa so od 90. let 20. st. stalno v porastu, tako v tujih raziskovalnih sferah (npr. najstarejši ameriško-angleški) kot med Slovenci, tako da *začetniških* problemov povezovanja strojnih/računalniških/matematičnih prijemov in od drugod prinešenih slovničnih konceptov, s čimer se soočam tudi v tej diplomii, ne gre jemati za zlo oz. neznanstvenost.⁹⁶

Sam kot študent sem izkušnje tega pristopa deloma zbiral znotraj rednega slovenističnega študija (pri P. Jakopinu, mentorju), deloma v okviru cehovskih dejavnosti in literature slovenskih jezikovnih tehnologov, ki izražajo lep porast: konference Jezikovne tehnologije (JT) na IJS, srečanja Jezikovnotehnološkega abonmaja pri SDJT, poročila o delu raznih drugih raziskovalnih centrov (SAZU, fakultete za računalništvo in elektrotehniko po slov. univerzah, razna podjetja, slavistika v Gradcu – B. Tošovič, P. Grzybek) in bolj ali manj javni slovenski korpusi. Naštevam jih, ker sem svoj pristop oblikoval bolj po slovenistiki, ki je (še) dobro začela kovati jezikoslovje na podlagi JT, kot po morda že bolj naprednih tujih jezikoslovjih, pri tem pa se zavedam svoje začetniškosti tudi pri zgledovanju po domačih dosežkih.

Pri oblikovanju svojega korpusa sem tako skušal slediti nekaterim iskanim značilnostim korpusov oz. ciljem, ki so v tej stroki vedno bolj dosegani: zbrati čim večji korpus, ga čim podrobneje jezikoslovno označiti, zbrati govorno gradivo oz. gradivo s čim širšega področja realnega sporazumevanja, ker so raziskave tega gradiva slovenščine še nove oz. neznanne, narediti gradivo javno dostopno, lahko uporabno, v standardnem zapisu in po možnosti vključljivo v referenčni korpus (ortografska transkripcija) in še kaj.

Izpostavil bi pa še eno značilnost korpusnega/računalniškega pristopa, ki izhaja iz njega samega in bogati tradicionalno jezikoslovje: odkrivanje novih jezikovnih struktur ter širjenje konsistentnosti v strukturalnem jezikoslovju. Računalniške analize so namreč v svoji omejenosti vendarle toliko objektivne, da zahtevajo dosledno definicijo analitičnih razredov, ki je dovolj kontrastivna in konsistentna: za razporeditev podatkov jasen in zaključen sistem oznak (npr. transkripcijskih), razredov, znakov (npr. za besedilne okoliščine) – četudi se en razred imenuje „zadnji“, „razno“, „brez“, „neznano“ (prim. Edwards 1995: 21–22). Ta tehnološka (konsistentnost; omogoča večjo obvestilnost oziroma nižjo entropijo) in analitična (kontrastivnost) zahteva posledično pogosto jezikoslovno teorijo (sistem) obogatita za kako novo spoznanje o realnem jeziku.

5.1. Zgledi govornih korpusov z govorom v ortografski transkripciji

Znani korpusi sledijo različnim namenom, različni so tudi prenosniki gradiva (v večini so pisni, pa tudi govorni/zvočni, redko vizualni) in vrste obdelave. Za zgled sem si vzel take **govorne korpus**e (angl. spoken language corpora, corpora of spoken discourse; Leech idr. 1995: 2), ki hranijo posnetke in ortografske zapise govora iz realnih govornih situacij – in sicer glede na dostopnost literature in sorodnost njihove korpusne ureditve z mojimi raziskovalnimi potrebami.

⁹⁶ Skeptičen pogled na računalniško jezikoslovje v slovenskem prostoru opisuje že Poniž (1974: 33), za področje govornokorpusnega jezikoslovja probleme sodelovanja jezikovnih tehnologov in tradicionalnih jezikoslovcev opisuje Sinclair (1995: 108), o odnosu do empiričnega pristopa širše v humanistiki oz. ožje v literarnih vedah pa razpravlja Hladnik 1995.

Od tujih je izbran British National Corpus (BNC), od domačih pa Učni korpus govornje slovenščine (UKGS) in korpus telefonskih pogovorov v turizmu Turdis-1; slednji zgolj kot zgled transkribiranja govora. Podrobneje so predstavljeni v treh sledečih podpoglavjih.

Po rešitvah urejanja gradiva iz teh korpusov se nisem zgledoval „suženjsko“, ampak so mi služile le kot študijska osnova; saj vendar nimam enakih ciljev uporabe kot omenjeni korpusi. Primerjave izbranih korpusov z mojimi rešitvami so vidne iz v poglavjih II.7.4 (za zapisa čistega besedila) in II.8.7 (za zapis okoliščin in metaoznaka besedil). Vodila pri oblikovanju lastnega govornega korpusa pa so bila tudi sistemska priporočila na podlagi drugih korpusov (Edwards 1995 in sploh ves prvi del zbornika Leech idr. 1995, pa EAGLES 1996), npr. po kriterijih:

- časovne in finančne omejitve obdelave,
- ročna – strojna obdelava,
- pisna narava zapisa – govornost izvornih podatkov,
- programska oprema,
- standardi,
- kvaliteta – kvantiteta,
- berljivost za zapisovalca oz. uporabnika.

In še en kriterij, izpostavljen v slovenski poziciji. Delo obdelave govora je že v osnovnih fazah zbiranja virov in transkribiranja zelo zamudno in obsežno, posebej za slovensko pozicijo, kjer je ta prva faza in analitična izraba govornih korpusov še v razvoju. Zato se hvaležno oziram po tujih zgledih in kljub samotarskemu ustvarjanju diplome ponujam svoje gradivo in skromne izkušnje v razpravo, da bi nekoliko prispeval k „združevanju raziskovalnih potencialov pri preseganju jezikovnih pregrad v okviru jezikovnih tehnologij naslednjih generacij“ (Kačič 2002).

Splošno stanje v Evropi je tako, da ima dandanes že vsak zahodnoevropski jezik vsaj en svoj govorni korpus, srednje- in vzhodnoevropski pa jim počasi sledimo. V naslednjem krajšem pregledu je za vsak tip korpusov navedenih le par zanimivih primerov (zahodnoevropski oz. angleški, srednjeevropski oz. slovanski in slovenski, kjer obstaja primer), več spletnih povezav do govornih korpusov pa je v poglavju VI.21.1.⁹⁷

- **samostojni govorni korpusi:**
 - Cambridge and Nottingham Corpus of Discourse in English (CANCODE , 17 mio. besed, v okviru Cambridge International Corpus, CIC, dopolnjen z 22 mio. besed),
 - nizozemski govorni korpus (Corpus Gesproken Nederlands, CGN, 9 mio. besed oz. 800 ur posnetkov; Oostdijk 2006),
 - Longman spoken American Corpus (5 mio.),
 - manjši korpusi (po)govorne bolgarščine iz 80. in 90. let (K. Aleksova, C. Nikolova 50.000, I. Mavrodieva 20.000 besed parlamentarnih debat; BGSpeech 2004),
 - slovenska UKGS, Turdis-1 (gl. naslednja podpoglavja);
- kot **del referenčnih/uravnoveženih korpusov** za potrebe najširših študij nacionalnih jezikov („mešanih“ – zapisi govora in pisna besedila skupaj, govorni del redko presega par odstotkov celote):
 - BNC (gl. naslednje podpoglavje),
 - The Bank of English (BoE, l. 2002 ima govorni del 20 mio. besed, a se korpus širi; njegov sestavni del The COBUILD Spoken Corpus je že l. 1994 imel 10 mio. besed, tradicijo zapisa pa že iz 80. let 20. st., ko je v projektu sodeloval John Sinclair, ki je v zgodnjih 60. letih oblikoval prvi računalniški govorni korpus, velik 220.000 besed – Sinclair 1995: 99)

⁹⁷ Literatura za korpusne angleščine, kjer ni posebej navedena: McEnery/Wilson 2001: 184–185, McCarthy 1998: 5 sl., kjer je tudi zgodovina manjših angleških govornih korpusov od 50. do 90. let 20. st.

- Češki nacionalni korpus (njegova dela praški in brnski govorni korpus skupaj 0,7 mio., kasneje dopolnitve; Kopřivová/Waclawicová 2005);
 - zgolj transkripcije brez posnetkov (parlamentarnih sej) pa imata že največja domača pisna/mešana korpusa, Nova beseda (27 mio.) in Fidaplus (2,04 mio.);
 - korpusi v različnih jezikih, ki zbirajo **medijski govor – t. i. „Broadcast-News“**:
 - domači zgled je SiBN (Žibert/Mihelič 2004 in Žgank idr. 2004),
 - hrvaški TV HRBCN (Martinčič Ipšić idr. 2004);
 - **stara govorna gradiva**, nekoč prepisana, zdaj v postopku povezovanja transkripcij z zvočnimi datotekami v računalniške korpuse oz. zbirke:
 - govorni korpus The London-Lund Corpus (izvorno 0,5 mio. besed, kasneje dopolnjen) kot polovica mešanega The Survey of English Usage (SEU),
 - Banca dati dell'italiano parlato (BADIP, v postopku je več korpusov, npr. Lessico di frequenza dell'italiano parlato – LIP Corpus, 490.000 besed; Schneider 2007),
 - Datenbank Gesprochenes Deutsch (DGD, 28 različnih podkorpusov iz 20. st., večinoma dialektoloških, Wagener 2004);
 - gradiva jezikoslovcev za **ožje analize govornega diskurza** (prehajajo od ortografske transkripcije k fonetični):
 - dialektološki in folkloristični posnetki Helsinki Corpus of English Dialects (245.000 besed), North Ireland Transcribed Corpus of Speech (NICTS, 400.000 besed);
 - otroški govor The Childes Databank (otroški jezik in jezikovne patologije), The Polytechnic of Wales corpus (POW, 61.000 besed), The Bergen Corpus of London Teenager Language (COLT, 0,5 mio. besed pogovorov najstnikov, Haslerud/Stenström 1995: 235)
 - tudi vsaka slovenska obsežnejša raziskava na tem področju ima za podlago posnetke in transkripcije govora (npr. Škofic 1991, Kranjc 1997: 314), le da gradivo zaradi časovne zamudnosti ni organizirano kot računalniški korpus in transkripcije ne vzporejene s posnetki;
 - **večjezični korpusi** z laboratorijsko posnetim govorom prevodov: domač primer je Multext-East V3 (Erjavec 2004: podstran Speech Files);
 - **govorne zbirke** (tudi *korpusi govora*, angl. speech corpora), zaenkrat predvsem za potrebe govornih tehnologij (GT), za domače zglede gl. Žganec Gros idr. 2003.⁹⁸
- Govorne zbirke za potrebe GT se po načelih lastne organizacije bistveno razlikujejo od govornih korpusov. Poglejmo primer biometričnega sistema za razpoznavanje govorcev (Dobrišek idr. 2006). Tudi taka raziskava ugotavlja istovetnosti govora osebe, a ne kot ta diploma na ravni idiolekta v besedah, marveč iz akustičnih značilnosti govora, ki so odvisne predvsem od ustroja posameznikovih govoril (zvočnost, barva, globina glasu, posebnosti – nosljanje, momljanje itd.). Zadostno gradivo je nekajminuten posnetek posameznikovega govora, ki ga samodejen razpoznavalnik govora razčleni v nize jezikovnih enot, mora pa biti govor karseda natančno, občutljivo posnet. Prenašanje izkušeni obdelave gradiva iz ene govorne zbirke v drugo niti ni kdovekaj odvisno od nacionalnega jezika govorca.

Ob dopolnjevanju seznama različnih govornih korpusov ob koncu diplome ugotavljam, da je v tujini na voljo še mnogo dobrih zgledov (kot velik korpus bi prišel v poštev BoE, kot eden sodobnejših in na raznovrstne načine označenih CGN, kot starejši zgled transkribiranja z dobro

⁹⁸ Slovenske govorne zbirke za GT je zanimivo primerjati zaradi njihove naprednosti v primerjavi z mednarodnimi GT, ki izhaja iz njihove specifične in napredne politike obdelave: iščejo najpreprostejšo obdelavo, se v veliki meri poslužujejo strojnega procesiranja, se ravnaajo po komercialnih potrebah in meddržavnih razpisih, na katerih so v stiku z najnaprednejšimi spoznanji, imajo smeje cilje raziskav (npr. strojno prevajanja govor-govor), ob tem pa širijo in prestopajo tradicionalne jezikoslovne okvirje.

tradicijo še COBUILD, pa še kak slovanski korpus za primerjavo), a je nadaljnja primerjava zahtevna, obsežna in presega okvir diplomskega dela.

5.1.1. BNC

„BNC je v izhodišču pisni korpus, čeprav dobrih 10 % transkribcije govora danes med vsemi /referenčnimi ali uravnoteženimi/mešanimi korpusi pomeni največji delež govornih besedil v korpusu.“ Njegova „korpusna dokumentacija /je/ za razliko od vseh drugih korpusov najbolj natančna in javno dostopna“ in zaenkrat „ostaja govorni del korpusa zaradi kakovosti podatkov nepresežen (Gorjanc 2005: 32, 37).

<i>ime</i>	British National Corpus (BNC) – zanima nas le njegov govorni del
<i>velikost</i>	90 mio. pisne in 10 mio. (700 oz. 630 ur govora iz 1200 ur vseh posnetkov) govorne britanske angleščine, od tega ok. 5 mio. besed spontanega govora
<i>čas ustvarjanja</i>	1991–1994
<i>ustvarjalci</i>	založbe Oxford University Press, Longman Group, Chambers Harrap; raziskovalna centra Oxfordske Univerze in Univerze Lancaster, British Library
<i>nameni uporabe</i>	v največji meri gradnja leksikografskih priročnikov omenjenih 3 založb, a tudi raziskave iz analize diskurza slovnice, splošnega jezikoslovja, jezikovni pouk, sociološke raziskave in še kaj; predvsem izbor posebnih jezikovnih vzorcev in pregled besedišča v konkordancah
<i>način zajema virov</i>	<i>demografsko izbran del</i> (4,21 mio.) sestavljajo vsakdanje konverzacije na podlagi statistike celotne britanske populacije izbranih posameznikov, ki so snemali svoje pogovore s prenosnim kasetnim snemalnikom 7 dni, ⁹⁹ <i>kontekstualno izbran del</i> (6,15 mio.) pa sestavljajo posnetki institucionalnega govora, v katerem se večina populacije ne pojavlja kot tvorec – po vnaprejšnji jezikoslovni kategorizaciji so bile izbrane 4 funkcijske zvrsti (enako 4 vrednosti našega podatka [32] <i>namen sporazumevanja</i> , gl. Tabela 26, str. 116, in poglavje II.8.7), količinsko enakovredno zastopane; zvočni posnetki do 1995 še niso bili vzporejeni s transkribcijami
<i>transkribcija oz. obdelava</i>	pretvorba v digitalne posnetke, razrez na besedilne/konverzacijske enote, sistematska transkribcija z izšolanimi prepisovalci (in poznavalci dialektov), transkribcije pregledane, široka ortografska transkribcija, malo prozodičnih in nič fonetičnih informacij, segmentacija: leksikalna/besedna enota, „poved“, izjava (speaker turn), besedilna enota; 17 metaoznaka okoliščin in obdelave na besedilno enoto in 10 metaoznaka za vsakega govorca; prenos v TEI konformno in preizkušeno (validated) SGML-strukturo, nato besednovrstno označeno (celoten korpus z naborom 61 oznak, delni podkorpus z naborom 170 MSD-oznak)
<i>prilagoditev posebnostim govora</i>	posebni sezname besed (za prilagoditev določevalnika govornemu delu korpusa in že sproti ob transkribciji ali ob prenosu v SGML: glavni leksikon (tudi za pisni del korpusa) in dodatna besedišča – sezname: sleng, tabuizirane besede, medmeti, besede z različno uporabo v govorni in pisni angleščini, krajšave, besede z izpuščenimi fonemi (na začetku ali koncu besede) in okrnjene besede
<i>programi obdelave</i>	Word Perfect za transkribiranje, Claws za določevanje besednih vrst, brskalnik po korpusu Xaira/SARA
<i>dostopnost na spletu</i>	na zgoščenki
<i>literatura</i>	http://www.natcorp.ox.ac.uk/ (2005, University of Oxford)
	Crowdy 1995, Garside 1995, Gorjanc 2005: 32–37

Naj povzamem tiste značilnosti obdelave, ki so mi bile zgled pri gradnji svojega korpusa:

- 7-dnevni model snemanju pogovora posameznika, ki sem ga poskušal izvesti tudi sam, a sem v končni meri prepisal le dva od posnetih sedmih dni;

⁹⁹ O avtentičnosti, pridobivanju govorcev za snemanje: varstvo osebnih podatkov zagotovljeno, snemanja se zaveda samo sodelavec korpusa, „problemi zaradi zavedanja snemanja so bili kvečjemu na začetku konverzacije, kasneje pa je ta naravno stekla. Govorci so lahko zavrnili vključitev posnetkov v korpus. Skupaj posnetih 700 ur govora, od tega vključenega 630 ur, kar je povprečno 8 ur posnetega oz. 5 ur uporabljenega govora na udeleženca. Povprečna dolžina korpusnega prispevka na govorca je bila 7000 besed (DBNC 1994: 13).“ Gorjanc 2005: 37. – Žal ni čisto jasno, koliko ljudi je bilo izbranih snemalcev, kolikšno pa je število vseh sodelujočih govorcev na posnetkih.

- rezanje posnetkov na besedilne/konverzacijske enote – hkrati širok zajem vsega, kar je posneto, ter razločevanje različnih okoliščin govora in izločanje nepotrebne tišine in slabih posnetkov;
- vzor transkripcijske in časovne treniranosti prepisovalcev (povprečno 750 zapisanih besed na uro dela) – glede na to, da sem hkrati označeval še časovne reference na posnetek, kar je podobno zamudno kot fonetično/prozodično označevanje, je ta vzor nedosegljiv – pri njegovi hitrosti bi 48.000 besed korpusa prepisal skoraj trikrat hitreje (moja hitrost je bila ok. 290 besed na uro, prim. Tabela 34, str. 126),
- taktika pregledovanja transkripcij – in izboljševanje transkripcij s pomočjo popravljenih napak;
- oblikovanje seznamov besed, ki so posebne v govornem jeziku – v BNC za potrebe prilagoditve besednovrstnega določevalnika, do koder v tej diplomii nisem prišel, sezname pa so pripravljene.

5.1.2. Turdis-1 in UKGS

Turdis-1 in UKGS sta slovenska korpusa, ki orjeta ledino pri zapisovanju in jezikoslovnem analiziranju spontanega govora, nastala ob doktorskih študijih. Prvi pomemben zaradi združevanja s stroko GT, drugi zaradi refleksije, s katero se ozira v splošne jezikoslovne raziskave in uravnoteženo združevanje s pisnimi korpusi (načeloma naj bi spadal pod korpus FIDAplus).¹⁰⁰

<i>ime</i>	korpus telefonskih pogovorov v turizmu Turdis-1	pilotski ali Učni korpus govornje slovenščine (UKGS)
<i>velikost</i>	za Turdis-1 izbranih 30 spontanih telefonskih pogovorov (106', 15.163 pojavnic, 2735 različnic, 2386 izjav, 44 govorcev, povprečna dolžina pogovora 3,5') iz celote 5 ur posnetkov	7 posnetkov pogovorov (skupaj 89'), 20 govorcev, dobrih 15.000 pojavnic
<i>čas ustvarjanja</i>	snemanje februar–junij 2004	posneto 2004, 3-mesečna obdelava 2004
<i>ustvarjalci</i>	Fakulteta za elektrotehniko, računalništvo in informatiko, Laboratorij za digitalno procesiranje signalov (Maribor)	Jana Zemljarič Miklavčič v okviru doktorskega študija in študijske izmenjave na Oddelku za kulturo, jezik in jezikovne tehnologije (AKSIS) na Univerzi v Bergnu, Norveška; pretvorba za zvočni konkordančnik Knut Hofland, isto mesto
<i>nameni uporabe</i>	analiza diskurza za potrebe priprave strojnega simultanelega prevajanja govora po priporočilih projekta LC-STAR	spoznavanje in razvijanje metod ustvarjanja, obdelave in uporabe govornega korpusa
<i>način zajema virov</i>	snemanje pogovorov na realni telefonski liniji (4 turistične agencije, 2 hotela, mestna turistična pisarna), klicatelji pa kličejo namenoma z vnaprej izbranimi temami, kanala snemana ločeno	z minidiskom snemanje realnih pogovorov, oportunistično zbranih z željo po čim večji raznolikosti govorcev, funkcij sporazumevanja ter drugih dejavnikov oz. strukture besedil
<i>transkripcija oz. obdelava</i>	prepisovanje po priporočilih EAGLES in prepisovanja govornih baz Broadcast News, prijemi programa Transcriber: ortografska transkripcija, deli besedila segmentirani, metaoznakev besedja ni; za govorce označuje spol, narečno skupino; zunanje okoliščine niso zabeležene, razen kar se tiče vzpostavljanja telefonske zveze	zglede EAGLES, BNC, BoE, CGN: ortografska transkripcija brez ločil in začetnic povedi, primerjava s pisnimi oblikami v jezikovnih priročnikih in oznake nestandardnih besed

¹⁰⁰ Na spletu (<http://nl2.ijs.si/~vicic/cwb/cwb.html>, nazadnje obnovljeno 3. 3. 2006) je prosto dostopen tudi „Pilotski korpus govornje jezika v vsakdanjih govornih položajih na slovenski Obali“, velik 1392 pojavnic. Nastal je leta 2005 v srednji šoli, avtorica je Jana Volk; sodeč po predstavitvi na spletu – http://www.pef.upr.si/MARA/STUDENTSKI_PROJEKTI/KorpusGovornegaJezika/Korpus_pilot.htm.

<i>ime</i>	korpus telefonskih pogovorov v turizmu Turdis-1	pilotski ali Učni korpus govorne slovenščine (UKGS)
<i>prilagoditev posebnostim govora</i>	strukturiranje pogovora, označitev tipičnih govornih diskurzivnih enot v sami transkripciji: diskurzni označevalci, ponavljanja, popravljanja, oporni signali, krajšave; fonetični zapis za nekatere besede, oznake neverbalnih govornih znakov (značilno za posnetek telefonskega pogovora)	razširjen ortografska transkripcija (znak za polglasnik v mašilih in medmetih), označitev nekaterih tipičnih govornih diskurzivnih enot in nekaterih podatkov o okoliščinah sporazumevanja (gl. II.7.4 in II.8.7); metaoznaka besedja ni
<i>programi obdelave</i>	Transcriber za prepisovanje	Transcriber in Praat za prepisovanje, Corpus Work Bench (CWB) za organizacijo metaoznake v korpus in za potrebe brskalnika po korpusu – zvočni konkordančnik
<i>dostopnost na spletu</i>	ni dostopno oz. ni podatkov	dostopno preko spleta z geslom
<i>literatura</i>	Verdonik 2006: 66–74, 163–167 in drugod	http://torvald.aksis.uib.no/talem/jana/s9.html Zemljarič Miklavčič/Stabej 2005, Zemljarič Miklavčič 2006 ¹⁰¹

Po zgledu Turdisa sem dejansko povzel malo, samo jezikovno-geografske oznake govorcev (poleg 7 narečnih skupin še 2 kategoriji – mariborski in ljubljanski govor), v primerjavi z nekaterimi njegovimi značilnostmi pa sem lahko zavzel distanco do nekaterih rešitev obdelave:

- odpoved pomožnemu fonetičnemu zapisu in nekaterim prozodičnim podatkom (vlečenje fonemov, vdih ipd.), ker ne sledijo potrebam diplome in ker nepregledno delujejo med besedilom – v partiturnem formatu transkripcije pa jih je mogoče pregledneje dodajati na poseben trak;
- telefonski pogovori kot gradivo se od govora v drugih medijih razlikujejo po kvaliteti posnetka in seveda po svojem mediju, kolikor ta določa druge okoliščine – prisotnost soudeležencev, časovna strnjjenost, nezmožnost kombiniranja z vidnim prenosnikom itd.;
- bogata razčlenitev diskurzivnih enot, še posebej značilnih za (spontani) govor (ponavljanja, ostaja zaželeno nadaljnja označitev tudi za gradivo v korpusu StAnId, tudi v povezavi z zaželenim besednovrstnim določevanjem.

UKGS, za katerega sem imel več literature in ob par stikih z ustvarjalko tudi več vpogleda v nastajanje, čeprav nimam dostopa do samega korpusa, mi je bil temeljni zgled in izhodišče zamisli, ciljev in dilem glede oblikovanja korpusa. Transkripcijske rešitve in metaoznake sem razširil in spremenil po svoje, malokatero zavrgel (npr. razlikovanje govornih situacij po moji oceni prekompleksnem besedilnovrstnem kriteriju formalno/neformalno), pomembna opora pa mi je bila ideja, da govorne besedne oblike brez znane ali uveljavljene pisne oblike med transkribiranjem primerjamo z jezikovnimi priročniki (slovarji, korpusi) oz. vsaj zamišljeno standardno pisno obliko.

5.2. Z računalnikom od transkripcij govora do njihove analize

Prenos jezikoslovnega dela na računalnik lahko ponudi nove možnosti obdelave in predstavitve gradiva, kar na splošno velja za korpus in zapis podatkov v njih. V tej diplomiji je mnogo različnih manipulacij z gradivom: urejanje zvočnih posnetkov, transkribiranje jezika, urejanje podatkov o okoliščinah sporazumevanja, pripisovanje metaoznake tako besedam kot besedilom ter analiza obojega (kvantitativna in kvalitativna), končno pa še predstavitev rezultatov. V naslednjih podpoglavjih so obrazložene možnosti in moj izbor izmed njih; podrobnejši potek dela s programi pa je opisan skozi tretji del diplome (III).

¹⁰¹ Disertacija raziskovalke o načelih gradnje govornega korpusa je bila obranjena konec avgusta 2007, ob oddaji te diplome pa je v knjižnicah ali na spletu še ni bilo zaslediti.

To delo z računalnikom pa ni nujno hitrejše in enostavnejše; že na začetku se izkaže za kar zamudno in zahtevno. Lažje je namreč prilagoditi se obstoječim standardom datotek ter obstoječi programski opremi in njenim sposobnostim, kot pa utemeljiti in vzpostaviti svoje lastne. Ker sem želel vztrajati pri svojih dolgi in specifični poti raziskav, je tako šlo kar nekaj časa za iskanje in izbor več računalniških programov (ki so nudili možnosti za moje cilje), za spoznavanje njihove uporabe ter gradnjo lastnih metod dela.

S to zamudnostjo se srečujejo vsi raziskovalci govora, zato je pomembno tudi, da si prizadevam gradivo pregledno urediti, obdelavo pripeljati do konca in rezultate primerno predstaviti in ponuditi drugim tudi v elektronski obliki kot jezikovni vir. Kljub trudu pa ni zagotovitve, da bo gradivo uporabno v elektronski obliki še dolg čas, ker standardi še niso utečeni in rok uporabnosti programske opreme niha v odvisnosti od njegovih razvijalcev (običajno kvečjemu desetletje); sam pa gradiva po diplomi ne bom več posodabljal. Stanovitnejše skrbništvo imajo jezikovni viri, združeni v mednarodne oz. institucionalne baze (npr. ELRA/ELDA, za druge in slovenske gl. Verdonik/Rojc 2004), a gradivo te diplome ni institucionalno umeščeno.

5.2.1. Obdelava zvočnih datotek

Izbiranje standardov elektronskega zvočnega kodiranja (o opremi za snemanje gl. poglavje III.10.1) ni bilo podvrženo kakšnemu poglobljenemu pretresu, pač zaradi zahtevnosti tega področja in finančnih omejitev glede opreme, programov. Na splošno pa velja nekaj, kar nam zelo pride prav: da je kodiranje zvoka za potrebe razpoznavanja človeškega govora (govornih znakov in paraverbalnih odtenkov govora: intonacije, „barve“ glasu ...; npr. razpon 400–3500 Hz) mnogo enostavnejše, robustnejše kot kodiranje zvoka za druge potrebe (glasba, zvoki izven človeškega govorno-slušnega spektra). Za naše potrebe se torej lahko poslužimo standardov (digitalne) kompresije zvoka, ki na podlagi izkušenj psihoakustike reducirajo človeku nepotrebne zvočne informacije; uporabljeni so sicer pogosto na področju telefonije in akustičnih naprav za množično uporabo (predvajalniki glasbe, RTV, računalnik).

Na splošno za računalniško urejanje posnetkov potrebujemo torej program (ali programe), ki zvok presnamejo (s pomočjo strojne opreme: avdio ali digitalna prenosna snemalna naprava, zvočna kartica na računalniku in kabli) v računalniško datoteko. Prva stopnja je podrobno kodiranje, ki običajno zavzame dosti pomnilnika (angl. *loseless compression*; formati/standardi AIFF, RIFF, WAV/PCM, FLAC idr.), program pa mora biti sposoben pretvorbe v enostavnejše kodiranje, ki zavzame manj pomnilnika (angl. *lossy compression*; formati/standardi MP3 (MPEG-1, Audio Layer 3), ATRAC, AAC, Ogg Vorbis, WMA idr.), in nazaj. Ko namreč uporabljamo različne naslednje programe za transkribiranje, ti običajno niso sposobni odpirati vseh formatov zvočnih datotek, ampak le nekatere (npr. znani program Praat le format .wav(e). Podobno obdelava in hranjenje manj kompresiranih zvočnih datotek računalniku zasede več delovnega in običajnega pomnilnika; problematična je tudi morebitna izmenjava preko interneta; primer sledi nižje spodaj.¹⁰²

Izbral sem preprost in prosto dostopen program Audacity (Mazzoni idr. 2004: A Free Digital Audio Editor. V), ki je deloma preveden v slovenščino. Posnetek shrani v svojem formatu (.aup), uporabna pa sta izvoza v formata .wav in .mp3; kvaliteto kodiranja je mogoče spreminjati po različnih standardih. Primer za posnetek, dolg 1h 3' 11":

Primer 15: Velikosti zvočnih datotek

<i>format</i>	<i>nastavitve formata</i>	<i>velikost na disku</i>
aup	hitrost 22050Hz, format 16 bit in interpolacije na osnovni ravni, 1-kanalno (mono)	339.648.512 B
wav	Microsoft 16 bit PCM	167.190.528 B
mp3	128 kbit/s (hitrost menda privzeto 44100 kHz)	60.661.760 B

¹⁰² O standardih zvočnega kodiranja povzeto iz Wikipedia 2007, angleška verzija.

Posnetki za sedem dni, ki sem jih zbral (74,5 ur, prim. Tabela 36, str. 129), v istem formatu .mp3 tako znesejo 4,15 GB, v začasnem formatu .aup pa 5,6-krat več, kar je bilo pri preprostem osebнем računalniku za obdelavo v prvi fazi leta 2004 kar problem – vseh posnetkov ni bilo mogoče hraniti na trdem disku računalnika. Tako preproste nastavitve formatov .aup in .mp3 so bile potrebne torej tudi zaradi zmogljivosti računalnika.

Ob nastavitvah kvalitete kodiranja zvočnih datotek pa je potrebno paziti tudi na opremo predvajanja: bolj kot zvočniki so bile koristne slušalke, ki školjkasto pokrivajo ušesa.

5.2.2. Standardi računalniškega zapisa govornih besedil

Pri izbiranju računalniškega zapisa svojega gradiva pri transkribiranju in naprej sem se srečeval z dilemami glede uporabe različnih računalniških standardov, ki jih delim v 4 skupine:

- 1)) kodirni nabori pisnih znamenj: Unicode UTF-16 in -8, ISO-8859-2, 16-bitni nabor znamenj programa EVA;
- 2)) nabori oznak podatkov, povezani v (opisne) jezike za označevanje dokumentov: markujeziki SGML, XML, XSLT ...;
- 3)) strukture dokumentov v jeziku markup oz. tipi dokumentov: dobro oblikovani (well formed), DTD, RELAX NG, XML Schema oz. vrste datotek po njihovih programih;
- 4)) standardi korpusov oz. sploh ureditve datotečnega sistema: TEI, (X)CES, zgledi posameznih korpusov.

Vodila pri izbiranju konkretnih standardov so mi bila:

- uporabiti standarde, ki bodo uporabni za strojno izmenjavo med programsko opremo in z drugimi raziskovalci;
- prilagoditi možnosti zapisa jezikovnih podatkov, ki jih ponujajo standardi, svojim raziskovalnim potrebam;
- izbrati tako programsko opremo – vmesnike (in od nje odvisne standarde), ki ob zgornjih dveh vodilih ustreza tudi mojemu uporabniškemu znanju (gl. naslednje poglavje).

Pri izbiri **kodirnih naborov pisnih znamenj** (za transkripcijske datoteke) sem bil od začetka vezan na programe: Elan in EXMARaLDA Partitur Editor (dalje EXM.) privzeto delata z naborom UTF-16. Pri izmenjavi med njima ni bilo problemov, za potrebe pretvorbe v .txt-format je bil uporabljen nabor UTF-8, Eva pa ga je prekodirala v svoj interni šestnajstbitni nabor. Tu je bilo potrebnih nekaj popravkov glede ločil (prim. Tabela 12, str. 89, in Tabela 14, str. 105): Eva ima svojo (drugačno) kodo za tropičje, podčrtaj, vezaj ..., potrebno jih je popraviti ročno ali v prečrkovalni shemi. Prav tako je bilo potrebno pri vsakem programu posebej definirati vlogo, ki jo imajo ključna ločila pri razlomu besedila na segmente (besede – tokenizacija; možne tudi segmentacije na druge enote). Nabori izven standarda konzorcija Unicode (npr. ISO-8859-2) ali prekodiranje znakov v entitete bi bili potrebni kvečjemu pri pretvorbah v besedila enostavnejših spletnih strani.

S prirejanjem **jezika za označevanje jezikovnih podatkov (markup)** se nisem posebej ukvarjal, tudi ne s tipom oz. strukturo dokumenta transkripcijskih datotek, ki so v korpusu zapisane v njem. Oboje je bilo določeno hkrati z izbiro obeh programov za transkribiranje (EXM., Elan), ki imata vsak svoj tip dokumenta v jeziku XML (več o tem jeziku Erjavec 2003: 62–63 in nadaljnje povezave tam), tudi Eva pa ima svojo strukturo datotek, podobno formatu markup. Oba transkripcijska programa omogočata pretvorbo v nekatere druge transkripcijske formate, tako v XML kot v drugih jezikih oz. običajnih tekstovnih datotekah z določeno logično trukturo. Kot je razvidno iz spodnje tabele, ima v sposobnostih pretvorb program EXM. več podlage; o njem gl. naslednje poglavje.

Tabela 8: Tipi dokumentov transkripcijskih datotek po programih Elan in EXMARaLDA

uporaba datoteke	izmenjava z EXM.		drugi formati	izmenjava z Elan	
	iz EXM.	v EXM.		iz Elana	v Elan
transkripcijske datoteke različnih programov	→ → → →	← ← ← ← ← ←	Praat (.textGrid) TASX (.xml) EXMARaLDA (basic/segmented .xml) Elan (.eaf) syncWRITER (.xml) HIAT-DOS (.dat) Shoebox (.txt) Transcriber (.trs)	← ←	→ → →
standardni transkripcijski formati	→ → → →	←	HIAT, DIDA, GAT, IPA (segmented .xml) CHAT (.cha) TEI (.xml) AIF (.xml)	←	→
drugi besedilni formati	→ → → →	←	XSLT pretvorba po lastni priredbi Simple EXMARaLDA (.txt) GAT (.txt) interlinear text (.xml) seznam segmentov (zlogi, besede, izjave; po transkripcijah HIAT, DIDA, IPA; .xml/.html) SMIL (.smil) Quicktime (podnapisi; .txt + .rt) (.txt) tradicionalna transkripcija, tabulatorsko ločena, interlinear text	← ← ←	
formati predstavitev datotek	→ → →		.html .rtf .svg (grafični format) .jpeg (grafični format)	←	
literatura	Schmidt 2006			Hellwig/Uytvanck 2006	

Vrednost programov za transkribiranje pa ni le v logičnem strukturiranju transkripcij v jezik markup, marveč tudi v vidnem predstavljanju podatkov v transkripcijski datoteki. Pri transkribiranju govora poleg zapisa fonemske strukture besed lahko označujemo še množico drugih podatkov (fonetika, prozodija, morfemi, kretalni jezik idr. okoliščine, oblikoslovne, besednovrstne, diskurzivne oznake, prevodi ...), ki jih program razčleni na kar najenostavnejše podatke in jih z besednimi oznakami (angl. *tags*) označene zakodira v (za človeka nepregledno) markup-strukturo, uporabnik pa podatke vpisuje in razbira preko vizualnega vmesnika, ki mora biti dobro vizualno strukturiran. Pri obeh omenjenih programih je vizualna strukturiranost transkripcije partiturni format (gl. poglavje II.6.1, pa tudi primer besedila 6pet09Dm v programu EXM, Slika 17, str. 135, za XML-strukturo datoteke za Elan in EXM. pa glej Priloga 8 in Priloga 9.

Raziskovalec se mora sicer tudi vizualnemu formatu prilagajati (no, program Elan ponuja tudi preklapljanje med različnimi vizualnimi vformati, tako pri vidnem vmesniku na monitorju kot pri iztisu transkripcij), nekoliko bolj proste roke pa ima pri tem, kar vanj vpisuje. Pri samem zapisu čistega besedila smo v dilemi, kako označiti posebnosti, ki veljajo le za nekatera jezikovna sredstva, npr. besede – *lahko se spet bližamo logiki markupu*. Pisava sama ima namreč podobne lastnosti kot markup – primerjaj dvodelne oznake v markupu (tags) ter oznaki začetka in konca povedi v pravopisu, velika začetnica in končno ločilo ipd. Primer: želimo označiti nestandardno (*nst*) besedo, po zgledu različnih transkripcij (prim. Tabela 18, str. 109) bi to lahko naredili takole:

- besedilo, posnemanje dvodelnih oznak (UKGS): `<nst>beseda</nst>`;
- besedilo, posnemanje enodelnih oznak, nanašanje na predhodno besedo (program Transcriber): `beseda[+lex=nst]`;

- besedilo, samostojno znamenje (zvezdica), ki simbolizira enodelno oznako, nanašanje na sledečo besedo (StAnId): **beseda*;

- markup, v tokeniziranem besedilu z oznako kot atributom: `<w type="nst">beseda</w>`.¹⁰³

V transkripcijski konvenciji StAnId tako besede niso tokenizirane in nimajo oznak po zgledu markupa (kot npr. UKGS), je pa transkripcija toliko sofisticirana, da je možno iz njenih posebnih oznak (pisnih znamenj – ločil, prim. Tabela 13, str. 93) podatke (pol)avtomatsko najti (prim. Tabela 40 in sledeča razlaga, str. 149). Sinclair (1991: 28–29) sicer odsvetuje take oznake, češ da so z vidika določevanja besed (z bolj ali manj strojnim določevanjem, s t. i. določevalniki) nepotrebne. A se jim ne odpovemo, ker jih potrebujemo najprej sploh za ugotavljanje posebnosti govora (in se to zdi hitreje narediti ob transkribiranju) in ker jih naknadno ni problem odstraniti, če je to potrebno.

Nazaj k strukturi oz. tipu dokumenta transkripcijskih datotek. Le-te vsebujejo tudi nejezikovne podatke ali vsaj reference nanje v drugih datotekah (časovne reference na zvočno datoteko, nekateri podatki o okoliščinah sporazumevanja so isti ali se ponavljajo za več besedil v korpusu: govorcji, oznake besedil). Te reference oz. meddatotečne odnose bi spet lahko strukturirali po naslednji vrsti standardov, za korpuse, pa tako daleč nismo šli. Mednarodni **korpusni standardi** so npr. (X)CES (Corpus Encoding Specification); pa bolj široki standardi, ki pokrivajo markup od različnih tipov datotek (besedilnih, leksikonov, do zbirk podatkov) do celotnega korpusa – TEI (Text Encoding Initiative). Kot ožji standardi bi lahko služili zgledi korpusov ali korpusnih orodij, ki konstruirajo vezne XML-datoteke. Tak primer ponujajo tudi pomožna orodja programa EXM. (CoMa, gl. naslednje poglavje).

V korpusu StAnId sem torej jezikovne podatke o besedilih ročno prenesel v eno samo datoteko zbirke podatkov, ki je dobro služila tudi analizam in pripravljanju izvlečkov. Lahko bi to zbirko podatkov verjetno nadgradili v glavo korpusa kot vezno datoteko, a to presega znanje raziskovalca. Nadaljnja pojasnila o urejanju korpusa in o določanju podatkov, specifičnem za ta korpus,¹⁰⁴ ter analize besedja, pa niso niti del nazadnje omenjene zbirke podatkov, ampak so preprosto predstavljena v tej diplomi.

Tabela 9: Pregled tipov datotek v korpusu in podatkov v njih

<i>tip datotek:</i>	<i>s podatki:</i>
zvočne datoteke .mp3, po ena datoteka za besedilno enoto korpusa	digitalizirani zvočni posnetki
XML-transkripcijske datoteke programov EXM. in Elan, po ena datoteka za besedilno enoto korpusa	zapisi tekoče spreminjajočih se govora, paraverbalnih znakov in okoliščin ob časovnih referencah na zvočno datoteko
ena sama datoteka – zbirka podatkov o korpusu, program Access	tabele podatkov o govoricah, o sporazumevalnih okoliščinah besedil, o govoricah v posameznem besedilu, o razmerjih med nerazrezanimi in razrezanimi transkripcijami in posnetki, o statistiki jezikovnih enot v besedilih; izbrani podatki za potrebe analiz (t. i. poizvedbe in poročila) in izbrani podatki za lažje vnašanje (t. i. obrazci)
analitične datoteke .eva, .txt	besedila, združena v posamezni datoteki za raven vsega korpusa ali podkorpusa; sezname različnic in statistikami besedja, sezname posebnosti v korpusu
tekstoven opis diplome ter gradiva za predstavitev	obrazložitev dela, obdelave gradiva, analiz, rezultatov
nasploh računalniške datoteke in mapna/direktorijska struktura	velikost in računalniška vrsta datotek, datumi nastanka in zadnjega urejanja datoteke, imena datotek/besedil ...

¹⁰³ Element `<w>` (beseda, dvodelna oznaka `<w>` `</w>`) in njegov atribut *type* (tip besede) sta iz naborov oznak TEI, ki pa imajo na zalogi še druge možne rešitve označitve našega primera z jezikom markup.

¹⁰⁴ Ti podatki so po standardih TEI opisani v t. i. glavi korpusa, `<teiHeader>`, v njegovih štirih elementih: `<fileDesc>`, `<encodingDesc>`, `<profileDesc>`, `<revisionDesc>` (Desc – description ‘opis’). Erjavec 2003: 68.

Da povzamem: Ta korpus sestavljajo relativno nepovezani tipi datotek, navedene v naslednji tabeli. Povezani so med seboj zgolj preko šifer besedil (temeljnih enot v korpusu) oz. imen datotek ter seveda v razlagah v tej diplomu. Standardov za oblikovanje korpusov nisem uporabljal, standarde za računalniški zapis preprostejših delov korpusa pa posredno v toliki meri, kolikor so jih uporabljali programi, ki sem jih izbral.

Mapna (ali direktorijska) in datotečna struktura korpusa je v končni meri odvisna od celotnega načrta njegove obdelave (ko le-ta še traja) in uporabe (po zaključku obdelave). Struktura korpusa StAnId se je tako med obdelavo neprestano spreminjala, sledila je pa samo svoji lastni organizaciji in nikakršnim tujim zgledom ali standardom (opisanim npr. v Romih 1998 in Erjavec 1997). Priprava korpusa za druge uporabnike je bila kvečjemu obstranski cilj, javna distribucija (internet) pa ni možna zaradi nerešenih avtorskih vprašanj.

Korpus je razmeroma ohlapen in „negovoreč“ (z izjemo zapisa tega diplomskega dela); dostop do podatkov je mogoče začeti pri datoteki zbirke podatkov in nadaljevati v določenih programih z analizami po želji uporabnika. Gradivo v korpusu je seveda mogoče dopolnjevati z nadaljnjimi analizami (npr. partiture: fonološki zapis, določevanje oblikoskladenjskih oznak, analize drugih jezikovnih ravnin in vidikov).

5.2.3. Programi za transkribiranje govora in obdelavo transkripcij

Po računalniških programih za obdelavo zvočnih datotek omenim še takšne za transkribiranje in označevanje govora; tesno so povezani s standardi iz prejšnjega poglavja, za to jih je nekaj že bilo omenjenih. Njihove osnovne sposobnosti so vizualni vmesnik transkribiranja med vtipkovanjem in XML-strukturo datoteke, predvajanje zvočne (pogosto tudi video) datoteke in navezovanje na njen časovni potek, vtipkovanje in oblikovanje transkripcij in (jezikoslovnih) oznak govora, manipulacija teh vpisov (kopiranje, iskanje in nadomeščanje vpisov, kombiniranje datotek itd.). Nekateri programi so bolj odprti za uporabniško definiranje (uporabniško določeni vnosi podatkov, bližnjice in imena ukazov, formati besedilnih, zvočnih in video datotek), drugi manj; odvisno od tradicije razvoja programa – običajno na kakšni univerzi ali javni raziskovalni ustanovi; komercialnih programov zaradi neproste dostopnosti nisem jemal v ozir.

V začetni fazi sem po poizvedbi in predstavitev zvedel za programa nizozemski Praat in francoski Transcriber, ki sta tipična za prepisovanje govora (ne pa tudi videa).¹⁰⁵ Praat bolj služi potrebam fonetičnih raziskav (histogramski prikazi stavčne fonetike, ustvarjanje umetnih signalov itd.), Transcriber pa grobi obdelavi golega besedila (iz radiodifuznih medijev, telefona; usmerjenost v tematsko, pomensko segmentacijo; njegova šibka stran je preprost stolpčni format transkripcije, kar omogoča samo interlinearni zapis stavčne fonetike in največ dva hkratna govorca).

Iskanje ustrežnejših programov pa sem se zaključil¹⁰⁶ ob nemškem EXMARaLDA Partitur Editor, verzija 1.3.3p1 (dalje EXM.) in nizozemskem Elan, verzija 2.6.3.¹⁰⁷ Pri prvem se me posebno pritegnili preprost način vpisovanja transkripcij (izvedba partiturnega formata iz tabele), številne nadaljnje možnosti urejanja (izvoz v Elan, v TEI-standardni zapis XML-formata, v RTF- in HTML-format) in možnost segmentacije transkripcij na podlagi transkripcijske konvencije, sorodne moji (HIAT – Halbinterpretative Arbeitstranskriptionen; štetje besed, povedi, govornik sklopov). Ta program tudi kot eden redkih v XML-datoteko transkripcije lahko vključi podatke o okoliščinah besedil in govorcih, te podatke pa je možno urejati in povezovati v

¹⁰⁵ Uporabljata in opisujeta ju Zemljarič Miklavčič 2005 in Verdonik 2006: 68. Izvorni podatki pa so v Boersma/Weenink 2007 za Praat in Barras 2006 za Transcriber.

¹⁰⁶ Povezave na spletne strani, kjer so navedeni še drugi programi za transkribiranje, so v poglavju VI.21.2.

¹⁰⁷ Za EXM., ki je del večjega projekta EXMARaLDA, gl. Schmidt/Wörner 2007. Avtor programa EXM. Partitur Editor je Tomas Schmidt, uporabljena verzija je z 12. 12. 2006. Za program Elan gl. Elan 2007 v seznamu literature. Avgusta 2007 sta na razpolago že novi verziji programov, in sicer EXM. 1.3.4 in Elan 3.1.0.

korpus s pomočjo dodatnih programov (CoMa in EXAKT, glej nižje). Program je še v fazi razvoja, ima torej zagotovljeno bodočnost, zasnovan pa je na sistematičnem preučevanju potreb transkribiranja.¹⁰⁸ Program Elan ima nekaj drugih sposobnosti, ki jih zaradi možnosti izmenjav transkripcij z EXM. lahko s pridom izkoristimo: različni vizualni prikazi transkripcij, izvoz transkripcij v datotečni format .txt, logična hierarhizacija trakov v partituri transkripcije in napredno iskanje po eni ali več transkripcijah hkrati (tudi npr. možnost iskanja po časovni razsežnosti govora) ter še kaj. Oba programa, EXM. in Elan, lahko podatke elektronske podatke izmenjujeta s programom Praat in z njegovo pomočjo prikažeta tudi sonogram govora, če je posnetek v formatu .wav. Druge lastnosti obeh programov in nekaterih drugih so primerjane v tujih člankih (npr. Rohlfsing idr. 2006, Sim/Kazi/Hong 2005).

Načeloma lahko transkribiramo tudi v preprost tekst (datotečni format .txt), ob tem pa zvočno datoteko predvajamo v preprostem programu za predvajanje glasbe in videa (cela paleta prosto dostopnih, npr. Real Player, Musicmatch Jukebox, Windows Media Player, BSPlayer). Njihova prednost so napredne možnosti krmiljenja posnetkov (previjanje naprej/nazaj, različne hitrosti predvajanja – v tem jih sicer povsem dosega Elan) in predvsem enostavno urejanje podatkov o posnetkih (vsebina, drugi zaznamki).

Za prepisovanje prav pride tudi naprednejša tipkovnica, in sicer dve njeni funkciji: vnašanje diakritičnih znamenj (privzete kombinacije/zaporedja tipk) in set tipk za predvajanje zvočnih/video posnetkov, ki ga zgoraj omenjeni predvajalniki v glavnem upoštevajo, programi za transkripcijo pa ne. Sam sem uporabljal tako tipkovnico – Logitech PS/2.

Z izborom programov za transkribiranje odpade uporabnikovo poglobljanje v urejanje XML-strukture datotek, saj XML-datoteke, ki niso vizualno prikazljive v teh programih, prav malo koristijo in so težko uporabne. Zaradi izmenjav transkripcijskih datotek med različnimi programi pa kdaj pride prav, da popravimo kakšno malenkost, logične strukture XML-datotek ne spreminjamo. Ker je vsaka transkripcija v korpusu besedil v svoji datoteki, je pa serijsko popravljanje hitrejše, če so ob tem vse transkripcije združene v eni datoteki (in jih kasneje razčlenimo nazaj). Prav pride kak XML-urejevalnik (npr. Oxygen, XMLSpy) ali zgolj - prikazovalnik (npr. Mozilla Firefox) ali pa napreden tekstovni urejevalnik.

Slednjega sicer potrebujemo za analizo besedja, predvsem kvantitativno, saj je nekaj označevanja možno opraviti že v programih za transkribiranje. Sam sem ostal pri slovenskem programu Eva, verzija 7.6 (Jakopin 2000), ki sem ga najbolj več, druge možnosti bi bile sicer UltraEdit-32, pa še celo MS Word itd. Partiture posameznih transkripcij so bile iz Elana v formatu .txt uvožene v Evo (pri tem so zavrženi odvečni podatki o časovnih referencah na posnetek in o paraverbalnem sporazumevanju) in tu najprej združevane v korpus in podkorpuse, nato pa razčlenjene v različne namene.

Eva je napol tudi že naslednja vrsta programa, ki ga potrebujemo za analizo korpusa, to je urejevalnik zbirke podatkov. Podatki, ki jih z njim urejam, so podatki o besedilih/transkripcijah (opis dejavnikov sporazumevanja, njihova jezikoslovna kategorizacija, njihove statistike in drugo urejanje). Ti podatki so običajno v korpusih del XML-strukture transkripcijske/besedilne datoteke (oz. povezani z besedilom, če je korpus v eni sami datoteki) in jih je mogoče pregledovati z iskalnikom po korpusu, del teh podatkov (statistike, urejanje korpusa) pa je zgolj manj pregledno opisan v spremljajoči strokovni literaturi. Taka urejevalnik korpusa in iskalnik (konkordančnik) sicer ponuja projekt EXMARaLDA (Schmidt/Wörner 2007) – CoMa (EXMARaLDA Corpus Manager) in EXAKT (EXMARaLDA Analyse- und Konkordanztool), a ju v končni fazi nisem uporabil. Razloga sta dva: prvi, ker se pri izmenjavi partitur med programi (EXM. : Elan : Eva) ti „izvenpartiturni“ podatki izgubljajo, in drugi, ker izbrani napredni urejevalnik zbirke podatkov (MS Access) omogoča podrobnejše rokovanje s podatki ter hkrati njihovo analizo in predstavitev. Uporaba tega urejevalnika je sicer zahtevna (in gre še za en program več, s tem še za eno vrsto datotek več), a je nadzor nad vpisom podatkov enostavnejši (maske za vnos podatkov, omejeni nabori vrednosti podatkov, vpis s pomočjo obrazcev, spustnih

¹⁰⁸ Možnost kontaktiranja po poštni listi, zasnova XML-strukture v Schmidt 2005.

seznamov in privzetih vrednosti itd.), možne pa so tudi najrazličnejše poizvedbe kombinacij in izborov podatkov (za potrebe ustvarjanja podkorpusov, statistik). Do orodij, kot sta Corpus Work Bench in Neva, nisem prišel.

Pri možnostih nadaljnje strojne obdelave bi seveda prišli do novih, dodatnih programov za označevanje transkribiranega besedila: določevalniki (angl. taggers), lematizatorji itd. Vendar bi njihova pomoč pripomogla le k polavtomatski obdelavi – učno gradivo, na katerem sloni obdelava teh programov, je (vsaj za slovenščino) še vedno predvsem pisno, algoritmi statističnega določevanja niso dodelani; označevanje bi v prvih fazah moralo potekati ročno. Pri primerjavah s pisnim gradivom pa bi moralo pisno in govorno gradivo biti usklajeno v transkripcijski konvenciji (tudi posebej dodelani za leme, pa s posebnim naborom oznak za besedne vrste in oblikoskladenjske kategorije, lahko celo posebne oznake za govorne elemente – ponavljanja, popravki, diskurzni označevalci ...). Nadaljnji problem pa je tudi združevanje programov za označevanje s programi za transkribiranje: ali zmorejo uporabljati iste datoteke (običajno v markup-strukturi) oz. kakšna je vmesna pretvorba za potrebe nadaljnjega strojnega označevanja.

Končne tehnične lastnosti in struktura korpusa so predstavljene v poglavju III.13.3.

II. TRANSKRIBCIJSKA KONVENCIJA STANID

Sledeče poglavje želi opisati, kako zapišemo vse za pričujočo raziskavo potrebne podatke. Najprej je to jezikovni vir z dejavniki sporazumevanja, potem njegove jezikoslovne in analitične metaoznake. Podatke načeloma razločujem po čim bolj preprostih kriterijih, posamezna oznaka se nanaša na čim bolj specifično lastnost sporazumevanja, tako da je predpostavka o gradivu čim manj (kot priporoča po svojih izkušnjah tudi Sinclair 1995: 106). Z drugimi besedami: ko opisujem sporazumevanje, želim uporabljati čim manj prototipičnih, po preprostih ali „zdravorazumskih“ predstavah ukrojenih opisov (izrazov, oznak, običajnih pomenov), ki so v svojem bistvu posploševalni in malo povedo o posebnostih govornega prenosnika. Podobno ugotavlja Skubic (2005: 82):

Za posamezno zvrst so potrebne precej nepredvidljive in spremenljive kombinacije funkcij, tako da jih je mogoče dejansko utemeljiti kot zvrst šele „zdravorazumsko“ /na tem mestu pojasnjuje, da je citat iz razprave Janeza Dularja iz 1979/. To pa nam seveda ne odstira nobenih novih vednosti in nam ne nudi metodološkega vpogleda v skrite mehanizme v ozadju jezikovnega vedenja. Kakšen status imajo *funkcije* stilov v jezikovni teoriji strukturalističnega funkcionalizma, ni bilo pravzaprav nikoli dokončno pojasnjeno. Semiotična teorija in aplicirana doktrina ostajata ločeni; zdi se, da se doktrina razvija *ad hoc* ob stiku z dejanskim jezikovnim gradivom, brez teorije.

Omenjeni strukturalistični manko poskuša preseči tudi ta diploma, res pa je, da dela *ad hoc*, ob stiku z gradivom ... Gradivo *obeh polov* (surovo-golo-čisto in metapodatki-oznake-analitično-jezikoslovno) se pojavlja na različnih ravneh oz. v različnih skupinah podatkov – klasifikacije izhajajo iz značilnosti gradiva in načina obdelave.

Tabela 10: Podatki v korpusu, stopnje njihove obdelave in pravila zapisa

10 osnovnih ravnin podatkov, od najkompleksnejše proti najpreprostejši:	6 stopenj obdelave gradiva (poimenovanje za potrebe razlikovanja datotek v korpusu)	pravila zapisa, ki jih je potrebno posebej določiti glede na namen raziskave	
dokumentacija obdelave celotne diplome	.6 <i>Informacije</i>	dokumentacija obdelave, gl. 0	
struktura korpusa (načrtovana/logična, računalniška/datotečna, vizualna ...),			
načrtovane (stilistične, kvantitativne) analize in statistike,	.5 <i>Analize</i>	zapis analiz, gl. I.4.3 in IV ¹⁰⁹	
besedila oz. sporazumevalne situacije, tudi njihove metaoznake	.4 <i>Besedila</i>	okoliščine sporazumevanja, gl. II.8	
nadbessedna segmentacija (povedi, izjave),	.3 <i>Transkribcije</i>		zapis govora (transkribiranje), gl. II.7
paraverbalni dodatni opis jezika,			
besedna/verbalna oz. morfo-fonografska razčlemba jezika,			
opis besedilnih okoliščin oz. sporazumevalnih značilnosti/dejavnikov,	.2 <i>Okoliščine</i>		
posnetek pogovora kot najbolj grob (sekundarni) podatek o sporazumevanju.	.1 <i>Snemanje</i>	(standardi snemanja, gl. III.10.1 in I.5.2.1)	

Štiri skupine pravil za zapis podatkov, navedene v skrajnem desnem stolpcu tabele, imajo skupno ime **Transkribcijska konvencija StAnId**; termin „transkribcijska konvencija“ je tu s pravil transkribiranja govora razširjen na širši zapis korpusnih podatkov za potrebe raziskave.

¹⁰⁹ Analize v tej diplomi so toliko specifične, da jih ni smiselno posebej obrazlagati njihovega zapisa znotraj transkribcijske konvencije, saj kot gradivo drugje ne bodo uporabljane. Sezname različnic (abecedni, frekvenčni, po posebnih znamenjih), besed, n-kratnic besed, povedi, pregledi besedil in govorcev, razni izračuni, grafi itd. so navedeni in obrazloženi v omenjenih dveh poglavjih in prilogah diplome. Za boljšo predstavo prim. tudi poglavje III.13.

V naslednjem podpoglavju je najprej razložen splošni nadzor nad podatki in mesta zapisa teh podatkov v računalniški strukturi, v nadaljnjih 4 podpoglavjih pa posamezna od skupin.

6. GRADIVO KOT ZBIRKA PODATKOV, NJEGOVA UREDITEV IN OBDELAVA

Praktična načela za takšen zapis, kot je v nadaljevanju razložen, izhajajo predvsem iz prakse ureditve korpusov in strojnega procesiranja jezikovnega gradiva (gl. poglavje I.5), v kateri jezikovno gradivo ni nič drugega kot množica podatkov, ki jo skušamo urediti v zbirko. Strojen način urejanja gradiva in kvantitativna načela analiz, po katerih se malokateri podatek izogne klasifikaciji, dosti pripomorejo h konsistentnosti podatkov (npr. omejeni nabori vrednosti za podatke o sporazumevalnih dejavnostih), seveda ob pogoju, da so posamezne kategorije in vrednosti podatkov dobro utemeljene.

Podatki so enostavno strukturirani tako:

<i>primer za ...</i>	<i>del gradiva</i>	<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa (program: tabela)</i>
omejen nabor vrednosti	soudeleženci kot sporazumevalni dejavnik	24) število soudeležencev	Število soudeležencev (ki jih besedilo predvidoma doseže) – ocena po eni od 6 kategorij 1 2 3–10 11–99 100– <u>neznano</u> .	Access: 08 BESEDILA
neomejen nabor vrednosti, a predpisano obliko vrednosti	govorci v besedilu	XOB) vloga	Socialna vloga posameznega govorca v konkretnem besedilu – enobeseden opis, npr. <u>prodajalka</u> .	Access: 05 govorce v besedilih

V tem poglavju smo vstopili v sistematski pregled tega, kar se dogaja s podatki o jeziku, ki jih izbiramo, transkribiramo, dopolnjujemo, urejamo ... V pričujoči diplomski skušam na različnih koncih te podatke *pregledno* prikazati še z drugih vidikov:

- *stopnje izvornosti virov* (gl. Slika 12, str. 125): kaj so primarni podatki (to ni transkripcija, niti ne posnetek), in katere njihove nadaljnje priredbe;
- *opis postopka priprave gradiva* (skozi vse poglavje III), kakor konkretno poteka pri tej diplomski;
- *zamudnost variant obdelave gradiva* (Tabela 35, str. 127), ki je važen dejavnik praktične uresničljivosti raziskave;
- *utemeljenost uporabe podatkov*, ki usmerja k stilometrični analizi (ves uvodni del naloge, I, in še posebej poglavje o slogu, I.4.3).

6.1. Partitura in drugi formati transkripcije

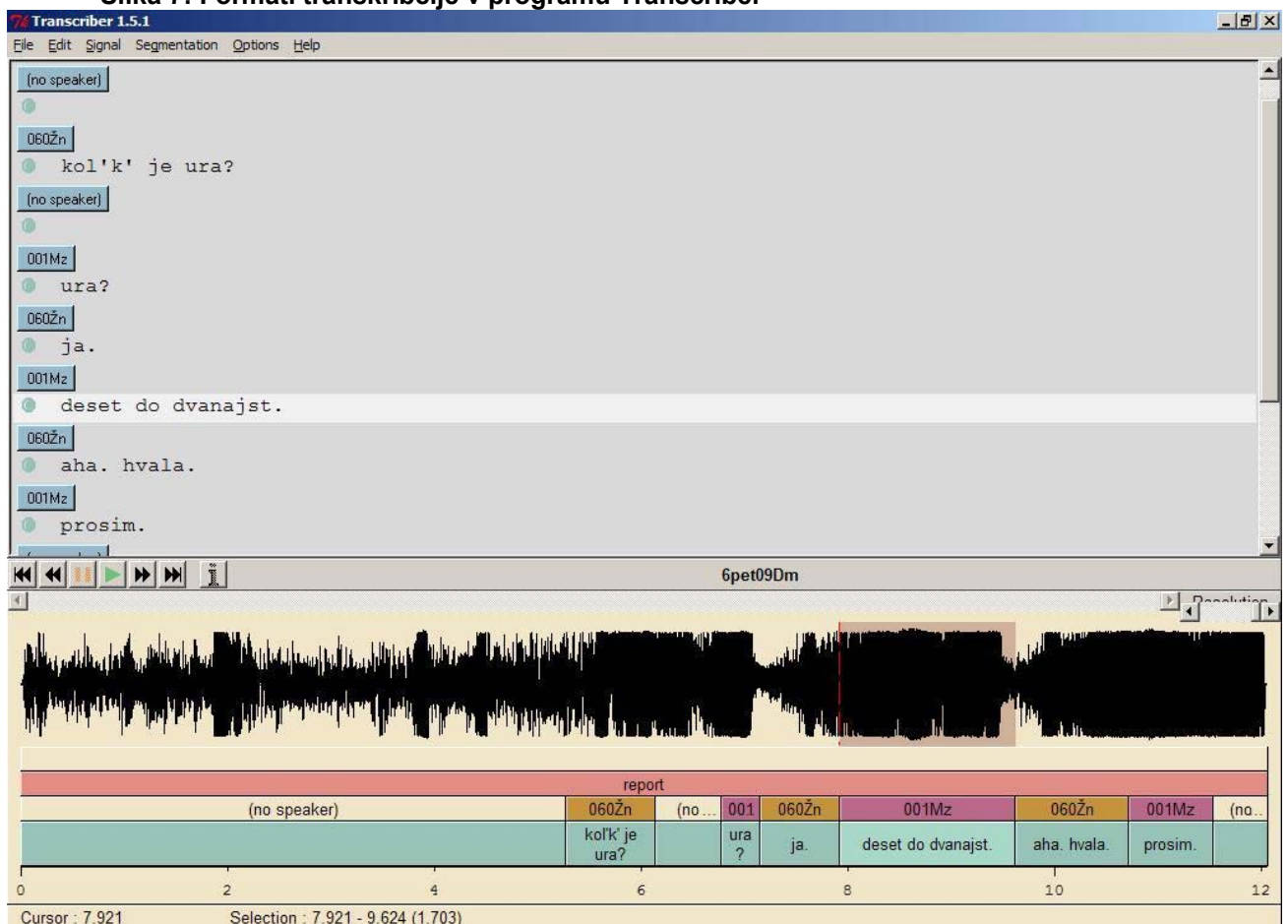
Osnovna skupina podatkov jezikovnih podatkov je transkripcija, zapis govora. Svoje mesto v računalniški obliki ima v telesu transkripcijske datoteke, kjer je strukturirana v jeziku markup, program za transkribiranje pa to XML-strukturo prevede v vizualno sliko. Računalniška transkripcijska datoteka pa ima poleg podatkov v *telesu* še podatke v *glavi*, ki vizualno niso prikazani hkrati z zapisom govora, ampak posebej.

Podatki transkripcije so logično pomembno določeni s pisnim oz. vizualnim prenosnikom (kanalom v smislu sporazumevalnega modela), v katerem so izraženi – govorimo o (vizualnih) **formatih transkripcije**. Ne glede na to, kakšna je vizualna uresničitev podatkov v formatu, od podatkov v glavi transkripcije jih loči to, da ne veljajo splošno (za vse besedilo), temveč se od začetka do konca transkripcije (besedila) *tekoče* spreminjajo.

V osnovi poznamo dva formata transkripcije, (navpičnega) besedilnega in (vodoravnega) partiturnega; pri prvem si izjave istega govorca sledijo od vrha proti dnu, pri drugem od leve proti desni¹¹⁰. Oba formata imata tri dele, ki jih imata drugače razmeščene: 1) oznake (sigle) govorcev, razporejene ob robu, 2) oznake časovnega in kvantitativnega poteka (npr. zaporedno število izjave) ter 3) vsebino sporazumevanja (jezikovni in parajezikovni znaki, opis okoliščin sporazumevanja).

Navpični besedilni format (gl. Slika 7 spodaj, zgornji del) je primernejši za zapis pogovorov z daljšimi govornimi pasażami posameznih govorcev ter nasploh za vpisovanje govora ob poslušanju. Enostavno ga je prikazati v tisku (prilega se običajnim razmerjem velikosti papirja v zlatem rezu, postavljenega navpično), težje pa je na njem pregledno dodajati metaoznake govora ter prikazati prekrivni govor (in nasploh vidik interaktivnosti soudeležencev, saj navpična izmenjava vrstic daje vtis o enakovrednih vlogah govorcev). Oznake govorcev, ki se ponovijo za vsako govorno vlogo govorca, in oznake časovnega poteka (na omenjeni sliki jih ni) se vrstijo na levem robu v navpični smeri.

Slika 7: Formati transkripcije v programu Transcriber



Vodoravni partiturni format (gl. spodnji del naslednje slike, Elan¹¹¹) omogoča predvsem preglednejše podajanje parajezikovnega dogajanja in okoliščin ter metaoznak. Ob „čistem“ besedilu, ki je še vedno središče tekočega sporazumevanja, dobi vsak opazovan spremljevalni

¹¹⁰ Namreč za evropske črkopise, ki imajo to smer pisanja/branja pisave. Običajna pa je ta smer tudi za pisanje not, po katerem je format dobil ime.

¹¹¹ Spodnji del, Slika 7 (Transcriber), ni pravi partiturni format, ker se oznake govorcev pojavljajo vodoravno. V osnovi pri tem primeru gre za navpični besedilni format, ki je le prekucnjen iz navpične v vodoravno dimenzijo, podatki pa niso prerezporejeni.

dejavnik svojo vrstico oz. trak zapisa. Kar trake (hkrati pa še zvočno datoteko) povezuje, je vzporedna umeščenost na vodoravno časovno os. Kaj zapišemo na posamezen trak in kako trake še drugače logično povežemo med seboj, pa je odvisno od našega izbora podatkov. Vsaka partitura pa ima še zapis imen trakov navpično na levi strani (začetku partiture; za vsakega govorca oz. trak le enkrat) in večina partitur še oznake časovnega poteka na vodoravni razsežnosti (lahko bi rekli: na abscisi). S slednjimi dobi vsak vpis v partituro avtomatično tudi časovno razsežnost. Zapisi časovnih referenc so nujni (enako njihovo upoštevanje pri branju partiture), ker linearna razsežnost zapisov besedila ni premosorazmerna z linearno časovno razsežnostjo (prim. izpuste predolgega besedila na spodnjem delu zgornje slike)!

Slika 8: Razdelana in hierarhizirana partitura med označevanjem besednih vrst, program Elan

Nr	Annotation	001Mz [BV]	001Mz [LEMA]
1	ura?	M	ura
2	deset	M	deset
3	do	G	do
4	dvanajst.	GP	dvanajst
5	prosim.	S	prosim

Če želimo partiturni format natisniti na papirju, ga moramo seveda razlomiti v bloke in le-te razporediti navpično (ponoviti oznake trakov ...; kot pri glasbenih notah), na računalniku pa posamezen trak besedila puščamo vizualno v eni vrstici. Programi za transkribiranje pa zmorejo iz XML-strukture transkripcijskih datotek prikazati še posamezne izseke, npr. zapise na posamezen trak ali trake v medsebojnih logičnih razmerjih (gl. zgornji del prejšnje slike, Elan) ali zapise predvajati hkrati z zvokom kot časovno omejene podnapise itd. Osnovni vizualni format zapisa pa moramo ne glede na možnosti preklopa prikazov v programih za transkribiranje izbrati glede na gradivo in namen obdelave (poleg zgoraj omenjenih so še druge izvedbe besedilnega in partiturnega formata, npr. stolpčni navpični; gl. Brinker/Sager 2001: 40 sl.). Za naše potrebe (v gradivu bolj prevladuje prosti govor, dodajati želimo besednovrstne oznake in leme) je bolj ustrezen partiturni format, priročnejši za vpisovanje vanj je bil program EXMARALDA Partitur Editor (gl. slike v poglavju III.11), za prenos v tekstualno obliko in morebitno dodajanje metaoznak pa program Elan (naslednja slika).

V partiturnem formatu **korpusa StAnId** je (kot tradicionalno) v središču pozornosti ubesedeno sporazumevanje, ortografsko zapisano na **trak [v]** (*v*-erbalno), prisojen posameznemu govorcev. Nebesedna sporazumevalna aktivnost posameznega govorcev je združena na svojem **traku, [pv]** (*p*-ara-*v*-erbalno), ki je drugi osnovni trak, pripisan govorcev. Vsaka beseda prvega traku pa lahko dobi še vzporedni metajezikovni enoti (*b*-esedno-*v*-rstno določitev in *lemo*), in sicer na tretjem **[BV]** in četrtem **[LEMA]** traku. Teh četveric je nato v partituri toliko, kolikor je govorcev. Samostojen trak, nepripisan kateremukoli od govorcev, imenovan **[Ozadje]**, pa opisuje tekoče spremembe splošnih okoliščin, ki jih ni mogoče zajeti v glavi; v partituri vsakega besedila je natanko en tak trak. Besedne trake je pri analizi mogoče enostavno ločiti od ostalih; tudi podatki o velikosti korpusa (v besedah) se nanašajo samo na besede iz teh trakov.

Zapis vsebine trakov je razložen v naslednjih poglavjih:

trak	poglavje
[v]	II.7 Pravila transkribiranja tekočega govora
[pv]	II.7.3 Zapis paraverbalnega
[Ozadje]	II.8.1 Tekoče spreminjanje okoliščin
[BV] in [LEMA]	V tej diplomii sta trakova samo predvidena kot možnost nadaljnje obdelave, dejanska obdelava gradiva pa ju ne vključuje.

Za začetku transkribiranja, na stopnji *3Transkribcije*, imamo „materinske“ transkribcije (partiture), ki jih na stopnji *4Besedila* ob upoštevanju okoliščin sporazumevanja razrežemo na „hčerinska“ besedila (še vedno partiture). Pri tem režemo partituro vedno na enotne časovne intervale, vse trake odrežemo ob isti časovni referenci. Ne režemo partiture recimo tako, da bi 2 sočasna pogovora, ki imata ločene govorce, dali vsakega v svojo partituro (da bi njune trake razrezali vodoravno in tako ustvarili novo partituro). (Primer je besedilo 6pet40Vg, kjer se hkrati s pogovorom dveh govorcev v osebnem stiku odvija še več besedil na radi, na katere se omenjena sogovorca sploh ne odzivata: reklame, glasba, kontaktna pogovorna oddaja.)

6.2. Druga mesta zapisa podatkov

V tem podpoglavju želimo nadaljnje podatke, ki so v korpusu poleg zapisa govora, strukturirati glede na to, *kako splošni so oz. kakšni so v razmerju med seboj, na katere druge podatke so vezani*. Ta **logična klasifikacija** se kasneje odrazi in spet spremeni v **mestu računalniškega zapisa podatkov**.

Posnetek oz. zvočna datoteka je temeljna enota korpusa v smislu primarnosti podatkov o jezikovnem gradivu; toda digitalni zvočni format je za naše namene analiz neuporaben. V tem jo nadomesti **transkribcijska datoteka** oz. struktura zapisa partiture v jeziku XML, kar smo opisali v prejšnjih poglavjih. Posnetek in transkribcijska datoteka sta si enotna ne samo v zapisu tekočega sporazumevanja, ampak tudi kot *enota gradiva* – besedilo, ki so mu *kot celoti* pripisane sporazumevalne okoliščine in nekateri drugi podatki. Te splošne podatke transkribcijske datoteke uvrščamo v „**glavo**“¹¹² **transkribcije** ali **posnetka**. To so:

- datotečni podatki:
 - ime datoteke,
 - podatki o vzporedni zvočni/transkribcijski datoteki in drugih izvendatotečnih referencah;
- dokumentacija obdelave posamezne transkribcije:
 - začetek prepisovanja,

¹¹² Imena kot glava, telo (korpusa, transkribcije/datoteke), baza govorcev, so tu mišljene kot oznake za skupine podatkov in so razmeroma abstraktne. Ko jih kasneje srečamo pri računalniški strukturi korpusa (npr. glava posnetka je konkreten del datoteke, ki se v jeziku markup začne u <head> in konča z </head>), pa so konkretnije. Takrat zajamejo manj podatkov, kot jih navajamo tukaj, ker so nekateri podatki umeščeni drugod v računalniški strukturi (ime mape, ime datoteke, podatki standarda datotečne izmenjave – npr. velikost datoteke v bajtih itd.).

o zadnje urejanje;

- okoliščine besedila (sporazumevalne situacije) – gl. poglavje I.3.1.

Naslednjo, splošnejšo skupino predstavlja organizacija podatkov, ki iz temeljnih enot (posnetkov in transkripcij) izvzame za analizo relevantne podatke za ves korpus ali njegov del – podkorpus: (pod)baze podatkov (npr. govorcev), analitične datoteke, poizvedbe itd. Kot primer preglednice vzemimo **bazo govorcev**. Njeni podatki, vezani na šifro govorca, so:

- biološki spol,
- starost,
- govorno poreklo,
- izobrazba,
- razmerje do izbranega govorca in
- poznanost oz. ime govorca (nekje skrito).

(V ta sklop podatkov, ki veljajo za ves korpus besedil, pa ne moremo uvrstiti podatka o vlogi govorca v *posameznem* besedilu.) Primer takšne baze podatkov je tudi leksikon besed (razbitje vsote vseh verbalnih trakov partitur) ali zvočni leksikon istih besed (izbor besed iz posnetkov), ki pa ni cilj našega dela.

Najbolj splošna skupina podatkov so tisti, ki se tičejo vsega diplomskega dela, vsega gradiva, vseh transkripcij. Če govorimo, da je gradivo kot enota besedilni korpus, ti podatki predstavljajo **glavo korpusa**; gl. poglavje 0. Kateri podatki so skupni za ves korpus, kateri pa le za posamezne datoteke, je odvisno od narave dela; v moji diplomi sem glavni obdelovalec gradiva sam, vsemu je skupna še snemalna naprava, pa programi za obdelavo, jezik gradiva je v večini slovenski. Mnoge podatke je težko pregledno navesti, ampak jih je enostavneje pojasniti v spremnem besedilu, kot to vlogo opravlja pričujoča diploma.

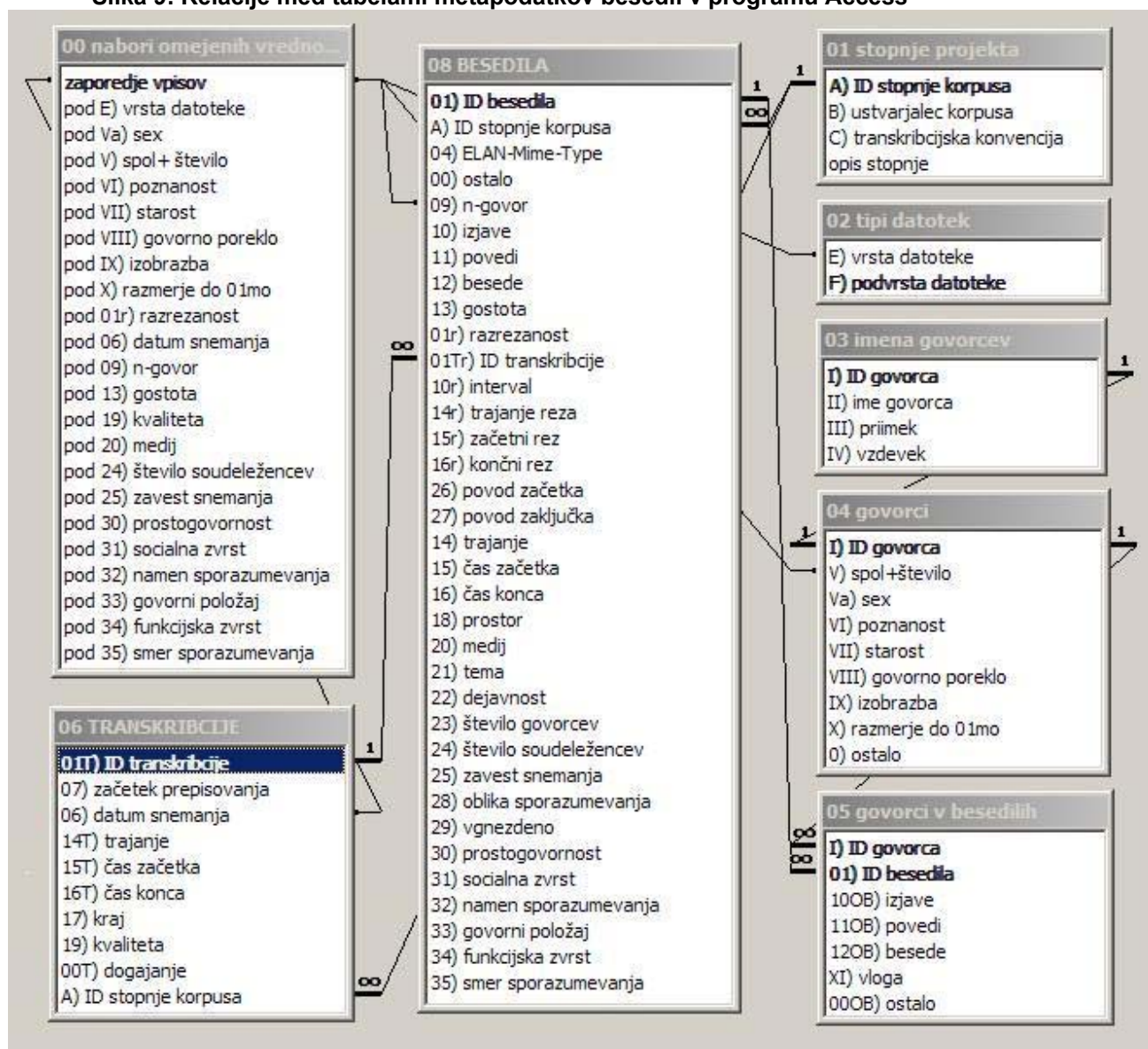
Od logične klasifikacije podatkov v korpusu se obrnimo k možnim mestom zapisa podatka v računalniški strukturi; le-ta so v tabelaričnih pregledih podatkov v II.8 in sledečih poglavjih (sporazumevalnih okoliščin, metaoznak besedil, podatkov o dokumentaciji obdelave gradiva) navedena v skrajnem desnem stolpcu.

Tabela 11: Možna mesta zapisa transkripcijskih podatkov

<i>mesto zapisa</i>	<i>obrazložitev uporabe</i>
partitura transkripcijske datoteke: trakovi ...	Gl. prejšnje podpoglavje.
glave XML-transkripcijskih datotek: EXMARaLDA Partitur Editor (EXM), delno Elan	Vanje je možno vpisati podatke o različnih delih transkripcije: okoliščine cele transkripcije, lastnosti posameznega govorca, lastnosti posameznega traku partiture, opomba k posamezni časovni referenci (točki v posnetku), opomba k posameznemu zapisu (celici, anotaciji) na traku, nabor metaoznak posameznega traku ter seveda logična organizacija cele datoteke. V korpusu od vsega tega uporabljam le: logična organizacija datoteke, zapis podatka o vzporedni zvočni datoteki (za celo transkripcijo), zapis šifer govorcev v seznam govorcev – te tri rabe so neobhodne ter nabor besednovrstnih oznak za trak [BV]. Ostale podatke, ki jih sploh vpisujem (to so za govorce in okoliščine celih transkripcij), pa zapisujem na enem mestu v zbirki podatkov – gl. dve vrstici nižje.
glave zvočnih datotek: .mp3	Vanje je možno zapisati s standardom določene (ime podatka, oblika vrednosti) podatke. Naše okoliščine sporazumevanja bi bilo logično sicer prisoditi tudi posnetkom kot temeljnim enotam in jih zapisati tukaj, a bi bili pri omejeni obliki zapisa in zamujeni z vnašanjem. Sem so bili vneseni nekateri podatki za stopnjo <i>.3Transkripcije</i> , na <i>.4Besedila</i> pa so bili že samo v zbirki podatkov.
zbirka podatkov kot datoteka v programu za urejanje podatkov: Access in tabele v njem	Osnovni podatki za stopnje <i>.3Transkripcije</i> in <i>.4Besedila</i> (dokumentacija obdelave, statistike enot), <i>.2Okoliščine</i> ter podatki iz glave korpusa so razporejene tu v preproste tabele in povezani med seboj (gl. Slika 9 spodaj). Z drugimi besedami – tu so na enem mestu zbrani podatki iz glav posnetkov in transkripcij, njihovih (pod)baz podatkov in iz glave korpusa. Vnos v te tabele in prikaz iz njih je mogoč tudi preko obrazcev (gl. poglavje III.12), priklic in preračunavanje podatkov opravljajo poizvedbe.

<i>mesto zapisa</i>	<i>obrazložitev uporabe</i>
standardni podatki o (računalniški) datoteki – datotečni podatki	Ti podatki so v osnovi dostopni iz podrobnega razvida datotek v brskalniku operacijskega sistema (npr. Explorer/Raziskovalec za Windows) in se jih ne splača prepisovati v zbirko podatkov, razen za potrebe izračunov: velikost datotek, datum nastanka in zadnje spremembe datotek, mesto datotek v pomnilniku oz. mapni strukturi korpusa, trajanje posnetkov, vrsta datotek. ¹¹³
dogovorjeni podatki , ki izkoriščajo mapno strukturo in računalniški format datotek	Gre za šifriranje – imenovanje map, datotek (isto ime besedil iste besedilne enote v različnih datotečnih formatih), pomožne medpone v imenih, imena in oštevilčitve podatkov, razvrstitev v mape (npr. transkripcijske datoteke in vzporedni posnetki v isti mapi) itd. (Mapna struktura = struktura direktorijev na pomnilniku.)

Slika 9: Relacije med tabelami metapodatkov besedil v programu Access



¹¹³ Žal ne poznam programa, ki bi v sebi združeval sposobnosti brskalnika in urejevalnika zbirke podatkov: da hkrati razmeščal datoteke na pomnilniku in za različne vrste datotek urejal podatke v njihovih glavah/lastnostih, npr. prenesel podatke o okoliščinah besedila iz transkripcijske datoteke v glavo posnetka. Bi pa tak program oz. standard prišel ustvarjalec korpusov še kako prav. Nekoliko v to smer deluje program CoMA (Corpus Manager) projekta EXMARaLDA, ki je sposoben na enem mestu modificirati le transkripcijske datoteke in njihove povezovalne datoteke – oboje v XML-jeziku; gl. poglavje I.5.2.2.

6.3. Šifriranje podatkov

Podatkov, predvsem tistih o sporazumevalnih okoliščinah transkripcij in besedil (gl. zgornjo sliko) je toliko, da se med njimi hitro izgubimo, katerega pozabimo ali zanj vstavimo napačno vrednost. Zato imajo podatki svoje stalno ime (v okvirčku) in opisano vrednost; oboje je podano tabelarično v poglavjih II.8 in 0 s podpoglavji ter v II.7.3. Podatki so v teh (pod)poglavjih razporejeni/kategorizirani na 6 načinov:

- 1)) začetna dva načina sta izražena z naslovi in mednaslovi:
 - # 1 podatkovna ravnina podatka (baza transkripcij, baza govorcev, baza besedil, govorcev v besedilih, metaoznake in analize, glava korpusa in obdelava gradiva);
 - # 2 mesto zapisa podatka v računalniški strukturi – desni stolpec v tabeli, npr. Access: tabela, EXM./Elan, datotečni standard, dogovorjeno; trak partiture: [pv], [Ozadje];
- 2)) naslednja dva načina sta izražena s šifriranjem imena podatka:
 - # 3 oštevilčitev podatkov – z abecedo za glavo korpusa, rimskimi številkami za bazo govorcev in arabskimi številkami za okoliščine in obdelavo transkripcij/besedil;
 - zaporedni podatki so si vsebinsko sorodni;
 - ista števila, imajo enako strukturo vrednosti in podoben podatek; primer 10, 10OB in 10r imajo vsi številsko strukturo vrednosti, njihovi podatki pa so: izjave besedila, izjave govorca v besedilu in obseg reza besedila po časovnih referencah v transkripciji;
 - # 4 pri nekaterih podatkih kombiniranje oštevilčitev s črkami za številko:
 - velike črke pomenijo eno od 8 stopenj obdelave korpusa (.3T-ranskripcije; OB – g-ovorcev (2Okoliščine) v .4Besedilih;
 - male črke: za podstopnjo -r-ezanje, za trakova p-araverbalno in o-zadje;
- 3)) definirane so vrednosti podatkov oz. njihov zapis:
 - # 5 bodisi je podan in razložen omejen nabor vrednosti (omejene vrednosti so podčrtane), npr. za V spol+število nabor moški | ženska | neznano | več (skupaj); gl. tudi seznam teh naborov levozgoraj – Slika 9 zgoraj;
 - # 6 bodisi je definiran zapis vrednosti: številka, omejen črkovni zapis, opis delov šifre itd.

Še to: potrebno je paziti pri ločevanju med vrednostmi podatkov neznano | brez | več in na to, katere kategorije podatkov pri besedilih/govorcih/itd. sploh nimajo podatka (ker ga ne rabim).

7. PRAVILA TRANSKRIBIRANJA TEKOČEGA GOVORA

Za transkribiranje pričujočega gradiva sem razvil svoja pravila, v primerjavi z zgledi drugih transkripcijskih konvencij in po svojih lastnih potrebah. Uporabljena **pravila transkribiranja** niso nič drugega kot slovenski pravopis, modificiran po načelih, nazadnje opisanih v poglavju I.2.2.3, in priročnik za vpis jezikovnih podatkov v korpusu (o razmerju do drugih podatkov predhodna in sledeča poglavja). Bila so ustvarjena tudi z mislijo, da bi bila lahko uporabljena v referenčnem korpusu (večprenosniškem, torej tudi pisnem – prav cilj možne (leksikografske) primerjave s pisnimi besedili jih bistveno določa). Pa nato še prilagojena praktičnim željam po enostavnosti, nezamudnosti ipd. Samoontroliranje pravil transkribiranja v gradivu je opisano v poglavju II.9.1, prikaz postopka transkribiranja pa je v poglavju III.11.

7.1. Razločevanje in zapis besedne (verbalne) strukture govora

Ortografski zapis govora v tej transkripcijski konvenciji je najenostavneje predstaviti skozi par osnovnih vodil:

- 1)) ves besedni govor, vključno z ubesedenjem medmetov (oz. mašil, zapolnjevalcev vrzeli ...), je pripisan enemu od govorcev na svoj trak v partituri;
- 2)) četudi je govor neprekinjen, so v zapisu razločene besede po oblikoglasni tradiciji, torej ločene z znakom za presledek;¹¹⁴
- 3)) črkopisni nabor transkripcije je ortografski (na poseben način rabljena gajica/slovenica), dopolnjujejo ga le nekatera izbrana oblikoskladenjska ločila (gl. Tabela 12 spodaj);
- 4)) pretvorba na pomenskorazločevalni ravni je fon(em)ografska, ne logografska ali celo prebesedenje/prevajanje na leksemski ravni: brez števčk, krajšav govora (kratic, simbolov in okrajšav), zvestoba razločevalnemu sistemu morfemov in besed v govoru, interpretacija fonov kot fonemov in pretvorba le-teh v črke po najbližji podobnosti s preglednico Zapisovanje glasov v SP²⁰⁰¹: § 1070;
- 5)) raba velikih začetnic le za lastna imena;
- 6)) pri dilemah zapisa besed so upoštevana Pravila v SP²⁰⁰¹, zapis je primerjan z javnimi normativnoslovarskimi priročniki, v primeru problematičnosti označen (zvezdica ali drug znak) in dokumentiran – gl. naslednje poglavje;
- 7)) kadar so besede tujega jezika izgovorjene z njegovim lastnim fonološkim sistemom, je za zapis uporabljena ortografija tega jezika (če jo zapisovalec seveda pozna), če pa s slovenskim fonološkim sistemom, pa je zapis podomačen – oba zapisa pa sta posebej označena; daljše pasaže v tujem jeziku niso zapisane;
- 8)) od skladenjskih ločil so izbrana štiri za označevanje meja povedi (gl. Tabela 14, str. 105), nadpovedne enote pa so definirane s partiturnim vizualnim formatom transkripcije, računalniško strukturo transkripcijske in drugih vrst datotek korpusa.

Vsa uporabljena ločila so stična vsaj na eni strani. Nobeno „oblikoskladenjsko“ ločilo ne zaznamuje posebne značilnosti govora (fonetike ...), vsa imajo metajezikovno oz. analitično vrednost.

Tabela 12: Izbrana ločila v transkripcijski konvenciji StAnId4

<i>znak</i>	<i>nabor= koda</i> ¹¹⁵	<i>primeri</i> ¹¹⁶	<i>pomen uporabe</i>	<i>morebitna dilema</i>
-	nedeljivi ¹¹⁷ vezaj (levo- in desnostičen) UTF-16= 002D, ISO-8859-2= 45, Eva= 658	*UeSBe-priključek *CeDe-jev *eMPe_tri- je ¹¹⁸ ja sraka je! pisana ... *letan'ce 'ma m- modro- bele. taka vrečka je *e od m- *Milerja.	Stični vezaj v oblikoslovni in sklad. rabi, posredno označuje pregibane kratice. Nedokončane oz. prekinjene besede.	Prva raba je povsem običajna in standardna po pravopisu. Od druge se loči po tem, da je lahko levo- in desnostična, druga pa vedno samo levostična.

¹¹⁴ Z uporabo presledkov je opravljena implicitna segmentacija besed (tokenizacija), kar je podobno kot nekakšen „implicitni markup“. Na tako ločene besede je potem na posebnih trakah partiture možno pregledno dodati metaoznake besed (oblikoskladenjske določitve in leme); ni recimo potrebno poleg traku „besedilo govorca“ imeti še tokeniziranega traku „besede govorca“ v partituri.

¹¹⁵ Desetiške srednjeevropske oznake (ISO-8859-2) veljajo tudi za program Eva, izjeme so v tabeli posebej zapisane.

¹¹⁶ Znak <|> ločuje primere izjav ali na besedo okrajšane primere med sabo.

¹¹⁷ Nedeljivi v smislu, da je tiste vrste znak, ob katerem program za urejanje besedil (ali funkcija tega programa) ne prelomi besede na dve besedi, ampak ga pojmuje kot del besede. Ko strojno tokeniziramo besedilo, v katerem so besede zapisane s pomočjo posebnih znakov iz zgornje tabele, si želimo, da se ti znaki skupaj s črkami, ki se jih stično „držijo“, ohranijo *kot del besede*. Zato je potrebno te posebne oznake v programih pazljivo definirati in razlikovati (s pomočjo kode) od „deljivih“ ločil (tudi non-character, nečrkovni znak, prim. opombo 45 na str. 30), čeprav so si z njimi na videz podobni ali enaki.

¹¹⁸ Vezaj je tukaj, kot bi bil v pisnem prenosniku (CD-ja, USB-priključek; prim. SP²⁰⁰¹: § 417 in 423).

<i>znak</i>	<i>nabor= koda</i> ¹¹⁵	<i>primeri</i> ¹¹⁶	<i>pomen uporabe</i>	<i>morebitna dilema</i>
'	(nedeljivi) opuščaj UTF-16= 0027, ISO-8859-2= 39	tak' z'aj pa *s'n! ja super () jaz mislim da so 'mel' razlog za to ...	V primerjavi s standardno pisno obliko na mestu opuščaja v govoru ni posameznega fonema, prim. SP ²⁰⁰¹ : § 468. ¹¹⁹	Možno na več nesosednjih mestih v isti besedi. Ne gre ga zamenjati z vezajem za nedokončane besede. Za dva in več sosednjih izpuščenih fonemov glej znak <*>.
–	podčrtaj oz. nedeljivi presledek UTF-16= 005F, ISO-8859-2= 95, Eva= 185	izjave po tekmi stran petsto_šestnajst. Feri_Horvat je predsednik državnega zbora?	Označevanje enakovrednih zvez pojavnice, ki skladenjsko delujejo kot ena beseda, ki ima lahko poseben pomen: števila, lastna imena.	
()	uklepaj in zaklepaj UTF-16= 0028 in 0029, ISO-8859-2= 40 in 41	() šlo da *nis'n 'mel kaj delat'. (je b'la) cel *keden megla *e *heheh ... *mm že dolgo časa nisem šel *tlele ... ne! sam' () ... ne boš šel po dežju!	Označuje pasaže popolnoma nerazločnega ali predvidevanega govora. Pomeni lahko eno ali več besed, nerazločne povedi in izjave pa so v celoti izpuščene. Lahko stoji na začetku ali sredi povedi, ali na koncu povedi, kjer za seboj potrebuje še končno ločilo. Da se povedi sploh zapiše, mora biti poleg znaka za izpust (in končnega ločila) v njej vsaj ena razločna beseda.	
=	enačaj UTF-16= 003D, ISO-8859-2= 61	(=014Mz_ime) ti se spomniš? (=009Žz_ime) a 'maš ti tvoj mobilni?	V prvotni transkripcijski fazi, ko imena govorcev še niso zakrita s šiframi, in kasneje olajša iskanje teh imen. Imena so poleg tega še v oklepajih.	
*	(nedeljiva) zvezdica UTF-16= 002A, ISO-8859-2= 42 (Za podrobno obrazložitev glej poglavji II.7.1.5 in II.7.1.6.)	<ul style="list-style-type: none"> • *Milerja 'Müllerja' • *e, *ee *agh • *pol 'potem' *kva 'kaj' • *mogel 'moral' 	ohlapna povezljivost s standardno obliko	
		*pol pa tam ko mu daš *print greš na *properties ...	običajna raba (tuje ortografije)	
		<ul style="list-style-type: none"> • *zdej *dones *sej 'saj' *prov 'prav' • *čak 'čakaj' *nam 'ne bom' • *bojo 'bodo' *sma 'sva' • *UeSBe-priključek 	jasna povezljivost s standardno obliko	
		• *pobojnih pobojev, *literalna pravljica	predizrazno nepopoln govor	

Zgornjih ločil sem se v transkripcijski konvenciji oz. modificiranem pravopisu poslužil, ker so zelo preprosto sredstvo označevanja. Poleg tega so bila merila izbora: potreba po rabi, raba v transkripciji se naj ne razlikuje od pomena ločil po pravopisu (zaradi lažje razumljivosti naključnemu uporabniku korpusa), enopomenskost/enofunkcijskost ločil, možnost enoznačne rabe v transkripciji (zaradi preprostejše strojne obdelave).

Potreba po rabi je v bistvu ena sama: *prikazati govor nepotvorjen v oblikoglasnosti, v besednem (leksikalnem) sistemu pa tako, da ga lahko čimbolj logično strukturiramo* (kot je to možno v leksikografiji pisnih besed). Ne gre torej enostavno za prikazovanje nestandardnosti posameznih elementov govora (fonov, morfemov, leksemov), ampak za beleženje razlik v oblikoglasnem (oblikorazločevalnem) funkcioniranju govornega in pisnega izraznega sistema.

Na pomenskorazločevalni ravni je troje razlik: izpust fonemov (zaradi fonetičnosti ali prekinjenega izgovora), „prekvalificiranje“¹²⁰ fonemov, dodajanje fonemov. Prekvalificiranje in

¹¹⁹ Ta paragraf pravopisa dostavlja: „V pogovornih in narečnih besedilih opuščaja ne pišemo: *mladga fanta; ne smeš oditi*.“ To navodilo nam ne brani zapisovati opuščaja pri transkribiranju govora; postavi nas le v pozicijo, da ima naš zapis status knjižnega, pravopisno standardnega besedila.

¹²⁰ „Prekvalificiranje“ fonemov tukaj pomeni, da je nek nestandardni alofon (v konkretnem/aktualnem izgovoru) fonema (namreč fonema standardne besedne oblike na tem mestu) interpretiran s transkripcijo kot osnovni/standardni alofon drugega fonema; npr. v primeru *don's (danes) je [o] nestandardni alofon fonema /a/, z zapisom s črko <o> namesto <a> pa je interpretiran za fonem /o/.

dodajanje fonemov prikažemo s črkami, izpuste pa z **oblikoslovnimi ločili** (to je takimi pisnimi⁷ znamenji, ki poleg črk sestavljajo besedo). Slednjih pravopis ne ponuja dosti; poleg zgoraj naštetega *vezaja* in *opuščaja* še *piko* (prim. opomba 29 na str. 21), *vejico* (§ 351–352) in *tropičje* (§ 410).¹²¹ A teh treh ločil nismo izbrali; vejice, ker je v takšni rabi le pri številkah, pike in tropičja pa ne, ker nista enopomenski in ker smo ju uporabili že za končni ločili. Njuno vlogo označevanja *opuščenih* glasov prav dobro opravi *opuščaj*, pri prekinjenih besedah pa to vlogo opravi levostični vezaj (prim. § 434).

Med oblikoslovnimi ločili nekatere transkripcijske konvencije uporabljajo oklepaje, s katerimi označijo neobvezne sestavine besed, ko jih primerjajo s standardno obliko.

Primer 16: Primera uporabe oklepajev v modificirani ortografski transkripciji

woll(e)n we (wir) nich(t) ein anderes Buch kucken (angucken) (ortografsko popravljena transkripcija; Brinker/Sager 2001: 48)

ənbəŋɡə → {ə}n[o]bən[e]ɡə | nubəŋɡə → n{u}[o]bən[e]ɡə (gre samo za navajanje primerov, ne tekočo transkripcijo; Zemljarič Miklavčič/Stabej 2006: 8. prosojnica)

Ta raba v tej diplomii ni uporabna predvsem zaradi olajšanja nadaljnje kvantitativne analize, pa tudi zaradi preglednosti; odprta rešitev pa je zapis primerljivih (standardnih) besed v partituri na vzporeden trak. Naša transkripcija oklepaje uporablja za označevanje slabo razumljenih besed, primerjavo s standardnim zapisom pa opravi s pomočjo drugih ločil.

Na kompleksnejših jezikovnih ravninah bi bilo za beleženje razlik potrebnih več ločil, pa še ta bi bila verjetno informativno prerevna. Prav pride izpostavno znamenje iz jezikoslovja, **zvezdica** „na levi strani pred besedo“, ki po SP²⁰⁰¹ „opozarja na vzpostavljeno (domnevno) besedo“ (§ 18) v praksi pa tudi na del besede oz. morfem. Kar je „domnevnega“ pri transkribiranih besedah, ki jih označuje zvezdica, je tako način zapisa;¹²² če pa je zapis besed pravopisno neproblematičen, ne vključuje zvezdice ali kakršnegakoli drugega izrednega ločila. *Zvezdica označuje razlike med pravopisno problematičnim zapisom in standardno (pravo)pisno obliko, ki so z različnih jezikovnih ravnin.*

Zadnje izredno rabljeno ločilo, podčrtaj oz. nedeljivi presledek, je vzet iz računalniške leksikografije, kjer pomaga izmed pojavnic izločati tiste besedne zveze, ki jih je smiselno obravnavati kot en leksem zaradi njihove pomenske (podatkovne, izrazne ...) zanimivosti: lastna imena, števila, elektronski naslovi, kode ipd.

Poleg v tabelah omenjenih znakov se pri strojni obdelavi priložnostno, v vmesnih fazah, uporabljajo še kakšni znaki. Primer: program EXMARALDA Partitur Editor ima vmesnik, ki besedilo v transkripciji razčleni na podlagi transkripcijske konvencije HIAT in prešteje njegova ločila, besede, povedi, izjave (gl. Slika 21, str. 140). Ker HIAT vzame opuščaj in vezaj sredi besede tako, kot da besedo delita na dve besedi, štetje za našo konvencijo StAnId4 ne deluje. Problem rešimo enostavno tako, da opuščaj začasno zamenjamo z znakom <□>, vezaj pa z <~>, ki ju program vzame kot črko oz. del besede.

7.1.1. Razdelitev pisnih znamenj po njihovem pomenu v transkripciji

Pri prenašanju govornega izraznega diasistema jezika v pisni izrazni diasistem jezika imamo nekaj problemov. (Več o tem v uvodu v poglavju I.2.1.) Ker je narava transkribiranja

¹²¹ Tu ne štejemo ločil, ki služijo kot diakritična znamenja v fonetičnih zapisih: stična resica za zaznamovanje naglašenosti, dvopičje za dolžino glasu; in tudi ne drugih matematičnih in drugih znanstvenih znamenj, ki so sicer logografsko znamenje za besedo ali del besede.

¹²² Primerljiva so tudi normativna znamenja iz SP, ki opozarjajo na neko razmerje jezikovnega pojava do standardnega jezika. Toda zvezdica je v računalniških naborih znakov bolj dostopna in bolj pogosta v občni rabi. Bolj svobodna je tudi interpretacija „domnevnosti“ pri zvezdici kot „prepovedanosti“, „nepravilnosti“ in „odsvetovanosti“ pri normativnih oznakah SP (prim. § 23).

deskriptivna, ker govor opisujemo in zapisujemo, pri uporabi ortografije ni problematičen govor, ampak njegov zapis. *Tudi ne primerjamo aktualnega govora v gradivu s standardno izreko ali nasploh z izrazno normo.* Le na najbolj ustrezen način želimo uporabiti glasovno in pomenskorazločevalno vrednost črk, ločil in pravopisnih načel. Zaradi tega med rezultati transkripcije lahko dobimo različne (tudi nestandardne, nepravopisne) zapise iste besedne oblike, seveda, ko jih gradivo na posnetkih sploh izpričuje; npr. jaz, *jast. Vsaki aktualni govorni obliki poskušam prisoditi primerljivo standardno pisno obliko.¹²³

Pravopisno problematične besede v zapisu enostavno ločimo od neproblematičnih prav po nekaterih od izbranih ločil, opisanih zgoraj. V tem smislu je vsak zapis, ki vsebuje eno teh izrednih ločil, zapisan v modificirani ortografski transkripciji in s tem sam po sebi nestandarden, problematičen.

Ta ločila tako izrazito delujejo kot metajezikoslovna oznaka. Je pa nasploh *že vsaka sama transkripcija metajezikoslovna oznaka*, saj „vsiljuje“ govornemu jeziku drug izraz¹²⁴. Pri nadaljnji metajezikoslovni sistemizaciji govornih enot pa smo bolj abstraktni, razlike med prenosnikoma/izrazoma se brišejo. Besede, problematične z vidika zapisa, niso več tako problematične z vidika besednovrstnega in oblikoskladenjskega določevanja ter z vidika določitve leme (ta bo podana v standardni pisni obliki), pomena itd. Nekaj dilem oz. mejnih primerov pa vedno ostane.

Nazaj k pravopisno problematičnemu zapisu govora, k besedam. Da bi bil njihov zapis (transkripcija) kljub razliki med govorno aktualizacijo in pravopisom čim bolj *konsistenten in zaradi tega v sebi standardiziran*, poskušam v sledeči tabeli in podpoglavjih *vse vrste zapisov* sistemizirati, primerjati med sabo in v razmerju do pravopisa.

V naslednji tabeli je razločenih 21 vrst zapisov. „**Vrsta zapisa**“ (transkripcije) nam tukaj pomeni podatek, ki ga transkripcijska konvencija omogoča o „čistem besedilu“ glede na svoje analitične cilje v diplomu. Te „vrste zapisov“ so opisane s pomočjo umestitev na jezikovno ravnino (levi stolpec; za oblikovno stran jezikovnega znaka pomembnejše ravnine so podrobneje razdelane), dalje sistema primerjav (desni stolpec) in ločila/znamenja, ki ga transkripcijska konvencija ponuja. 21 vrst zapisov lahko poenostavimo na 5 sistemov primerjav, ki so podrobneje opisani v naslednjih podpoglavjih. „Problematični zapisi“, o katerih sem pisal tri odstavke nazaj, predstavljajo le dve od teh skupin.

5 sistemov primerjav – skupin podatkov o čistem besedilu govora, ki jih omogoča transkripcijska konvencija StAnId4:

- 1)) analitična izpostavitvev podatka – izpostavitvev nekaterih vsebinskih (za raziskovalca zanimivih) delov govora;
- 2)) predizrazno nepopoln govor in nepopolno razumevanje – zapisi, ki govorijo o motnjah v prvi in zadnji fazi sporazumevanja (izven problematike strukturiranja izraza jezikovnih znakov);
- 3)) neproblematičen zapis – v zapisu (fonografski izrazni strukturiranosti) govornega izraznega sistema s pisnim ni razlik;
- 4)) problematični zapisi: jasna povezljivost s standardno pisno obliko – problematičnost prenosa govorne v pisno izrazno strukturo;
- 5)) problematični zapisi: ohlapna povezljivost s standardno pisno obliko – problematičnost prenosa govorne v pisno izrazno strukturo.

¹²³ Priročnik, v katerem je bila preverjena standardna oz. uveljavljena pisna oblika, je bil samo SSKJ v elektronski obliki. Bolj poglobljena raziskava bi vključila še modernejše jezikovne priročnike: SP²⁰⁰¹, Seznam slovenskih besed (SSKJ, BSJ in korpus Nova Beseda) in korpus Fidaplus.

¹²⁴ S tem ga izrazno prekategorižira. Daje mu nov izraz/obliko, eno izmed dveh bistvenih sestavin/polov znaka (vsebina in oblika).

Tabela 13: Skupen pregled znamenj transkripcije po jezikovnih ravlinah in pomenu v transkripciji

<i>JEZIKOVNA (pod)RAVNINA izpostavljenega podatka</i>		<i>primer</i>	<i>VRSTA ZAPISA: opis podatka, ki ga transkripcija izpostavlja</i>	<i>LOČILO /ZNAMEN JE</i>	<i>pomen v transkripciji: SISTEM PRIMERJAV, ki izpostavi podatek</i>
<i>kognitivna</i>	<i>a.1</i>	taka vrečka je *e od m- *Milerja.	nedokončane besede	- (levostičen)	predizrazno nepopoln govor
	<i>a.2</i>	*pobojnih pobojev, *literalna pravljica	napačna izgovorjava, lahko tudi disleksija	*	predizrazno nepopoln govor
<i>kognitivna (pa tudi besedna, besednozvezna)</i>	<i>a.3</i>	() šlo da *nis'n 'mel kaj delat'. (je b'la) cel *keden megla *e *heheh ...	nezanesljivo razumevanje ene ali več besed	() , (besede)	nepopolno razumevanje
<i>fonološka</i>	<i>b.4</i>	*zdej *dones *sej (saj) *prov (prav)	„prekvalificiranje“ posameznega fonema (drug fon v fonemski funkciji; gl. opomba ¹²⁰)	*	jasna povezljivost s standardno obliko
<i>grafofonemska (ortografska v ožjem smislu)</i>	<i>c.5</i>	aha. bom jaz vzel.	zapis standardnih fonemov	25 malih črk gajice/slovenice	neproblematičen zapis
	<i>c.6</i>	taka vrečka je *e od m- *Milerja.	zapis podomačenega izgovora tuje/prevzete besede	*	ohlapna povezljivost s standardno obliko
	<i>c.7</i>	*pol pa tam *ko mu daš *print greš na *properties ...	zapis tujega izgovora s tujo ortografijo	*, tuja ortografija	neproblematičen zapis
<i>morfemska (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)</i>	<i>č.8</i>	*pr'su (prišel) *neki (nekaj)	„prekvalificiranje“ vsaj dveh sosednjih fonemov (gl. opomba ¹²⁰)	*	jasna povezljivost s standardno obliko
	<i>č.9</i>	*jest (jaz)	dodaten fonem	*	jasna povezljivost s standardno obliko
	<i>č.10</i>	tak´ z´ aj pa *s´ n!	izpuščen fonem	'	jasna povezljivost s standardno obliko
	<i>č.11</i>	*nes (danes) *čak (čakaj) *nam (ne bom)	opuščanje vsaj dveh sosednjih fonemov, tudi skržitev več besed v eno	*	jasna povezljivost s standardno obliko
	<i>č.12</i>	*e, *ee (polglasnik-mašilo) *agh *ova	nestalne oblike medmetov (njihovo daljšanje in ponavljanje; bližanje zapisa dejanskemu izgovoru)	*	ohlapna povezljivost s standardno obliko
<i>besedotvorna (morfološka – oblikotvorna)</i>	<i>d.13</i>	*bojo (bodo) *sma (sva)	drug fonem ali sklop fonemov v morfemski funkciji	*	jasna povezljivost s standardno obliko
	<i>d.14</i>	*eMPe_tri-je ja sraka je! pisana ... *letan´ce´ma m- modro-bele.	navadni vezaj	- (obojestičen)	neproblematičen zapis
	<i>d.15</i>	*UeSBe-priključek *CeDe-jev *eMPe_tri-je	izgovor kratice (logogram za besedno zvezo v obliki besede)	* + velike črke za foneme, ki označujejo okrajšane besede	jasna povezljivost s standardno (pisno) obliko
<i>leksikalna (pomenska)</i>	<i>f.16</i>	Boris Njegoševa (vrečka od) *Milerja	lastna imena po pravopisu	velike začetnice	analitična izpostavitve podatka (tudi neproblematičen zapis)

JEZIKOVNA (pod)RAVNINA izpostavljenega podatka	primer	VRSTA ZAPISA: opis podatka, ki ga transkribcija izpostavlja	LOČILO /ZNAMENJE	pomen v transkribciji: SISTEM PRIMERJAV, ki izpostavi podatek	
	f.17	petsto_šestnajst Feri_Horvat	večbesedna števila in lastna imena	–	analitična izpostavitvev podatka
	f.18	(=014Mz_ime) ti se spomniš?	skriti podatki govorcev	(=podatek)	analitična izpostavitvev (skritje) podatka
	f.19	*sprobal' *komot ('udobno') *pol ('potem') *kva ('kaj')	neknjižni/nestandardni leksemi – besede	*	ohlapna povezljivost s standardno obliko
	f.20	samo vozni red bi še *mogel 'met pa bi b'lo.	neknjižni/nestandardni pomen besede	*	ohlapna povezljivost s standardno obliko
nadstavčna	f.21	Gl. poglavje II.7.2.	eno medstavčno in 4 končna ločil	:: ? ! ...	Gl. poglavje II.7.2.

V naslednjih podpoglavjih so najprej za vsako vrsto zapisa dani primeri iz korpusa, sledi pa jim pojasnilo o razlogu za tako obliko transkribcije in še kakšna dilema glede transkribiranja.

7.1.2. Analitične izpostavitve podatkov

Kot sem višje zgoraj že omenil, da je vsa transkribcija že metajezikovna oznaka jezika, v bistvu velja tudi, da je vsaka od vrst zapisov v transkribciji tudi analitična izpostavitvev podatka. Pa vendar želi to podpoglavje opozoriti le na tista transkripcijska znamenja, ki izpostavljajo zanimivo vsebino govora kot jezikovnega znaka – gre za izpostavitvev določenih pomenov. Zanimivi so za našo diplomu, kdo drug bi izbral druge: lastna imena po pravopisu večbesedna števila in lastna imena ter prikrivanje podatkov govorcev.

leksikalna r.	f.16	lastna imena po pravopisu	velike začetnice
Boris Njegoševa taka vrečka je *e od m- *Milerja.			

Zgornja vrsta zapisa bi sodila tudi v kategorijo *neproblematičen zapis* (gl. spodaj II.7.1.4), saj so velike začetnice zapisane po določilih iz pravopisa (§ 34–115). Dilema je le pri zapisu izrazov posebnega razmerja in spoštovanja (§ 110–114), veroslovnih in bajeslovnih imenih (§ 47, 48) ter alegoričnih poosebitvah (§ 49); zapisovalcu iz govora ni razvidno posebno razmerje govorca, ki ga sicer v pisnem izraznem sistemu izraža velika začetnica. (Nasploh govor podatka o lastnosti imen ne izraža; le izjemoma za posebna razmerja v stavčni fonetiki npr. z barvo glasu; ali s pomočjo telesnih gibov vzporedno z govorom.)

V korpusu je tudi transkribcija obreda svete maše in znotraj njega molitev Marijinih litanij. Nekatera imena so pisana z veliko (*Jezus_Kristus, naš Gospod, Bog Oče, Sin in Sveti_Duh*), nazivi Marije v litanijah (*devica, mati, kraljica* in drugi) so z malo, tudi pridevnik *božji* sem pisal z malo začetnico. Čeprav – raba teh nazivov je pogosto podvržena pisnemu prenosniku (branje) in v verskih knjigah (zapisovalec nima istih kot govorcev, bralci) so ta imena skoraj brez izjeme pisana z veliko začetnico; torej bi se dalo ta imena/izraze/nazive pisati z veliko, pa to posebnost obrazložiti na paraverbalnem traku partiture [pv].

Pozor: prve besede v povedih v transkripcijski konvenciji nimajo velike začetnice (ni po § 28–34)!

Vrsta zapisa je umeščena na leksikalno ravnino zato, ker se informacija, skrita v razlikovanju velike začetnice ne nanaša samo na posamezno črko, ampak na vso besedo ali celo preko nje – na besedno zvezo – gl. naslednji primer.

leksikalna r.	f.17	večbesedna števila in lastna imena	<_>
*Sacrum_*romanum_*imperium_*nationis_*germaniae. dvanajst_deset do trinajst_petdeset. *Bufonovo mrežo je zatresel Boštjan_Cesar. kvalifikacije za svetovno prvenstvo na straneh petsto_petnajst do petsto_devetnajst.			

Pri večbesednih lastnih imenih je spet motiv ohranitev podatka o *posameznih* pojavih. Dodatno znamenje, podčrtaj, skuša preseči problem interpretacije, da z malo zapisane besede teh imen niso rabljene občnoimensko. Hkrati podčrtaj ohranja večbesedno lastno ime kot pomensko-izrazno celoto v fazi razloma besedila na besede. Omogoča možnost oblikovanja leksikona lastnih imen – popolnih imen). Za primerjavo sestavin lastnih imen z občnoimenskim besedjem in za vključitev posameznih besed v občnoimenski leksikon pa je podčrtaj zamenjan z dvema znakoma (podčrtaj + presledek + podčrtaj; ali pa je lahko zadnji besedi večbesednega imena namesto na začetku na koncu dodan podčrtaj), da po strojnem razbitju besedila na seznam besed podčrtaj za besedo še vedno označuje njen kontekst v konkretnem gradivu.

Podobno velja pri pisanju večbesednih števil, ki jih sicer po pravopisu pišemo narazen (§ 485). Brez podčrtaja pišemo besedne zveze števil, v katerih so vmes števniske besede: datumi ali opis ure (*e po mene pride tri četrt na devet. | lokalni potniški vlak v Zidani_Most Senovo odpelje ob sedemnajsti uri petdeset minut s sedmega tira.).

Pri opazovanju stavčnofonetične podobe besed sem bil kdaj v skušnjavi, da bi tudi kakšno pogosto stalno besedno zvezo, ki ima en sam naglas (včasih v diskurznoanalizični literaturi imenovano fonetična beseda), zapisal s podčrtajem: <v_*redi> ali celo kot eno besedo <vredi> Vendar sem ohranil običajni zapis: <v *redi>.

Tudi glede krajšanj več zaporednih besed v navidez eno besedo (<ne bom> → <ne 'om>; slednji zapis je bil za take primere uveljavljen) ali prenosa naglasa z naglašene beseda na naslonko (ne bóm → nê bom) je bil nagib, zapisati jih s podčrtajem. Zaradi konsistentnosti in odmika od fonetične predstavitve govora pa je bil podčrtaj prihranjen le za zgoraj omenjena večbesedna števila in lastna imena.

leksikalna r.	f.18	skriti podatki govorcev	<=>
(=014Mz=ime) ti se spomniš?		(=009Žz=ime) a 'maš ti tvoj mobitel?	

Podatke govorcev (imena, priimke, naslovi in drugi kontaktni podatki ...) je etično skriti zato, da pri morebitni javni uporabi korpusa ne pride do zlorab teh osebnih podatkov. Zakrivanje teh podatkov identitete govorcev sicer povsem ne zakrije (prim. opombo 146), saj druge stavčnofonetične značilnosti zagotavljajo dovolj prepoznavno sled posameznega govorca.

Ko slišani govor med transkribiranjem na podlagi te „sledí“ pripisujemo posameznim govorcem na traku, nam imena govorcev med besedilom (klicanja po imenu, vzdevkih ...) delo precej olajšajo. Zato v prvi fazi transkribcije ti podatki govorcev še niso zašifrirani, so pa označeni z enačajem in dani v oklepaj, izmišljen primer je <(=Štefka)>. V fazi zakrivanja podatkov govorcev jih je tako v besedilu možno najti strojno z iskanjem tega znaka. Dilema je, ali podatke govorcev zakriti tudi drugih besedilih, kot govorci sami sodelujejo; v korpusu tako daleč nisem šel.

Zakritje imena sestavljajo poleg enačaja in oklepaja šifra govorca, podčrtaj in oznaka podatka (ime, priimek, podatek; svojilno – za svojilne pridevnike). Ker je tako zakrivanje podatkov v besedilu zaradi iskanja šifer govorcev kar zamudno, je možna tudi krajša varianta: na mesto šifre govorca je lahko dan drug znak <#>; v tem primeru se izgubi informacija o usmerjenosti govornega dejanja k nekemu naslovniku, kar pa je bolj važno za kvalitativne kot kvantitativne raziskave. Šifriranje govorcev v oznaki posameznega traku in v bazi govorcev pa je vneseno v vsakem primeru.

7.1.3. Predizrazno nepopoln govor in nepopolno razumevanje

Zapisi te skupine govoriijo o motnjah v prvi (faza zamišljanja govora pred fazo ubesedovanja) in zadnji (dekodiranje besedila pri naslovniku) fazi sporazumevanja in so zato izven problematike razlik v strukturiranju izraza med pisnim in govornim izraznim sistemom.

kognitivna r.	a.1	nedokončane besede	- (levostičen)
taka vrečka je *e od m-	*Milerja.	ahm tam je črta vprašanje kak' če bo če bo omejit-	... ah posku- ... ne mislim da *pol ne mor's levo zavit'. od srede *os'najstega do tisoč ... -najststo *os'nšt'r'deset_*šeset ... izgubili smo na naum- neumen način .

Pri nedokončanih oz. prekinjenih besedah (tudi „požrtih“) standardna pisna oblika obstaja kvečjemu za popolno obliko, le-ta pa nam ni gotovo znana (in s tem tudi lema ne). Le včasih lahko sklepamo o njej, če se pojavi v kontekstu, npr. s popravkom, ponovitvijo. Besednovrstna določitev prekinjene besede je lahko kvečjemu izredna (npr. z oznako K – kratica, kot v pisnem korpusu, Jakopin/Bizjak 1997: 521), oblikoskladenjska določitev ni mogoča (zaradi izražanja oblikoskladenjskih podatkov v končnem delu besede, ki je izpuščen), iz sklepanja pa ni smiselna. V nekem smislu so take prekinjene besede podobne nedokončanim in prečrtanim besedam v rokopisu.

Pri zadnjem primeru je bil razlog prekinitve napačno glasovno ubesedenje (v primerjavi z bralno predlogo). Razlog je isti kot pri primerih v naslednji tabeli, a se dodatno označevanje z zvezdico ne zdi potrebno, prav tako bi prekinjeno besedo težko kategorizirali.

Obstajajo tudi dileme uporabe levostičnega vezaja pri nekaterih podobnih primerih. Vezaj na začetku besede pomeni samo nedosledno izgovorjeno besedo (zaradi trenutno pogojene nepopolne artikulacije, npr. do tisoč ... -najststo_*os'nšt'r'deset_*šeset ...; izmišljen primer še -bro jutro), ni pa ga možno uporabiti za leksikalno običajno krajšane besede: *nes ('danes'; torej ne <-nes>). Tudi pri besedah, sestavljenih iz ponavljanja zlogov (npr. opevanje, mrmranje; *lalalalalalalala), ali pri originalno okrnjenih besedah (v besedilu pesmi; *alelu aleluja) nu levostičnega vezaja, marveč samo zvezdica.

kognitivna r.	a.2	napačna izgovorjava (zarekanje)	<*>
zamolčanim žrtvam	*pobojnih	pobojev. nehala se je *literalna	pravljica z 'aj se pa nadaljuje nogometna. ja. ker je oni postal *e Pahor je *poslal

Zgornji primeri so v besedilu redki. Konvencija jih ne izpostavlja zato, ker bi bili pomembni za analize, želi jih samo označiti. (Nekatere konvencije oz. redakcije transkripcij govora take primere in medmete, ponavljanja ipd. označijo zato, da jih pri določenih analizah lahko umaknejo in dobijo lepše povedi.) Označba celo nekoliko presega raven same besede, saj je šele iz razumevanja konteksta razvidno (no, včasih tudi iz spremenitve ritma govora, skoraj jecljanja), da je beseda napačno izgovorjena. Pri razlomu besedila na seznam posameznih besed bi tak kontekst umanjkal in informacija o napaki bi bila izgubljena; z oznako (zvezdico) pa ni; oznaka raziskovalca usmerja nazaj v kontekst.

kognitivna (pa tudi besedna, besednozvezna r.	a.3	nezanesljivo razumevanje ene ali več besed	() , (besede)
() šlo da *nis'n 'mel kaj delat'. (je b'la) cel *keden	megla *e *heheh ... () nakazano *jel'? () (=003Mz_ime) () deseti?	ja. si pa točno na tisti dan naredil.	

Med transkribiranjem sem posamezne odseke poslušal tudi po večkrat, da sem jih lahko natančno razumel in zapisal. Nerazumljive dele sem izpustil (brez oznake ali z oznako <()>, če je iz iste povedi vsaj še ena razumljiva beseda), nejasno razumljive pa posebej označil z oklepajem. Posebnih funkcij za odstranjevanje šumov iz posnetkov nisem uporabljal.¹²⁵

Zapisati čim več, četudi je govor slabo razumljiv, je smiselno zato, da zajamemo tudi govor v slabo slišnih situacijah, ob hrupu, govoru (pre)več govorcev hkrati, saj se le v takšnih okoliščinah govor tako oblikuje, kot običajno ne (to mnenje posebej izpostavlja Sinclair 1995: 101–102). Je pa v nekaterih situacijah razločevanje govora in govorcev res problematično: najtežji primeri so javni prostori, kjer se giblje in pogovarja množica ljudi: skoncentrirani pešci na hodnikih, prometnih postajah, v restavracijah, predavalnicah med odmori ipd. – sliši se samo žlobudranje. Govora mimoidočih pešcev nisem zapisoval, če ni razločna vsaj ena poved

¹²⁵ V govornih tehnologijah pa take tehnike uporabljajo (Vlaj idr. 2000.).

mimoidočega pešca; podobno pri pogovorih sopotnikov na vlaku, kjer razločnost močno moti hrup vlaka med vožnjo.

Torej, ni vedno za slabo razumevanje govora kriv dejavnik, ki vpliva na samo aktualno sporazumevanje, ampak tudi sam slab posnetek – premajhna ločljivost snemalnika, mikrofona, zvočnega kodiranja datoteke itd. Poleg tega mora zapisovalec posebno paziti na ponavljanja, napačne začetke, mašila ipd., saj jih kot poslušalci običajno naravno preslišimo, ker za nas nimajo pomenske vrednosti. In še eno „poslušalno“ navado sem ob transkribiranju odkril pri sebi kot poslušalcu: manj kot je neka beseda razločno slišljiva, večja je verjetnost, da bom njeno fonografsko podoba prepoznal kot popolno (npr. „slišiš“ nereducirane vokale, kjer so reducirani) – kot da bi jo nadomestil z idealno, priučeno podobo iz svojega abstraktnega spominskega leksikona.

7.1.4. *Neproblematičen zapis*

V tej skupini so tiste vrste zapisa, kjer med aktualnim izgovorom v gradivu in možnostjo zapisa s slovenskim standardnim pravopisom ni razlik v fonografski izrazni strukturiranosti – ni dilem glede tega, s katerimi črkami naj zapišemo govor. S tega vidika bi lahko govorili o osrednjem izraznem sistemu oz. o prepletu pisnega in govornega izraznega sistema na pomenskorazločevalni ravnini ene same slovnice (prim. poglavje I.2.1.2).

grafofonemska (ortografska v ožjem smislu) r.	c.5	zapis standardnih fonemov	25 malih črk gajice/slovenice
aha. bom jaz vzel. ja. si pa točno na tisti dan naredil.			

Pa vendar – pisni izrazni sistem pri nekaterih vrstah zapisa vseeno dodaja informacije, ki jih v govoru ni. Primera sta vezaj (glej tabela, vrsta zapisa d.14) in že omenjena velika začetnica (f.16), ki ju tu ne navajamo ponovno.

grafofonemska (ortografska v ožjem smislu) r.	c.7	zapis tujega izgovora s tujo ortografijo	ortografija tujega jezika (če jo zapisovalec pozna) in <*>
*Sacrum_*romanum_*imperium_*nationis_*germaniae. *pol pa tam *ko mu daš *print greš na *properties ... ja. ni tak' ko' računalnik *ko bi stisnil *back pa bi te vrglo tja nazaj. *bek ja. ja ka' pa pomeni *wellness? tam je *velnes center.			

Vrsta zapisa iz zgornjih primerov sledi načelu, da kadar so besede tujega jezika izgovorjene z njegovim lastnim fonološkim sistemom, je za zapis uporabljena ortografija tega jezika (če pa s slovenskim fonološkim sistemom, pa je zapis podomačen – primeri za to so nižje spodaj, vrsta c.6). Daljše pasaje (daljše povedi in nadpovedni sklopi) v tujem jeziku niso zapisane. Oznake so takšne, da ima vsaka beseda zvezdico ne glede na vrsto ortografije, na traku [pv] pa je še oznaka tuje. Latinske, angleške, italijanske in nemške besede prepisujem v njihovi ortografiji, ker jo obvladam. Izpuste transkribiranja govora drugih tujih jezikov zaradi nepoznavanja njihovih ortografija oz. fonoloških sistemov je tudi mogoče zabeležiti na trak [pv], vendar takšnih primerov v transkribiranem korpusu ni.

Kar je zares težko, je pri posameznih besedah razločiti, kdaj je kaka beseda zares prevzeta s podomačenjem in kdaj ne – z vidika grafofonemskega zapisa gre predvsem za prepoznavanje govorne podomačenosti, ki je pravzaprav odvisna od aktualne fonetične podobe v aktualnem sporazumevanju; dileme so tudi kaki srbohrvatski ć, đ ipd.). Tudi če je kdaj kaka daljša besedna zveza v tujem jeziku izgovorjena bolj s slovenskim kot tujim fonološkim sistemom oz. fonetiko, je prav iz tega tujega sosledja razvidno, da gre za tuj jezik – in prepis bo tudi v tuji ortografiji. Zadnja dva primera iz gradiva (*back/*bek, 1ned04Vg, in *wellness/*velnes, 6pet74Vg) sta redka primera, ko se v govoru blizu skupaj pojavita po jeziku različna fonološka izgovora.

Dilema je tudi zapis dobro poznanih imen, besed in besednih v slovenščini, ki smo jih navajeni pisati nepodomačeno v ortografiji – s tujo ortografijo, primer: v *Studiju_*City v ponedeljek zvečer (). (1ned64Eg). Kakorkoli že, besede imajo zvezdico.

7.1.5. Problematični zapisi: jasna povezljivost s standardno pisno obliko

Sistem standardnih pisnih oblik besed včasih pozna **podobne možnosti zapisa** (lahko bi skovali termin „blizuizraznice“), kar očitno izhaja prav iz variantnosti v govoru. Prim. iz SSKJ par živijo/živio in druge, navedene v nižje komentarju pod primeroma č.10 in č.11. Tako smo tudi pri transkribciji pogostih (po)govornih besed, ki so blizu standardnemu zapisu, v skušnjavi, da bi njihov zapis kar uvedli kot standarden (brez zvezdice ali opuščaja), npr. pri živ'jo | *pol 'potem, kasnejē | *te 'potem; no' | *ko 'kjer' | *ahm | *e | tak' 'tako'. Vendar bi bilo to dejanje, ki ne sodi v fazo neposrednega transkribiranja govora. Če smo s posebnimi ločili te zapise označili, smo ohranili informacijo, da v transkribciji kasneje *morda* uveljavljeni zapis (npr. živjo namesto živ'jo, agh namesto *agh) še ni bil standarden. Prav tak podatek je tudi lema, pripisana vsaki besedi v fazi lematizacije. Pregled pogostejših problematičnih zapisov besed je predmet zaključnega dela naloge, gl. IV.14.1, uveljavljanja standardnih zapisov za govorne posebnosti pa se ne lotevam in tudi ni v pristojnosti zapisovalca.

Transkribiranje tako poteka tako, da so (še)le v zaključni fazi raziskave diplome opravljene primerjave, ki spodbujajo k novim pogledom na govorni jezik in ves jezik kot notni sistem sploh; prim. v zaključnem delu naloge preglednice nedokončanih besed, redukcij, prevzetih besed, govornih vs. pisnih oblik, kratic, medmetov, leksikalnih nestandardnosti ipd. Spodnje tabele pa so manj obsežne, dajejo zgolj ilustrativne primere in pojasnjujejo dileme, ki se rojevajo ob problematičnih zapisih govora.

Dilema pa je pogosto že pri uvrstitvi predmeta problematičnosti zapisa na posamezno jezikovno ravnino. Spodnji primeri so vsi znotraj tradicionalne „morfološke“ jezikovne ravnine; skušajo biti razvrščeni podrobneje znotraj nje: od funkcioniranja posameznih fonemov do istih pojavov (razločevalnih enakovrednic, izpustov/redukcij, dodajanj) pri sklopih fonemov, do sklopov fonemov v morfemski funkciji in do sestavljanja besed (zapis izgovora kratic).

fonološka r.	b.4	„prekvalificiranje“ fonemov (drug fon v fonemski funkciji; gl. opomba ¹²⁰)	<*>
*zdej *dones *don's *sej *prov *kej *une *s'n		zdaj danes (2x) saj prav kaj ene sem	

Preden se odločimo za zapis „prekvalificiranega“ fonema, res dobro poslušamo in premislimo. Npr. če gre le za drugače „zategnjen“ vokal (malo širši ali ožji, bolj zaokrožen ali svetel od standardnega ipd.) in bi lahko rekli, da je na polovici v poziciji med dvema standardnima osnovnima alofonomoma v samoglasniškem izgovornem trikotniku, ga raje zapišemo s standardnim fonemom. Npr. pri nekaterih štajerskih/koroških narečnih končnicah: marsikje se je težko odločiti med zapisoma <-iga> in <-ega>; v večini nejasnih, mejnih primerov je bil uveljavljen zapis <-ega>. Ali pa razlika med <saj>, <*sej>, <sa'> in <*se'> za izgovorjave veznika saj.

Pogostejši mejni primer je reduciran in v zapori končnega nosnika spremenjen glagol biti v prvi osebi ednine: zapisa <*s'n> oz. <*sen> za standardni <sem>. Dilemi sta, ali je ta končni [-n] fonem ali zgolj alofon in ali reducirano zvočno središče besede, nosilca naglasa zapisati kot reduciran standardni vokal (z <-e->) ali kot zlogotvorni n (z <-n>). Pri ločevanju končnega (alo)fona v gradivu je skoraj gotovo [-n] pred sledečim [s-], [š-] in zgolj običajno na koncu povedi ali pred daljšim premorom; [-m] pa je skoraj gotovo pred sledečimi glasovi, izgovarjanimi z ustnicami [m-, b-, p-]. (Nekateri govorniki pa dosledno izgovarjajo [-m].) Glede zapisa zvočnosti omenjene besede reši dilemo primerjava s polglasnikom v standardnem izgovoru besede sen 'sanje; spanec': pri naši besedi je središčna zvočnost besede/zloga izrazito v sklopu nosnika [-n], lahko je tudi vlečena. Pri prej opisanem izgovoru iz gradiva je torej uveljavljen zapis [*s'n] po transkripcijski konvenciji. Pretvorbo izgovorjave /m/ v [n], a ne zlogotvorno, pa je mogoče opazovati tudi pri primeru os'najst 'osemnajst' ipd.

morfemska r. (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)	č.10	izpuščen fonem	<'>
dobr' mor'te bit' 'mel' ra' *sprob'al' z'aj vi'iš *nih ne 'om		dobri morate biti imeli raje *sprobali zdaj vidiš enih ne bom	

Kakor je razvidno iz zadnjih dveh primerov, opuščaj ne pomeni le popolne redukcije vokala, ampak tudi izpuščene soglasnike. Vsem transkripcijskim rabam opuščaja pa je lastno to, da zaznamujejo tak izpust fonema, pri katerem sosednji fonemi besede jasno obstanejo kot v primerljivi standardni pisni besedi.

Dileme v zvez z izpusti pri soglasnikih so samo v nekaterih primerih, kjer je opuščaj samo modifikacija običajnega zapisa fonema z dvočrkjem: če je <-lj-> zapis za mehčani/jotirani /l/, potem je <-l'-> na mestu te prej omenjene standardne variante/okolja fonema zapis za nemehčani/nejotirani/ /l/; npr. pr'pel'ala 'pripeljala'.

Največ dilem glede uporabe opuščaja pa je vendarle v primerih redukcij vokalov. Merilo za presojanje redukcije vokalov je število zlogov in le pri zmanjšanem številu zlogov bo uporabljen opuščaj na mestu enega od vokalov/fonemov po primerljivi standardni pisni besedi. Delne redukcije iz zapisa niso razvidne; izjemoma le, če gre redukcija v takšno smer, da je zapisana s črko za „prekvalificirani“ fonem, kot je pojasnjeno višje.

Problem je tudi označiti redukcijo tako ali drugače zapisanega polglasnika¹²⁶ ali „prekvalifikacijo“ drugega vokala v polglasnik. Najpogostejša primera sta predpona {pri-} in zaimek/veznik kar. Prekvalifikacija /i/ v polglasnik prvem primeru je zapisana <pr'->, delna redukcija /a/ v polglasnik v drugem primeru pa ne more biti razvidna iz zapisa <kar> (<k'r> bi bil nepravilen zapis po konvenciji), ugotovimo jo lahko le iz poslušanja posnetka.

Včasih nas lahko zmedejo primeri iz, ko so uveljavljeni v pravopisnem slovarju knjižnega jezika „bližnji pari“ (podobni primeri, ki se v relevantnih primerih razlikujejo samo v enem reduciranem vokalu) brez posebne sistematizacije: ko/kot, prek/preko, skozi/skoz, zgledati/izgledati, kratki nedoločnik ipd.

Nekatere redukcije in izpuste je prav težko opaziti in zapisati v skladu z načeli konvencije: tule → t'le, tulele → t'lele (napačna zapisa bi bila <*tle> in <*tlele>; iztočnici <tle> in <tlele> ima celo SSKJ!). Drugi primeri: ko → k', kaj/kar → ka' (ta imata še druge podobne variante: *k, *ko; glej proti koncu naslednjega podpoglavja in opombo¹⁵⁸), tako → tak' idr.

5) morfemska r. (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)	č.9	dodaten fonem	<*>
*jest vi'iš tu nam je *črez udar'lo ko je tu *kapal ne ... Proletarskih je prva *pol *mor'mo *jit' pa po *keri?			jaz čez iti

Ti primeri so redkejši, v glavnem gre za soglasnike (t. i. protetične idr.) in njihova razločitev ni težka.

Pa potegnimo zadnje tri vrste zapisov skupaj („prekvalificiranje“, izpust in dodatek fonemov; glej še spodaj primere za iste dogodke s sklopi fonemov). Dilemi sta dve: prva, ali je sploh pravilno uporabiti fonetični ortografski zapis za nestandardno fonemsko varianto; in druga, kje postaviti meje pri variacijah alofonov fonema (v najširšem smislu, npr. več stopenj redukcije vokalov, različno (so)delovanje govoril med artikulacijo). Ob primeru različnih izvedb leksema 'domá': v prvi dilemi, ali zapisati [da:ma] z <doma> ali z <*dama>; v drugi dilemi, katere od primerov iz besedila [da:ma], [d̥a:ma], [d̥o'ma:] zapisati z <doma>, katere pa z <*dama>.

Zakaj bi bil fonetični ortografski zapis nepravilen za nestandardno fonemsko varianto? Ker glas še vedno funkcionira kot fonem v fonemskem sistemu svoje (recimo narečne) variante jezika, čeprav bi sodil pod drug fonem v drugem fonemskem sistemu (recimo standardnem – zbornem). V določeni govorni situaciji, v kateri nastane sporočilo – besedilo, vedno uporabljamo za sporočanje le en fonemski sistem, četudi uporabniki obvladajo več fonemskih sistemov

¹²⁶ Brez znak ob r/l/m/n ali z <e>; nekdanji polglasnik v narečjih, ki so dajala besedje slovenskemu pisnemu jeziku, pa še z a/i.

(diglosija/dvojezičje). Znotraj omejene količine sporočanja (gledano z vidika linearne časovnosti govornega sporočanja) ne pride do mešanja enega in drugega fonemskega sistema oz. do različnih rešitev istih pomenskorazločevalnih nalog – da bi npr. v eni povedi v istem pomenu imeli uporabljeni govorni besedi [ˈda:ma] in [dɔˈma:].¹²⁷ Pa vendar smo v jezikovnih zbirkah, kot so govorni korpusi, k temu prisiljeni, saj soočamo mnoge fonemske sisteme (pravimo: diasisteme slovenskega jezika, ne da bi posebej reflektirali druge vidike kot sinhronega, statičnega). Soočeni so v eni standardni pisavi, ki mora biti neobhodno fonetična (fon(em)ografska), saj lahko le tako prikažemo tudi razlike v oblikoglasju, kor jih ponuja zbrano govorno gradivo. Primerjava bi bila mnogo natančnejša ob uporabi specializiranega fonetičnega črkopisa (npr. IPA), a ima tudi ortografija kljub svoji omejenosti (na 25 črk in v tej transkripcijski konvenciji izbrana ločila) svoje prednosti – prim. poglavje I.2.2.3.

S tem pa je napol odgovorjeno tudi že na drugo dilemo. Meja zapisa bo postavljena od primera do primera, od artikulacije do artikulacije, in vedno znova primerjana z:

– naborom alofonov standardne črke (za fonem) v isti pomenskorazločevalni vlogi (v isti besedi),

– najbolj podobnim osnovnim alofonom (izmed osnovnih alofonov/glasovnih vrednosti standardnih črk za foneme)

– ter z možnostjo, da se ji prisodi vrednost popolne redukcije/izpusta glasu. Najprej omenjena možnost je kombinatorično bolj verjetna od druge, odločitev za tretjo pa je lažja od razločanja prvih dveh. Sistematskih dilem pri ortografskem naboru črk je mnogo manj kot pri specializiranem fonetičnem, tako da naloga v praksi ni tako težavna, kot se zdi tukaj opisana.

Prim. SP²⁰⁰¹: § 215–225 (Načela za obravnavo pisav pri prevzemanju besed).

morfemska r. (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)	č.8	„prekvalificiranje“ vsaj dveh sosednjih fonemov (gl. opomba ¹²⁰)	<*>
*pr'šu *neki		prišel, nekaj	

Pri „prekvalificiranju“ vsaj dveh sosednjih fonemov gre najpogosteje za redukcijo v izglasju in glasovno zlivanje končajev. Le-to je do neke stopnje predvideno tudi pri knjižnem izgovoru, predvsem prim. deležniki na -l, šibkost zadnjih nenaglašanih zlogov. Zato je pri transkripciji z zvezdico in drugačno fonemsko realizacijo zapisanih karseda malo „prekvalificiranih“ (reduciranih, v manj glasov zlitih) končajev, večina je zapisana kar s standardnimi končaji. Nestandardni končaji v primerih kot *pr'šu (ˈprišel), *neki (ˈnekaj) morajo biti res izrazito reducirani (kratki, brez dvoglasnosti) in izrazito podobni posameznim alofonom drugih fonemov (v konkretnih primerih /u/ in /i/).

morfemska r. (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)	č.11	opušcanje vsaj dveh sosednjih fonemov, tudi skrčitev večih besed v eno	<*>
*nam *čaki *spodi *nazodno		ne bom čakaj spodaj nazadnje	

Te izpuste je lažje transkribirati in označiti kot izpuste posameznih fonemov, ker so obsežnejši in opaznejši, saj vsebujejo različne glasove (samo- in soglasnike). Vendarle imata tudi SSKJ in SP²⁰⁰¹ kar nekaj primerov „bližnjih parov“, od katerih je eden z izpuščenim sklopom fonemov – taki primeri so v spodnji preglednici brez zvezdic. Vsi primeri okrajšanih različic besed (v preglednici desno od puščice) so vzeti iz gradiva korpusa. Pri nekaterih gre hkrati še za „prekvalificiranje“ fonemov.

¹²⁷ Po tem načelu nezdržljivosti fonemskih sistemov bi torej veljalo, da lahko ves govor zapišemo z nemodificirano ortografsko transkripcijo – da vso fonetično variabilnost *prevedemo* v fonografsko najbližji standardni zapis, brez označevanja redukcij, daljšanj osnov, dodatnih zapor ipd. pojavov. V tem primeru bi bila še vedno uporabna (informativna) posebna ločila, ki bi označevala, da je fonematika aktualne besede drugačna kot v zapisu; npr. zapis <*zdaj> za aktualni izgovor [ˈzdej]. Vendar je to diametralno nasprotni princip uporabe izrednih ločil kot v pričujoči transkripcijski konvenciji.

<i>okrajšani začetki besed</i>	<i>okrajšane sredine besed</i>	<i>okrajšani konci besed</i>
bomo → *mo danes → *nes ne bomo → ne *mo seveda → *veda	drugega → drug'ga → *druga jabolko → *jabko katera, kateriga, katerih → *kera, *keriga, *kerih koliko → *kek narediti → *naret' nobeden → *noben počakajte, počakaj → *počajte, *počaj pojdite → *pojte vidiš → vi'š → *viš	čakaj → *čak noter → not skoraj → *skoro, *skor skupaj → skup spodaj → spod tak(o) → *ta (redok primer)

Čeprav bi lahko te redukcije tudi zapisali z opuščajem (enim, torej <zmer> (zmeraj); dva opuščaja za dva izpuščena fonema ali tri za tri itd. bi učinkovali nepregledno), jih raje zapišemo z zvezdico (torej <*zmer>). S tem poudarimo, da pojav presega samo pomenskorazločevalno ravnino in se približuje morfemski („pomenskogradbeni“) ravnini. Pri zgornjih primerih še ne gre sistemske oblikospreminjevalne (paradigmatske) ali besedotvorne spremembe na morfemski ravnini, pri spodnjih primerih pa gre za take.

besedotvorna r. (morfološka – oblikotvorna)	d.13	drug fonem ali sklop fonemov v morfemski funkciji	<*>
*bojo *sma *drugiga *dreve *zunam *vjutro		bodo sva drugega drevesa zunaj zjutraj	

Na prvi pogled fonetična sprememba glede na pisno obliko (pojavnico), dejansko pa se problematični fonem pojavlja še kje v paradigmi leksema (sem, *sma, smo; *bojo, *vejo 'vedo', *jejo 'jedo', *grejo 'gredo') ali na istih paradigemskih mestih drugih leksemov (*tokiga, *dva 'ndvajs'tiga, *totiga, *varhniga 'vrhnjega' ...) ali izpust daljšane osnove (*dreve) ali kot poseben del besedotvorne podstave (*zunam, *vjutro).

besedotvorna r. (morfološka – oblikotvorna)	d.15	izgovor kratice (logogram za besedno zvezo v obliki besede)	<*> in velike črke za foneme, ki označujejo okrajšane besede
*UeSBe-priključek *CeDe-jev *eMPe_tri-je *Ka *Ka je cesarsko-kraljevi v avstrijski polovici *K *KR je v ogrski je kraljevi *Ka *und *Ka cesarski in kraljevi. mi 'mamo pa *e strica pa teto Jožico pa *Božeja ne pa 'maš tisto *O *Ž pa to ne ...			USB-priključek CD-jev mp3-je

Rešitev zapisa je samosvoja, vendar je so nje razvidne začetnice (ali drug del) besed, katerih anagram ali krajšava je kratica kratic – velike črke, in dodatni fonemi (samoglasniki) s pomočjo katerih je izgovorjena kratica. Zvezdica pride bolj prav pri razločevanju posameznih črk, vezaj pa je samo pri pregibanih kraticah običajno po pravopisu.

7.1.6. Problematični zapisi: ohlapna povezljivost s standardno pisno obliko

grafonemska r. (ortografska v ožjem smislu)	c.6	zapis podomačenega izgovora tuje/prevzete besede	<*>
taka vrečka je *e od m- *Milerja. *fudzbal *fajn *čao		Müllerja fooball fine ciao	

Primerjava s situacijami, ko tuje besede zapišemo s tujo ortografijo, načelo ločevanja in nadaljnji zanimivi primeri so v poglavju II.7.1.4.

Govorjene prevzete besede se zdi smiselno prepisovati fonetično ne le zaradi cilja razločevati fonetični sistem izgovora in oblikoglasnost govora, ampak tudi zato, da se izognemo spremenljivosti njihovih pisnih oblik. Pisne oblike prevzetih besed se namreč pogosto menjujejo hitreje kot jezikovni priročniki, ki se posvečajo pisnemu jeziku. Fonetični zapis (pa čeprav z ortografskim črkopisom) pa daje ustrežnejšo zgodovinsko obliko prevzete besede. Važna pa je prav zvezdica, ki opozarja na to, da bi pravopisna oblika besede lahko bila drugačna. Prim. še SP²⁰⁰¹: § 215–225 (Načela za obravnavo pisav pri prevzemanju besed).

morfemska r. (v ožjem smislu; sklopi fonemov, ne še nujno v funkciji morfema)	č.12	nestalne oblike medmetov (njihovo daljšanje in ponavljanje; bližanje zapisa dejanskemu izgovoru)	<*>
*e, *ee (polglasnik-mašilo) *agh *ova *mmhm *mm *hehe, *hehehehe *jah *ou *ss *tc			

Ali so ti govorni pojavi res bolj oblikoglasno nestanovitni kot druge besede in je to vzrok njihove nemorfemiziranosti, neverbaliziranosti? Saj nam primeri iz drugih tabel kažejo, da tudi besede (besedne oblike) variirajo v svojih izgovornih variantah. Nekaj sorodnih pojavov ni zaslužio zvezdice, ker so v pisnem jeziku dokaj pogosti in očitno kdaj tudi razlog zavestne, priučene jezikovne rabe (iz pisnega prenosnika preneseni v govornega): ah | aha | aja | au | eh | ej | hm | jah | jo | joj | ma | mhm | nak | no | m'rš (marš) | oh | ojej | uh | evo itd. Praksa iz literarnih besedil je taka, da se skuša različne vzklike, mašila kar najnatančneje ubesediti, ker imajo pomemben literarni učinek, zato npr. v seznamu medmetov BSJ najdemo tudi bolj „eksotična“ mašila oz. medmete. Npr. vlečenje vokalov, posebno težko je zapisati tleske. Če na take redkeše realizacije naletimo v transkripciji, se nanje vendarle zdi vredno opozoriti z zvezdico, tako kot tiste brez zvezdice pa bodo vse razvidne tudi iz besednovrstne oznake (za) medmet.

Pri sledenju fonetični obliki se vendarle držim naslednjih omejitev. Pri vlečenju glasov so možne tri stopnje zapisa, tj. z eno (trajanje 1–2 sekundi), dvema (2–4 sek.) ali tremi črkami (nad 4 sek.) za glas. Tretja stopnja je zgolj predvidena kot rezervna možnost, zelo redka. Razni vzdih, neslišen smeh ipd. niso zapisani v govorno vrstico, če niso zvočno izraziti; v takem primeru so zapisani samo v paraverbalno vrstico govorca v partituri (gl. poglavje II.7.3). Načeloma se lahko z dvema ali tremi črkami zapiše tudi kak podaljšan glas sicer verbalizirane besede.

Pri ponavljanju zložitnih neverbalnih govornih pojavov (običajno smeh) se zapiše toliko znakov, kot je črkam najbližjih glasov. Dilema pisati te zloge skupaj ali narazen (npr. s podčrtajem <*he_he> ali vsak zlog posebej s svojo zvezdico <*he *he>) pa je rešena na prvi način – zloge pišemo skupaj – torej samo <*hehe>!

Potrebno je tudi da med transkribiranjem stanovitno razlikujemo podobne si besede: *e/*ahm (za polglasniško mašilo), *mhm (pritrditev)/*nahm (odklon). Pri zadnjem paru bi pravo razliko pomagal odkriti šele zapis intonacije besed (pri pritrditvi rastoč, pri odklonu padajoč), saj je sama artikulacijska razlika med obema začetnima nosnikoma zelo majhna, artikulacija nadaljevanja pa je tako skoraj enaka.

leksikalna r. (pomenska)	f.20	neknjižni/nestandardni pomen besede	<*>
*mogo		'moral' (morati)	

Gre za redke besede, pri katerih je v aktualni rabi samosvoj pomen, ki se drži vseh pregibnih oblik leksema.

leksikalna r.	neknjižni/nestandardni leksemi – besede	<*>
*sprob'al *komot *pol *kva *glih, *lih *ful *itak *totega *te *siguren *kao *cuker *ne'e	'preizkusil' 'udobno' 'potem' 'kaj' 'ravno(kar)' 'zelo' 'tako ali tako' 'tega' 'potem' 'prepričan' 'menda' 'sladkor' 'ne'	

Pri teh primerih zvezdica zgolj označuje to, da zapisa leksema ni bilo najti v konzultiranem jezikovnem priročniku (SSKJ). Ne gre za normiranje leksema ali kakršnokoli zvrstno umeščanje (kot s slovarskimi kvalifikatorji), pač pa za opozorilo, da je v gradivu uporabljen zapis samosvoj, nepreverjen v drugem jezikovnem gradivu/priročniku. Še dve skupini primerov, oblikovanih kot leme: prva so leme besed iz gradiva, ki so bile najdene v SSKJ (tam imajo zvečine kvalifikatorje „pog.“, pa tudi „nižje pog.“, „ekspr.“; druga skupina so leme besed, ki jih ni bilo najti v SSKJ. Niso pa to vsi mejni primeri iz gradiva.

baje cuker čik čohati deci fejs frajer fiksen hec, hecati, hecen itak kanta kelnar kila (kilogram) koštati lonati magari matrati mularija nalimati nobel nonstop not (noter) običen opasen paradajz penkalo piksna plac porajtati probati pukel rampa ratati; <i>vendar</i> *gratati serbus, servus siguren skup spod stoprocentno šansa šment šoder šparati špilati tle, tele vkup zmatrati zmerom zrihtati žleht župa
*bog, *boga, *bogo (<i>prisl. *bogo; pomen 'reven, ubog, usmiljenja vreden'</i>) *cajt *cimer 'sostanovalec' *cukrati *čao *dolgcajt *drekarija *faliti *fajn *fitnes *folk (ljudje) *fora 'šala' *frac *gužva *hiphop *kafe *laufati *lušten *majster *mobitel *nadrilati (<i>čprav SSKJ ima drilati</i>) *pocrkniti *porihtati *pucati *ru'zak, *rukzak *skapirati *smotan *spucati *sprobati *spumpati *štuklati *švoh *valda (<i>ali *val'da?</i>) *vun *ziher *zramati

Paziti je potrebno še pri nestandardnostih oz. govornih posebnosti, ki segajo preko meja besede (skladenjske nepravilnosti, pogovorne modifikacije frazemov itd.) – teh zvezdica ne označuje in transkripcija podatkov o njih ne daje.

7.2. Segmentacija – jezikovne enote tik nad besedno ravnino

Razpoznavamo različne dejavnike govora, ki jih z zapisom hkrati prestrukturiramo v vizualni/pisni prenosnik. Govor „forenzično“ ločujemo po govornih, razločamo nebesedno od tega, kar lahko zapišemo s črkami – kot besede. Ločujemo – segmentiramo pa tudi enote jezika med seboj, le da so elementi ločevanja v pisnem prenosniku (tako kot neverbalna fonetika v govornem prenosniku) deležni manjše pozornosti: najosnovnejše ločilo, presledek, je prazno, nekatere logične besedne zveze sploh niso posebej razločene, druga ločila večjih skladenjskih enot kot stavek, poved so med redkejšimi znaki zapisa.

Problema razločevanja segmentov jezika pri zapisu govora sta dva: kje postaviti meje segmentov in kako jih zapisati. Ena skrajnost je segmentiranje govora na več definiranih fonetičnih ravneh, na vsaki s svojimi oznakami (npr. posebej za kadenco, posebej za pavze, posebej za manjše intonacijske enote, za naglase, dolžine glasov, za hitrost itd; ali oznake za absolutne vrednosti intonacije itd.). Druga skrajnost pa je preprost nabor pisnih znamenj, ki beležijo le tista fonetičnih znamenja (ali sklope znamenj), ki ustrezajo pomensko najverjetnejši interpretaciji govora, jo okrepljujejo. Od prve k drugi možnosti nas vodi tradicija pravopisa in dejstvo, da se paraverbalna govorna znamenja pri ločevanju pomena sporočila med seboj dopolnjujejo, ker nimajo tako jasne pomenoskorazločevalne moči kot fonemi.

Slabost druge možnosti pa je, da ne ponuja enoznačne pretvorbe v obeh smereh. Če bi nas na podlagi pisnega znamenja po drugi skrajnosti zanimalo, katera stavčnofonetična znamenja zamejujejo segment govora na označenem mestu, tega ne moremo jasno dekodirati. Sploh pa nas lahko pri ločilih, če jih vzamemo po pravopisu, zmedejo še njihove potencialne neskladenjske vloge.

Kakršnakoli znamenja za neverbalno (stavčno) fonetiko je pri potrebah transkripcij torej potrebno dobro utemeljiti. Nedvoumno nam služijo samo za prepoznavanje segmentov govora (gl. naslednja tri podpoglavja), za odkrivanje stavčnofonetičnih vzrokov, ki dejansko stojijo zadaj, pa je potrebno v vsakem posameznem primeru poslušanje vzporednega posnetka.

7.2.1. Izjave in premolki med njimi

V pričujočem korpusu besedil je govor posameznika, ki ga predstavlja zapis na verbalnem traku partiture, v osnovi sestavljen iz izjav in (daljših) premolkov, ki se izmenjujejo. Vsaka **izjava** je definirana kot relativno neprekinjen sklop govora, zapisanega z besedami. In izjave in premolki so na traku razločeni vizualno kot interval traku (celica v vrstici tabele), meje intervalov pa imajo zabeleženo vrednost na časovni osi; meja med eno in drugo enoto ima eno samo časovno vrednost. **Premolki** med izjavami niso absolutno časovno definirani; trajajo pa približno vsaj 1–3

sekunde. Premolka vedno sestavlja en sam interval traku (ena celica), izjavo pa lahko tudi več zaporednih intervalov traku (več zaporednih celic z govorom – govorni dogodki), med katerimi ni intervalov premolka. S tako razporeditvijo lahko dosežemo natančnejše časovne umestitve govornih dogodkov.

Naslednja nižja stopnja jezikovnih segmentov so povedi; izjavo sestavlja vsaj ena poved.

Do krajših premolkov oz. prekinitev izjav in njihovega nadaljevanja (ki je pojmovano kot nova izjava) prihaja pogosto ob prekrivnem govoru in izmenjavah govorne vloge med govorcema. Seveda je, časovno gledano, možnih več hkratnih izjav pri prav tolikih govorcih. Tudi posamezni t. i. oporni signali so lahko pojmovani kot samostojne izjave, če tvorijo samostojno poved med premolki.

Primerjava s tujo terminologijo (v programu EXMARaLDA) je takšna:

- **dogodek** (časovno umeščen govorni dogodek ali premolka; vizualno kot **celica**, najmanjši interval traku) = *event* = *Ereignis*;
- **govorni dogodek** tukaj pomeni govor, vizualno omejen na celici traku, časovno kot interval traku, sestavljajo pa ga samo cele povedi, vsaj ena (v programu Elan = *annotation*);
- **segmentacijska veriga** = (časovno neprekinjeno zaporedje dogodkov – celic) = segment chain = *Segmentkette*;
- **izjava** (segmentacijska veriga govornih dogodkov; neprekinjen sklop govora med premolki) = *speaker contribution*;
- **poved** = *utterance* = *Äußerung*. (Schmidt 2006: strani H-1 do H-3).

Slika 10: Nadbesedna segmentacija v transkripciji

The screenshot shows the EXMARaLDA Partitur-Editor interface. The main window displays a transcription with time markers and phonetic symbols. Red annotations are overlaid on the transcription, identifying specific segments: 'GOVORNI DOGODEK' (speech event), 'PRAZNI DOGODEK' (empty event), 'PREMOLK' (pause), 'IZJAVA' (utterance), and 'POVED' (utterance). The transcription includes the following text: 'jaz *nis'n ... pozabim čisto *s'n pozab'la.', 'za? za poroko? bo m- ... (=025Mz_ime) do naslednjč pr'nesel seznam *te bomo pa terjali.', 'za poročno darilo. pozab'la *s'n!', 'jaz *nis'n ... pozabim čisto *s'n pozab'la. n', and 'a si ti *naja dva ()? aja.' The interface also shows a menu bar with options like File, Edit, View, Tier, Event, Timeline, Format, Segmentation, and Help.

7.2.2. Meje povedi

Ker lahko tako govorni sklop kot izjavo sestavlja več povedi, je potrebna še dodatna označitev meja med povedmi. Začetki povedi niso posebej označeni (velike začetnice so rabljene zgolj pri lastnih imenih). Začetek izjave pomeni hkrati začetek povedi, če pa znotraj izjave eni povedi sledi druga, pomeni oznaka za konec prve povedi hkrati oznako za začetek druge. Velike začetnice za oznako začetka povedi namreč niso uporabljane.

Pri označevanju koncev povedi pa sem se odločil za provizorična, posebno rabljena končna ločila: pika, vprašaj, klicaj, tropičje. Ta ločila ločujem nedvoumno *med sabo* (samo) na podlagi **intonema**¹²⁸ povedi.

¹²⁸ Intonem – „intonacija v svoji pomenskorazločevalni vlogi:“ intonemi 1 (pripovedna), 2 (vprašalna) in 3 (vzključna, poudarjalna inotnacija). Toporišič 1992: ESJ: **intonem**: 62.

Kot končno ločilo pa lahko *vsako* od njih pomeni še katerega naslednjih **stavčnofonetičnih sredstev za konec povedi**: zaključek govora in premolk, predhodna upočasnitev hitrosti govora, skladijska zaokroženost ali nenadna prekinitev skladijske zveze brez nadaljevanja ali popravljanja, izrazita sprememba intonacije v katerokoli od smeri in prekinitev intonacijske verige, določena sprememba v jakosti, npr. stišanje glasu; pravzaprav *prekinitev katerega koli (daljšega) stavčnofonetičnega pojava*, vezanega na pomensko zaokrožen sklop govora.

Omenjeni stavčnofonetični pojavi pri dopolnjevanju verbaliziranega pomena delujejo v sklopih: jasne meje za konec povedi ni mogoče navezati le na enega od njih. Tudi poteku intonacije, čeprav je dosledneje prikazan v razlikovanju končnih ločil, pri razpoznavanju mej povedi ni dana prednostna vloga. Razlog je prav pomenska vrednost stavčnofonetičnih pojavov: **poved je po definiciji pomensko zaokrožena in samostojna enota sporočila** in jo lahko nasploh zaokrožujejo vsa stavčnofonetična sredstva, v konkretnem primeru pa le (ne)katera od njih.

Tako je pri transkribiranju potrebno paziti na nekatere „skladijske prelome“: prekinjene besede (označene z levostičnim vezajem, gl. spodaj), medmeti in nerazumljiv govor (oz. oznaka zanj, <()>) še ne pomenijo sami po sebi nujno tudi konca povedi. Če ga pomenijo, potrebujejo za seboj končno ločilo. Ali ga pomenijo ali ne, pa je razloženo iz poslušanja posnetka in reflektiranja verbalnih in paraverbalnih govornih sredstev, ki zaokrožajo pomen povedi, in primerjanja le-te enote z ostalim sporočilom oz. kontekstom.

Poleg končnih ločil sem izjemoma uporabil še eno skladijsko ločilo, ki ne pomeni konca povedi: dvopičje. Je redko, rabljeno nedvoumno samo na meji stavki, kjer gre za napoved premege govora (med spremnim stavkom in dobesednim navedkom); ker navednic med ločili v transkribiciji nisem uporabljal, lahko odkrivamo vsaj premi govor ob spremnih stavkih s pomočjo dvopičja. Stavčnofonetično se običajno pokriva s padajočo ali obviselo intonacijo. Ni pa dvopičje zapisano v primerih, ko govorec le napove posamezen pojem, ga izpostavi, torej ko ne gre za mejo med celimi stavki.

Dvopičje, pika, vprašaj in klicaj so levostični, tropičje pa nestično.

Tabela 14: Izbrana skladijska ločila v transkribijski konvenciji StAnId4

<i>znak</i>	<i>ime nabor: koda</i> ¹¹⁵	<i>pomen uporabe – morebitna dilema Končno ločilo označuje:</i>	<i>primeri iz transkribicije</i>
.	pika UTF-16=002E, ISO-8859-2 (10)= 46	pripovedno intonacijo povedi. Ne označuje kratic!	še tole ven vzamem. otroški sirup.
?	vprašaj UTF-16=003F, ISO-8859-2 (10) = 63	vprašalno intonacijo povedi. Vedno samo enkrat zaporedoma.	*kva naj *pol nared'm?
!	klicaj UTF-16=0021, ISO-8859-2 (10) = 33	vzklično oz. poudarjalno intonacijo povedi. Vedno samo enkrat zaporedoma.	pa nič ne poveš pa to!
...	tropičje UTF-16= 2026, ISO-8859-2 (10) = 133, Eva= 593; enotna koda, ne pa sestavljeno iz 3 kodnih znakov – 3 pik!	intonacijo, ki je po poteku nekje vmes med pripovedno in vprašalno: intonacija ni toliko jasna in tako nakazuje zavestno pomensko nedokončanost povedi in izpust, elipso. Ta nezaokroženost misli pogosto spodbudi po nekoliko premolka ali istega govorca k novi podobni misli ali (v dialogu) sogovorca k nadaljevanju nedokončane misli (in k novi povedi). govorimo o soledju, npr. med spremnim stavkom in dobesednim navedkom (tudi v govoru, v pisavi pa bi na takem mestu sicer stalo dvopičje).	ja zdaj 'mam en ... dober mesec se že *ne'e brijem ja ...
:	dvopičje UTF-16= 003A, ISO-8859-2 (10) = 58	Označuje mejo med spremnim stavkom in sledečim premim govorom. Samo na meji med stavki – ne označuje napovedi/izpostavitve pojmov znotraj stavka.	*te *s'n rekel: joj ne več kopat' da ne bo vam noter v glavni kanal udar'lo.

Tudi pri zapisovanju meja povedi so nastajale dileme. Pri vprašalnih členicah na koncu povedi vprašaj ni načeloma sledeče ločilo; le pri pomenski vprašalnosti z izrazito vprašalno kadenco je vprašaj; pri povednosti z vprašalno kadenco tropičje, brez vprašalne kadence pika. Primeri:

ja. saj je *gljih () mogoče ni b'lo stoprocentno ne? veš to so sam' govoril' ne? | a veš jaz *s'n si pa *pol dopovedovala: pa saj so javna predavanja pa saj se lahk' gre ne ... | samo s tem pomeni da bo zmerom v greznici neka' not ne.

Tudi pri napačnih začetkih se pojavlja dilema, ali je to ena poved skupaj s pravilnim nadaljevanjem ali gre za prekinjeno poved. Konec povedi je označen, če napačnemu začetku sledi daljši premor in je naslednja poved skladijsko in pomensko zaokrožena; seveda tudi, če je naslednja poved pomensko povsem samosvoja. Če pa premora ni in bi bilo nadaljevanje brez besed pred napačnim začetkom pomensko ali skladijsko samo zase pomanjkljivo, je napačni začetek z nadaljevanjem povezan v eno poved. Enako tudi, če ni napačni začetek nič drugega kot jecljanje oz. zarečena beseda. Primera:

seveda je pa tiste luknje *nis'n od- ... tam *nis'n *odkapal pa *nis'n dol zazidal. | nimaš kam drugam drekarije speljat' če ka- se kanalizacija pokvari ...

Pri natančnem opazovanju stavčnofonetičnih situacij, v katerih „po občutku“ zapišemo klicaj, lahko ugotovimo, da njegova raba nikakor ni odvisna le od poteka intonacije, kot smo zapisali zgoraj. Nekako ne pomeni gibanja intonacije gor ali dol (kot pika, tropičje in vprašaj), ampak nenadno, skokovito spremembo intonacije (ki je zaradi hitrosti manj opazna, bolj opazno je hkratno povečanje glasnosti). Če je omenjeni skok združen z vprašalno/rastočo intonacijo, bi bilo končno ločilo <!> povsem na mestu (samo klicaj bi nam potem pomenil skok ob povedni/padajoči intonaciji); zaradi poenostavljanja ločil pa ostanemo samo pri klicaju za vse „skoke“.

7.3. Zapis paraverbalnega

Pod paraverbalno zapisujemo vse za raziskovalca relevantne stavčnofonetične informacije o tekočem govoru, ki jih nismo mogli zapisati na verbalni trak. Zapisi na tem traku niso osnovni predmet raziskave v tej nalogi, zato jim je posvečene manj pozornosti, informacije so manj razdelane. Kako uredimo trak v partituri s programom EXMARaLDA Partitur Editor, je opisano v poglavju III.11., strukturiranost paraverbalnih podatkov glede na opise sporazumevalnih situacij pa je podana v tabeli spodaj.

Kljub obrobni v raziskavi so vpisi na trak [pv] klasificirani v kategorije, kar zagotavlja konsistentnost opisa govora in omogoča celo neko vrsto kvantificiranja. Oznake (spodaj v tabeli) so pravzaprav že kategorialne določitve, ki pa so vse v transkripciji dopolnjene z opisom, ki je nepredvidljiv.¹²⁹

Dodaten problem je trajanje, časovna razsežnost teh podatkov, ki je z vpisom v partiturni format avtomatsko prisojena, a skoraj nepotrebna. Pri podatkih iz vrstice [Ozadje] (naslednje podpoglavje) je važna le kontekstualna umestitev, pojmovani so kot *trenutne vrednosti*, relevanten je le začetek njihove umestitve na časovni poltrak, njihovo medsebojno prepletanje ni problematično. Za podatke iz traku [pv] pa je važna njihova povezanost z besedilom govorca, njihov začetek in konec glede na kontekst. Medtem ko so običajno le neverbalni govorni znaki trenutni, se lahko drugi podatki raztezajo celo čez več izjav govorca. Zato so razbiti na začetno in

¹²⁹ Ta zapis podatkov spominja na poznejši računalniški markup: paraverbalna vrstica in vrstica Ozadje sta *elementa* transkripcije, ki dobivata različne *attribute* – kategorije podatkov, vsak atribut ima *vrednost* – opis podatka; edino pri atributu medij je nabor vrednosti definiran oziroma omejen.

končno oznako po zgledu jezika markup, le začetni pa sledi tudi vrednost; ker ima vsak podatek svojo časovno umestitev v partituri, oklepaji oznak niso potrebni. Po zaključku podatka je smatrano, da so paraverbalne okoliščine navadne, kot pred začetkom oznake oz. kot je opisano v glavi posamezne transkripcije/besedila. Torej:

Tabela 15: Pregled oznak na traku [pv] v partituri

začetna oznaka: <i>primer</i> : končna oznaka	<i>opis vrednosti</i>	<i>primerljiv podatek okoliščine</i>	<i>mesto zapisa</i>
12p) nvg:	Neverbalni govorni znaki (ni časovno razsežno). Kratak opis govornega znaka ali šuma.	Access: 08 BESEDILA: 12) besede	partitura: trak [pv]
11p) izg: petje izg	Izgovorjava besed na izreden način glede na fonetiko okolja . Opis v sam. oz. glagolnikih (<i>smeh, petje, žvižganje, vzdih, zehanje, kašelj, pogrkovanje ...</i>).	Access: 08 BESEDILA: 11) povedi	partitura: trak [pv]
Dp) pren: branje pren	Prenosnik , ki je na novo uporabljen (ne gre za izvorno prosti govor). Kratak opis v obliki glagolnika (<i>branje, recitiranje</i>) in opisa vira (npr. <i>pren: branje s teleteksta</i>).	Access: 01 stopnje projekta: D) prenosnik Access: 08 BESEDILA: 30) prostogovornost	partitura: trak [pv]
Čp) tuje: latinščina tuje	Tuj jezik (za neslovenščino, ki velja sicer kot privzet jezik gradiva). Jezik označen kot samostalnik na -ina.	Access: 01 stopnje projekta: Č) jezik	partitura: trak [pv]

Pozor: v partituro so vpisani le taki „izustki“, ki so pomenljivi, so torej jezikovni (to mora biti razvidno iz njihovega konteksta na posnetku). Če torej v partituri na traku [pv] naletimo na opis <kašelj>, to pomeni, da govorec zakašljajal (se odkašljajal, izkašljajal ...) z nekim sporazumevalnim namenom – da bi pritegnil pozornost, prevzel govorno vlogo, poudaril nekaj v govoru, izrazil s tem medmetom nek (konotativni) pomen ...

Tudi obliko opisov posameznih pojavov sem skušal tipizirati; nekateri opisi se med oznakami ponavljajo. V spodnji preglednici so vsi opisi iz gradiva.

Tabela 16: Tipi opisov oznake nvg: s traku [pv] v partituri transkripcije

<i>izg/nvg:</i>	<i>samo nvg:</i>	<i>samo izg:</i>	<i>pren:</i>
smeh [smejanje] (pritajen) zehanje	kašelj [kašljanje, od-/iz-, kihanje] mrmranje (zanikanje) oponašanje ... (npr. konjskega hrzanja) pogrkovanje riganje tlesk (negativno) vzdih (npr. ob pretegovanju) vzklik žvižg [žvižganje, požvižgavanje] (polglasno)	pačenje petje (mrmranje, npr. lestvice) s polnimi usti šepet [šepetanje] zlogovanje zvišan/znižan glas	branje (npr. s teleteksta, iz časopisa) recitiranje (po spominu)
Pomen oklepajev v tabeli: [druge variante, zajete v prvi] (dopolnitve)			

Oznaka **nvg:** na traku [pv] označuje pomenljive neverbalne (brez zapisa v grafofonemsko podobo: žvižganje, kašelj; tihi smeh (pritajen, nezveneč), vzdih in zehanje; lahko tudi mrmranje, kadar je polglasno) in polverbalne govorne znake (gl. spodnjo tabelo: medmeti, mašila; z možnostjo zapisa v približno grafofonemsko podobo na traku [v]: glasen smeh, vzdih in zehanje, mrmranje z zanikanjem ali pritrditvijo). Gre za relativno neprekinjene zapise (hkrati na trakih [v] in [pv] ali samo na traku [pv]), ki se ne izmenjujejo z jasnim verbalnim govorom istega govorca (na traku [v]).

Oznaka **izg:** na traku [pv] pa označuje s paraverbalno izvedbo opremljen (obogaten/obremenjen) govor, ki na traki [v] sega čez več besed, te besede pa niso le približne grafofonemske ubeseditve polverbalnih jezikovnih znakov.

Tabela 17: Primeri gradiva za posamezne opise ob oznaki nvg:

<i>grafofonemski zapis [v]</i>	<i>opis v vrstici [pv]</i>	<i>izrazite variacije zapisa [v]</i>
*mhkrghm	nvg: pogrkovanje	
*hehe	nvg: smeh	*hehem *e *hehehe *ejo- *ho ... *ehehe *hahaha *hem
*mrkhm	nvg: kašelj	*khrhmhm *khrhmkhrhmhm *khrhmmhmh *akihaa
*uah	nvg: zehanje	*huehehehehueha *hua *hehehehehehehe *aah *ojej
*nahm.	nvg: mrmranje zanikanje	
*mmmha	nvg: vzdih pri pretegovanju	
*eijuvhohoa!	nvg: oponašanje konjskega hrzanja	
*bevkapelvepev	nvg: oponašanje lajanja - bevskanja	
*tc	nvg: tlesk negativno	
*tharam	nvg: vzdih	*hmh *agh *ova *aagh
*vuv	nvg: vzklik	
ko rano zjutraj se preb- ... ne vem! -buja prebujam se še jaz. *tatatatatatata *grhm *tatatatatahm *taum *tananata.	(Primer petja in zlogovane verbalizacije, ko zmanjka besedila; tu ne gre več za neverbalni govorni znak, marveč za medmete!)	

Če predstavljajo ti neverbalni znaki eno samo poved, ima ta lahko različne intonacije in tako tudi različna končna ločila, ki so v zgornjih primerih izvezeta.

7.4. Primeri zapisov govora iz drugih transkripcijskih konvencij

Razvoj načel lastne transkripcijske konvencije je potekal ob študiju zgledov konvencij korpusov s podobnimi raziskovalnimi cilji oz. referenčnih korpusov (gl. I.5.1). Domača zgleda sta bila Učni korpus govorne slovenščine (UKGS, Jana Zemljarič Miklavčič) in korpus telefonskih pogovorov v turizmu Turdis-1 (Darinka Verdonik), tuji pa British National Corpus (BNC) ter posamezen članek o korpusu COUBILD (Payne 1995), korpusu Bergen v okviru projekta korpusa COLT (Haslerud/Stenström 1995) ter površno priročnik transkripcijske konvencije HIAT za računalniško transkribiranje (Rehbein idr. 2004); za literaturo prvih treh gl. spodnjo tabelo. Verjetno najustreznejše izhodišče, priporočila raziskovalne skupine EAGLES za govorna besedila (EAGLES 1996), ki jih upoštevata oba domača zgleda, so bile reflektirana šele naknadno, dobro pomagalo pa je bil tudi nemški priročnik Linguistische Gesprächsanalyse: eine Einführung (Brinker/Sager 2001: 21–56). V nadaljevanju želim prikazati nekaj primerov skupnih značilnosti in razlik med transkripcijskimi praksami StAnId, UKGS, Turdis in BNC, celovitejša primerjava pa tu nima mesta.

Skupne značilnosti:

- vsi uporabljamo modificirano ortografsko transkripcijo; najbližje pravopisu je BNC, Turdis ima dopolnitev nestandardnih besed z oznako [+lex=*], ali fonetični prepis po SAMPA (in oznako [+pron=*] ali [+pron=i^{sg}]);
- med besedami so presledki, številke so zapisane z besedo (BNC ?), zapisi lastnih imen imajo veliko začetnico, začetka povedi ne označuje velika začetnica razen pri BNC;
- imena (podatki – naslov, telefon) govorcev so zakrita z nadomestno oznako, imena javno znanih oseb in krajev pa nespremenjena.

Razlike so prikazane v spodnji tabeli. Iz primerjave našega načina zapisa z ostalimi je razvidno, da sem se po ostalih precej zgledoval: izhodišče mi je bil BNC (s svojimi načeli nezamudne transkripcije, vmesnimi in končnimi ločili; tudi UKGS se precej zgleduje po njem), v bolj natančno razčlenitev fonetične in prozodične plati pa sta sta me nagnila slovenska zgleda. Povsem samosvoja pa je v našem zapisu priredba zapisa za partituro, razdelitev podatkov po trakovih za različne dejavnike sporazumevanja.

Tabela 18: Primerjava transkripcijskih rešitev za govor: StAnID, Turdis, UKGS, BNC

<i>StAnId</i>	<i>element transkripcije</i>	<i>Turdis</i>	<i>UKGS</i>	<i>BNC</i>
vodoravni zapis – partitura, trakovi [v], [pv], [Ozadje] – EXMARALDA, deloma ločila po HIAT	0) FORMAT ZAPISA, transkripcijska konvencija	navpični zapis, oznake v eni liniji – Transcriber, tudi oznake v oglatih oklepajih povzete po njem; zglede tudi po BNSi	navpični zapis, oznake v eni liniji – Transcriber, deloma partitura Praat, oznake v lomljenih oklepajih samostojno oblikovane po zgledu markupa oz. TEI in EAGLES 1996	navpični zapis
1) ČISTO BESEDILO – 1.a) BESEDNO				
*e *ee *agh *ova *mmbm	zapis medmetov, mašil ...	eee hm mbm aba ab eh eeb nnn	ø m om mbm ab ej	hm erm er
taka vrečka je *e od <u>m</u> -*Milerja.	nedokončana beseda	samo malo da <u>pa</u>	bese< z> oziroma bese-	bese=
() šlo da *nis'n 'mel kaj delat'. (je b'la cel *keden meglja *e *bebeb ...	slabo razpoznavni govor, nezanesljiva transkripcija	oznaka [+pron=pi] mislim ka(k) ...	<neraz>, <?>besedilo</?>, negotovo razpoznavanje govorca <??>besedilo</??>	<unclear>, za daljše od 5 sekund <unclear>(15), za nerazpoznanega govorca sigla <?> kot oznaka govorca, za zapisovalcu neznane besede <??>beseda</??>
*CeDe-jev *eMPe_tri-je	kratica	črkovano: ~GMX, izgovorjeno kot ena beseda: @FERI	<okr>beseda</okr>	za črkovanje velike črke s presledki: <i>That's spelt H I N X T O N</i> ; za črkovane kratice enako: <i>V A T</i> ; za izgovorjeno kot besedo pa bez presledkov: VAT
1.b) PARABESEDNO (fonetika, prozodija)				
–	pomožni fonetični zapis	s simboli SAMPA: tudi [t/ u:t] [+pron=*], kakšne [k/ a:SnE] [+pron=izg]	–	–
oznaka <u>izg</u> : (izgovor) + opis; za bran/recitiran ipd. tekst oznaka <u>pren</u> : (prenosnik) + opis : <i>branje, recitiranje</i> itd.; trak [pv]	prirejenost besednega izgovora	za bolj glasno oz poudarjeno: #do# širi ure; za podaljšane glasove [:] za črko za fonem	<singing>besedilo</singing>, <smeh>besedilo</smeh>, za poudarjeno <shift=poud>besedilo</>, za branje besedilo</br>	<shouting>besedilo </>, <whispering>, <laughing>, <yawning>, <singing>, <mimicking>
oznaka <u>mg</u> : (neverbalni govor) in opis; trak [pv]; pri vokaliziranih lahko zapis na trak [v]	neverbalni govorni znaki	smeh [rire], kašelj [tɕ], vdih [i], izdih [e]	<nv>laughing</nv>	<nv>cough</nv>, sneeze, laugh, yawn, whistling; ob dolžini prek 5 sekund trajanje v oklepaju: <nv>laugh</nv>(10)
oznake <u>šum</u> , <u>signal</u> : in kratek opis zvoka; trak [Ozadje]	zvočni signali, šum, hrup v ozadju	[indicati], za daljše funkcija <i>background</i>	(listanje papirja)	(eating biscuits)

<i>StAnId</i>	<i>element transkripcije</i>	<i>Turdis</i>	<i>UKGS</i>	<i>BNC</i>
poved in izjava osnovni enoti	2) NADBESEDNA SEGMENTACIJA 2.a) SKLADENJSKE ENOTE	osnovna enota izjava	osnovna enota izjava	osnovna enota poved
' - * = :	<i>uporabljena (obliko)skladenjska ločila</i>	–	–	, ' (opuščaj za posesivnost in okrajšave)
daljša pasaja niso zapisane, opomba na traku [Ozadje], krajše <*> za vsako besedo in oznaka <u>tuje</u> ; na traku [pv]	<i>govor v tujem jeziku</i>	[lang=Engl-]coffee break[-lang=Engl]	<tj: norv>besedilo</tj>	–
končna ločila . / ? ... na koncu povedi	<i>končna ločila, segmenti izjav</i>	... za nedokončane izjave, ? za rastočo intonacijo; govorna vloga lahko razdeljena še na segmente	za vprašalno intonacijo <shift=vpr>besedilo</>	. ! ?
podatek <u>vgenzdeno</u> in kratek opis vrste in števila umeščenih delov, ki bi lahko bili tudi samostojna besedila	<i>del besedila</i>	sekcije: uvod, jedro, zaključek	–	–
2.b) STAVČNA FONETIKA				
– (razen če se slučajno prekrivajo s koncem povedi)	<i>premolki znotraj govornih vlog</i>	[.]	<pavza>, za 5 sekund in daljše <pavza>(5)	... za krajše, za daljše pavze pa v sekundah: ... (10)
– (razviden iz partiture)	<i>prekrivni govor</i>	posebne oznake [1], [2], [P] na začetkih/koncih prekrivanja	[besedilo]	[besedilo dveh govorcev] {hkrati še besedilo tretjega}
	<i>krajši prekrivni govor, oporni signali</i>	za oporne signale: [+overlap_mhm]		
2.c) PRAGMATIČNO-DISKURZIVNE ENOTE				
–	<i>popravljanja</i>	[N-]/[+N]	–	–
–	<i>ponavljanja</i>	[+lex=ipon], [+lex=bipon]	<repet>besedilo</repet>	med ponovljenimi besedami je vejica kot pri naštevanju
gl. - za nedokončano besedo, končno ločilo ..., * pri negotovi verbalizaciji oz. nestandardnem zapisu	<i>drugo (zarečene besede, nedokončane izjave, oporni signali, mašila)</i>	[+pron=lapsus], diskurzivni označevalci [+lex=marker]	nestandardna besedna oblika <nst>beseda</nst>	nestandardne besede so preverjene v uveljavljenih slovarjih, umeščene v liste in ortografsko zapisane, variacije v izgovorjavi niso prikazane
	LITERATURA	Verdonik 2006, prim. tudi Žgank idr. 2004	Zemljarič Miklavčič/Stabej 2005, Zemljarič Miklavčič 2006	Crowdy 1995

Primerjava rešitev za zapis sporazumevalnih okoliščin oz. dejavnikov je v poglavju II.8.7.

8. ZAPIS SPORAZUMEVALNIH OKOLIŠČIN

V spodnjih podpoglavjih so podrobno predstavljena pravila za zapis oziroma opis sporazumevalnih okoliščin (za teoretični uvod o teh gl. poglavje I.3), pa še nekaj o obdelavi in metaoznakah samih transkripcij in besedil. Za kakšne podatke gre, je razvidno iz naslova podpoglavij in nadaljnjih naslovov tabel. V tabelah so nato v levem stolpcu imena podatkov kot oznak okoliščin, v naslednjem stolpcu proti desni pa način opisa posameznega podatka. V skrajnem desnem stolpcu je nato še podano mesto zapisa podatka v datotečni strukturi korpusa; večinoma gre za razne tabele iz zbirke podatkov programa Access (izsek iz najobširnejše tabele je prikazan na str. 116, Slika 11).

Kako poteka postopek vpisovanja med samo obdelavo gradiva, gl. poglavje III.12.

8.1. Tekoče spreminjanje okoliščin

Kot pove že naslov, so na traku [Ozadje], enem samem skupnem za partituro posamezne transkripcije/besedila, opisane tekoče okoliščine sporazumevanja. Okoliščine so sicer označene/opisane med metaoznakami besedila (statično, glej naslednja podpoglavja), a ti opisi vendarle ne zadostujejo. Važni so opisi tekočih sprememb okoliščin tudi zaradi določanja meja besedil pred fazo razreza transkripcij na besedila; razlogi za razrez pa so vendarle bolj razvidni iz baze metaoznak besedil kot iz traku [Ozadje].

Kako uredimo trak v partituri s programom EXMARaLDA Partitur Editor, je opisano v uvodu poglavja III.11. Zapisi na tem traku niso osnovni predmet raziskave v tej nalogi, zato jim je posvečene manj pozornosti kot recimo traku [v], informacije so manj razdelane. Opis okoliščin vendarle poskuša biti tipiziran s pomočjo oznak (gl. spodnjo tabelo), ki so dopolnjene z opisi v kratki povedi.

Tabela 19: Pregled oznak in primeri opisov na traku [Ozadje] v partituri transkripcije

oznaka: primer iz gradiva	<i>opis vrednosti</i>	<i>primerljiv podatek okoliščine</i>	<i>mesto zapisa</i>
00o) razno: <i>Govorki na zadnjem sedežu se kdaj pa kdaj tudi tibo, nerazumljivo pomenkujeta.</i>	O važnih trenutnih okoliščinah, dejanjih, ki niso zajeti v drugih kategorijah podatkov.	Access: 08 BESEDILA: 00 ostalo	partitura: trak [Ozadje]
10o) signal: <i>Zvonjenje mobilna - dohodni klic.</i>	Negovorni zvočni znak, a pomenljiv, v trenutku pojava ni pripisan posamezniku.	Access: 08 BESEDILA: 10 izjave	
15o–16o) rez= 1ned01P0 1ned01P0 =rez	Oznaki začetnega / končnega reza besedila, dvodelni (na stopnji .3Transkripcije).	Access: 08 BESEDILA: 15r–16r začetni/končni rez	
18o) prostor: <i>Hoja po mestu.</i> prostor: <i>Na stranišče.</i>	Menjave prostora nosilca mikrofona.	Access: 08 BESEDILA: 18 prostor	
19o) šum: <i>Vožnja po slabši cesti, pogovor govork na zadnjem sedežu je nerazumljiv.</i>	Negovorni, nepomenljivi šumi, hrup; sprememba razločnosti posnetka.	Access: 08 BESEDILA: 19 kvaliteta	
22o) dejavnost: <i>Odlaganje prtljage v prtljažnik avtomobila.</i>	Spremembe osrednje dejavnosti soudeležencev, v kateri sodeluje nosilec mikrofona.	Access: 08 BESEDILA: 22 dejavnost	
23o) navzočnost: <i>Govorka pokuka skozi vrata, za njo se oglasi še druga.</i>	Navzočnost govorcev, če se med besedilom spremeni (zaradi njihovega gibanja ali menjave medijev).	Access: 08 BESEDILA: 23 število govorcev Access: 08 BESEDILA: 20 medij	

8.2. Metaoznake transkribcij

Podatki za te enote so večinoma prepisani iz dnevnika snemanja. Nekateri podatki (06), (07), (17), (19)) bodo kasneje uporabljeni še na ravni *.4Besedila*.

Tabela 20: Podatki o transkripcijskih enotah na stopnji *.3Transkribcije*

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
(00T) dogajanje	Dogajanje – aktivnost soudeležencev, socialna situacija, ostalo. Opis v kratki povedi.	Access: 06 TRANSKRIBCIJE (EXM.: Meta information: Comment)
(01T) ID transkribcij	Ime transkribcije, nanj se navezujejo ostali podatki transkribcije, dogovorjeno pa tudi ime transkripcijske datoteke in posnetka. Iz 4 delov: dan + zap. št. minidiska + zap. št. posnetka na minidisku (+ rezanje posnetka na minidisku).	dogovorjeno ime Access: 06 TRANSKRIBCIJE (EXM.: Meta information: Transcription name)
(02T) datoteka	Osnovna transkripcijska datoteka stopnje <i>.3Transkribcije</i> – pot na disku, ime in tip datoteke.	dogovorjeno ime in mesto (v mapi <i>StAnId\3Transkribcije+4\3Transkribcije</i> , podmapa dne: <i>1nedelja, 6petek</i>)
(03T) posnetek	Izvendatotečne reference transkribcije na zvočno datoteko stopnje <i>.3Transkribcije</i> – pot na disku, ime in tip datoteke.	dogovorjeno ime in mesto (v mapi <i>StAnId\1Snemanje</i> , podmapa določenega dne: <i>1nedelja, 6petek</i>) (EXM.: Meta information: Referenced media file)
(06) datum snemanja	Datum ter ime dneva v tednu; omejenih 7 oz. izbrani 2 možnosti.	Access: 06 TRANSKRIBCIJE
(14T) trajanje	Trajanje vzporednega posnetka. (Neposneti deli kot na stopnji <i>.4Besedila</i>) v to tabelo niso vpisani.) Podatek v <i>h:mm:ss</i> (brez pike na koncu).	Access: 06 TRANSKRIBCIJE
(15T) čas začetka, (16T) čas konca	Dnevni čas začetka in konca snemanja. Podatek, zaokrožen na minuto (v <i>(h)h:mm</i>).	Access: 06 TRANSKRIBCIJE
(17) kraj	Navedba krajevnega imena kraja (naselja), v katerem se nahaja nosilec mikrofona (ne pa npr. tvorec sporočil preko telekomunikacij). (Enak za vsa hčerinska besedila na stopnji <i>.4Besedila</i> .)	Access: 06 TRANSKRIBCIJE
(19) kvaliteta	Kvaliteta oz. kompresija posnetka pri snemanju na minidisk Sony. Ena od 3 kategorij: <i>SP LP2 LP4</i> . (Enaka za vsa hčerinska besedila na stopnji <i>.4Besedila</i> .)	Access: 06 TRANSKRIBCIJE
(A) ID stopnje korpusa	Ena od 8 stopenja projekta, zgolj priklic že obstoječega imena – privzeta vrednost <i>StAnId.3Transkribcije</i> . S pomočjo reference v tej povezavi so dostopni še vsi podatki za posamezno stopnjo, kot so opisani v glavi korpusa – gl. 0.	Access: 01 stopnje projekta > 06 TRANSKRIBCIJE (EXM.: Meta information: Project name)

Podatki o urejanju transkribcij ((07), (08T)), tudi zapisani v tabelo 06 TRANSKRIBCIJE, so opisani v poglavju 0.

8.3. Baza govorcev

Podatki o govornikih so izbrani in zapisani tako, da na tem mestu veljajo za katerokoli besedilo v korpusu; četudi isti govorec nastopa v različnih vlogah v več besedilih, ima v tej bazi govorcev le eno šifro in podatke, ki zanj veljajo v vseh besedilih. Podatki, ki se za govorca menjujejo od besedila do besedila, pa so zapisani kasneje, gl. II.8.6.

Kot „govorec“ ni mišljena samo posamezna človeška oseba, ampak tudi vir zaokroženega sklopa enoznačnega govora (mejni primeri: več oseb govori isto besedilo, predvajanje posnetkov, avtomati). Baza govorcev spada pod stopnjo *StAnId.2Okoliščine*.

Tabela 21: Podatki o govornih na ravni celega korpusa, stopnja .2Okoliščine

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
I) ID govorca	Šifra govorca na stopnji <i>StAnId.4Besedila</i> . Nanjo so vezani vsi ostali podatki o govornih, tako na ravni korpusa kot posameznega besedila. Sestav šifre iz 3 delov (5 znakov skupaj): zap. št. pojavitve v korpusu (tromestna) + spol (M Ž N-eznan spol V-eč skupaj; gl. podatek V) + poznanost (z-nan, j-avno znan, a-vtomat, n-eznan; gl. podatek VI). Primeri: 001Mz, 002Žz, 096Nn, 021Vn, 089Ža, 045Mj	Access: 04 govornici EXM.: Speakertable: Abbreviation ¹³⁰ Elan: Tier Attributes: Participant
II) ime govorca III) priimek IV) vzdevek	Dejansko osebno lastno ime ali mnemotehnični vzdevek (<i>Ime, Priimek; Vzdevek</i>) govorca, zapisano v partituri le na stopnji <i>StAnId.3Transkribcije</i> , v Accessu pa v skriti tabeli podatkov. Za neznanca (nepoznano ime) je uporabljeno podobno kodiranje kot za šifre (iz 3 delov, 4 znaki): zap. št. neznanca (brez ničle spredaj, eno- ali dvomestna) + (M Z N V; Z za ženske, ne ž!) + (Z J A N; same velike črke za razliko od osrednjih šifer) + za presledkom poljubna mnemotehnična oznaka (npr. <i>novinar, bralec, nedeljnik, zbor, zbor 2. glas</i> itd.). Te oznake neznancev iz EXM., stopnja <i>3Transkribcije</i> , so potem vpisane v bazo imen podatkov: šifra pod II) ime, mnemotehnična oznaka pod IV) vzdevek.	Access: 03 imena govornic EXM.: Speakertable: Abbreviation
V) spol+število Va) sex	Biološki spol in število govorca – 4 možnosti, prve tri pomenijo enega govorca: <i>moški ženska neznano več</i> (skupaj). ¹³¹ (V EXM. še dodatno označeno tudi iz omejenega nabora vrednosti <i>m-ale f-emale u-nknown</i> – slednje tudi za več govornic.)	Access: 04 govornici EXM.: Speakertable: Sex
VI) poznanost	Merilo tega podatka je vir poznanosti govorca osrednjemu govorniku 001Mz=zapisovalcu (oz. ustvarjalcu/uporabniku), kar potrjuje vpis imena govorca v bazi govornic (nasploh, ne samo v konkretnem besedilu; konkretnost odražajo drugi podatki: medij, št. govornic, metaoznake besedila ...). Podatek X) je le podrobna razdelava teh vrednosti: <ul style="list-style-type: none"> • <i>znan</i> – znano je vsaj ime ali priimek govorca, med govornic in 01Mz je vsaj minimalen odnos zunaj tega besedila; pod X) podrobnejše: <i>isti družina prijatelji znanec</i>. • <i>javno znan</i> – poznano je vsaj ime ali druga „nalepka“ tega govornika; vendar je odnos do govornika samo enostranski, govornik nima do 01Mz nikakršnega odnosa zunaj tega besedila; pod X) obvezno podatek <i>javno spoznan</i>. Običajno gre za govornice, ki tudi v konkretnem besedilu nastopajo preko telekomunikacijskih javnih množičnih (neosebni) medije – (R)TV, zvočnik). • <i>avtomat</i> – vir govora nima identitete živega govornika in ni personalno avtoriziran, npr. avtomatski odzivnik (nemnožični), posneta ponavljajoča se javna obvestila (lahko množična); pod X) <i>neznanec</i>. Sem ne sodijo npr. posnetki glasbe (privatno poslušanje zgoščenk) in reklame. • <i>neznan</i> – brez podatkov, govornik ne ustreza nobeni od prejšnjih treh kategorij; pod X) <i>neznanec</i>. 	Access: 04 govornici
VII) starost	Biološka starost govornika – ena od 8+2 kategorij <i>–6 7–15 16–19 20–26 27–39 40–49 50–65 66– neznano več</i> .	Access: 04 govornici
VIII) govorno poreklo	Geografsko govorno poreklo po tradicionalni slovenski dialektološki razdelitvi na 7 narečnih skupin, dodani sta dve največji mesti – ena od 9+2 kategorij <i>kor. štaj. pan. dol. gor. rovt. prim. LJ MB neznano več</i> .	Access: 04 govornici
IX) izobrazba	Gre za končano izobrazbeno stopnjo govornika – ena od 5+2 kategorij (<i>še</i>) <i>brez OŠ SŠ VŠUNI PODIPL neznano več</i>	Access: 04 govornici

¹³⁰ Vsak govornik ima v EXM. poleg uporabniško določene okrajšave (Abbreviation) še avtomatsko določeno enoznačno kodo, uporabljeno v markupu datoteke (SPKn; SPK=speaker, n=naravno število).

¹³¹ Razločitev hkratnega govora recitacij na partituri: Enotna šifra govornika je določena za tisti govor (recitacije, petje), ki ga nekaj govornic (ki jih je toliko, da jih ni mogoče ločiti na podlagi poslušanja) govori brez variacij. Vseeno pa se lahko vzporedno zapiše isto besedilo (govorna vloga) pri posameznem govorniku (pod njegov osebno šifro; pod pogojem, da ga pravilno določimo), a le pri tistem, ki ga je mogoče razločiti za večino časa recitacije in ki variira glede na osnovno besedilo recitacije.

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
X) razmerje do 001Mz	Socialno razmerje poznanosti do izbranega govorca – ena od 6 kategorij isti družina prijatelj znanec javno spoznan neznanec. Gre za podkategorije glede na vrednosti podatka VD) poznanost.	Access: 04 govorce
0) ostalo	Opis kakih relevantnih ostalih podatkov, ki jih ni mogoče zajeti v zgornjih kategorijah; v kratki povedi.	Access: 04 govorce (EXM.: Speakertable: Comment)
D) jezik	Jeziki, ki jih uporablja govorec (ne zapisujem!): First Language ..., Second Language ... – možnost izvire okrajšave jezika iz bogatega standardnega seznama. Kot privzeto velja, da je jezik govorcev slovenščina.	(EXM.: Speakertable: Languages used)
XIO) prisoten v besedilih	Število besedil, v katerih govorec v korpusu nastopa. Podatek služi zgolj statistični predstavitvi. Gl. tudi Tabela 30, str. 118.	Access: 04 govorce
100) izjave, 110) povedi, 120) besede	Vsota vsakega absolutnega podatka 10–12)OB) izjave/povedi/besede za vsa besedila v korpusu, kjer govorec nastopa.	Access: 04 govorce

Na stopnji *.3Transkribcije* so teoretično vsi podatki razen I), na stopnji *.4Besedila* pa ni podatkov II–IV), so pa zajeti v I). Podatki 0), D) (oz. II–IV)) in Va) so v EXM. vedno ponujeni, ostale pa je možno posebej uporabniško določiti.

Govorce, ki imajo pri podatku V) spol+število vrednost več, lahko imajo pri podatkih VII–IX) vrednost več, pri VI) in X) pa običajno vrednost N-eznan oz. neznanec (Z-nan/znanec le, če so v bazi govorcev znani vsi člani, ki govorijo).

8.4. Besedila – poimenovanje in rezanje

Naslednji podatki nastanejo v fazi rezanja transkripcij in posnetkov (enote gradiva na stopnji *.3Transkribcije*) po kriterijih za meje besedil (gl. poglavje I.3.2) na „besedila“ in njim ustrezne posnetke (enote gradiva na stopnji *.4Besedila*). Podatki iz druge spodnje tabele služijo samo zapletenemu postopku rezanja (njegov opis gl. v III.12.3), kasneje pa potrebujemo le redke izmed njih – 14–16).

Tabela 22: Poimenovanje (šifriranje) „besedil“ v korpusu, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
01) ID besedila	Ime besedila, pa tudi transkripcijske datoteke ter njenih variant in vzporednega posnetka. Iz 5 delov (skupaj 8 ali 9 znakov): zap. št. in dan (1ned, 2pon ...); zap. št. besedila dneva (01–99); n-govor (P M D V – prim. podatek 09)) + oznaka gostote besedila (g-osto t-edko 0 m-inimalno n-eznano (=neposneto) – prim. podatek 13)); datoteke, ki nastanejo ob vmesni fazi rezanja, preden so „zlepljeni“ v cele datoteke, imajo na koncu kot 9. znak še enega od znakov ± _ (prim. podatek 01r)).	dogovorjeno ime Access: 08 BESEDILA (EXM.: Meta information: Transcription name)

Tabela 23: Podatki o rezanju in časovni razsežnosti „besedil“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
01r) razrezanost	Status besedila v tabeli glede na celovitost v medstopnji rezanja – ena od 3 možnosti: celo del zlepljeno. Podatek se odraža tudi v 01) IDbesedila: ± _ na zadnjem mestu govori o necelovitem besedilu, delu, ki ima zlepljeno varianto: isto ime brez + _ na koncu.	Access: 08 BESEDILA
01Tr) ID transkribcije	Ime materinske transkripcije dela besedila; zgolj priklic že obstoječega imena.	Access: 06 TRANSKRIBCIJE > 08 BESEDILA
10r) interval	Glede na zaporedje časovnih referenc v partituri materinske transkripcije v EXM: Od nnnn do nnnn.	Access: 08 BESEDILA
14) trajanje	Trajanje vzporednega posnetka oz. neposnetega dela (za P-razno + n-eznano) – podatek v h:mm:ss (brez pike na koncu).	Access: 08 BESEDILA datotečni podatki: trajanje

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
14r) trajanje reza	Trajanje razrezanega dela – podatek v s(sss) (brez pike na koncu) zaradi lažjega preračunavanja; privzeta vrednost 0.	Access: 08 BESEDILA
15) čas začetka, 16) čas konca	Dnevni čas začetka in konca snemanja – podatek, zaokrožen na minuto (v (h)h:mm).	Access: 08 BESEDILA
15r) začetni rez, 16r) končni rez	Vrednost časovne reference začetka/konca reza v materinski transkripciji v EXM, zaokrožena na sekundo; privzeta vrednost 0.	Access: 08 BESEDILA
26) povod začetka, 27) povod zaključka	Povod za začetni rez posnetka (dogajanje prej) oz. za zaključni rez posnetka (dogajanje potem). Opredelitev po enem od 6 kriterijev za mejo besedila in dopolnitev s kratkim opisom do 150 znakov: prostor (18), navzočnost (18), (20), (23), (24), dejavnost (22), kontinuiteta (govora pri enakih okoliščinah), razločnost (19), vir (14–16); za delna besedila ter prvo in zadnje v dnevu). Podatek 26) je enak kot podatek 27) za predhodno besedilo.	Access: 08 BESEDILA

8.5. Besedila in njihove metaoznake

Razen dogovorjenih podatkov, ki ne potrebujejo zapisa, so vsi podatki iz tega poglavja zapisani v tabeli 08 BESEDILA v programu Access (gl. Slika 11 spodaj); poleg ID besedila, na katerega so kot na ključ navezani ostali podatki, jih je še 34. Zapis vseh teh podatkov pa je možno tudi posebej uporabniško zastaviti v EXM., v kolikor ni že predviden v programu samem.

Na sama besedila se nanašajo še nekateri podatki njihovih materinskih, nerazreznih transkripcij, ki jih je mogoče priklicati s poizvedbo programa Access iz tabele 06 TRANSKRIBCIJE; to so 06), 07), 17) in 19).

Vseh podatkov o posameznem „besedilu“, ki imajo svojo šifro, je 43.

Tabela 24: Datotečni podatki „besedil“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
02) datoteka	Osnovna transkripcijska datoteka stopnje .4Besedila – pot na disku, ime in tip datoteke.	Dogovorjeno isto ime in mesto (mapa StAnId\3Transkripcije+4\4Besedila)
03) posnetek	Izvedatotečne reference transkripcije na zvočno datoteko stopnje .4Besedila – pot na disku, ime in tip datoteke.	transkripcijske datoteke in posnetka. (EXM.: Meta information: Referenced media file) (Elan: Linked Media Files: Master Media: "Yes" in druge določbe; avdio-datoteka mora biti v isti mapi kot vzporedna transkripcijska datoteka .eaf).
04) ELAN-Media-File	Podatki za pretvorbo v EXM>Elan. Podatek o posnetku – .mp3-datoteka v isti mapi (= podatek 03)).	Access: 08 BESEDILA (EXM.: Meta information: User defined attributes: ELAN-Media-File)
05) ELAN-Mime-Type	Podatki za pretvorbo v EXM>Elan. Privzeta vrednost audio/* (čeprav EXM. ponuja druge 4 vrednosti; edina pravilna za .mp3 bi bila audio/mpeg).	Access: 08 BESEDILA (EXM.: Meta information: User defined attributes: ELAN-Mime-Type)
00) ostalo	Ostale relevantne opombe: nečloveške okoliščine, daljši šumi, hrup ipd. Opis v kratki povedi.	Access: 08 BESEDILA (EXM.: Meta information: Comment)
A) ID stopnje korpusa	Ena od 8 stopenja projekta, zgolj priklic že obstoječega imena – privzeta vrednost StAnId.4Besedila. S pomočjo reference v tej povezavi so dostopni še vsi podatki za posamezno stopnjo, kot so opisani v glavi korpusa.	Access: 01 stopnje projekta > 08 BESEDILA (EXM.: Meta information: Project name)

Podatki o imenu besedila, o transkripcijski datoteki in posnetku so v prejšnjem poglavju, podatki o urejanju besedila so v poglavju 0.

Tabela 25: Podatki o neodvisnih okoliščinah „besedil“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
(18) prostor	Kratek opis prostora (soba, del ulice, krajine); v primeru trajnega gibanja pa tudi opis gibanja (<u>vožnja</u> <u>hoja</u>). Nanaša se na nahajanje nosilca mikrofona, ne drugih oseb.	Access: 08 BESEDILA

Podatki o kraju so prevzeti od materske transkripcije, tisti o časovni umeščenosti in razsežnosti posnetka pa so navedeni v prejšnjem poglavju o rezanju besedil.

Slika 11: Izsek podatkov iz tabele 08 BESEDILA, Access

01) ID besed	20) medij	21) tema	22) dejavnost	23) število govorcev	24) število prisotnih	28) oblika sp	29) vloga	30) prostost	31) socialni	32) namizni	33) govornik
+ 1ned71Dg	osebni stik	vodovod, začete	Pogovor o preteklih p	2	2	pogovor pre		prosto	pok. pog	ZPČO	zasebn
+ 1ned72Em	sam zase	milo	Večerna higiena pred	1	1	samogovor		prosto	nar.	ZPČO	zasebn
+ 6pet01Dg	osebni stik	vstajanje, zajtrk	Jutranji pogovor ob ka	2	2	pogovor ob z		prosto	pok. pog	ZPČO	zasebn
+ 6pet02Dr	mobitel	delitev hrane za	Pogovor po mobitelu,	2	3-10	pogovor po r		prosto	pok. pog	ZPČO	zasebn
+ 6pet03Dr	osebni stik	neznani zlicatelj	Pogovor - ugibanje o	2	2	pomenek		prosto	pok. pog	ZPČO	zasebn
+ 6pet04Er	telefon	pot domov	Pogovor po telefonu.	1	3-10	pogovor po t		prosto	pok. pog	ZPČO	zasebn
+ 6pet05Dr	osebni stik	neznani klicatelj	Pospravljanje po zajtr	2	2	pomenek		prosto	pok. pog	ZPČO	zasebn
+ 6pet06P0-			Pomivanje posode.	0							
+ 6pet06Pn	neznano	neznano	Pomivanje posode, u	0	2	neznano	nezna	neznano	neznano	nezna	neznano
+ 6pet06Pn+			Umivanje zob, urejanj	0							
+ 6pet07Er	sam zase	predme Zgodovi	Samoučenje arhivistil	1	1	polglasno (s citati i		prosto z r	knj.-pog.	INFO	zasebn

Tabela 26: Podatki o soudeležencih kot okoliščinah „besedil“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
(21) tema	Tema/vsebina sporazumevanja/besedila. Kratek opis ene ali več glavnih tem; do 200 znakov oz. <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). Pri krajših besedilih (gl. podatek (12) besede) je oznaka nezanesljiva.	Access: 08 BESEDILA
(22) dejavnost	Nanaša se na dejavnost (opravki, socialna situacija), v kateri je soudeleženec nosilec mikrofona, ne pa dejavnosti brez njega pri drugih osebah (četudi nosilec mikrofona ni govorec v besedilu, pa je to nekdo drug). ¹³² Kratek opis do 150 znakov.	Access: 08 BESEDILA
(23) število govorcev	Absolutni numerični podatek (se ujema s številom šifer govorcev v posameznem besedilu); privzeta vrednost 0.	Access: 08 BESEDILA
(XI) prisotni govorniki	Našteti govorniki v besedilu. Navedene šifre govorcev. 100 znakov.	Access: 08 BESEDILA
(24) število soudeležencev	Število soudeležencev sporazumevanja (vseh govorcev+poslušalcev+ naslovnikov+prejemnikov, katerih se predvidoma dotakne sporazumevanje v trenutku časa snemanja). Ocena po eni od 6 kategorij 1 2 3-10 11-99 100- <u>neznano</u> . Če je v besedilu več pogovorov oz. kakšni medklici in napovedi, so v oznaki omenjeni udeleženci (po)govora, ki prevladuje. ¹³³	Access: 08 BESEDILA
(25) zavest snemanja	Zavest o snemanju med samim sporazumevanjem (govori tudi o spontanosti govora). Ena od 2 kategorij <u>da</u> <u>ne</u> ; privzeta vrednost <u>ne</u> . Če je nosilec mikrofona edini udeleženec ali je št. udeležencev neznano (kar je pogosto pri besedilih P0), potem DA; sicer leti oznaka na ostale govorce in je <u>ne</u> . ¹³⁴	Access: 08 BESEDILA

¹³² Strogo označuje le početje, ne pa tudi teme, prostora, vrste pogovora. Ta omenjena razširitev opisa se običajno zgodi, ker ta podatek vpisujemo že ob rezanju besedil, omenjene tri podatke pa kasneje.

¹³³ Šteti so tudi razni predvajalniki in avtomati (gl. opombe govorcev), namreč kot govorniki; koliko se pa šteje poslušalcev avtomatov, je odvisno od deleža govora, ki ga ima avtomat v besedilu, in od medija, ki ga avtomat uporablja. Primer: napoved odhoda vlaka na polni postaji potnikov preko zvočnika v besedilu, kjer je edini govorec postajni napovednik, bo soudeležencev 100-; napoved postaje preko zvočnika v polnem vlaku v besedilu, kjer je ta napoved minorna v primerjavi s pogovorom dveh potnikov, bodo število udeležencev 3-10 (3 govorniki).

¹³⁴ Edini primer, kjer se sogovorec zave snemanja, je 1ned57Dg, kjer je snemanje razkrito med pogovorom.

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
31) socialna zvrst	Socialno-geografska zvrstnost jezika po prirejeni tradicionalni slovenski razvrstitvi – ena od 4+2 kategorij <u>zbor. gov.</u> <u>knj.-pog.</u> <u>pok. pog.</u> <u>nar</u> <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). Kadar iz besedila ni razvidna (prekratko besedilo), je označena zvrst, ki je najobičajnejša (materna) za govorce – prednost imajo socialne zvrsti dlje od knjižne.	Access: 08 BESEDILA
32) namen sporazumevanja	Namen/funkcija sporazumevanja – ena od 4+2 kategorij: <u>INFO</u> <u>JINST</u> <u>POSL</u> <u>ZPČO</u> <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). ¹³⁵	Access: 08 BESEDILA
33) govorni položaj	(Socialni) govorni položaj (tip sporazumevanja glede na medij in št. soudeležencev) – ena od 2+2 kategorij: <u>javni</u> <u>zasebni</u> <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn).	Access: 08 BESEDILA
35) smer sporazumevanja	Smernost sporazumevanja (ali ima govorec odziv poslušalca). Ena od 4 + 1 kategorij: <u>brez</u> (za P0 in govor samemu sebi) <u>enosmerno</u> (eden ali več govorcev, kjer govor nikogar ni odgovor prejšnjemu) <u>1,5-smerno</u> (sporazumevanje je dvosmerno, a so posneti/zapisani le govorci, brez odgovora naslovnikov na njihov govor) <u>dvosmerno</u> (vsaj dva govorca zapisana, govor vsaj enega naslovnika je odgovor prejšnjemu govorniku) <u>neznano</u> (= za -Pn).	Access: 08 BESEDILA

Tabela 27: Podatki o sporazumevalnih sredstvih „besedil“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
20) medij	V besedilu prevladujoči medij, ki ga uporablja ali ga je deležen nosilec mikrofona, ne druge osebe. Ena od 9 kategorij: <u>sam zase</u> <u>osebni stik</u> <u>telefon</u> <u>mobitel</u> <u>TV</u> <u>radio</u> (javni) <u>zvočnik</u> <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn).	Access: 08 BESEDILA

Prenosnik in kod/jezik sta kot privzeta opisana v podatkih glave korpusa, gl. poglavje 0.

Tabela 28: Podatki o statistikah jezikovnih enot v „besedilih“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
09) n-govor	Tip sporazumevanja glede na število govorcev (ni isto kot število udeležencev). Ena od 4 kategorij <u>Prazno</u> <u>Monolog</u> <u>Dvogovor</u> <u>Večgovor</u> , ki se mora po začetnici ujemati v 7. znaku ID besedila. Oznaki <u>dialog</u> , <u>polilog</u> namenoma nista uporabljeni zato, ker ne označujeta le števila govorcev (pa še tu <u>dialog</u> ne pomeni dobesedno <u>dvogovor</u>), temveč tudi smer sporazumevanja (<u>dialog</u> v nasprotju z <u>dvogovorom</u> predpostavlja odgovarjanje na govor, torej dvosmerno sporazumevanje). Resda uporabljam oznako <u>monolog</u> (možen bi bil tudi izraz <u>enogovor</u>), a ta ne pomeni dominanc enega govornika, ampak zgolj enega samega govornika v besedilu. Najpogosteje je edini govorec v besedilu izbrani nosilec mikrofona (<u>001Mz</u>), ni pa vedno.	Access: 08 BESEDILA
10–12) izjave /povedi /besede	Absolutni podatek iz štetja v EXMARALDA; privzeta vrednost 0. Za načela členitev izjav in povedi gl. II.7.2, za izvor statistik vseh teh enot pa III.12.4.	Access: 08 BESEDILA
13r) gostota	Ocenitev časovne strnjene in količine besed v besedilu ob rezanju. Ena od 5 kategorij vrednosti: <u>gosto</u> (enakomerno gost govor v posnetku, dosti nad 15 besed) <u>redko</u> (redki govor v posnetku, daljši premori, nad 15 besed) <u>0</u> (niti ena beseda na traku [v] transkribirana) <u>minimalno</u> (transkribiranih do 15 besed, gostota ni važna) <u>neznano</u> (transkribiranih ni nič besed, večina besedila je neposnetega), ki se mora ujemati v 8. znaku ID besedila.	Access: 08 BESEDILA
13) besede/s	Gostota „besedila“: izračun količine besed na sekundo vzporednega posnetka; na eno decimalo. Preglednejši bi bil morda preračun na minuto, a kaj ko nekatera besedila niso dolga niti minuto in bi bil tak podatek zanje malo smiseln. Izračunano iz podatkov 12) in 14r).	Access: 08 BESEDILA

¹³⁵ Izobraževanje-INFOrmiranje | Javni-INSTitucionalni govor | POSLovno-poklicno | Zabava-Prosti Čas-Osebni | brez (= za -P0) | neznano (= za -Pn). Kot JINST je npr. označen obred svete maše.

Tabela 29: Jezikoslovne metaoznake „besedil“ in sobesedila, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
[28] oblika sporazumevanja	Običajna, vsakodnevna oznaka za obliko sporazumevanja. Opis z eno besedo: <u>recitacija</u> , <u>pomenek</u> , <u>pogovor</u> , <u>dogovarjanje</u> , <u>monolog</u> , <u>nagovor</u> , <u>prepir</u> , <u>predavanje</u> ... (bolj kompleksna, obširna kot je govorna situacija, bolj oznaka pomeni dejavnost kot jezikoslovno oznako govornobesedilne oblike). Privzeti vrednosti <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). Za pogoste primere je bila razvita tristopenjska lestvica: <u>pogovor</u> – bolj resne obdelave tem, trden socialni stik; <u>pomenek</u> – sporazumevanje ob delu, na kratko o temah; <u>klepet</u> – ob brezdelju, lahka obdelava tem. Pri krajših besedilih (gl. podatek [12] besede) je oznaka nezanesljiva.	Access: 08 BESEDILA
[29] vgnezdno	Kratek opis vrste in števila daljših v besedilu umeščenih besedil, ki bi lahko bila tudi samostojni. Privzeti vrednosti <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn).	Access: 08 BESEDILA
[30] prostogovornost	Ocena količine prostogovornosti v besedilu (nasproti recitacijam, popolnemu sledenju predlogam = navajanju, branju). Ena od 5 (lestvica)+2 stopenj: <u>prosto</u> <u>prosto z navedki</u> <u>prosto in navedki</u> <u>navedki s prostim</u> <u>navajanje</u> <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). Kakšen je prenosnik navajanja (branje z raznih predlog, recitiranje/deklamiranje), prim. trak [pv], oznaka <u>prenj</u> , in tabelo 05 govorcev v besedilih.	Access: 08 BESEDILA
[34] funkcijska zvrst	Prevladujoča funkcijska zvrstnost jezika v besedilu po tradicionalni (Toporišičevi) razvrstitvi – ena od 4 kategorij: <u>UM</u> (-etnostni) <u>ST</u> (-rokovni) <u>P</u> (-raktično)- <u>S</u> (-porazumevalnu) <u>PU</u> (-ublicistični) <u>brez</u> (= za -P0) <u>neznano</u> (= za -Pn). ¹³⁶	Access: 08 BESEDILA

Značilen sistem vrednosti zgornjih podatkov imajo PRAZNA (-P0 = prazno/brez govorcev, 0 besed; brez zapisanega verbalnega govora) in NEZNANA besedila (Pn = prazno/brez govorcev, neznane morebitne besede; taka brez posnetka, ID pa imajo, da se opozori, da je bilo snemanje izpuščeno).

Za P0-besedila imajo podatki [20–21] in [28–35] vrednost brez, [10–13] in [23] pa vrednost 0. Pa Pn-besedila imajo podatki [13], [20–21] in [28–35] vrednost neznano, [10–12], [14r–16r] in [23] vrednost 0, [25] vrednost ne, brez vrednosti pa imajo podatke [01r], [10r], [02–04], in [07–08].

8.6. Vloge govorcev v besedilih

Osnovna podatka, [01] ID besedila in [1] ID govorca, sta referenci na podatke besedil in govorcev, hkrati pa skupaj služita kot primarni ključ, ki se mu pripisujejo naslednji podatki. Vse podatke je možno tudi uporabniško določiti v EXM in jih pripisati posameznemu traku v partituri transkripcije.

Tabela 30: Podatki o govorcih v „besedilih“, stopnja .4Besedila

<i>ime podatka</i>	<i>opis vrednosti podatka</i>	<i>mesto zapisa</i>
[XOB] vloga	Eno- ali dvobeseden opis govorceve socialne vloge v aktualni sporazumevalni situaciji; npr. <u>prodajalka</u> . Ne gre zamenjati z odnosom do nosilca mikrofona (podatek X).	Access: 05 govorcev v besedilih
[10OB] izjave, [11OB] povedi, [12OB] besede	Absolutni podatek iz štetja v EXMARaLDA; privzeta vrednost 0. Za načela členitev izjav in povedi gl. II.7.2, za izvor statistik vseh teh enot pa III.12.4.	Access: 05 govorcev v besedilih
[00OB] ostalo	Opombe o mediju, prenosniku, jeziku, posebnih paraverbalnih značilnostih (prim. trak [pv]); kar pač ni mogoče zapisati pod samega govorca.	Access: 05 govorcev v besedilih

¹³⁶ Razvrstitvi se vidi, da je narejena na podlagi knjižnih (tiskanih, pisni prenosnik) besedil za potrebe normiranja jezika. Tako so pri razvrščanju v tem govornem korpusu dileme pri uvrstitvi nekaterih besedil (maša kot UM, reklame kot PU), uvrstitev večine besedil v kategorijo PS pa ne daje večjih možnosti za podrobnejše razlikovanje funkcijskosti v korpusu.

Obstajajo tudi statistične vsote zgornjih podatkov za ves korpus: za vsako besedilo, kateri govorniki nastopajo v njem (XI prisotni govorniki); za vsakega govornika, v koliko besedilih v korpusu nastopa (XIO prisoten v besedilih) in koliko enot (izjav, povedi, besed) spregovori v vsem korpusu (gl. Tabela 21).

Zanimiv primer so avtomati kot govorniki: v besedilih, kjer prevladujejo, so metaoznake njihovega govora (besedil) stalne – iste. Tak primer je govorec 089Za, posnetek napovedi postaj na vlaku, ki nastopa v nekaterih besedilih od 6pet47Dr do 6pet61Dm.

8.7. Primeri iz drugih korpusov

V nadaljevanju želimo z našim sistemom podatkov oz. zapisa sporazumevalnih okoliščin primerjati še dva: iz korpusov BNC in UKGS (kot primerjavo smo ju predstavljali že v poglavjih I.5.1 in II.7.4). Za BNC je bil vir članek soustvarjalca (Crowdy 1995), dosti pa je njegovih opisov tudi v slovenski strokovni literaturi (Zemljarič Miklavčič 2004, Gorjanc 2002 in 2005). Za UKGS do samega korpusa nisem imel dostopa, podatki iz literature pa so različni, kar odraža preizkušanje različnih rešitev ob delu; od tod tudi nezanesljivost podatkov v spodnji tabeli; Zemljarič Miklavčič 2005 in 2006, Zemljarič Miklavčič/Stabej 2005).

Pri obeh omenjenih korpusih moramo od dokumentiranosti sporazumevalnih okoliščin/dejavnikov (kar nas izrecno zanima zaradi primerjave z našim sistemom) ločiti vnaprejšnje kriterije za izbor gradiva t. i. referenčnega ali uravnoteženega korpusa.¹³⁷ (UKGS ni sestavljen po njih, gradivo je zbrano bolj oportunistično; Zemljarič Miklavčič/Stabej 2005: 231.) Praktična razlika v strukturiranosti podatkov o okoliščinah sporazumevanja med obema vidikoma ni velika,¹³⁸ razlika je le v konkretnosti in časovni fazi pri nastajanju korpusa: **kriteriji za uravnoteženje korpusa** so sestavljeni pred nastankom korpusa in slonijo na sociolingvističnih ocenah in namenih uporabe korpusa, **kontekstualni podatki besedilnih enot korpusa** pa so konkretni podatki, gradivo samo oz. dokumentacija realnega sporazumevanja.

V našem korpusu (StAnId) imamo le te druge podatke: izhajajo iz dnevnika snemanja v neizbranih okoliščinah in iz modela razlikovanja, ki je dokaj splošen, nekoliko pa tudi podrejen ciljem diplome. Pri strukturiranju teh podatkov v transkripcijski konvenciji pa sem si vendarle pomagal z uravnoteževalnimi kriteriji korpusa BNC, tako za *demografsko* (le govorniki kot okoliščine) kot *kontekstualno* (drugi vidiki okoliščin) komponento – ker so ti kriteriji po namenih korpusa BNC in kot sociolingvistična študija dovolj široki, splošni, da pokrijejo tudi naše cilje razčlenitve okoliščin sporazumevanja. Pa tudi – v domači literaturi je malo take pregledne sociolingvistične teorije ali analize, ki bi lahko služila potrebam računalniške oz. kvantitativne lingvistike – v njej npr. moramo zaradi razloma *celega* besedila na pogostostni seznam besed imeti natančne oznake dejavnikov sporazumevanja, ki se nanašajo na vse besedje še tako nepredvidljivega primerka govora. Primer zgledovanja po uravnoteževalnih kriterijih referenčnega korpusa so 4 vrednosti podatka 32 namen sporazumevanja (gl. spodnja tabela), ki so vzete iz načrta kontekstualne komponente za korpus govorne slovenščine (prim. Zemljarič Miklavčič 2004: 516).

¹³⁷ Za BNC so 3 uravnoteževalne oznake (čas, predmetnostno področje, medij; Gorjanc 2005: 32) in nadaljnje neobvezne: „velikost vzorca (v besedah), velikost besedila (v besedah), naslov, avtor in založnik, datum nastanka, historiat besedila od nosilca avtorskih pravic do korpusa, naklada, nivo zahtevnosti besedila, spol avtorja, starost avtorja, avtorjeva etnična pripadnost, domicil avtorja in ciljna publika (glede na starost, spol in zahtevnost) (DBNC 1994: 7)“ Gorjanc 2005: 34. Koliko se ti podatki nanašajo na govorni del korpusa, ni jasno.

¹³⁸ Pa tudi: menda je v teoriji korpusnega jezikoslovja težnja, da naj bi se namesto termina *referenčni korpusi* začelo uporabljati oznako *uravnoteženi korpusi*, namreč da bi se oznake referenčen ne razumelo kot ‘normodajen’ v jezikoslovnih raziskavah. Gotovo cilj korpusnega jezikoslovja nasploh ni delati niti normodajne niti uravnotežene korpuse, ampak dovolj *dokumentirane* (z vidika okoliščin za besedila/sporazumevalnega produkt/ jezikovne znake), da je *glede na potrebe* različnih raziskav možno iz njih pripraviti *uravnotežene podkorpuse* (izbore).

Tabela 31: Primerjava transkripcijskih rešitev za opis sporazumevalnih okoliščin besedil: StAnID, UKGS, BNC

<i>StAnId</i>	<i>element transkripcije</i>	<i>BNC</i> ¹³⁹	<i>UKGS</i> ¹⁴⁰
3. OKOLIŠČINE – 3.a.1) BAZA GOVORCEV, njihove oznake			
podatki <u>II</u> ime, <u>III</u> priimek, <u>IV</u> vzdevek, a zakriti uporabniku korpusa; podatek <u>VI</u> poznanost govori o dokumentiranosti imena v korpusu oz. o javni poznanosti	<i>ime govorca</i>	first NAME: samo osebno ime	–
<u>V</u> spol+število: <u>moški</u> <u>ženska</u> <u>neznano</u> <u>več</u> (skupaj)	<i>spol</i>	GENder: <u>m</u> <u>f</u> <u>neopredeljeno</u>	spol: <u>m</u> <u>f</u>
<u>VII</u> starost in ena od 8+2 kategorij <u>–6</u> <u>7–15</u> <u>16–19</u> <u>20–26</u> <u>27–39</u> <u>40–49</u> <u>50–65</u> <u>66–</u> <u>neznano</u> <u>več</u>	<i>starost govorca</i>	AGE: absolutna vrednost; (kot demografski kriterij izbiranja snemalcev pa 6 skupin: 0–14 15–24 25–34 35–44 45–59 60+)	starost: v letih (v načrtu pa 3 ali 4 starostni razredi: <u>–17</u> <u>18–30</u> <u>30–59</u> <u>60+</u>) leto rojstva: številka
<u>VIII</u> govorno poreklo, 9+2 kategoriji: <u>kor.</u> <u>štaj.</u> <u>pan.</u> <u>dol.</u> <u>gor.</u> <u>rov.</u> <u>prim.</u> <u>LI</u> <u>MB</u> ¹⁴¹ <u>neznano</u> <u>več.</u>	<i>pokrajina oz. materna socialno-geografska zvrst jezika</i>	LANguage variety: šifra narečne skupine; DIAlect: šifra narečja;	regija: <u>ljubljska</u> , <u>gorenjska</u> , <u>dolenjska</u> (v načrtu pa razdelitev na 5 narečnih skupin)
–	<i>geografska pripadnost</i>	(demografski kriterij izbiranja snemalcev pa ima 4 geografske skupine: <u>jug</u> <u>osrednji del</u> <u>sever</u> <u>neopredeljeno</u>)	–
<u>IX</u> izobrazba, 5+2 kategorij: (<u>še</u>) <u>brez</u> <u>OS</u> <u>SS</u> <u>VŠUNI</u> <u>PODIPL</u> <u>neznano</u> <u>več</u>	<i>izobrazba</i>	age left EDUCation: starost v letih	izobrazba: <u>visoka šola</u> , <u>osnovna šola</u> , <u>univerza</u> , <u>neznano</u> (v načrtu pa 3 izobrazbeni razredi)
–	<i>poklic</i>	OCCupation: opis	poklic: opis, npr. <u>profesor</u> , <u>administr. del.</u>
–	<i>družbeni razred</i>	SOCial class: šifra družbenega razreda; (demografski kriterij izbiranja snemalcev pa ima kategorije po JICNARS, glede na poklic nosilca gospodinjstva)	–
<u>X</u> razmerje do 001Mz: 6 kategorij <u>isti</u> <u>družina</u> <u>prijatelj</u> <u>znanec</u> <u>javno spoznan</u> <u>neznaneec</u> ; <u>XOB</u> vloga (v posameznem besedilu)	<i>odnos do drugih</i>	RELationship to others: npr. <u>brother2</u> , številka se nanaša na sogovorca	odnos: –
3.a.2) SOUDELEŽENCI (kot dejavnik besedila, oznaka za besedilo!)			
<u>21</u> tema: kratek opis ene ali več glavnih tem	<i>tema</i>	topic domain classification system (3 podatki): SUPerfield, SUBject field, INDividual topic: opis	–
<u>32</u> namen sporazumevanja: 4+2 kategoriji: <u>INFO</u> <u>JINST</u> <u>POSL</u> <u>ZPČO</u> <u>brez</u> <u>neznano</u>			
<u>22</u> dejavnost: kratek opis	<i>dogajanje</i>	ACTivity: opis	–

¹³⁹ Kode so tričrkovne, velike črke v opisu sestavljajo kodo; pri govorcih je tričrkovni kodi dodana zaporedna številka govorca: npr. za ime NAM1, NAM2 itd. Podatki o demografskih kriterijih izbiranja snemalcev iz Gorjanc 2005: 36.

¹⁴⁰ V oklepaju pa je bodoča načrtovana razdelitev na vrednosti, napovedana v Zemljarič Miklavčič/Stabej 2005: 231.

¹⁴¹ Tudi korpus Turdis pri tem podatku govorca razločuje 7 narečnih skupin, ki jim dodaja posebni kategoriji za mariborski in ljubljanski govor.

<i>StAnId</i>	<i>element transkripcije</i>	<i>BNC</i> ¹³⁹	<i>UKGS</i> ¹⁴⁰
23) število govorcev: absolutno število	<i>število govorcev</i>	AUDience: absolutno število	št. govorcev: številka
24) število soudeležencev: 6 kategorij 1 2 3-10 11-99 100- <u>neznano</u>			novi govorcev: številka
25) zavest snemanja, 2 kategoriji <u>da</u> <u>ne</u>	<i>zavest o snemanju, spontanost</i>	SPoNtaneity: opis s šifro, npr. 3	<u>posneto skrivaj</u> <u>z vednostjo</u>
31) socialna zvrst jezika 4+2 kategoriji: <u>zbor. gov.</u> <u>knj.-pog.</u> <u>pok. pog.</u> <u>nar</u> <u>brez</u> <u>neznano</u>	<i>socialno-geografska zvrstnost</i>	–	–
33) govorni položaj 2+2 kategorij: <u>javni</u> <u>zasebni</u> <u>brez</u> <u>neznano</u>	<i>(socialni) govorni položaj</i>	–	<u>govorni položaj</u> (pripisan posam. govorcui): <u>formalni</u> <u>neformalni</u> , tudi <u>7</u> : <u>javni</u> <u>zasebni</u> ;
35) smer sporazumevanja, 4 + 1 kategorija: <u>brez</u> <u>enosmerno</u> <u>1,5-smerno</u> <u>dvosmerno</u> <u>neznano</u>	<i>interaktivnost oz. smer(i) sporazumevanja med govorcev besedila</i>	–	(v literaturi je podatek, da so besedila v razmerju monolog : dialog (polilog) = 6 % : 94 %, čeprav sta v vsakem besedilu vsaj 2 govorcev)
3.b) SPORAZUMEVALNA SREDSTVA			
20) medij, 9 kategorij: <u>sam zase</u> <u>osebni stik</u> <u>telefon</u> <u>mobitel</u> <u>TV</u> <u>radio</u> (javni) <u>zvočnik</u> <u>neznano</u>	<i>medij</i>	–	<u>medij</u> : <u>osebni stik</u> <u>telefon</u> <u>radio</u> <u>TV</u>
3.c.1) NEODVISNE OKOLIŠČINE			
06) datum snemanja (snemanja: dan v tednu in mesecu, leto), dnevni 15) čas začetka in 16) čas konca posnetka	<i>datum in čas snemanja</i>	DATE of recording: dan, mesec, leto; TIME of recording: (začetka?)	–
14) trajanje v <i>h:mm:ss</i> in v <i>s(sss)</i>	<i>trajanje snemanja</i>	DURATION: minute.sekunde	<u>trajanje</u> : v <i>mm:ss</i>
17) kraj (naselje), 18) prostor (soba)	<i>prostor snemanja</i>	REGION; COUNTY (okrožje), LOCATION (naselje), SETTING (soba)	opis: <i>univerza, studio, dom, urad, skejtpark, delovno mesto ...</i>
za vse minidisk; 19) kvaliteta (snemanja)	<i>snemalna naprava</i>	RECORDING DEVICE	–
3.č) JEZIKOSLOVNE METAOZNAKE BESEDIL IN STATISTIKE			
09) n-govor, 4 kategorije: <u>prazno</u> <u>monolog</u> <u>dialog</u> <u>večgovor</u>	<i>statistike enot v besedilni enoti</i>	(v izbrani literaturi ni podatka, verjetno pa je dana podatek o velikosti v besedah)	<u>št. enot</u> (besed): število
10) izjave, 11) povedi, 12) besede, 13r) gostota govora, 13) besede/s			
28) oblika sporazumevanja, kratek opis: <i>recitacija, pomeneke, pogovor, monolog, nagovor, prepir, predavanje ...</i>	<i>oznaka za stalno obliko sporazumevanja, vrsta oz. žanr besedila</i>	TYPE of discourse: eno- ali dvobeseden opis	<u>vrsta</u> (besedila): opis: <i>intervju, okrogla miza, multilog med sodelavci, dialog med prijateljema ...</i>
29) vgnazdeno: opis v besedilu umeščenih besedil			
30) prostogovornost, 5 (lestvica)+2 kategoriji: <u>prosto</u> <u>prosto z navedki</u> <u>prosto in navedki</u> <u>navedki s prostim</u> <u>navajanje</u> <u>brez</u> <u>neznano</u>			
34) funkcijska zvrst jezika, 4 kategorije: <u>UM</u> <u>ST</u> <u>PS</u> <u>PU</u>	<i>funkcijska zvrstnost</i>	–	–

Primerjani trije korpusi imajo pri dokumentiranju okoliščin besedil nekatere skupne rešitve, ki v zgornji tabeli niso posebej navedene:

- govorce in besedila imajo svoje šifre – identifikacije (BNC še neobvezno možnost *naslov besedila*);
- (nacionalni) jezik besedila oz. govorca ni posebej označen, saj gre za gradivo izbranega jezika (UKGS ima v poskusni fazi podatek (nacionalni) **prvi jezik**);
- gotovo še kak podatek o obdelavi gradiva (snemalec, zapisovalec, datum obdelave ...), a pri UKGS in BNC v tej izbrani literaturi o tem ni zapisa.

Različne rešitve so podane v spodnji tabeli, tu pa naj omenim le, da imata StAnId in BNC posneto gradivo razrezano po določenih kriterijih na besedila, pri UKGS pa so besedila že sami izvorni posnetki v celoti.

9. DOKUMENTACIJA OBDELAVE GRADIVA IN KORPUSNE STRUKTURE

Nazadnje še o najkompleksnejši skupini podatkov. Dokumentiranje obdelave gradiva in korpusne strukture pomaga predvsem med samo obdelavo, ko na njeni podlagi ovrednotimo učinkovitost dela (npr. hitrost transkribiranja) ter urejenost gradiva (preglednost, praktične rešitve nasproti teoretično zastavljenim, kompatibilnost datotek in programov, priprava gradiva za tuje uporabnike) ter sprejemamo nove odločitve, spremembe. Pri manjšem korpusu in manjšem obsegu dela, kot je v mojem primeru (en ustvarjalec, en namen izrabe gradiva), imajo manj pomena kot pri nekajkrat večjih korpusih, ki so rezultat timskega dela.

Kolikor so to podatki iz glave korpusa, gre za podatke z najvišje stopnje obdelave gradiva, za stopnjo *StAnId.6Informacije*, če pa podatki iz glav transkribcij, besedil in ostalih datotek, pa gre za druge ustrezne stopnje.

Tabela 32: Podatki o obdelavi in strukturi korpusa, stopnja .6Informacije

<i>ime podatka</i>	<i>obrazložitev vrednosti podatka</i>	<i>mesto zapisa</i>
A) ID stopnje korpusa	Sestavljeno iz imena projekta/korpusa + imena stopnje (predvidenih 6 stopenj obdelave za potrebe razločevanja gradiva): <i>StAnId</i> . + <i>.1Snemanje</i> <i>.2Okoliščine</i> <i>.3Transkribcije</i> <i>.4Besedila</i> <i>.5Analize</i> <i>.6Informacije</i>	Access: 01 stopnje projekta (EXM.: Meta information: Project name)
B) ustvarjalec	<u>Rok Bečan</u> za vse gradivo.	Access: 01 stopnje projekta
C) transkripcijska konvencija	<i>StAnId4</i> za vse gradivo.	Access: 01 stopnje projekta (EXM.: Meta information: Transcription convention)
Č) jezik	Podatek o etničnem/narodnem jeziku gradiva. Privzeto mišljena <u>slovenščina</u> .	dogovorjeno Variacije na partituri na traku [pv]; glej tudi trak [Ozadje] in Access: 08 BESEDILA: 00 ostalo.
D) prenosnik	Podatek o prenosniku/kanalnu gradiva. Privzeto (prosto) <u>govorni prenosnik</u> .	dogovorjeno Variacije na partituri na traku [pv].
E) vrsta datoteke	Računalniški tip datoteke (vpis brez pike): <u>xml</u> <u>caf</u> <u>ctf</u> <u>mp3</u> itd.	datotečni podatki: pripona Access: 02 tipi datotek
F) podvrsta datoteke	Medstopnja obdelave datotek (vpis s piko spredaj in zadaj): oznake za skupne datoteke, razrezane (+/- v imenu), interpolirane časovne reference, z metaoznaki, brez njih ...	datotečni podatki: ime Access: 02 tipi datotek
06) datum snemanja	Datum ter ime dneva v tednu; omejenih 7 oz. izbrani 2 možnosti.	Access: 06 TRANSKRIBCIJE
07) začetek prepisovanja	Datum. Podatek materinske transkribcije velja za vsa hčerinska besedila na stopnji <i>.4Besedila</i> .	datotečni podatki (za stopnjo <i>.3Transkribcije</i>): datum nastanka Access: 06 TRANSKRIBCIJE
08) zadnje urejanje, 08T) zadnje urejanje	Datum, različen za materinske transkribcije in hčerinska besedila.	datotečni podatki: datum spremembe

9.1. Preverjanje tekočih zapisov v partituri

Pri uporabi gradiva moramo računati s subjektivnimi napakami, ki so pred tem nastale pri zapisovanju ali vpisovanju samega gradiva kljub transkripcijski konvenciji. Najmanj jih je tam, kjer jih objektivno reducira sam način vnosa – za podatke okoliščin sporazumevanja preko programa Access (prim. poglavje III.12).

Po transkribiranju govora pa še vedno ostajajo pogostejše napake oz. nedoslednosti. Skušal sem se jim izogniti na različne načine: s strukturiranjem problematičnih zapisov (kot poglavje II.7); ter s posebno fazo transkribiranja, v kateri se pozornost prepisovalca usmeri samo na pravilnost zapisa in še posebej na končna in posebna ločila; končno z analizo ugotovljenih površnosti. Spodaj dajem nekaj primerov, ki bi jih seveda bilo lažje razumeti ob hkratnih posnetkih, a naj zadostuje nekaj konteksta – konkordanc.

Nekaj napak je bilo tudi posledica razvoja transkripcije še med samim prepisovanjem – s kontrolo je bilo potrebno vpeljati zaključne načine zapisa: npr. prej *uesbepriključek, po novem *UeSBe-priključek; prej nestična končna ločila, po novem stična razen tropičja; prej na traku [pv] oznaka [medij], zdaj na traku [Ozadje] oznaka [navzočnost] (npr. za začetek/konec pogovora po mobitelu).

Popravljanje napak je potekalo na dveh točkah obdelave gradiva. Med samim transkribiranjem so nastajale spodnje tabele, na podlagi njihovih izkušenj so bile pregledane materinske transkripcije po koncu transkribiranja za vsak posamezen dan. Drugo, zadnje popravljanje napak je bilo izvršeno v začetni fazi analiz, ko so bile vse transkripcije že razrezane na besedila in v programu Eva združene v korpus v eni sami datoteki. Takrat so bili popravki vneseni v to skupno datoteko (iz katere so nastale analize) ter v tista razrezana besedila, v katerih so bile napake (v programu Elan) – predvsem je šlo za zapise posameznih besed. V materinske transkripcije pa se drugo popravljanje ni vračalo.

9.1.1. **Napake zapisa na verbalnih trakovih**

Tabela 33: Primeri površnega transkribiranja

<i>kontekst s površnim primerom</i>	<i>pravilen zapis</i>	<i>obrazložitev napake</i>
*ee *pal pa*s'n si večer *s'n vzela zdravila	*ee pa *s'n	Pomotoma vstavljen presledek med ločilo in besedo. Pozabljen presledek.
vi's tukaj sem mislil da je da je odcep	vi'is (videti)	Pravilen zapis je zelo približan aktualnemu izgovoru, v katerem je [i] v besedi nekoliko podaljšan in je v zapisu to transkribirano kot dva /i/, med katerima je glede na standardno obliko izpuščen /d/. Možen zapis bi bil tudi <*viš> (kar pomeni, da sta izpuščena 2 fonema, /d/ in njemu sledeči /i/ glede na standardno obliko), a ta zapis bi bil za konkretni primer fonografsko manj natančen.
*e *kol'k' vas je vas je sedem al' vas je več ?	kol'k' (koliko)	Zvezdica ni potrebna, saj ne gre za izpust sosednjih fonemov.
1) *dons 'mate pa samo ukaze 2) ja ja ... bom *pr'su bom *pr'su ...	*don's (danes); *pr'su (priti)	1) Nepravilni zapis spregleduje izpust /ə/, standardno za konkretni primer zapisanega z <e>. Zvezdica opozarja na to, da je nestandardni alofon [o] fonema /a/ interpretiran s transkripcijo kot standardni alofon drugega fonema /o/; v tej funkciji zvezdica ne more nadomestiti tudi opuščaja za izpuščeni /ə/. 2) Podobno.
1) da ne boš pol presenečen no 2) tak z'aj pa *s'n !	*pol (in pika na koncu) tak'	1) Pozabljena zvezdica pri pogostih problematičnih besedah in pozabljeno končno ločilo, sploh za medmeti na koncu izjave. 2) Pozabljen opuščaj.
*ahm nisem niti kaj posebej razmišljal	ahm (ahm)	Ta medmet je dokaj pogost in ima stanovitven zapis, ki ga je najti v več priročnikih, torej je zvezdica nepotrebna. Tudi konkretni izgovor ustreza tej obliki.

<i>kontekst s površnim primerom</i>	<i>pravilen zapis</i>	<i>obrazložitev napake</i>
*lej tole ti kar dam	'lej (glej)	Uporaba zvezdice besedo interpretira kot „leksikalno standardnost, samoniklost“ (skladenjskofunkcijsko kot medmet), glede na nepojavljanje oblike <lej> v jezikovnih priročnikih pa kaže besedo s pomočjo opuščaja raje interpretirati kot <glej> (skladenjskofunkcijsko glagol) z izpustom /g/.
daj poglej n- nisem *sigurn	*siguren (*siguren)	Problematična beseda je izgovorjena 3-zložno, jedro zadnjega zloga v izgovoru [ə] je potrebno tudi zapisati z <e>. Zvezdica ostane zaradi leksikalne nestandardnosti besede.
ampak v b- ... se da se no .	ampak	Izpuščena črka pri tipkanju.
a ta 'ma *ee *e *uesbepriključek za na računalnik?	*UeSBe-priključek	Pozabljeno pisanje oblikoslovnih vezajev (po pravopisu) in pozabljeno lastno načelo konvencije o pisanju kratic.
() .	(prazno)	Zapisana nerazložna poved, čeprav ni v njej razpoznavna niti ena beseda.
Nadaljevanje tabele gl. v naslednjem podpoglavju.		

Razne druge napake: 2 ali več presledkov zaporedoma namesto enega, preveč povedi v eni celici (partituro je v takih primerih težje obdelovati in pregledovati, interpolirane časovne reference so manj natančne), pri pogostih neknjižnih/nestandardnih besedah (npr. *te, *pol, *ahm) pozabljena zvezdica ali pri kaki neproblematični zvezdica dodana (npr. aha, mhm, probati, spod, not, skup, ma), pozabljeni podčrtaji pri večbesednih lastnih imenih – sploh kjer je več velikih začetnic, dvopičja pri vseh napovedih (in ne le pri napovedi citatov stavkov), vprašaji kar za vsemi vprašalnimi členicami idr.

9.1.2. Napake zapisa na trakovih [pv] in [Ozadje]

<i>kontekst s površnim primerom</i>	<i>pravilen zapis</i>	<i>obrazložitev napake</i>
nvg: polglasno požvižgovanje nvg: požvižgovanje nvg: izkašljanje nvg: pritajen smeh izg: s smehom :izg izg: mrmranje pojoč nvg: žvižga izg: pojoče izg: smejoč	nvg: žvižg polglasno nvg: žvižg nvg: kašelj nvg: smeh pritajen izg: smeh :izg izg: petje nvg: mrmranje :izg nvg: žvižg izg: petje izg: smeh	Razne napake na traku [pv]: dvodelne oznake niso sklepčne (en del manjka), opisi niso tipizirani itd.

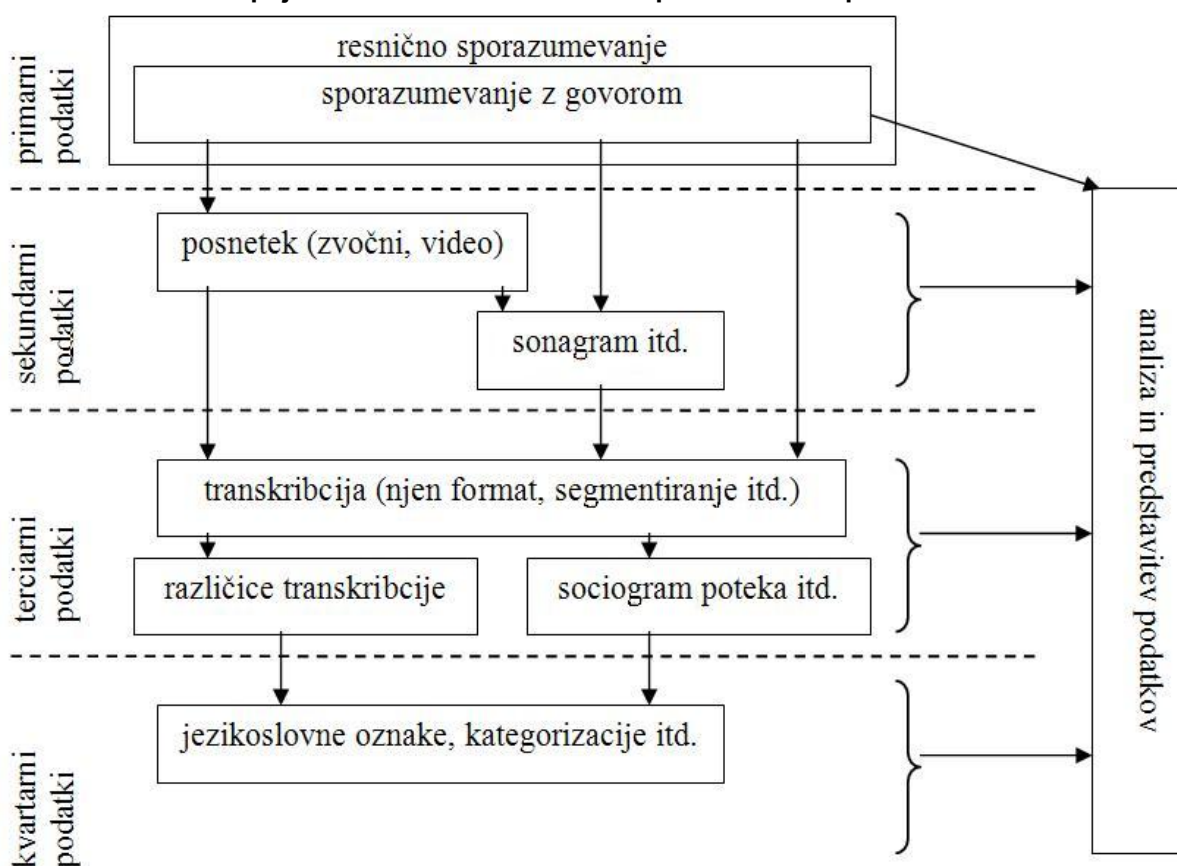
Razne napake na traku [Ozadje]: Opisi niso v lepih povedih, opisi so brez oznak, prvi in zadnji razrezani del razrezane transkripcije nimata oznake rez=/=rez, oznake rezov niso v svojih celicah

III. POTEK OBDELAVE GRADIVA

Ta tretji del diplome želi predvsem preprosto opisati potek dela pri obdelavi gradiva od začetka do konca diplome, v primerjavo z drugimi možnostmi obdelave gradiva se ne spušča, to sta opravila prejšnja dela diplome. Sledeča podpoglavja so razporejena približno tako, kot je potekalo zaporedje dela, nekoliko pa so koraki združeni glede na to, kateri podatki so v ospredju ene od 6 stopenj raziskave (za vsebino stopenj gl. Tabela 35 spodaj).

Iz postopkovnega vidika dela je dobro ohraniti zavest, da imamo z več obdelave vedno večji raziskovalni vpliv na podatke; proti koncu dela se večata delež abstrahiranja in možnost, da izgubimo izpred oči realno gradivo: resnično vsakdanje sporazumevanje iz preteklosti. To ponazarja naslednja slika:

Slika 12: Stopnje konkretnosti/abstraktnosti podatkov v korpusu¹⁴²



K sliki morda še dodam opombo, da primarni podatki, na katerih gradijo naslednje stopnje dela, niso le besedilo in njegovi verbalni in paraverbalni deli, ampak tudi okoliščine sporazumevanja. Analiza sama kot podatek pa je vedno eno stopnjo abstraktnosti višje kot njeno gradivo.

Z več stopenj obdelav (istih) osnovnih podatkov se večja tudi seštevka časa obdelave osnovnega podatka. Nekatere vmesne stopnje obdelave so neizogibne, čas pa lahko prihranimo z manj podrobno obdelavo ali z različnimi polavtomatskimi računalniškimi pomagali (katerih priprava je pri delu na kratek rok vendarle prezamudna), npr. v Evo vključen določevalnik (POSTagger, prim. Jakopin/Bizjak 1997: 523) ali spustni sezname o okoliščinah sporazumevanja v Accessu. Za marsikateri govorni korpus je zanimiv podatek, koliko časa obdelave pride na minuto posnetega govora, vendar so ti podatki neprimerljivi, če ne moremo primerjati tudi

¹⁴² Delno povzeto po Brinker/Sager 2001: 41.

kvalitete posnetkov, števila govorcev, natančnosti transkripcijskih konvencij in nadaljnje obdelave ter še česa. Gorjanc¹⁴³ navaja, da je za minuto posnetega govora potrebnih 6 ur dela, pri korpusu UKGS pa je bilo za golo transkribiranje brez teoretičnih odločitev potrebnih okoli 67 minut dela na minuto posnetka (preračunano po Zemljarič Miklavčič/Stabej 2005: 235).

Sam sem recimo za golo transkribiranje pri utečenem načinu dela (brez obdelave posnetkov in metaoznačevanja besed ter besedila) porabil:

Tabela 34: Zamudnost transkribiranja v korpusu StAnId

<i>enota stopnje .3Transkribcije</i>	<i>dolžina posnetka</i>	<i>število govorcev, razločnost</i>	<i>število besed</i>	<i>čas prepisanja</i>	<i>čas obdelave za posnetek</i>
ned_3+pon_1MD_3	76 '	18 govorcev (16 s TV, 2 v osebni stiku), večinoma razločno	5017	17 ur	13 ' za 1 ' posnetka oz. 295 besed na uro dela
ned_2MD_2	37 '	3 govorcev, vožnja v avtu slabo razločno, a redek govor	1079	3 ure 45 '	6 ' za 1 ' posnetka oz. 288 besed na uro dela

Dodatno delo, razvijanje zamisli in njihovo izpeljavo, je težko realno ovrednotiti; priprava na to diplomu in izdelava s korpusom sta trajala okoli 18 mesecev vsakodnevnega dela med delavniki, vključeno pa je tudi samostojno učenje priprave govorenega korpusa.

V naslednji tabeli je po zamudnosti ocenjena obdelava gradiva, razporejena po stopnjah obdelave. Tega, kar je bilo izpuščeno in je še možno dodati, je mnogo večja množica od nakazanih stvari (zadnje tri vrstice tabele). Bolj podroben opis dejanske obdelave pa je v naslednjih podpoglavjih. V njih po kratkem uvodu z navezavo na uvodne dele naloge sledi opis postopka ter osnovni rezultati dela in slike programov, pripomočkov. Datotečne statistike vsega korpusa so v zadnjem podpoglavju, III.13.

¹⁴³ Gorjanc 2005: 32, opomba 16; tam citira „Wallace et al. 1991: 70–71“, ki ga pa v seznamu literature na koncu Gorjančeve knjige ni.

Tabela 35: Zamudnost obdelave gradiva po stopnjah obdelave

<i>faza</i> <i>širina dela</i>	<i>.1Snemanje</i>	<i>.2Okoliščine</i>	<i>.3Transkribcije in .4Besedila</i>	<i>določevanje besednih vrst in lematizacija (opuščeno)</i>	<i>.5Analize</i>	<i>.6Informacije (predstavitev)</i>
<i>groba osnova</i>	oprema snemanja in presnemavanja	zapiski v dnevnik snemanja	nerazrezane transkribcije, tu in tam s časovnimi referencami; ortografska transkribcija	osnovni nabor besednih vrst in lematizacija	statistika besed, povedi, izjav, govorcev, besedil	zapis diplome, prilog; elektronski arhiv dela
<i>kratkoročno zamudna sistematizacija dela</i>	izbor kvalitet snemanja in digitalne kompresije	kategorizacija podatkov o okoliščinah (trakovi, sezname govorcev, vrednosti podatkov)	protokol transkribiranja oz. dodelava lastne transkripcijske konvencije; hierarhizacija trakov v partituri	pomožni leksikon najpogostejših besednih oblik in zglede problematičnih določitev; primerjava s strojnim določevalnikom (SAZU)	statistika za vsakega posameznega govorca in besedilo	urejen elektronski arhiv dela, arhiv programov in literature (tudi spletnih povezav)
<i>vsebinsko koristne širitve</i>	rezanje posnetkov po mejah besedil	organizacija teh podatkov v zbirki podatkov (Access)	razrez transkripcij na besedila po merilih za meje besedil; uporaba v različnih formatih (xml/eaf/txt); kontrola transkribcije	pomožne oznake govornih posebnosti, ki sledijo iz transkripcijske konvencije	pogostostni sezname besednih vrst, lem in oblik (za podkorpuse)	predstavitev primerov na internetu, prosojnicah
<i>vsebinsko luksuzne širitve: če bi bil čas</i>	na posnetkih brisanje podatkov o govornih (imena ...)	zapis teh podatkov tudi na datoteke posnetkov in datoteke besedil EXMARaLDA za uporabo v CoMA	razširitev in prilagoditev transkripcijske konvencije strojnim sistemom (npr. HIAT); nadomestitev imen govorcev v tekstu s šiframi	razširitev nabora oznak (v oblikoskladenjske določitve – MSD)	statistike besednih vrst, lem in oblik za posamezne govorce in besedila	priredba za konkordančnik EXAKT; objava dela korpusa na internetu, članek
<i>opuščeno zaradi obsega, prezahtevnosti</i>		vpis okoliščin naravnost v XML-strukturo transkripcij ali lastni DTD	zapis s podrobnimi časovnimi referencami; podrobnejši opis fonetike (npr. naglasi)	xml-definicija MSD-določevanja za EXMARaLDO; prilagoditev določevalnika v EVI; izvoz podrobnejše določitve iz Eve v Elan, EXM. (XML-format)	sociometrične, fonetične, druge kvantitativne analize; primerjave z drugimi korpusi	priredba za drug brskalnik po korpusu (NEVA, ASP32) in primerjava z drugimi korpusi
<i>a priori ni predmet raziskave</i>	govorni leksikon besed; videoposnetki	podrobnejše ankete o govornih, posnetek situacije sporazumevanja	fonetična transkribcija, razčlemba na diskurzne enote oz. posebnosti (npr. ponavljanja)	višje ravni označevanja: dependency treebank, pomensko oz. leksikografsko razvrščanje ipd.	govornotehnološke analize	institucionalizacija korpusa, spodbuda k nadaljevanju dela

10. SNEMANJE IN PRIPRAVA ZVOČNIH DATOTEK

Za zajem govorno-slušnega sporazumevanja je bilo izbrano snemanje z mikrofonom. Posneto naj bi bilo realno in naravno govorno sporazumevanje v slovenščini. Skupni imenovalac vseh okoliščin tega sporazumevanja naj bi bila ena oseba, nosilka mikrofona. Snemanje naj bi trajalo brez prekinitve (razen v času, ko je oseba minimalno aktivna – ko npr. spi), obdobje, v katerem naj bi bila posneta predvidena količina sporazumevanja (100.000 besed), pa je bilo vnaprej ocenjeno na 7 dni. Dejansko je snemanje potekalo od nedelje 10. 10. do sobote 16. 10. 2004.

Z izborom načina snemanja oz. medija (posameznik, na njem mikrofona) sem omejil tudi socialno okolje, v katerem zajemam gradivo. Ker se nosilec pojavlja malo v javnih okoljih, je več zasebne komunikacije, kot če bi bil mikrofona na fiksni točki ulice ali v neki ustanovi ali na drugi osebi, ki s bolj pojavlja v javnih situacijah ...

Snemal sem torej sam, kar je kasneje bistveno olajšalo pretvorbo posnetkov v računalniške datoteke, transkribiranje (prepoznavna govorcev) in opisovanje okoliščin sporazumevanja. Distanco do snemanja je bilo navidez težko vzpostaviti zaradi skrbi za snemanje (prilagajanje kvalitete posnetka, zagotavljanje napajanja in pomnilnika – minidiskov), a ocenjujem, da je bila vendarle v dobri meri vzpostavljena s pomočjo poskusnih snemanj in navsezadnje z navajenostjo na neprestan potek snemanja. Soudeleženci v zasebnih situacijah so bili povprašani o dovoljenju za snemanje teden prej ali kasneje (z nedoločno napovedjo časa snemanja), za javne situacije (formalne, uradne, na javnih mestih) dovoljenja za snemanje nisem predvidel. Že od začetka pa je bilo predvideno prikrivanje podatkov govorcev (imen ipd.) tako v glavi partiture kot v samih besedilih.

Količina in kakovost zajetih sporočil (to je katerih koli govorno-slušnih rezultatov sporazumevanja, ki jim je bil snemalec izpostavljen) sta najprej omejeni z zmogljivostjo snemalnih naprav, nadalje pa s predelavo posnetkov, s transkribiranjem in (drugim) označevanjem.

10.1. Snemanje

Za snemanje je bil izbran preprost mikrofona podjetja (Sony ECM-TS 125) velikosti „muhice“ za dvokanalno snemanje (stereo L – levo in R – desno). Imel je tudi ojačevalec na napajanje, ki pa ni bil potreben zaradi ojačevanja v minidisku-snemalniku. Mikrofona je bil skrito nameščen pod zapognjen ovratnik moške srajce, usmerjen proti sprednjemu delu telesa, kabel pa pod srajco napeljan do pasu, kjer je bil snemalnik.

Snemalo se je na 80-minutne minidiske (snemalnik Sony Net MD Walkman MZ-N710 MDLP). Pri snemanju je bila prilagajana kvaliteta snemanja s pomočjo funkcije MDLP. Robustnejše kot je bilo snemanje, daljši posnetek je bilo možno napraviti na en sam minidisk in dalj časa je snemanje trajalo (porabilo je manj napajanja). Možnosti so bile tri:

vrsta kompresije	največje trajanje posnetka	akumulator NH-10WM	baterija AA = LR6	NH-10WM + LR6
		snemanje	snemanje	snemanje
SP (L+R)	80'	7,5 h	12,5 h	26 h
LP2	160'	9,5 h	16 h	34 h
LP4	320'	13 h	19 h	41 h

Zmogljivost napajanja pri *predvajanju* iste posamezne kvalitete je 2,1–2,6-krat daljša.

Kombinirane so bile vse 3 možnosti (pri napajanju pa kombinacija obeh virov), največ pa se je seveda skušalo uporabljati način LP4; bistvene razlike pri kvaliteti posnetka govora ni. Časa za napajanje akumulatorja pa tudi ni bilo, saj je ponoči, ko ni bilo snemanja, potekalo presnemavanje iz minidiskov na računalnik.

Poleg prilagajanja kvalitete snemanja, preverjanja pomnilnika in napajanja je bilo med snemanjem potrebno skrbeti za označevanje posnetkov, izbor časa snemanja (v resnici se ni snemalo ves dan) in pisanje dnevnika snemanja – vse te obveznosti je opravljal snemalec na samem. Za označevanje je zadostoval pritisk na gumb in na minidisku je bil vnesen pisk in začetek novega posnetka; označbe so se zgodile ob menjavanju dejavnosti, prostorov (predvideno vnaprej ali tik za spremembo). V dnevnik je bila zabeležena zaporedna številka minidiska in posnetka na njem; za vsak posnetek pa kvaliteta snemanja, dolžina posnetka, umeščenost v dnevni čas ter dogajanje v posnetih situacijah (prostor, dejavnost udeležencev, udeleženci). Primer z okrnjenimi imeni udeležencev:

Primer 17: Zapis v dnevnik snemanja

oznaka posnetka	kompresija	trajanje	dnevni čas	dogajanje, udeleženci
cet-5+pet-1MD=3*	LP4	0:56:56	19.50 – 20.45	kuham večerjo, večerjam (gres); pri sostanovalcu T. na obisku dekleti M. in N.; pijejo mošt, jedo pečene kostanje, gledajo filme na računalniku, sostanovalec A. prinese domov 2 skakača (živalci)

*Oznaka posnetka pomeni, da gre za 5., zadnji minidisk v četrtek, ki je bil dalje uporabljan tudi še v petek, in za 3. posnetek na njem.

Posnetki dejansko ne potekajo neprestano od jutra do večera, ampak so vmes tudi izpusti. Do njih je prišlo bodisi nehote zaradi malomarnosti pri snemanju (zmanjkalo pomnilnika ali napajanja; tu gre za manjše izpuste, par minut) bodisi hote, ko je bil snemalec sam in tiho (zjutraj med vstajanjem in zvečer pred spanjem, med dnevom, ko je bil sam v hiši ali tiho v čitalnici; izpusti nekaj deset minut). Tudi za izpuste so bili narejene opombe v dnevnik snemanja in tako je kljub izpustom sklop posnetkov zadosten vir za sociogram aktivnosti posameznika v enem tednu.

V naslednji preglednici je podana statistika posnetkov po dneh; samo posivljena dneva sta bila naknadno transkribirana in vključena v korpus:

Tabela 36: Pregled posnetkov po dneh

dan	nedelja	ponedeljek	torek	sreda	četrtek	petek	sobota	skupaj
zap. št. MD + (posnetki) + dolžina	1 (5) 319' 2 (4) 323' 3* (3) 215'	1* (1) 29' 2 (6) 271' 3 (8) 193' 4* (2) 101'	1* (7) 162' 2 (5) 196' 3 (5) 141' 4 (4) 167'	1 (5) 197' 2 (5) 122' 3 (4) 192' 4* (4) 231'	1* (3) 80' 2 (4) 138' 3 (3) 106' 4 (4) 157' 5* (7) 158'	1* (2) 100' 2 (9) 219' 3 (6) 168'	1 (7) 271' 2 (3) 323' 3 (1) 96'	22 MD, 117 posnetkov
skupaj dolžina	14 h 17'	9 h 54'	7 h 46'	12 h 22'	10 h 39'	8 h 7'	11 h 30'	74 h 35'

Zvezdica pomeni, da so na istem minidisku posnetki iz dveh zaporednih dni.

10.2. Predelava posnetkov – prenos na računalnik

Pri tej verziji snemalnika digitalen prenos posnetkov na računalnik (preko USB-vhoda) še ni bil mogoč, novejši Sonyjevi snemalniki pa to možnost že imajo.¹⁴⁴ Tako je presnemavanje posnetkov z minidiskov na računalnik potekalo analogno, v realnem času posnetkov. Da bi se prihranilo pri nakupu minidiskov, je prenos posnetkov na računalnik potekal ponoči, med spanjem, preneseni posnetki so bili z minidiskov zbrisani in minidiski ponovno uporabljeni.

Prenašanje – presnemavanje je potekalo s pomočjo navadne zvočne kartice in programa Audacity, kot je opisano v poglavju I.5.2.1. Posnetki so bili shranjeni komprimirani v formatu .mp3 (en minidisk – en posnetek), deloma shranjeni na trdem disku deloma preneseni na zgoščenke zaradi slabih pomnilniških kapacitet računalnika.

Zaradi zamud pri študiju je bilo rezanje teh dolgih posnetkov na krajše posnetke (117, gl. zgornjo tabelo) opravljeno šele leto kasneje, jeseni 2005, tokrat že na boljšem računalniku.¹⁴⁵ Meje krajših posnetkov so bile določene s pomočjo označb – zvočnih signalov iz časa snemanja. Nekateri krajši posnetki so bili združeni, nekateri daljši (daljši od poldruge ure) pa še naprej razrezani; kot časovno mesto razreza je bil s predhodnim poslušanjem izbran trenutek, ko se sporazumevalna situacija spreminja. Končno število te stopnje razreza je po dnevih tako bilo:

¹⁴⁴ Poleg tega novejši Sonyjevi snemalniki poznajo tudi snemanje v načinu „Hi-MD“, kjer je možno na običajen minidisk v digitalni obliki shraniti do 240 MB podatkov, na poseben Hi-MD pa do 1 GB podatkov in s tem tudi daljši posnetek.

¹⁴⁵ Ta računalnik je bil uporabljan do konca, brez problemov: Intel Pentium 4 CPU 3.00 GHz, 2,92 Ghz, 512 MB RAM-a, operacijski sistem Windows XP Home Edition 2002, trdi disk 150 GB. Celoten korpus, vključno z uporabljanimi specifičnimi programi, je bil nameščen v svoji mapi neposredno na disku D:.

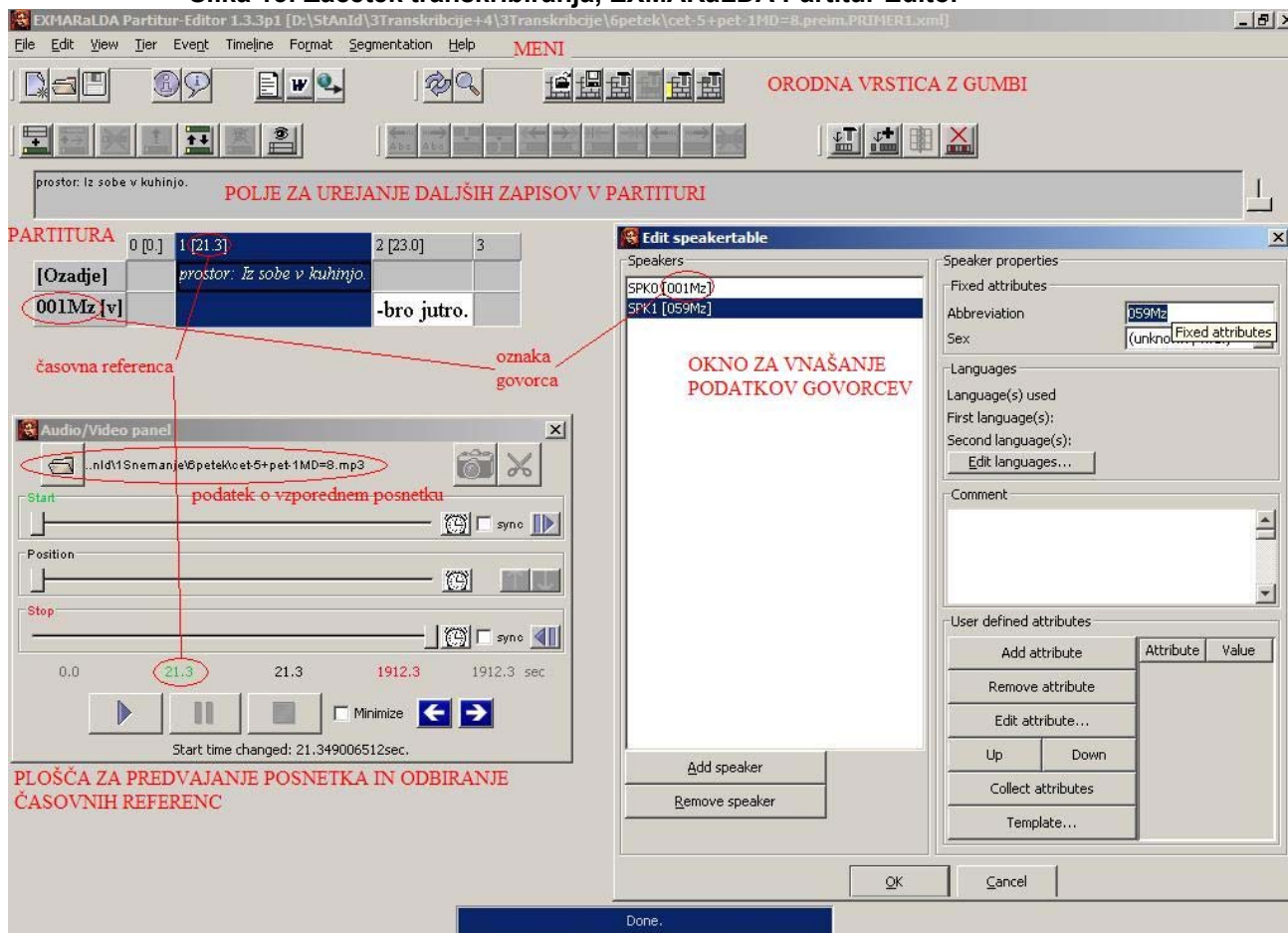
nedelja 16, ponedeljek 16, torek 16, sreda 16, četrtek 18, petek 17, sobota 14, skupaj 113 posnetkov.

Prvotni namen je bil transkribirati ves teden, transkribiranje se je začelo s prvim dnevom, nedeljo. Zaradi obsežnosti dela je bil načrt transkribiranja skrajšan na 2 dneva. Da bi dobili čim večjo pestrost sporazumevalnih situacij, hkrati pa še iste govorce, sta bila izbrana dan vikenda in dan delavnika; na podlagi kvalitete posnetkov in vsebine dogajanja je bil poleg nedelje izbran še petek. Na podlagi vsakega posnetka je nastala transkripcija, za ta dva dneva torej 16+16=32 transkripcij (na stopnji *3Transkripcije*).

Na stopnji *4Besedila* so bili posnetki razrezani (in zlepljeni) še tretjič, in sicer po merilih za meje besedil (poglavje I.3.2.1). Tako je v končnem korpusu razdelanih 149 posnetkov za izbrana dva dneva (153 besedilnih enot manj 4 neposnete). V prvotni zasnovi je bilo zamišljeno, da bodo v posnetkih izbrisani podatki govorcev (podatek poiskan v transkripciji, nato pa s pomočjo programa Audacity izbrisan še na posnetku), a je bilo to zaradi zamudnosti izpuščeno. Bi pa to bilo potrebno storiti v primeru predaje gradiva v druge roke zaradi vsaj delnega zakrivanja identitete govorcev.¹⁴⁶

11. ZAPIS „ČISTEGA“ BESEDILA IN OSTALEGA V PARTITURI

Slika 13: Začetek transkribiranja, EXMARaLDA Partitur Editor



Ko smo ustvarili transkripcijsko datoteko in poznamo v grobem že sporazumevalne dejavnike besedila, se začne finejše delo, podrobnejše transkribiranje govora na podlagi

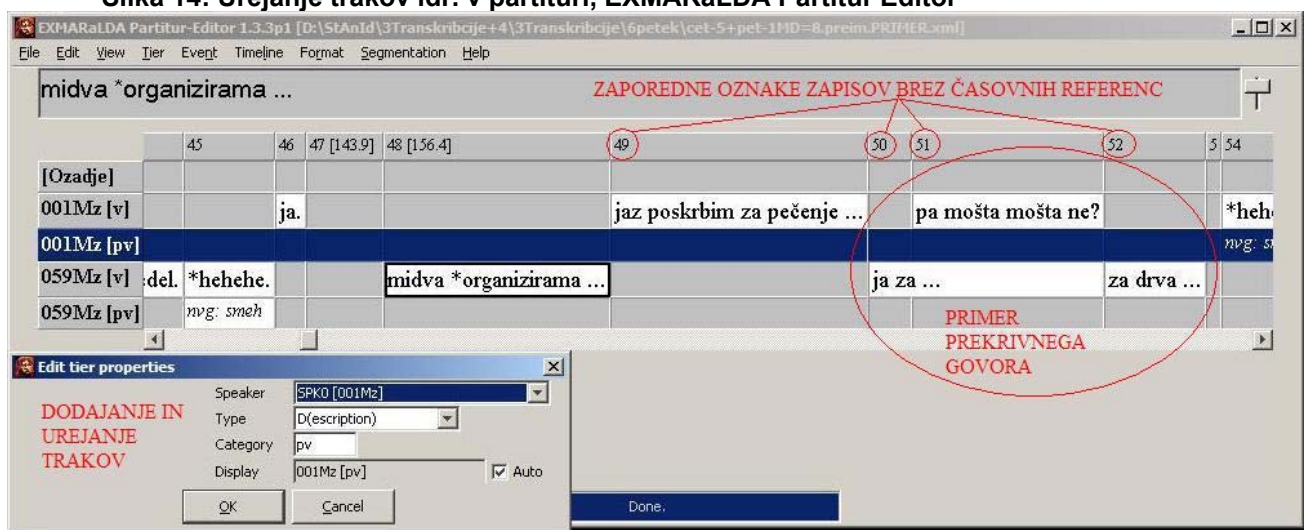
¹⁴⁶ Po današnjih metodah razpoznavanja identitete govorcev pravzaprav niti ne potrebujemo več imen ali drugih podatkov v govoru. Primer takšne raziskave je Dobrišek idr. 2006.

predvajanja zvočnega posnetka. Poleg kvalitete posnetka pride tu do izraza jezikovnorazločevalna zmožnost zapisovalca, transkribtorja, delo je opravljano ročno. Kar ga dela polavtomatskega, so pripomočki za predvajanje posnetka (zvočniki/slušalke, slike signala, krmiljenje predvajanja – prav pride tudi upočasnjevanje) in pripomočki za zapisovanje (od tipkovnice preko monitorja do programa za transkribiranje na računalniku).

Transkribiranje poteka v izbranem programu EXMARaLDA Partitur Editor (dalje okrajšano EXM.), gl. tudi poglavje I.5.2.2 ter zgornjo in naslednjo sliko. Po odprtju nove transkripcije se zanjo najprej vnese podatek o vzporednem posnetku. Po predvajanju posnetka zapisovalec vnese govor (na verbalni trak [v] k posameznemu govorniku) in ga segmentira na povedi in izjave.

Vsakega govornika je potrebno najprej vpisati v seznam govornikov in ga s tega seznama izbrati ob dodajanju posameznega traku na partituro. (Za več gl. pojasnila k podatkom govornikov, poglavje II.8.3.)

Slika 14: Urejanje trakov idr. v partituri, EXMARaLDA Partitur Editor



Opise paraverbalne plati sporazumevanja vnesemo prav tako h govorniku, in sicer na trak [pv], druge opise okoliščin pa na trak [Ozadje]. Traki imajo tudi logične oznake: za [v] oznaka *T*(*transcription*, 'prepis'), [pv] in [Ozadje] *D*(*description*, 'opis'), trakova z metaoznaki [BV] in [LEMA] pa oznako *A*(*notation*, 'metaoznaka'). Pri strojni obdelavi se te oznake (T, D, A) odražijo npr. tako, da so pri segmentaciji (štetje besed) razčlenjeni le trakovi tipa T, pri iskanju v konkordančniku pa lahko iskanje omejimo na določen tip traku.

Celice partiture se širijo avtomatsko v desno, vsebino in razsežnost posamezne celice pa lahko premikamo levo in desno po istem traku partiture. Govorne in paraverbalne zapise ter okoliščine sproti ali naknadno časovno umestimo s pomočjo gumbov na predvajalni plošči, torej brez ročnega vpisovanja. Če želimo transkribirati hitreje, časovnih umestitev (referenc) ne delamo za vsako celico (stolpec), ampak redkeje; manjkajoče časovne reference je program sposoben vpisati sam z interpoliranjem. Tako vpišemo časovne reference le na začetke in konce premolkov in izjav, katerih dolžina izstopa v govoru (so ali izjemno kratki ali izjemno dolgi). Tako površno vpisanih referenc kasneje seveda ne moremo uporabiti za kakšne govornotehnološke namene; hitreje kot v EXM. pa jih je mogoče popravljati v Elanu. Za primer segmentov govora v EXM. gl. Slika 10, str. 104, za partituro v Elanu pa Slika 8, str. 84.

Subjektivno je že razločevanje besed, pri katerem se nekako začne razumevanje govora, ki je osnova za zapisovanje, še težje pa je doseči konsistentnost zapisa. Čeprav je v pričujočem korpusu zapisanemu vzporeden posnetek, s pomočjo katerega je možno vse iz zapisa in še marsikaj nezapisanega preveriti, je transkribirano vseeno zelo važno za kvantitativne analize. Transkribiranje po določenih pravilih je za zapisovalca kot uporabljanje pravopisa v tujem jeziku

in mora biti priučeno, ozaveščeno. Če so pravila zapisa (s terminom: transkripcijska konvencija) podobna obstoječemu pravopisu, se jih zdi dokaj lahko naučiti, težko pa je pri njihovem ozaveščanju vztrajati. Pravila transkribiranja tekočega govora so v našem primeru opisana v *Transkripcijski konvenciji StAnId*, gl. poglavje II.7.

12. ZAPIS SPORAZUMEVALNIH OKOLIŠČIN

Okoliščine sporazumevanja ob samem snemanju niso bile tako sistematsko dokumentirane (po naprej predvidenih kriterijih) kot so zdaj v tej fazi zapisa; za natančnejšo analizo bi želeli še kakšen podatek. Bil pa je voden dnevnik snemanja (Primer 17, str. 129), okoliščine so razvidne iz poslušanja posnetkov (in iz tekočih zapisov v transkripcijah) in spominja se jih tudi zapisovalec korpusa, saj je bil v njih soudeležen; iz teh 3 virov sem zajemal ob zapisovanju okoliščin.

Za potrebe načrtovanih analiz je smiselno sistematično zapisati tudi sporazumevalne okoliščine. Kako smo logično strukturirali njihove vrednosti, je opisano v poglavju II.8 (tam gl. tudi za obrazložitev rabe podatkov, ki jih v spodnjih podpoglavjih navajam le s šifro – številko in imenom). Toda kako jih zapisati za potrebe strojne kvantitativne analize? Sledeče podatke bi lahko pripisovali vsaki datoteki posebej (posnetku, transkripcije v različnih formatih), v kolikor datotečni formati to omogočajo, a je to delo je mogoče poenostaviti in podatke bolje nadzorovati v programu MS Access, posebej na enem mestu (prim. poglavje I.5.2.3).

Stopnja obdelave korpusa *.2Okoliščine*, v kateri dopolnjujemo zbirko podatkov v Accessu, ni časovno enotna stopnja. Dnevnik snemanja je nastajal hkrati s stopnjo *.1Snemanje*, prvi vpisi v zbirko podatkov potekajo med stopnjo *.3Transkripcije*, glavnina po stopnji *.4Besedila*, malenkosti še kasneje, podatki pa so največ uporabljeni med stopnjo *.5Analize*.

Tudi samo snovanje zbirke podatkov in postopka vpisovanja vanjo je nastajalo kar precej časa, je pa zato samo vpisovanje in pridobivanje podatkov (z iskanjem, filtri in poizvedbami) toliko hitrejšo in lažje. Vsi podatki iz korpusa (ne samo sporazumevalne okoliščine) so bili razbiti v kar najpreprostejše skupine in razporejeni v preglednice (tabele), povezane med seboj z relacijami (Slika 9, str. 87). Vrednosti podatkov so bile oblikovane na različne načine, npr.: daljši zapisi za podatka 00 in 0 ostalo; omejitev dolžine opisa na 150 znakov – 22 dejavnost; vnosna maska za 01 IDbesedila (9 znakov = številka + 3 majhne črke + 2 števki + velika + mala črka + neobvezen znak +/-); številski zapis za podatke 10–13 izjave/povedi/besede; časovni format zapisa (hh:m ali h:mm:ss) za nekatere časovne podatke – 15–16 čas začetka/konca, 14 trajanje (posnetka). Vnos je bil najbolj omejen (in s tem zagotovljene njegova konsistentnost, pravilnost in hitrost vpisa) pri nastavitvi privzetih vrednosti (npr. 05 ELAN-Mime-Type) in omejenih seznamih vrednosti podatkov, ki se jih izbira iz t. i. spustnih seznamov (gl. vsi podatki v tabeli *00 nabori omejenih vrednosti* na zgornji sliki relacij. Za vsak podatek je zabeleženo tudi navodilo za vpis podatka, ki se ob premiku kurzorja v polje za vnos podatka prikaže v vrstici stanja (spodnji rob okna programa Access).

Postopku vpisovanja okoliščin so bili prirejeni t. i. obrazci, preko katerih je mogoče v določeni fazi aktualne podatke za eno ali več tabel hkrati vpisovati in razbirati. Imena obrazcev se od imen tabel ločilo po tem, da imajo za številko podčrtaj, npr. *05_besedila*.

Slike obrazcev in dela v programih so v naslednjih podpoglavjih; na njih nastopajo podatki o isti materinski transkripciji <pet-2MD=1>, o njenem hčerinskem besedilu <6pet09Dm> in o nekaterih od nastopajočih govorcev, npr. <060Žn>.

12.1. Preglednica transkripcij – stopnja *.3Transkripcije*

Manjša skupina podatkov je zabeležena za vsako transkripcijo, ki ima ime po posnetku; npr. <ned_1MD_2> in <cet-5+pet-1MD=8> sta prvi dve transkripciji vsakega posameznega dneva. Ko so po stopnji *.3Transkripcije* transkribirani vsi posnetki, je za vsakega v obrazec

prepisana skupina podatkov (glej obrazec *01_transkribcije*, Slika 15) iz dnevnika snemanja, iz datotečnega standarda (razvidnega v programu Raziskovalec za operacijski sistem Windows XP) pa sta prepisana še podatka o začetku prepisovanja (datum nastanka transkripcijske .xml-datoteke v EXM.) in trajanju posnetka. Vseh skupin vpisov je toliko, kot je posnetkov oz. transkripcij na stopnji *3Transkribcije*, torej 32. Iz teh zapisov, zbranih v tabeli *06 TRANSKRIBCIJE*, kasneje zajema še obrazec *03_razrez*.

Slika 15: Obrazec za vpis transkripcij, Access

01T) ID transkripcije	bet-2MD=1
06) datum snemanja	15. 10. 2004, petek
07) začetek prepisovanja	5. 6. 2007
14T) trajanje	0:21:25
15T) čas začetka	11:44
16T) čas konca	12:06
17) kraj	Ljubljana
19) kvaliteta	LP4
00T) dogajanje	Peš proti žel. P., mimoidoča vprašata za uro, nakup kajzarice in rogljička v pekarni, na bankomat v avli z blagajnami.

Zapis: 19 od 32

12.2. Preglednica govorcev

Vsak govorec ima dve identiteti, ime in šifro; s prvim (podatki II-IV) ime govorca/priimek/vzdevek) nastopa v transkripcijskih datotekah stopnje *3Transkribcije*, z drugim, I) ID govorca, pa na stopnji *4Besedila*. Zato obstajata tudi dve preglednici o govoricah, *03 imena govorcev* in *04 govorci*; prva je dostopna samo ustvarjalcu korpusa, druga, v kateri so zapisani še ostali analitični podatki o govoricah, pa splošno. S

tem so do neke mere zaščiteni osebni podatki resničnih govorcev, ki sodelujejo v korpusu. Tudi na obrazcu *02_govorci* (Slika 16) so podatki ene in druge preglednice ločeni; spodnji, pod črto, so iz preglednice *04 govorci*.

Slika 16: Obrazec za vpis govorcev, Access

03 I) ID govorca	060žn	II) ime govorca	28ZN
IV) vzdevek	mimoidoča	III) priimek	
04 I) ID govorca	060žr	IX) izobrazba	neznano
V) spol+število	ženska	X) razmerje do 01mo	neznanec
Va) sex	f	0) ostalo	Mimoidoča na ulici, ki vpraša za uro.
VI) poznanost	neznano		
VII) starost	neznano		
VIII) govorno poreklo	neznano		

Zapis: 60 od 103

Do konca transkribiranja posnetkov so govorce označeni torej z imeni, neznani govorce pa s pomožno šifro in mnemotehnično nalepko-vzdevkom, npr. na zgornji sliki 28ZN mimoidoča (za logiko teh šifer in ID govorca gl. Tabela 30, str. 118). Ko so vsi posnetki transkribirani, začnemo pri prvem (časovno zaporedje dnevov v tednu in nato znotraj posameznega dne) šifrirati govorce. Ime na glavi traka v transkribiciji nadomestimo s šifro, šifro vpišemo v obrazec 02_govorci, imena v zgornji del, ostale podatke o govorcih (iz dnevnika snemanja ali iz zapisovalčevega spomina) pa v spodnji del obrazca. Transkripcijsko datoteko shranimo z medpono <.preim.> , da ne zberemo starih podatkov o imenih. Ko bodo tako preimenovane transkribicije razrezane na besedilne enote, bodo v njih samo še šifre govorcev, le-teh pa ne bo potrebno vpisovati v vsako besedilo posebej.

Pri šifrah govorcev posebno pazimo na zaporedje števil; pomaga nam številka posameznega vpisa na spodnjem robu obrazca, ki jo Access avtomatično prikazuje. V korpusu za dva dni so 103 različni govorce.

12.3. Rezanje gradiva na besedilne enote oz. besedila

Posnetke in transkripcijske datoteke, katerih meje so bile nekonvencionalno in nereflektirano določene med snemanjem (ko ima snemalec to čas neopazno storiti, odvisno tudi od kapacitete pomnilnika) in presnemavanjem (velikost posnetkov), želimo zdaj po določenih kriterijih razločiti na besedila: razrezati in v nekaterih primerih tudi zlepiti. (Za kriterije gl. poglavje I.3.2.1.) Rezultat tega postopka je praktični predpogoj, da lahko potem **posameznim besedilnim enotam korpusa – „besedilom“** pripisujemo sporazumevalne okoliščine in posamezna besedila združujemo v (pod)korpuse za potrebe analiz. Iz „materinskih“ transkripcijskih datotek in posnetkov s stopnje 3Transkribicije dobimo „hčerinske“ na stopnji 4Besedila.

Postopek rezanja potrebuje kar nekaj dela: poznavanje vsebine transkripcij (sporazumevalnega dogajanja) in odločanje o mejah, zapis tega v skupno bazo podatkov in udejanjenje v rezanju ter zlepjanju transkripcijskih datotek in posnetkov. Delo poteka v programih EXM., Audacity in Elan. Da pa se pri tem ne izgubimo med množico „sestavnih delov“, dosti pripomore prav obrazec 03_razrez v Accessu, v katerem so zabeleženi vsi potrebni podatki o rezanju.

Vsaka enota na stopnji 4Besedila ima svoj vpis (vrstico) v tabeli 08 BESEDILA v Accessu, med seboj pa jih ločujemo na podlagi 3 vrednosti podatka 01r) razrezanost: celo | zlepljeno | del. Vseh besedilnih enot za dva dni je 192: 135 celih in 18 zlepljenih (tem 153-im so nato pripisane sporazumevalne okoliščine) ter 39 delov (ti po rezanju niso več zanimivi za analize, uporabo).

Nazadnje velja še opozoriti na to, da so v cele in zlepljene transkripcijske datoteke zapisane tudi opombe o tistih delih dneva, za katere posnetkov ni (snemanje prekinjeno ali opuščeno).¹⁴⁷ Če tako izbranih 153 enot stopnje 4Besedila razvrstimo po vrsti (od jutra proti večeru posameznega dneva), imamo s transkripcijami pokrito neprekinjeno obdobje dneva:

- za nedeljo 72 „besedil“ pokrije čas od 8.40 do 22.59 – 14 h 19';
- za petek 81 „besedil“ pokrije čas od 9.20 do 23.29 – 14 h 9';
- skupaj za 2 dni torej 28 h 28'.

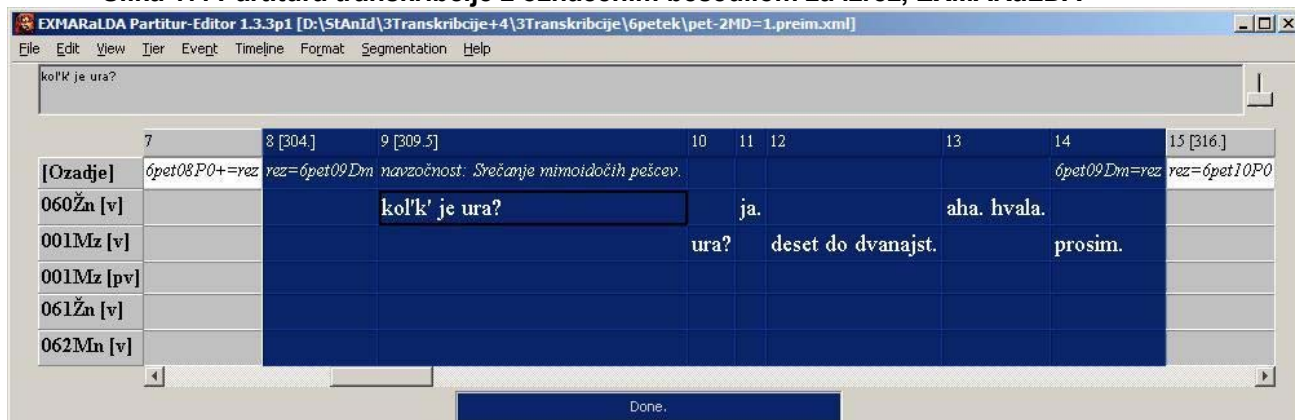
¹⁴⁷ Cele take transkribicije nimajo vzporednega posnetka (v celem korpusu so 4), zlepljene take transkribicije (le en njihov del je neposnet) pa imajo vzporedne posnetke (za neposneti del je umetno ustvarjen signal za tišino; takih je v korpusu 7, podatek o neposnetem delu je zapisan v 0) *ostalob*).

12.3.1. Zamišljanje rezov oz. besedil

Zamišljanje rezov – meja med besedili poteka med samim transkribiranjem, ob koncu transkribiranja posnetka pa poteka vpis posameznega reza hkrati v transkripcijsko datoteko in Access.¹⁴⁸ Kako so označeni vpisi v transkripcijskih datotekah (stopnja *3Transkribcije*):

- začetek besedila označuje oznaka `<rez=šifra-besedila>`, konec pa `<šifra-besedila=rez>` na partiturnem traku [Ozadje]; npr. `<rez=1ned01P0>` in `<1ned01P0=rez>` za prvo besedilo v korpusu; pozor: v oznaki `<rez=/=rez>` ni dvopičja kot pri drugih oznakah na trakih partiture;
- oznaka reza (začetnega in končnega) v vrstici [Ozadje] je v svoji celici partiture, pa četudi je potrebno to celico ali celotno časovno točko v partituri narediti na novo; časovni interval (trajanje) te celice je umetno zaokrožen na okoli eno sekundo;
- časovna točka reza je zaokrožena na sekundo (brez decimalnih mest) in vnesena ročno – ukaz *Timeline>Edit Timeline Item*; ta zaokroženost pride kasneje zelo prav – pri vnašanju podatka o trajanju v Access, pri označevanju intervalov za rezanje in zlepljanje posnetkov, pri zamikanju časovnih referenc partiture;
- kasneje so kot rezi v partituro (zlepljenih besedil) vpisani še neposneti deli.

Slika 17: Partitura transkribcije z označenim besedilom za izrez, EXMARaLDA



Ko v oznake rezov vpisujemo šifre bodočih besedil, s tem besedila že metaoznajuemo – posebno važni so zadnji 2–3 znaki šifre, ki se morajo kasneje ujemati v podatkih preglednice besedil. Npr. šifra `<6pet08P0+>` je sestavljena tako: `<6ped08>` (pomeni 8. besedilo šestega dne, petka), `<P>` (podatek `9) n-govor` bo **prazno**, brez govorcev), `<0>` (podatek `13r) gostota` govora bo `0`) in `<+>` (podatek `01r) razrezanost` bo **del**; namreč zaključni del bodočega zlepljenega posnetka, prvi del ima oznako `<->`).

Zamišljanje reza je zaključeno z vpisom podatkov rezanja v obrazec *03_razrez* v Accessu (Slika 18). V zgornjem delu obrazca so prikazani stari podatki iz preglednice transkribicije za posamezno materinsko transkribicijo, v tabelo v spodnjem delu obrazca pa vnašamo podatke o vseh hčerinskih rezih (na spodnji sliki so vpisani že vsi razrezani deli in celo zlepljeno besedilo – `6pet08Pn` ter `6pet12P0`). Na spodnji tabeli so zapisana polna imena podatkov o rezanju besedil, ki so na spodnjem delu obrazca na sliki nekoliko okrnjena.

¹⁴⁸ Zgornji postopek je bil uveljavljen pri prepisovanju drugega dne, petka, pri prvem dnevu pa je bilo rezanje narejeno šele, ko je bil prepisan ves dan, ker je bilo izoblikovanje načel rezanja mogoče šele ob večji količini rezov. Za prvi dan, nedeljo, načela mogoče niso tako dosledno rabljena – npr. prvi posnetek, dolg 20", je izbran zase, čeprav bi ga zaradi kratkosti mirno lahko združil s sledečim besedilom.

Tabela 37: Podatki o rezanju na besedila v obrazcu 03_rarez in viri teh podatkov

podatek o izrezanem besedilu	od kod ga dobimo
01) ID besedila	ob rezanju ga ustvarimo (na podlagi vsebine transkripcije)
09) n-govor, 13r) gostota, 01r) razrezanost	se ujemajo z deli šifre besedila – s podatkom 01) ID besedila
01Tr) ID transkripcije	ime materinske transkripcije – pred razrezom
22) dejavnost	del dogajanja, opisanega za materinsko transkripcijo v 00T) dogajanje
26) povod začetka, 27) povod zaključka	utemeljitev reza po teoretičnih merilih za meje besedil
00) ostalo	dopolnitev predvsem v kasnejših fazah; pri zlepljenih besedilih so naštetih sestavni deli
15r) začetni rez, 16r) končni rez	časovni referenci začetka in konca reza na materinskem posnetku
10r) interval	številski referenci začetnega in končnega stolpca (točke časovnega traku) partiture materinske transkripcijske datoteke v EXMARaLDI
14r) trajanje reza, 14) trajanje	najprej trajanje izrezanega posnetka v sekundah (izračunano iz razlike 15r) in 16r)), nato pretvorjeno še v minute in ure
15) čas začetka, 16) čas konca	glede na dnevni čas začetka/konca materinske transkripcije (15T) in 16T)) preračunani časi hčerinskih izrezanih besedil

Slika 18: Obrazec za vpisovanje in odbiranje razreza (primer hčerinskega besedila označen), Access

01) ID bese	09) n-go	13) gc	01r) ra	01Tr) ID trans	22) dejavnost	26) povod zač	27) povod zak	00) ostalc	15r) z	10r) interval	16r) k	14r) t	14) traj	15) č.	16) ča
6pet08P0+	prazr	0	del	pet-2MD=1	Pripravljanje nahrbrtr	dejavnost: spr	navzočnost: v		0	od 0 do 7	304	304	0:05:06	11:44	11:49
6pet08Pn	prazr	ne:	zlep	pet-2MD=1	Preoblačenje, stran	dejavnost: kon	navzočnost: v	Zlepljeno	0		0	1864	0:31:06	11:18	11:49
6pet08Pn-	prazr	ne:	del	pet-2MD=1	Preoblačenje, stran	dejavnost: kon	dejavnost: spr		0		0	1560	0:26:00	11:18	11:44
6pet09Dm	dvogc	mit	celo	pet-2MD=1	Srečanje na ulici.	navzočnost: vp	navzočnost: o		304	od 8 do 14	316	12	0:00:12	11:49	11:50
6pet10P0	prazr	0	celo	pet-2MD=1	Hoja skozi mesto.	navzočnost: od	prostor: vstop		316	od 15 do 18	891	575	0:09:35	11:49	11:59
6pet11Vr	večgc	red	celo	pet-2MD=1	Čakanje na nakup ii	prostor: vstop	dejavnost: zal		891	od 19 do 65	1038	147	0:02:27	11:59	12:02
6pet12P0	prazr	0	zlep	pet-2MD=1	Hoja v stavbo žel. p	dejavnost: zak	prostor: pristo	Zlepljeno.	0		0	329	0:05:29	12:02	12:08
6pet12P0-	prazr	0	del	pet-2MD=1	Hoja v stavbo žel. p	dejavnost: zak	vir: konec pos		1038	od 66 do 71	1285	247	0:04:07	12:02	12:06
*									0		0	0			

Ta obrazec je uporabljan še v nadaljevanju, ko iz njega pri serijskem rezanju posnetkov in transkripcij zajemamo podatke, kje napraviti reze.

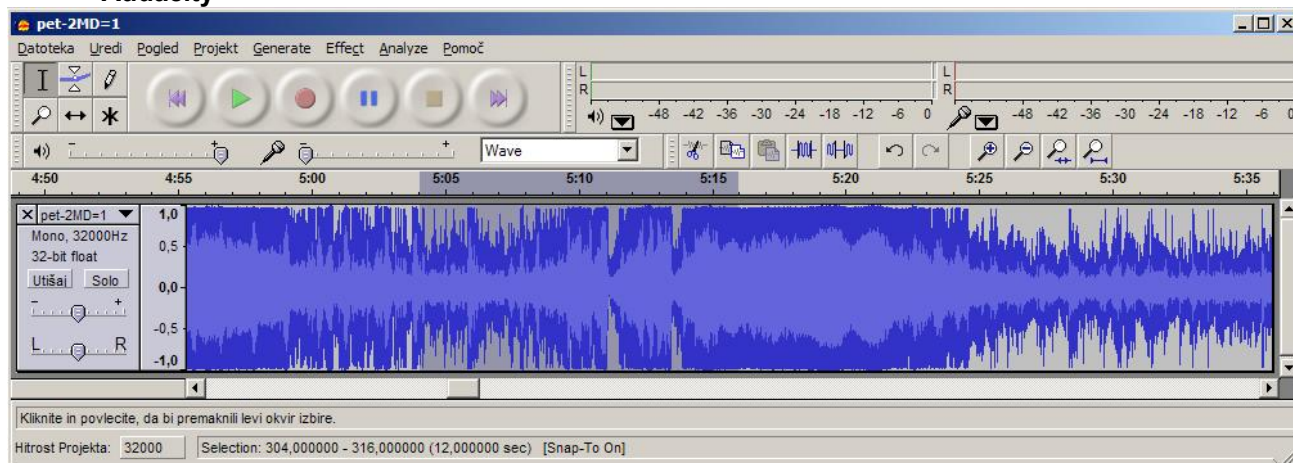
12.3.2. Rezanje posnetkov

Rezanje posnetkov poteka v programu Audacity (Slika 19), hkrati pa je na ekranu prikazan še obrazec *03_rarez* iz Accessa. Rezal sem vse materinske posnetke enega dne naenkrat, da je delo potekalo bolj serijsko, hitreje, postopek je predstavljen spodaj v alinejah. Za rezanje in zlepljanje petka sem porabil 2 h 2' (ok. 100" na besedilo; 100 celih + zlepljenih + delov). Postopek:

- uvozimo nerazrezan mp3-posnetek (*Datoteka>Open*);
- nastavitve: povečana skala posnetka, merjenje intervalov v sekundah (*Pogled>Set selection format: sec*), zaokrožanje intervalov na cele sekunde (*Uredi>Snap to: Snap on*);
- selekcija intervala posnetka za izrez: iz Access odčitamo začetni in končni čas reza, v Audacity označimo interval na spektrogramu; v vrstici stanja (na dnu okna programa) odberemo začetni, končni čas in čas trajanja označenega – vse preverimo v Accessu;
- mp3-izvoz (*Datoteka>Export Selection As MP3 ...*), za ime izrezanega posnetka damo vrednost podatka 01) ID besedila;

- zlepljanje posnetkov (za več gl. spodaj, poglavje III.12.3.5): začetni del shranimo kot običajen rez, drugega (ali dva druga, če sta) pa ne, ampak ga (ju) le prilepimo prvemu delu in vse skupaj shranimo kot zlepljen posnetek; če je del zlepljenega posnetka tišina (<Pn> v šifri dela besedila), najprej prilepimo kakršenkoli posnetek v odgovarjajočem trajanju, nato pa isti interval spremenimo v signal za tišino (*Uredi>Silence*).

Slika 19: Spektrogram posnetka transkribcije z označenim intervalom za izrez besedila, Audacity



V programu Audacity je možno iz posnetkov izbrisati tudi v govoru omenjena imena govorcev ali druge njihove osebne podatke, če jih želimo zakriti. To naredimo še na materinskih posnetkih, časovno mesto imena pa razberemo iz nepreimenovane materinske transkribcije v programu EXM. (da ni potrebno preposlušati celih posnetkov; podatki govorcev so v transkribcijah označeni z enačajem). Postopek je tak: v EXM. poiščemo časovno referenco imena oz. podatka in nato s poslušanjem označimo njegove natančne meje v Audacity, nato označeni signal spremenimo v tišino (*Uredi>Silence*); nato tak materinski posnetek pred rezanjem posebej shranimo, da ne izgubimo originalnega posnetka. Zaradi zamudnosti postopka brisanja podatkov govorcev iz posnetkov še nisem opravil.

12.3.3. Rezanje transkribcij

Razrezovanje materinskih transkribcij poteka v programu EXM., hkrati pa je na ekranu prikazan še obrazec *03_razrez* iz Accessa, iz katerega zajemamo podatke o rezih. Tudi to rezanje sem opravljal serijsko, za ves dan naenkrat. Za boljšo predstavbo gl. Slika 17 zgoraj; označeni del materinske transkribcije je moder. Postopek:

- v nerazrezani materinski transkribciji označimo interval transkribcije, ki bo izrezan; čas zaključka reza je desno od zadnje označene časovne točke – stolpca v partituri;
- pretvorimo označeni del v novo transkribcijo (*Edit>Selection to new*);
- v novi transkribciji pobrišemo nepotrebne trakove (prazne, na katerih ni besedila oz. oznak; vedno pa ostane trak [Ozadje]) – pazljivo, ker postopka ni mogoče razveljaviti;
- shranimo pod novim imenom – vrednost podatka 01 ID besedila;
- za neposneta besedila (<Pn> v šifri, za cele in dele), ustvarimo transkribcijo: vpišemo 4 stolpce (časovne reference) in celice na trak [Ozadje]: v prvo celico oznako začetka reza in začetni čas 0, v drugo celico oznako dejavnost: s kopijo podatka 22 dejavnost iz Accessa in začetni čas 1, v četrto celico samo zaključni čas trajanja neposnetega dela (ponavadi je na minuto okroglo), v tretjo, predzadnjo celico pa oznako zaključka reza s časom, ki je za eno sekundo manjši kot dolžina neposnetega dela.

12.3.4. Uskladitev razrezanih posnetkov in transkribcij

Razrezane transkribcije na tej točki rezanja še niso poravnane z razrezanimi posnetki (ampak še z materinskimi). Za to je potrebno v strukturi posamezne transkripcijske datoteke urediti dvoje: kot vzporedno zvočno datoteko nastaviti izbrani razrezani posnetek in zamakniti vse časovne reference za isto vrednost (namreč zmanjšati jih tako, da bo vrednost začetne časovne reference v transkripcijski datoteki enaka 0).

Drugega koraka urejanja program EXM. ne zna, zato si pomagamo s programom Elan. Z uvozom besedil v ta program hkrati posamezno transkripcijsko datoteko deloma pripravimo na nadaljnje stopnje obdelave (izvoz transkribcij v .txt-format za obdelavo v programu Eva in izvoz nazaj v EXM. za prikaz transkribcij v formatih .rtf in .html; gl. poglavje III.13).

Izvoz v Elan:

- v EXM. odpremo razrezano besedilo in ga izvozimo za Elan: (*File>Export >Elan*);
- v vmesniku za pretvorbo pustimo isti vzporedni posnetek kot za materinsko transkripcijo (tako bo odpiranje v Elanu brez problemov), kot vrednost podatka *ELAN-MIME-type*¹⁴⁹ pa vpišemo audio/*;
- shranimo pod istim imenom kot originalni .xml-zapis; nova datoteka dobi pripono .eaf.

Urejanje v Elanu (podatke za delo zajemamo hkrati iz obrazca *03_razrez* v Accessu):

- odpremo .eaf-datoteko izbranega besedila in izberemo zanjo nov, razrezan posnetek (*Edit>Linked Files ...>Add*) in izbrisemo starega (*Remove*), oboje potrdimo z ukazom *Apply* v oknu *Linked Files*;
- v prikazu transkribcije *Grid* (mreža) izberemo trak [Ozadje] in dobimo preglednico časovnih referenc začetnega in končnega reza;
- zamaknemo vse časovne reference transkripcijske datoteke .eaf (*Edit>Shift All Annotations ...*): iz Accessa odštejemo vrednost časovne reference začetnega reza (podatek [15r]; v sekundah, pomnožim v milisekunde); preverimo spremembo v pogledu *Grid* – podatek *Begin Time* za začetni rez mora biti zdaj ničelne vrednosti;
- popravimo podatke trakov v transkripcijski datoteki .eaf¹⁵⁰ (*Edit>Change Tier Attributes*): *Tier Name*: <ID govorca + presledek + [oznaka traku]>, npr. <001Mz [v]>, *Participant*: <šifra govorca>, npr. <001Mz>;¹⁵¹ preostale nastavitve pustimo, kot so (*Parent Tier, Linguistic Type, Default Language*); zapremo, shranimo.

12.3.5. Zlepljanje mejnih delov materinskih transkribcij v besedila

Ker materinski posnetki (in po njih transkribcije) niso bile ločene po načelih za meje besedil, se tu in tam zgodi, da konec predhodnega in začetek naslednjega materinskega posnetka po naših kriterijih za besedila sestavljata enoto. Torej je potrebno dela, ki smo ju pridobili iz materinskih transkribcij, še zlepit; včasih pa je med njima še tretji del za zlepljanje: čas brez posnetka (v šifrah besedil jih označujeta črki <Pn>). Po merilih za meje besedil pa običajno velja,

¹⁴⁹ Oznaka MIME (Multipurpose Internet Mail Extensions) pomeni standard za izmenjavo podatkov (komunikacijski protokol), in sicer za tiste dele e-pošte, ki se razlikujejo od teksta v znakovnem naboru US-ASCII. V Elanu je uporabljena kot oznaka večpredstavnostnih datotek (video, zvočnih); audio/* je najustreznejša vrednost za zvočni format .mp3.

¹⁵⁰ Zato da bodo oznake govorcev in trakov pri uvozu nazaj v format .xml programa EXM. urejene; pri izmenjavi med programoma oz. formatoma (oba imata XML-strukturo, čeprav ima Elanov format končnico .eaf) prihaja do nekaterih napak.

¹⁵¹ Za stalni trak v transkribcijah so končne nastavitve v Elanu takšne: *Tier Name*: [Ozadje], *Participant*: (brez).

da so skupaj zlepljeni lahko samo deli, ki imajo enako število govorcev (oz. iste govorce; torej P+P (brez govorcev), M+M, D+D, V+V) in enake druge okoliščine sporazumevanja.

Postopek zlepljanja posnetkov je že opisan zgoraj (poglavje III.12.3.2), transkripcije zlepimo v Elanu, nato pa moramo zlepljena besedila še vpisati v Access. Postopek:

- Elan, priprava za zlepljanje za posamezne transkripcijske datoteke delov: za vsak posamezni del izberemo ustrezen zlepljeni posnetek in popravimo oznake trakov, nato pa zamaknemo: prvemu delu je potrebno odšteti časovno referenco njegovega začetnega reza, drugemu delu (ki ima običajno časovno referenco začetka reza 0) pa *prišteti* dolžino (trajanje) prvega dela (pazimo, če dolžina ni okrogla na sekunde – to vidimo iz pogleda *Grid*, trak [Ozadje] za prvi del v Elanu); za morebitni tretji del (ki ima običajno časovno referenco začetka reza 0) prav tako prištejemo vsoto dolžin (trajanj) prvih dveh delov; vsak del posebej shranimo;

- Elanu, zlepljanje po dveh delov naenkrat: izberemo dela in skupno datoteko (*File>Merge Transcriptions ...*), nato še ustrezne trakove; nato preverimo rezultat, dopišemo v transkripcijo morebitna pojasnila in shranimo zlepljeno besedilo.

Vpis zlepljenih besedil v Access, v obrazec *03_razrez*:

- zlepljeno besedilo vpišemo pod materinsko transkripcijo prvega dela, za tista zlepljena besedila, ki imajo prvi del <P_n>, pa pod materinsko transkripcijo drugega dela;

- poimenovanje zlepljenih: če časovno v zlepljenem besedilu prevladuje del <P_n> nad <P₀>, ima šifra zlepljenega dokumenta oznako <P_n> (gl. primer 6pet08P_n, Slika 18 zgoraj);

- ostale podatke iz rezanja skombiniramo iz podatkov delov, prazni pa ostanejo podatki [10r], [16r] in [16r]; trajanje v minutah še preverimo v zapisu datotečnega standarda pri zlepljenem posnetku v Raziskovalcu.

Zlepljenih besedil je malo in je njihova obdelava nezamudna, čeprav zapletena; za vpisovanje zlepljenih dokumentov petka (9) sem porabil 50'.

12.4. Metaoznake besedil – zapis okoliščin sporazumevanja

Nazadnje sem opravil zapis vseh preostalih metaoznak in/oz. sporazumevalnih okoliščin besedil. Podatki in njihove opredelitve, tj. vrednosti, so razloženi v poglavju II.8.5, spodaj pa je razložen najprej vpis statistik enot za posamezne govorce v besedilu in kasneje za cela besedila. Metaoznake so vpisane samo za cela in zlepljena besedila (dele pa prepoznamo po dodatnem zadnjem sestavnem znaku v [01] ID besedila: <+> ali <->).

12.4.1. Preglednica govorcev v besedilih

Vloga in količina govora istega govorca v različnih besedilih (sporazumevalnih situacijah) sta različna in oboje želimo označiti tudi za potrebe naše raziskave. Teh dveh podatkov ne moremo zapisati niti v preglednico besedil niti v preglednico govorcev, zato pa Access ponuja možnost kombinirane preglednice. Vlogo govorca določi raziskovalec sam (subjektivno) in jo ročno vpiše preko obrazca *04_besedila* v Access (gl. sliko v naslednjem poglavju). Ko so preko tega obrazca vneseni podatki o vseh govorcih v vseh besedilih, zgleda izsek iz preglednice takole:

Slika 20: Preglednica govorcev v besedilih (označen primer besedila), Access

l) ID govorca	01) ID besedila	10OB) izjave	11OB) povedi	12OB) besede	XI) vloga	00OB) ostalo
059Mz	6pet04Mr	12	21	61	sin	
001Mz	6pet05Dr	5	7	27	sostanovalec	
059Mz	6pet05Dr	5	5	14	sostanovalec	
001Mz	6pet07Mr	44	53	245	sam	
001Mz	6pet09Dm	3	3	5	mimoidoči	
060Žn	6pet09Dm	3	3	6	mimoidoči	
061Žn	6pet11Vr	8	9	30	prodajalka	
062Mn	6pet11Vr	2	3	11	kupec	
001Mz	6pet11Vr	16	22	62	kupec	
063Žn	6pet13Dg	4	5	9	prodajalka vozovnic	

Zapis: 170 od 287
Referenca iz tabele 04 govorcei.

Messages
Tier [Ozadje]
Segmentation Event 3 segments of name e
Tier 060Žn [v]
Segmentation SpeakerContribution_Utterance_Word 7 segments of name HIAT:ip 4 segments of name HIAT:u 3 segments of name sc 6 segments of name HIAT:w
Segmentation SpeakerContribution_Event 3 segments of name sc 3 segments of name e
Tier 001Mz [v]
Segmentation SpeakerContribution_Utterance_Word 5 segments of name HIAT:ip 3 segments of name HIAT:u 3 segments of name sc 5 segments of name HIAT:w
Segmentation SpeakerContribution_Event 3 segments of name sc 3 segments of name e

Slika 21: Obrazec štetja segmentov besedila, EXMARaLDA

Partiturni format transkripcije in njena struktura v jeziku markup pa omogočata, da lahko za vsakega posameznega govorca *strojno* preštejemo jezikovne enote v njegovem govoru. Za nas to naredi program EXM., ki prešteje izjave, povedi, besede, ločila in celice na trakih govorcev.¹⁵² Za nas so uporabna le štetja prvih treh enot: izjav <sc>, povedi <u:> in besed <w:>; gl. slika levo.

Iz slike razberemo za trakove [v] tako govorce kot statistike osnovnih enot (ostale trakove, [pv] in [Ozadje], pustimo ob strani). Iz njega prepisujemo govorce v Access, obrazec *04_besedila*, po vrsti, kot nastopijo v transkripciji. Opis vloge govorcev v besedilih in statistika jezikovnih enot sta opravljena seveda le za tiste besedilne enote korpusa, ki sploh vsebujejo govor.

Preden pa opravimo štetje in rezultate štetij prepisemo v Access, moramo „zlepljena“ besedila uvoziti iz Elana („celih“ ni potrebno) in narediti manjšo zamenjavo ločil.

Postopek uvoz zlepljenih besedil iz Elana v EXM.in priredba poteka takole:

¹⁵² EXMARaLDA to štetje opravi le na trakih, označenih kot Transcription, ne Annotation ali Description; in sicer po pravilih transkripcijske konvencije HIAT, ki jih je prilagojen tudi StAnId. Možne pa so še statistike drugih enot, če so te vpisane po drugih transkripcijskih konvencijah (DIDA, GAT, CHAT, TEI, IPA) ali po lastno določenih segmentacijsko-statističnih pravilih.

- uvoz: *File > Import: Elan*; v vmesniku pustimo prazne vse 4 možnosti predelave partiture;
 - popravimo podatke govorcev: v *File > Speakertable* izbrisemo govorca <[]> (za trak [Ozadje] ga ne rabim); v *Tier > Tier Properties* preverimo nastavitve za posamezen trak (običajno je potrebno popraviti podatek *type* za traka [pv] in [Ozadje] v vrednost Description in podatek *display* na Auto);
 - uredimo vzporedno zvočno datoteko (če je potrebno): *File > Meta information > Referenced media file*;
 - uvoženo zlepljeno datoteko še shranimo.
- Zamenjava znakov, ki ovirajo statistiko jezikovnih enot (*Edit > Replace in events ...*):

<i>prvotni znak</i>	<i>zamenjava</i>	<i>razlog zamenjave</i>
tri pike (znaki) za tropičje	tropičje kot en znak	posamezna pika v HIAT-u pomeni končno ločilo; tri pike bi pomenile 3 povedi
<'>, <~>	<α>, <~>	HIAT bere prvotna znaka kot mejo med besedama (kot še presledek), njuni zamenjavi pa kot del besede

12.4.2. Preglednica besedilnih enot

Preglednica vseh metaoznak besedil vsebuje toliko podatkov (36 stolpcev), da ni naenkrat prikazljiva na ekranu. Na tej točki vpisovanja sporazumevalnih dejavnikov in drugih metaoznak so nekateri stolpci že izpolnjeni; preostale neobdelane vnašamo bodisi naravnost v tabelo bodisi preko zadnjega obrazca.

Vpis preko tabele:

- privzete vrednosti za vsa besedila, ki se ne spreminjajo: ker so za to stopnjo za vsa besedila enaka (podatek [A] ID stopnje korpusa, vrednost StAnId.4Besedila) ali ker služijo ob prenosu podatkov v drug program (v EXM. oz. CoMa – podatek [04] ELAN-Mime-Type, vrednost audio/*);
- privzete vrednosti, ki pred vpisom predvidene kot najpogostejše in jih je kasneje možno spremeniti: vsi številski podatki (na vrednost 0), [25] zavest snemanja (ne), [32] namen sporazumevanja (ZPČO), [34] funkcijska zvrst (PS);
- privzete vrednosti za določene tipe besedil: prazna besedila z gostoto 0 <-P0> imajo vrednost brez pri podatkih [20–21], [28–34]; prazna in neposneta besedila <-Pn> imajo vrednost neznano pri podatkih [13], [20–21], [28–34].

Preostale podatke o besedilih vpisujemo preko obrazca *04_besedila*. V njem ni vseh podatkov iz preglednice besedil, je pa prikazana preglednica govorcev, ki nastopajo v posameznem besedilu. Pred zadnjo fazo vpisovanja ima obrazec takole podobo, vneseni so tudi že podatki o posameznih govornikih v besedilih:

Slika 22: Obrazec za vpis metaoznak besedil pred ročnim vnosom preostalih oznak, Access

The screenshot shows a software window titled '04_besedila'. It features a table with the following data:

01) ID besedila:	0) ID govorca	10OB) izjave	11OB) povedi	12OB) besede	XI) vloga	00OB) ostalo
6pet09Dm	001Mz	3	3	5	mimoidoči	
01r) razrezanost:	060Žn	3	3	6	mimoidoči	
		0	0	0		

Below the table are several input fields and dropdown menus:

- 20) medij: [dropdown]
- 24) število soudeležencev: [dropdown]
- 25) zavest snemanja: [ne] [dropdown]
- 30) prostogovornost: [dropdown]
- 31) socialna zvrst: [dropdown]
- 32) namen sporazumevanja: [ZPČO] [dropdown]
- 33) govorni položaj: [dropdown]
- 34) funkcijska zvrst: [PS] [dropdown]
- 35) smer sporazumevanja: [dropdown]

At the bottom, there are navigation controls and a status bar showing 'Zapis: 105 od 192'.

Vpis preostalih metaoznank besedil preko obrazca:

- vpis na podlagi podobrazca: statistike jezikovnih enot posameznih govorcev iz podobrazca so seštete in vnesene v podatke [10–12], tudi število govorcev iz podobrazca je vneseno kot vrednost podatka [09];
- vpis na podlagi branja transkripcijske datoteke besedila, ki je hkrati odprta v programu EXM.: najprej podatki, ki zahtevajo ročen vpis ([18], 21), 28–29) ter [00]), nato pa podatki, pri katerih vrednost izberemo iz spustnega seznama (zadostuje vtipkati prvo črko izbrane vrednosti; [20], 24–25), 30–34)).

Slika 23: Obrazec za vpis metaoznank besedil, izpolnjen, Access

The screenshot shows the '04_besedila' form in Microsoft Access. It features a table with columns for speaker ID, transcript ID, and various counts. Below the table are several form fields for entering transcription details, such as start and end points, activity, location, and theme. On the right side, there are dropdown menus for selecting transcription parameters like 'osebni stik', '3-10', 'ne', 'prosto', 'pok. poq.', 'INFO', 'zasebni', 'PS', 'dvosmerno', 'brez', 'dvosmerno', 'enosmerno', 'neznano', and '1,5-smerno'.

l) ID govorca	10OB) izjave	11OB) povedi	12OB) besede	XI) vloga	00OB) ostalo
001Mz	3	3	5	mimoidoči	
060Žn	3	3	6	mimoidoči	
*	0	0	0		

13. PRIPRAVA ANALIZ

Obdelava gradiva se je zaključila z naslednjo stopnjo, strojno analizo transkripcij, urejenih kot osnovne besedilne enote korpusa, in z analizo okoliščin, pripisanih tem enotam. Sami rezultati analiz so predstavljeni v poglavju IV ter v prilogah na koncu diplome, v naslednjih treh podpoglavjih pa je predstavljen opis postopka.

13.1. Delo s transkripcijami – besedili

Od tam, kjer smo osnovne besedilne enote korpusa nehali obdelovati v programu Elan (zadnji popravki po razrezovanju materinskih transkripcij), naprej obdelujemo predvsem „polna“ besedila (dalje samo: besedila), in sicer iz njihove partiture samo trakove [v] (s transkripcijo govora). Takih besedil je v korpusu 115 (v njih 287 trakov [v]), ostalih 38 „praznih“ besedil pa je ostalo na stopnji obdelanosti transkripcijske datoteke z vzporednim posnetkom in metaoznankami besedil. Analize besedil so potekale v 4 programih (v tem redu: Elan, Eva, MS Excel, Elan, EXMARaLDA Partitur Editor), rezultati pa so bili vključeni v diplomu s programom MS Word. Poglejmo delo po posameznih programih.

1)) Elan (1.)

Pred koncem obdelave v programu Elan so bili iz vsakega besedila izvoženi [v]-trakovi v standardno .txt-datoteko (*File>Export As>Interlinear Text*), ki ima le nekaj vrstic oz. izgleda kot partitura. Nastavitve: odključani le [v]-trakovi, pokazane oznake trakov (ime govorca, oznaka [v], npr. <001Mz [v]>), časovne reference skrite, posamezen trak kot neskončna vrstica (neprelomljena!), kodiranje UTF-8. Besedilo (govorni dogodki) na trakih je razporejeno tako, da

je navpično prikazan naenkrat le govor na enem traku – prekrivni govor, kjer bi v partituri morale stati besede enega govorca nad besedami drugega, žal tu ne pride do izraza.

2))Eva

Tu so bile .txt-datoteke pretvorjene v format .eva in zaporedoma, kot si časovno sledi posneto sporazumevanje, dodana vsa besedila korpusa v skupno datoteko. V tej skupni datoteki ima vsako besedilo toliko vrstic (neskončnih, neprelomljenih), kolikor je v njem [v]-trakov, in še eno začetno vrstico, v kateri je ime besedila. Najdaljša vrstica korpusa, [v]-trak govorca 003Mz v besedilu *1ned25Vg*, znaša 22.539 znakov. Skupna datoteka ima 402 vrstici in v pomnilniku obsega 1,9 MB. Na naslednji sliki vidimo levi zgornji začetek te datoteke:

Slika 24: Začetek skupne korpusne datoteke, program Eva

```

◆1ned02Dm.eaf
001Mz [v] tak' z'aj pa *s'n! dobro jutro!
002Žz [v] tak' ... dobro jutro!
◆1ned04Vg.eaf
001Mz [v] *vijadva sta tud' vstala ne? ob osmih ...
002Žz [v] () ata je že
003Mz [v]
◆1ned05Mm.eaf
001Mz [v] človeško življenje je večno iskanje miru. *papapapapapa
◆1ned06Dm.eaf
001Mz [v] adi jo se vidimo!
002Žz [v] adi jo!

```

Skupna datoteka je bila najprej izhodišče raznih popravilanj:

- zamenjave ustreznih ločil naše transkripcije in nabora UTF-8 za potrebe obdelave v Evi (nedeljivi vezaj, podčrtaj v nedeljivi presledek, tropičje),
- drugo popraviljanje transkripcijskih napak,
- preimenovanje podatkov govorcev v besedilu z ustreznimi šiframi.

Za zadnja dva popravka so bile narejene zabeležke, v katerem besedilu je potrebna sprememba in mesto te spremembe, in po koncu popraviljanj in preimenovanj so bile zabeležke sortirane po besedilih in v ustreznih transkripcijskih datotekah preko programa Elan bili narejeni še zadnji popravki.

Za raven korpusa so bile s pomočjo uporabe ukazov programa Eva iz nepreimenovanega korpusa narejene nebesedne statistike: štetje znakov, dolžine in pogostosti povedi ter ustrezni grafi.

Besedne analize so bile narejene bodisi iz nepreimenovanega bodisi iz preimenovanega (podatki govorcev v besedilu) korpusa: štetje pojavnic, različnic, sortiranja, izračun deležev različnic, graf rasti besednega zaklada, število posameznih znakov v besedah (npr. opuščajev). Glede na različne potrebe seznamov različnic (prim. abecedno razvrstitev, Priloga 3, in seznam najpogostejših različnic v podkorpusu prostega govora, Priloga 4) so bila prilagojena pravila sistemske tabele razloma na besede (npr. znaki <*'|=|(|)> kot del besede po pravilih transkripcijske konvencije StAnId). Za razlom besedila na besede sta bila uporabljena ukaza menija *Data>Right side routines>Break text into items*, ali *Standard* (ki se odziva na omenjeno sistemska tabelo) ali *Dictionary* (redkeje – za izločitev besedila, zapisanega v oklepajih, kar pomeni negotovo transkripcijo).

Besedne analize so bile narejene bodisi za cel korpus bodisi za podkorpus besedil; deli slednjega (besedila ali posamezni [v]-trakovi) so bili predhodno izbrani na podlagi statistik sporazumevalnih okoliščin v programu MS Access. Za vsak podkorpus oz. gradivo iz slogovnih primerjav so bile shranjene 3 datoteke: nerazlomljen podkorpus, seznam različnic s pogostostmi in deležem v podkorpusu ter 20 najpogostejših različnic z deležem v korpusu; prvi dve v formatu .eva, zadnja v formatu .txt, ki je bil nadalje obdelan v programu MS Excel.

3))MS Excel

Ta program je bil uporabljan le za izdelavo seznamov (različnic po pogostostih) za potrebe slogovnih primerjav (za boljšo predstavo gl. npr. Primerjava 2, str. 159). Najprej so bili za vsa gradiva v posamezni primerjavi uvoženi sezname 20 najpogostejših različnic z njihovimi deleži v gradivu. Da bi lahko izračunali povprečne deleže za vse različnice v primerjavi, so bili manjkajoči podatki (redkejše različnice z njihovimi deleži) ročno poiskani v seznamih vseh različnic posameznega gradiva v Evi. Nato je bila vsaki primerjani različnici primerjave izračunana razlika do povprečne pogostosti, različnice, ki se niso pojavile v vseh gradivih posamezne primerjave, pa so dobile posebno oznako (BP ali pp). Nazadnje so bili s pomočjo sortiranja v zapis diplome preneseni sezname najpogostejših 20 različnic za gradiva in označene petere različnice, katerih delež v posameznem gradivu najbolj odstopa od povprečnega deleža različnice v primerjavi.

4))Elan (2.)

V programu Elan je bilo v posamezne datoteke transkripcij potrebno vnesti še zadnje popravke napak in šifrirati podatke govorcev v besedilih (ukaz iskanja in zamenjave). Podatke govorcev, ki pa se pojavljajo na drugih trakovih ([Ozadje], [pv]) v transkripcijskih dateotekah in ki jih v verbalnem delu korpusa ni bilo potrebno popravljati, sem popravil s pomočjo iskalnika v Elanu, ki deluje kot preprost konkordančnik (*Edit>Search multiple .eaf files*). Išče po transkripcijskih datotekah, ki smo jih nastavili, kot rezultat pa prikaže: govorni dogodek (časovno omejen interval traku, angl. annotation), v katerem se nahaja iskani znakovni niz, in njegove časovne reference (začetni in končni čas ter dolžina); oznako traku in datoteke; predhodni in naslednji govorni dogodek. Klik na besedilo nas prestavi v partituro posameznega besedila, kjer lahko poslušamo okolico zadetka. Rezultate je možno shraniti kot .txt-datoteko, v kateri so podatki ločeni s tabulatorji.

Konkordance v diplomi, gl. Priloga 5, so narejene s pomočjo tega iskalnika in nekoliko poenostavljene. Iskati je možno po pravilih naprednega iskanja znakovnih nizov na podlagi jezika java (angl. regular expression search, Hellwig/Uytvanck 2006: 128 sl.), na katerem sloni Elan. Osnovni iskalnik v Elanu (za posamezno transkripcijsko datoteko) pa ponuja poleg tega precej naprednejše možnosti iskanja: iskanje po kombinacijah trakov, iskanje po časovnih referencah itd. (npr. lahko poiščemo vse primere prekrivnega govora, ki trajajo manj kot 1 sekundo).

5))EXMARaLDA Partitur Editor (EXM.), CoMa in EXAKT

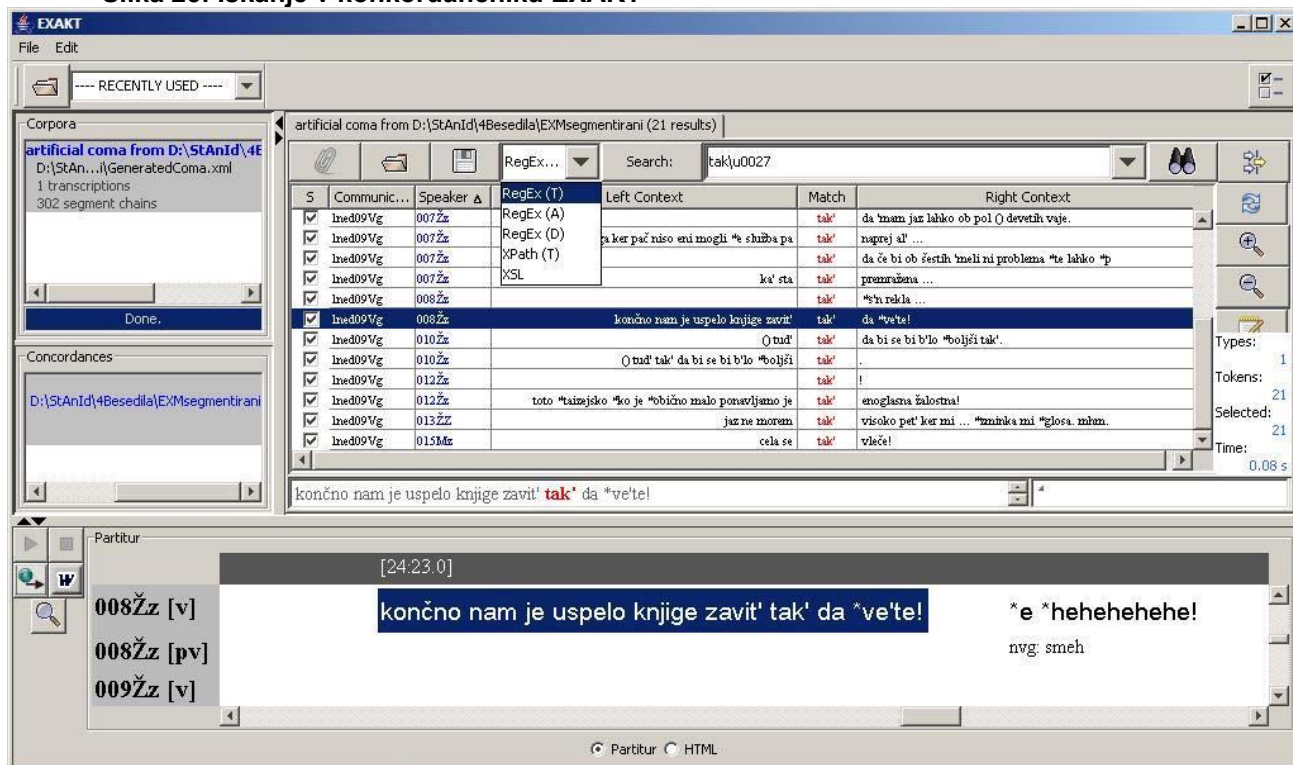
Besedila oz. transkripcijske datoteke lahko iz Elana uvozimo še v EXM. (kako, gl. poglavje III.12.4.1), npr. da besedilo lahko izvozimo kot partituro v format .html ali .rtf – za primer slednjega gl. Priloga 10. Lahko pa datoteke priredimo (avtomatska sprememba XML-strukture iz „basic“ v „segmented“ transcription) za nadaljnjo obdelavo: najprej za urejevalnik korpusov CoMa in nato za konkordančnik EXAKT¹⁵³ – gl. naslednja slika.

Konkordančnik EXAKT ima isti iskalni jezik kot Elan, prav tako omogoča nastavitve datotek, po katerih iščemo, in preskok iz konkordanc v partituro, na podlagi katere lahko predvajamo posnetek. Nekoliko naprednejši je v tem, da to iskanje po več datotekah razločuje tri

¹⁵³ V letu 2007 je pri projektu izšla nova verzija iskalnika po korpusu, imenovana EXAKT (EXMARaLDA Analyse- und Konkordanztool), verzija 0.2. (Schmidt/Wörner 2007). Predhodnji iskalnik z zelo podobnimi lastnostmi, le nekoliko manj sposoben, pa se je imenoval Zecke (Ziemlich einfaches Konkordanztool für EXMARaLDA).

tipe trakov v partituri (transkribcije, opisi, oznake– T, D, A), da je izbere datotek (podkorpuse) možno shraniti, da je levo in desno okolje iskanega niza možno širiti in ožiti, da je možno prikazati metainformacije o besedilu (če smo jih zapisali v transkripcijsko datoteko) ter da je rezultate možno shraniti v formatu XML.

Slika 25: Iskanje v konkordančniku EXAKT



13.2. Izkoriščanje podatkov o okoliščinah sporazumevanja

Podatke, ki so bili pripisani vsaki besedilni enoti korpusa v programu MS Access, obdelujemo prav tam. (Možnost prenosa podatkov v transkripcijske datoteke formata XML, ročno ali preko programa CoMa, projekt EXMARALDA, sem opustil, tako da so nekateri podatki, predvideni samo za potrebe tega prenosa, odpadli kot nepotrebni.)

Preostale važne podatke iz različnih tabel (npr. 153 celih in zlepljenih besedil, od teh dalje pa le 115 „polnih“) je bilo preprosto mogoče izbrati s poizvedbo. Nadaljnje poizvedbe pa so omogočile statistike okoliščin sporazumevanj (deleži posameznih okoliščin po njihovem deležu besed v korpusu), razne izračune (npr. podatek [13] besede/s – gostota besed na sekundo v posameznem besedilu) itd. – prim. tabele v poglavjih IV.15 in IV.16.

Za podatke, katerih vrednosti ob vpisu niso bili kategorizirani (npr. [28] oblika sporazumevanja), je bilo potrebno vpise kategorizirati naknadno – v omenjenih tabelah so ti podatki posivljeni. Najprej je bila narejena kopija tabele, nato vpisane vrednosti zamenjane s kategoriziranimi in nazadnje je bila narejena poizvedba, ki je izračunala statistiko novih vrednosti, ki so predstavljene v tabelah. Podatke bi bilo v programu Access možno enostavno tudi serijsko popravljati, zamenjevati, primerjati med seboj v križnih tabelah itd.

Narejene poizvedbe niso služile le statistikam, ampak tudi izboru podkorpusev. Rezultat izborov v Accessu je bil seznam šifer besedil (ali tudi zgolj govorcev v besedilih), na podlagi teh pa je bil v skupni datoteki korpusa v Evi ročno narejen izbor besedil podkorpusev.

13.3. Končna mapa in datotečna struktura

Ko smo z obdelavo gradiva pri kraju, lahko govorimo o končni mapni (direktorijski) in datotečni strukturi korpusa, saj se ta ne spreminja več; nastajajo le že nove raziskovalne obdelave in predstavitve. Končno mapno strukturo želim predstaviti zaradi pregleda korpusa v računalniški obliki (kot je razvidno sicer na zgoščenki, Priloga 12, in na spletni predstavitvi dela) in tudi zato, da bo jasno, kaj na zgoščenko oz. v gradivo za druge uporabnike ni vključeno, pa je ob diplomi vendarle nastalo. To moje interno gradivo je v naslednji preglednici posivljeno.

Tabela 38: Mapa in datotečna struktura korpusa, interno in javno

<i>mapa</i>	<i>velikost</i>	<i>podmape, tip in število datotek, opis</i>	
\1Snemanje	6,9 GB	8 podmap: 7 za vsak dan snemanja v celem tednu, 1 mapa za nerazrezane posnetke, po enega s posameznih diskov; 122 datotek: 1 od tega tekstovna – popis posnetkov, ostalo .mp3 datoteke	
\2Okoliscine*	14,1 MB	*datoteka StAnId.mdb (zbirka podatkov o okoliščinah)	
\2Okoliscine** (drugo)	14,7 MB	** 4 podmape s poskusi zbirk podatkov, varnostnimi kopijami ter poskusi podatkov o okoliščinah v programu CoMa; 18 datotek (.mdb, .odb, .xml, izvozi zbirk podatkov v druge formate jezika markup)	
\4Besedila\3Transkribcije	1,7 MB	3 podmape z materinskimi transkribcijami po obeh dnevih (nedelja, petek), vsemi nepreimenovanimi in za petek preimenovanimi; 52 datotek .xml (struktura programa EXM.)	
\4Besedila\Elan	3 MB	po 149 transkripcijskih datotek .eaf in 149 njim vzporednih .mp3;	v podmapi primer transkribcije z metaoznaki besed
	1,4 GB		
4Besedila\Elan-EXM 4Besedila\EXMsegmentirani	1,9 MB 1,8 MB	transkripcijske datoteke be4sedil, iz Elana uvožene v EXM. in v drugi mapi segmentirane za potrebe konkordančnika EXAKT; dvakrat po 115 datotek .xml	
4Besedila* (drugo)	394,6 MB	* 3 mape s 24 podmapami, v katerih so nepreimenovane transkribcije iz EXM. takoj po razrezu (.xml), iz Elana (.eaf), ter poskusi transkribiranja še v programih Praat in Transcriber (datoteke teh dveh programov in vsi formati, v katere morejo vsi 4 programi izvažati transkribcije); in zvočne datoteke .mp3 in .wav; skupaj 429 datotek	
\5Analize\korpus	16,9 MB	5 podmap z 51 datotekami (večinoma .eva, nekaj .txt, .gif in .jpg); preimenovana in nepreimenovana korpusna datoteka ter razne statistike (različnic, znakov, povedi), primeri konkordanc (.txt, .eva, .doc) in histogrami	
\5Analize\podkorpusi	13,6 MB	68 datotek (.xls – Excel, .eva ter dve slikovni); podkorpusi in sezname različnic za potrebe slogovnih primerjav	
\5Analize\EvaPriredbe	9,9 KB	9 datotek z makroji in spremembami sistemskih datotek v Evi	
\5Analize\Doloevanje	180 KB	3 podmap s 15 datotekami: poskusi POS-tagiranja (.eva) in priprave transkripcij nanj (.eaf, .etf)	
\6Informacije*	5,5 MB	* datoteka z zapisom diplome (.pdf)	
\6Informacije** (drugo)	18,3 MB	** 6 podmap, 75 datotek (tekstovnih in slikovnih); pomožni dokumenti in slike pri nastajanju diplome	
\literatura	719 MB	263 podmap, 4709 datotek s študijskim gradivom, korpusnimi zgledi (npr. MultextEastV3) in literaturo v elektronski obliki	
\programi	393 MB	195 podmap, 2130 datotek; uporabljeni in preizkušani programi za transkribiranje, obdelavo zvočnih in tekstovnih datotek ter obdelavo in brskanje korpusa (brez MS Office) ter njihov zgledi in literatura	
\izbraniposnetki (le v javnem delu)	41,2 MB	za 5 izbranih besedil v korpusu po 3 tipi transkripcijskih datotek (.xml, .seg.xml in .eaf) in vzporeden posnetek .mp3; 1 primer transkribcije v .rtf	
\izbraniprogrami (le v javnem delu)	96,6 MB	osnovne instalacijske datoteke in navodila za uporabo programov Elan, EXM., CoMa in EXAKT	
javni del korpusa	194,6 MB	gl. Priloga 12	
interni del korpusa	9,8 GB		

IV. ANALIZE

Sledeče analize korpusa potekajo na 5 ravneh, različno kompleksnih, ki smo jih napovedali že v poglavju I.4.3.3: korpus; podkorpusi, izbrani po okoliščinah; govor govorcev, govor govorca proti sobesedilu in vloge govorca med seboj. Na vsaki ravni začenjam s pregledom značilnosti gradiva (okoliščin in jezikovnih sredstev), nato pa sledi slogovna (stilometrična) analiza jezika po izbranih primerjavah. Pregledi in analize so sproti komentirani, nekaj komentarjev se nanaša na daljše preglede, ki so v prilogah; povezava ugotovitev je podana v poglavju V.

14. KORPUS KOT CELOTA

Oznaka „korpus kot celota“ se nanaša na podatke o besedilnih enotah, ki zajemajo dva dneva sporazumevanja izbranega govorca z njegovim okoljem, nedeljo in petek. Ostalih 5 dni v tednu, za katere so sicer bili narejeni posnetki in dnevnik snemanja, v korpus ni vključenih. Še enkrat velja opozoriti, da ta korpus kot celota ne predstavlja uravnoteženo (po)govornega jezika, ker je zasnovan sociometrično in je „govoren“ v prenosniškem smislu. Prosto govorno sporazumevanje sicer v njem predstavlja večino, a ima neprosti govor (branje, recitiranje, petje) pomemben delež, ki pride do izraza v jezikovnih statistikah.

Za pregled tehničnih karakteristik korpusa (velikosti datotek, mapna struktura) gl. prejšnje poglavje (III.13.3), v naslednji tabeli pa so podani osnovni podatki o besedilnih enotah v korpusu. Podrobneje bodo ti podatki razdelani na nižjih ravneh primerjav, kjer pa se bodo nanašali *samo na „polne“ besedilne enote*, to je vse tiste, ki vsebujejo vsaj eno transkribirano besedo govora – to je *korpus v ožjem pomenu – zbirka besedil*.

Tabela 39: Pregled korpusa po št. govorcev, dnevih in deležu prostega govora

BESEDILNE ENOTE	nedelja			petek			korpus (dneva skupaj)			
	besedil	besed	posnetki	besedil	besed	posnetki	besedil	besed	delež	posnetki
<i>večgovor</i>	18	21.302	4:53:39	14	15.614	2:41:54	32	36.916	76 %	7:35:33
<i>dvogovor</i>	17	5.367	1:26:00	25	4.673	1:25:34	42	10.040	21 %	2:51:34
<i>monolog</i>	22	496	3:48:07	19	1.058	2:11:24	41	1.554	3 %	5:59:31
POLNO SKUPAJ	57	27.165	10:07:46	58	21.345	6:18:52	115	48.510	100 %	16:26:38
<i>0 govora</i>	15	0	4:06:11	15	0	2:25:47	30	0	0 %	6:31:58
<i>neposneto</i>	0	0	0:00:00	8	0	5:23:11	8	0	0 %	5:23:11
PRAZNO SKUPAJ	15	0	4:06:11	23	0	7:48:58	38	0	0 %	11:55:09
SKUPAJ	72	27.165	14:13:57	81	21.345	14:07:50	153	48.510 ¹⁵⁴	100 %	28:21:47
[30] PROSTOGOVORNOST	(za preostali kategoriji glej: <i>brez</i> → <i>0 govora</i> , <i>neznano</i> → <i>neposneto</i> ; seštevek spodnjih kategorij pa je enak kot v POLNO SKUPAJ)									
<i>prosto</i>	38	12.417	4:18:55	48	19.229	4:23:46	86	31.646	65 %	8:42:41
<i>prosto z navedki</i>	4	9.137	1:45:05	3	1.902	1:35:57	7	11.039	23 %	3:21:02
<i>prosto in navedki</i>	3	82	1:47:57	3	193	0:03:18	6	275	0,6 %	1:51:15
<i>navedki s prostim</i>	4	4.740	1:53:29	0	0	0:00:00	4	4.740	10 %	1:53:29
<i>navajanje</i>	8	789	0:22:20	4	21	0:15:51	12	810	2 %	0:38:11
GOVORCI	novih = različnih 58			različnih 49, novih glede na nedeljo 45 (4 isti za oba dneva)			moških 49, žensk 48, več hkrati 5, neznano 1; skupaj 103			

¹⁵⁴ Podatka o količini besed in povedi v korpusu sta dvojja, saj izhajata iz analize v dveh različnih programih, a razhajanja niso prevelika. Po obdelavi z EXMARALDO je v korpusu 48.510 besed in 8.337 povedi (gl. 4 podpoglavja), po obdelavi z Evo (gl. poglavja od IV.15 naprej) pa 48.398 besed in 8.423 povedi. Glede na ta razhajanja je ocena velikosti korpusa StAnId **48.450 ± 80 besed**.

Najprej beseda o „praznih“ besedilnih enotah korpusa, ker jih kasneje ne bomo več omenjali. Podatki o njih (šifre besedil, datoteke) so vendarle ohranjeni, ker so bile posnete, za vsako je bila ustvarjena celo vzporedna partitura (transkripcijska datoteka), na kateri pa ni nič drugega kot trak [Ozadje] z oznakami rezov in opisi šumov ter dogajanja, razvidnega iz posnetkov. Na posnetkih je morda kaj sporazumevanja, a je premalo razločno za verbalno transkripcijo ali pa gre le za daljše pasaje v tujem jeziku (npr. popevke v angleščini z radia). Po dolžini posnetkov so te enote korpusa povprečno daljše od „polnih“ besedil, kar je posledica pravil za meje besedil; čeprav je v „praznih“ besedilih 0 % besed korpusa, pokrivajo za oba dni skupaj 12 ur oziroma 42 % posnetkov.

Med „polnimi“ besedili korpusa je izrazit velik delež besedja, ki ga pokrijejo besedila, v katerih sodelujejo vsaj 3 govorniki (večgovor), glede na razmerje s svojo dolžino posnetkov je govor v njih tudi bolj zgoščen kot v besedilih dvogovorov in monologov. Ko bodo sestavljeni razni podkorpusi za analize sloga, bo najugodnejše vključiti ta besedila. Monologi ne predstavljajo dolgih govorov, ampak je govor v njih zelo redek; po eni strani gre za posnetke dialogov po telefonu, kjer je slišati le enega govornika, po drugi za posnetke situacij, v katerih je nosilec mikrofona sam, po tretji pa za posamezne ujete povedi neznanih govorcev (na ulici, napovedi postaj na vlaku ipd.).

V zgornji tabeli je izpostavljen podatek oz. ocena o deležu prostega govora, ker je prav prosti govor eno od najbolj zanimivih gradiv govornih korpusov. Za prosti govor bi lahko vzeli prvi dve kategoriji (posivljeni), ki skupaj predstavljata 88 % besedišča korpusa, besedila v njih pa so razni vsakdanji pogovori, večinoma v osebnem stiku, nekaj pa po telefonu. Pregled besedil ostalih treh kategorij po podatku 30) prostogovornost:

- 526 besed navajanja predstavlja poslušanje slovenskih popevk, ostalo so RTV-reklame in javna obvestila na vlaku in železniški postaji;
- 3756 besed navedkov s prostim predstavlja obred nedeljske maše (ki prihaja do izraza v pogostostnih seznamih besedišča korpusa zaradi značilnih besedil: litanije, pesmi), 928 besed predstavlja TV-dokumentarec, ostalo je par verzov pesmi v prepletu s pogovorom;
- v kategoriji prosto in navedki so prav tako odlomki pesmi in javna obvestila na železniški postaji.

Govorcev je presenetljivo veliko, povprečno na enega pride 471 besed, a kot bomo videli kasneje v njihovih podrobnejših statistikah, večina govorcev spregovori izrazito malo: 70 govorcev spregovori manj kot 100 besed (skupaj 3,7 % korpusa), 25 govorcev spregovori vsak 101–1000 besed (skupaj 18,7 % korpusa), le 8 govorcev pa vsak več kot tisoč besed oz. skupaj več kot tri četrtine korpusa. Med slednjimi 8 in 4 tistimi, ki nastopijo v obeh dnevih, trije sestavljajo družino (mati, oče, sin) in skupaj spregovorijo 54 % besed v korpusu (za njihovo primerjavo gl. poglavje IV.16.1).

Oba dneva se po količini besed in številu govorcev bistveno ne razlikujeta. Petek ima več obdobj brez posnetkov, posledično ima (transkribiranih) manj besed, a ima vseeno več prostega govora.

14.1. Pregled jezikovnih elementov in transkripcijskih posebnosti

Preden se usmerimo v slogovno analizo jezika, bi bilo škoda, da ne bi kratko pregledali par najenostavnejših statistik jezikovnega gradiva v korpusu – vsega, kar je zbrano v polnih besedilih na trakovih [v] v partituri.

14.1.1. Pisna znamenja

Pred očmi moramo imeti, da razločujemo transkripcije govora, in ne izvorno (za)pisanega jezika. Marsikatero enoto so pomembno pogojene z logiko transkripcije (npr. za pisna znamenja napovedano v poglavju II.7.1.1, Tabela 13) in njihove statistike potrebujejo pozornejšo interpretacijo kot statistike pisnih besedil. Ko govorim o črkah, s tem podajam statistiko transkripcij in ne statistiko fonemov govora, pisnih ločil v govoru ni, v transkripciji pa mi pomenijo simbole za povedno intonacijo (končna ločila), primerjavo s širokim sistemom standardnih pisnih besednih oblik (opuščaj, zvezdica) ter analitično izpostavitvev podatkov (podčrtaj, veza) itd.

Tabela 40: Pregled črk in pisnih znamenj v jezikovnem gradivu korpusa StAnId

<i>skupaj velikih in malih črk</i>		<i>od tega velike začetnice</i>		<i>primerjava deležev črk po padajoči lestvici</i>			<i>črka, pisno znamenje</i>		
<i>absolutno število</i>		<i>absolutno število</i>		<i>StAnId</i>	<i>SSKJ</i> ¹⁵⁵	<i>BSJ</i> ¹⁵⁶	<i>absolutno število</i>		
a	23.643	A	59	a	12,8	10,7	10,3	ü	2
e	20.020	E	29	e	10,9	9,8	9,5	w	6
o	15.451	O	65	o	8,4	8,9	8,3	x	2
i	12.798	I	43	i	7,0	9,2	9,6	y	13
n	10.422	N	36	n	5,7	6,7	7,4	OBLIKOSKLADENJSKA LOČILA	
t	9.649	T	73	t	5,2	5,1	6,5	'	4967
j	9.364	J	87	j	5,1	3,9	3,1	*	4702
s	9.008	S	174	s	4,9	5,1	4,7	-	475
l	8.683	L	107	l	4,7	4,6	4,5	_	328
m	8.069	M	253	m	4,4	2,7	2,3	KONČNA LOČILA	
r	7.939	R	102	r	4,3	5,7	6,6	.	4201
k	6.979	K	114	k	3,8	3,9	4,5	...	2310
p	6.733	P	77	p	3,7	3,6	3,5	?	1132
d	6.144	D	86	d	3,3	3,3	2,4	!	692
v	5.984	V	57	v	3,3	4,1	3,8	:	92
b	3.843	B	132	b	2,1	1,6	1,6	DRUGE OZNAKE	
z	3.682	Z	39	z	2,0	2,5	2,0	◊ – <i>nerazločno</i>	po 890
u	3.207	U	16	u	1,7	1,8	1,7	(<i>negotov zapisi</i>)	po 230
h	2.965	H	27	h	1,6	0,8	0,8) (– <i>podatki govorcev</i>	po 182
š	2.495	Š	16	š	1,4	0,9	1,1	() <i>SKUPAJ</i>	po 1302
č	2.464	Č	10	č	1,3	1,6	1,8	=	182
g	2.223	G	124	g	1,2	1,6	1,3		
c	1.079	C	48	c	0,6	0,8	1,6		
ž	888	Ž	4	ž	0,5	0,9	0,7		
f	346	F	27	f	0,2	0,2	0,6		

Vseh črk v verbalnem gradivu korpusa je 184.101 (54 različnih – po 25 malih in velikih slovenice ter 4 druge). Specifične **črke tujih črkopisov** zastopajo tuje besede s tujo izgovorjavo vred; a je obojih, tujih črk in besed, komaj za vzorec, ostale črke tujih besed se skrijejo med slovenskimi. Nabor tujih črk je pri naši transkripcijski konvenciji načeloma odprt (če gradivo sploh izpričuje tuja fonografska razmerja), praktično pa omejen s zapisovalčevo razgledanostjo po neslovenskih fonetičnih pisavah. Statistika **velikih črk** pove malo; ker smo jih uporabljali predvsem za lastna imena, se bodo najpogostejše velike črke pokrivalo z najpogostejšimi imeni in

¹⁵⁵ Statistika črk za SSKJ je povzeta po Jakopin 1995 (342 in 347): vseh črk v SSKJ je 17.517.300 (v 3.343.700 pojavnih velikem gradivu), med njimi 271.200 naglašeni in posebni. Kolikšna točno je vsota črk za izračun deleža posameznih črk in iz katerega gradiva (katerih delov gesel; verjetno iz celotnih gesel, vključno s kvalifikatorji), ni razvidno. Po primerjavi z drugo takšno statistiko, izbranimi vzorcema disertacije (Jakopin 2002, Slika 7: Porazdelitev najpogostejših 25 črk v obeh vzorcih skupaj; velikost gradiva je 3.129.354 pojavnih), bi mogli oceniti (druga statistika je grafična), da sta Jakopinova vzorca primerljiva med seboj in dajeta referenčno sliko za pisni jezik.

¹⁵⁶ BSJ, vseh črk je 1.716.209 (iz gradiva 178.457 iztočnic), gl. Tabela 1 na str. 1001.

bo statistika povsem odvisna od gradiva; za <M> tako imeni Marija, Maribor in njune besedne oblike.

Več o črkah pove **primerjava deležev posameznih črk** korpusa s SSKJ, vzorcem, ki ga imamo zaradi večje velikosti za reprezentativnejšega. (Za resnejšo primerjavo pa bi potrebovali še kakšno organsko bolj enovito primerjavo; druga primerjava, BSJ, ima prav tako svojo logiko.) V tabeli je posivljenih 5 največjih razlik, ki jih poskušam razložiti. Manjši delež <i> v korpusu je posledica redukcij vokala /i/ v govoru (oblike glagola *imeti*, kratki nedoločnik, predpona {pri-} in še kaj), mislimo pa si lahko, da ima tudi SSKJ večji delež <i> kot realno gradivo zaradi nedoločniških iztočnic in razlag glagolov. Večji delež <m> in <h> v korpusu je posledica zapisovanja mrmranja – mašil (<mhm> in druge oblike). Za izstopajoča deleža <r> in <a> ne najdem utemeljitve.

Vseh pisnih znamenj v korpusu je 205.789; med njimi je torej 89,5 % črk in 10,5 % ostalih pisnih znamenj. *Presledke*, ki običajno imajo najpogostejši delež v statistikah znamenj korpusov, je zaradi partiturnega zapisa transkripcij težje prešteti. Lahko jih definiramo znotraj izjav govora enega govorca, težje pa pri menjavanju govornih vlog med govorniki in pri hkratnem govoru. In še zato, ker dajejo samo informacijo o transkripciji, manj pa o samem govoru (tu bi bil pomembnejši podatek o premolkah), podatka o presledkih ne navajam.

14.1.2. Besede

Nabor nečrkovnih pisnih znamenj je omejen, ker so pisna znamenja natančneje določena s transkripcijsko konvencijo. Tako njihova statistika tudi več pove oz. omogoča hitro „rudarjenje“ ali sondiranje podatkov. Seznam različnic v korpusu je na podlagi teh znamenj razdelan na koncu diplome, gl. Priloga 3. Ta abecedni seznam ni podan kot enota za ves korpus, ampak so iz skupnega seznama po naslednjem zaporedju izbrane skupine (vsaka besedna oblika tako nastopa le enkrat):

Tabela 41: Pregled različnic korpusa po oblikoskladenjskih ločilih in pomožnih znamenjih

<i>skupina</i>	<i>pojavníc</i>	<i>različnic</i>	<i>opomba o statistiki znamenj, prim. zgornja tabela</i>	<i>pomen znamenja, gl. Tabela 13, str. 93</i>
1)) kratice in posamezne črke	73 (74 besednih oblik)	44	razpoznava po zvezdicah in velikih črkah; brez velikih črk v angleščini, ki so v gradivu, npr. *I 'jaz'	d.15
2)) imena in podatki govorcev, v samem korpusu zakodirani s šiframi (brez pogostosti)	182 (184 besednih oblik)	50	prepoznani po enačajih (182) in oklepajih; praktično še po velikih začetnicah	f.18
3)) druga imena, tudi večbesedna	1.202 (1473 besednih oblik)	677	osebna, veroslovna, naselbinska imena, naslovi pesmi ...; tudi večbesedno	f.16 (brez f.18)
4)) preostale večbesedne zveze (brez podatkov govorcev, imen in kratic)	32 (76 besednih oblik)	28	razpoznava po podčrtajih (vseh podčrtajev 328); večinoma gre za števila, primeri še za: spletni naslov, oznaka časa	f.17 (brez f.16 in f.18)
5)) (kognitivno) okrnjene besedne oblike: nedokončane in sredi besede začete besedne oblike	443+1	228+1	razpoznani po vezaju na začetku ali koncu besede; in ena sama beseda z vezajem v sredini	a.1; za d.14 le ena beseda; brez d.15
6)) preostale besede z zvezdico	4.468	1.067	Seznama sta razporejena kataloško, krajše besede sledijo daljšim, npr. najprej <i>aba</i> ·113, nato <i>ab</i>) ·1, nazadnje	a.2, b.4, c.6, č.8–9 in.11–12, d.13, f.19–20

<i>skupina</i>	<i>pojavníc</i>	<i>različnic</i>	<i>opomba o statistiki znamenj, prim. zgornja tabela</i>	<i>pomen znamenja, gl. Tabela 13, str. 93</i>
7)) preostale besede	41.573	6.420	ab·18. V obeh seznamih so besede po znamenjih: a.3 (razpoznava po oklepajih), ¹⁵⁷ ; besede z opuščajem: č.10.	

V zadnji skupini besed nisem podrobneje razdelal besed, označenih z zvezdico in opuščajem, saj je to že izpeljava naloge v nezastavljeno smer. Pojavnic v korpusu, ki imajo zvezdico, je toliko kot zvezdic samih, torej 4.702 (9,7 %), pojavnic z opuščajem pa je nekoliko manj kot opuščajev samih, 4.798, ker imajo nekatere besede več opuščajev: 161 pojavnic/47 različnic ima po dva opuščaja (5 najpogostejših s pogostostmi so: dr'gač' 23 | kol'k' 22 | št' r'najst 20 | tol'k' 19 | 'met' 10), 4 besede pa imajo 3 opuščaje (pr'jat'1'a | tri'ndvajs't'ga | *št'r'npe'cset | -najststo_*os'nšt'r'deset_*šeeset).

Zanimivo je še vprašanje, koliko različnic bi bilo v vsem korpusu manj brez posebnih transkripcijskih znamenj:

Tabela 42: Variiranje števila različnic glede na upoštevanje posebnih ločil transkripcije

vsa znamenja po <i>privzetem</i> načinu štetja: pojavnice besednih oblik v oklepajih – negotov zapis – in v s podčrtajem označenih večbesednih zvezah so prištete ostalim; podatki govorcev pa šifrirani, vsi deli posamezne šifre kot ena pojavnica: 48.398 pojavnic	različnic: 8.384	razlika v št. različnic
samo brez <'> (enofonemski izpusti)	8.266	- 118
samo brez <*> (različne oznake)	8.208	- 176
samo brez <-> (nedokončane besede in običajni vezaji)	8.323	- 61
samo brez <=> (podatki govorcev razlikovani od drugih podatkov, imen) → 48.397 pojavnic	8.380	- 4
brez <= * -> → 48.397 pojavnic	7.990	- 394

Izračun razmerij različnih besed v zapisu govora brez vseh nečrkovnih oznak (zadnja vrstica) pokaže, da se z vidika kvantitativne ocene skoraj ne splača uporabljati posebnih oznak, če nas zanima le številčno izražena pestrost zaklada besednih oblik. Bi pa bilo potrebno oceniti kvalitativno, koliko je vrednih 394 različnic, ki se izgubijo z zakritjem posebnih ločil, oz. več kot desetini vseh besed, ki imajo zvezdico ali vsaj en opuščaj.

Kvalitativna analiza posebnih različnic bi si pomagala z njihovimi konkordancami. Konkordance 2 zanimivih besed, ki so vzbudile pozornost med transkribiranjem, so v prilogah (Priloga 5) na podlagi gradiva prvega dne, nedelje:

- večfunkcijska beseda *ko¹⁵⁸ 'kaj' (tudi kot <ka'> z manj zaokroženim alofonom fonema /a/), 'ker', 'kjer', veznik 'da', 'ki' (in njegovi odvisni skloni, npr. 'v kateri');
- narečna pogosta beseda *te 'tedaj', 'potem'.¹⁵⁹

Jezikovna ravnina samih besed v korpusu bo obravnavana v sledečih poglavjih s stilometrično analizo (najpogostejše različnice), tukaj pa komentiram le sezname najpogostejših dvojčkov, trojčkov, četverčkov in peterčkov pojavnic (gl. Priloga 6), ki v katerih je potencirano izražena pestrost besednega zaklada. Vzorec našega korpusa je namreč zelo majhen, zato se že enakih besednih zvez 2 besed najde precej malo, pri trojčkih, četverčkih in peterčkih pa nasploh prevladujejo besedne zveze iz navajanih besedil (branih, petih, recitiranih na pamet). Iz seznamov v prilogi so razvidni celotni refreni kakih 6 pesmi in 3 molitve. Kljub temu, da se ti sezname za tako majhno gradivo v začetku zde malo smiselni, nam pomagajo preveriti gradivo z vidika prostogovornosti in pestrosti sporazumevanja – razkrivajo stalne vzorce sporazumevanja.

¹⁵⁷ Besedam, ki so v transkripciji v oklepajih, kar pomeni negotov zapis (230 teh oklepajev) sledi v tem seznamu oklepaj.

¹⁵⁸ Skrajšani 'kot' je zapisan <ko'>, poleg tega pa <ko> (brez zvezdice) tako ali tako pomeni veznik. Prim. tudi Zemljarič Miklavčič/Stabej 2006, 8. prosojnica: k = „k[er], k[jer], k[ə], k[ar]“.

¹⁵⁹ Podobno še *pol 'potem'. Zanimivo bi bile tudi konkordance različnice ne – nikalnica ali vprašalna členica. Pri slednji rabi je bila pri transkribiranju pogosto dilema, ali ob rastoči intonaciji hkrati s členico nakazati konec povedi. Ker se ta različnica pojavi že samo v nedeljo 1273-krat, z njenimi konkordancami priloge ne obremenjujem dodatno.

14.1.3. Povedi

Podobno kot pri n-terčkih besed so za majhno velikost našega korpusa borne raziskave povedi (za naslednje komentarje gl. Priloga 7). Najpogostejše povedi so zanimive zato, ker prinašajo značilne elemente prostega govora, ki jih je v tem majhnem korpusu relativno dosti več kot v kakšnem mnogo večjem pisnem korpusu (prim. recimo najpogostejše povedi iz leposlovnih in časopisnih besedil v korpusu Nova beseda, Jakopin 2003: 151–150, Tabela 8). Ti značilni elementi prostega govora, ki jih najdemo med najpogostejšimi povedmi, so:

- velika večina so **oporni signali** (npr. 4 najpogostejši povedni: ja. | mhm. | aha. | okej; in 4 najpogostejši vprašalni: a res? | ja? | kaj? | aja?),

- nato še **mašila** – v kolikor niso to že sami oporni signali (3 najpogostejši z „obviselo“ intonacijo ja ... | *ahm ... | aha ... ter s 1 pripovedno intonacijo *e ja.)

- in krajši **pozdravi** (npr. adijo. | dober dan. | zdravo! | zdravo.).

Kot med omenjenimi najpogostejšimi besednimi n-terčki se tudi med najpogostejšimi povedmi najdejo takšne iz bolj specifičnih ustaljenih sporazumevalnih vzorcev oz. neprostega govorjenja, ki kažejo na posamezen sestavni del korpusa:

- deli molitev (odgovor litanij in molitvenih obrazcev: prosi za nas! | amen.)

- in pesmi (aleluja aleluja aleluja. | zakaj star song se še vrti?)

- in deli opevanja: (tri štir'! | *lalalalalalalala. | *mihoho *mihoho *mihoho *mihoho.).

Vseh povedi v korpusu je 8.423,¹⁶⁰ od tega 6.810 različnih. Kljub problematičnosti oz. samosvojesti koncepta povedi v govoru in odvisnosti meja povedi od transkripcijskih načel (razložena so v poglavju II.7.2.2) lahko rečemo naslednje:

- prevladujejo povedi s padajočo, pripovedno intonacijo, zaključene s piko (50 %);
- kot posebnost govora jim sledijo povedi, ki sem jih zapisal s tropičjem: hote ali nehote nedokončane, z „obviselo“ intonacijo (27 %);

- skoraj kot v stilu padanja kvadratne funkcije sledita še deleža povedi z vprašalno (vprašaj; 13 %) in vzklično intonacijo (klicaj; 8 %);

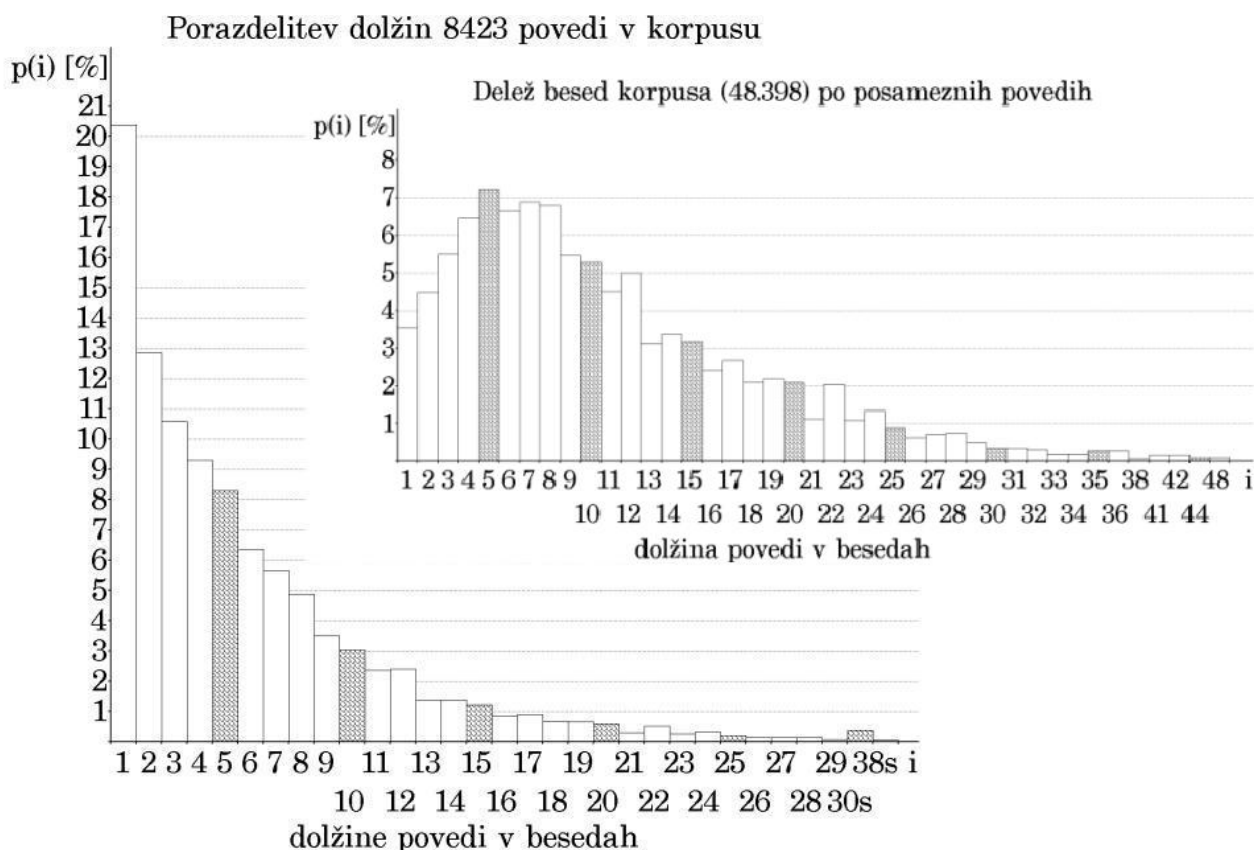
- najmanj je takšnih povedi s padajočo intonacijo, ki napovedujejo in so pomensko nezaokrožene, zaključene z dvopičjem (92 oz. 1 %). Za absolutne pogostosti prim. statistiko končnih ločil, Tabela 40 zgoraj.

Povprečna dolžina povedi je 30 znakov (vključno s presledki in ločili) oz. 5,8 besede. Pri 103 najkrajših povedih iz priloge je zanimivo, da so vse vrste povedi enakomerno zastopane: pripovedne 31 %, vzklične 27 %, obviselo (tropičje) 23 %, vprašalne 19 %; možen je torej preprost sklep, da je večina vzkličnih in vprašalnih povedi v govoru (podpovprečno) kratkih. Pri devetih najdaljših povedih pričakovano prevladujejo neprosto govorjene povedi (brane), pri prosto govorjenih, ki po dolžini bistveno ne zaostajajo, pa gre za pripovedi ali razlage, deloma kar strokovne.

Priložena grafa prikazujeta **porazdelitev povedi različnih dolžin** v besedah (obe abcisi). Levi graf prikazuje, da so enobesedne povedi daleč najštevilčnejše med vsemi dolžinami, desni graf pa, da največji delež besed korpusa pokrijejo 5-besedne povedi. Zaradi večje preglednosti grafov je posivljen stolpec vsake pete povedi in na prvem grafu združene vrednosti povedi z dolžinami 30–36 besed („30s“) ter dolžin 38, 41, 42, 44 in 48 besed („38s“). Za številčne vrednosti deležev s slike gl. Priloga 7, njen konec.

¹⁶⁰ Kot samostojne povedi je tu prišteti 92 spremnih stavkov, zaključenih z dvopičji – četudi pomensko niso nujno bogati, so pa pomensko logično zaokroženi, imajo svojo izrazito pripovedno intonacijo.

Slika 26: Primerjava dolžin povedi v korpusu: številčnost in pokritje besednega zaklada korpusa



14.1.4. Besedila

115 polnih besedil korpusa je v povprečju dolgih 421 besed. Ta podatek pa zakrije dejstvo, da je večina besedil izrazito kratkih – kot je razvidno iz spodnje tabele, je 88 besedil krajših od 500 besed¹⁶¹ in skupaj predstavljajo le 10,3 % korpusa (dejansko je med temi 88 najdaljše besedilo dolgo 342 besed, naslednje daljše besedilo pa je dolgo 516 besed).

Tabela 43: Korpus po velikosti besedil v jezikovnih enotah

podatek	velikostna kategorija:				delež korpusa v besedah v % [št. besedil]					
10) izjave	1-5	1,8 [49]	6-10	0,8 [14]	11-50	10 [26]	51-100	12,8 [9]	100<	74,5 [17]
11) povedi	1-5	0,9 [46]	6-10	0,8 [11]	11-50	5,4 [25]	51-100	7,4 [9]	100<	85,6 [24]
12) besede	1-10	0,4 [35]	11-100	2,6 [35]	101-500	7,3 [18]	501-1000	20,8 [13]	1000<	68,9 [14]
13) besed/s	≤0,5	2,4	>0,5 in ≤1	7,9	>1 in ≤2	68,2	2<	21,6		

Podajanje kolikostnih oznak za posamezne podkorpuse bo, kot je običajno za korpus, tudi vnaprej podajano za velikosti podkorpusev *v besedah*, mestoma preračunano v deleže glede na celotno velikost korpusa (48.510 besed). Podatka o dolžini besedil v izjavah oz. povedih sta zgoraj dodana le kot zanimivost, ki jo omogoča gradivo.

Zanimiv je tudi **podatek o gostoti besed govora** glede na posnetek besedila ($\frac{13}{s}$ besed/s), čeprav velja, da se posnetek običajno ne začne s prvo besedo in ne konča z zadnjo besedo transkripcije – dejanske gostote besedil so večje od tukaj izračunane oziroma odvisno različne od

¹⁶¹ 500–600 besed znaša recimo ena gosto tipkana A4 stran diplome, vključno z opombami.

tempa govora tudi znotraj besedil. Skoraj 90 % besedil ima gostoto govora več kot 1 besedo v sekundi, najgostejše po teh izračunih pa je besedilo *1ned17Vg* s 3,73 besede na sekundo (306 besed, 7 govorcev, traja 1' 22").

14.2. Slog govora v korpusu kot celoti

Prav bi bilo, da bi osrednjo analizo naloge, *stilometrično*, opravili tudi za korpus kot celoto, za največjo količino gradiva, ki jo imamo na voljo. A sloga ni smiselno pripisovati *nadbosedilu*, ki nima notranje koherentnosti ali vsaj raziskovalno utemeljene uravnoveženosti (zaokroženost korpusa je po zunanjih okoliščinah sporazumevanja: dva dneva).

Največji zanimivi podkorpus je tako **podkorpus prostega govora**, njemu v primerjavo pa lahko postavimo (pod)korpus, izbran po podobnem kriteriju – nasprotno prenosniško in/ali funkcijsko zvrst jezika. Ker je podobnih kvantitativnih raziskav govora še malo, izbiramo po pisnih korpusih (leposlovje, publicistika, strokovni jezik), in še tu je malo objav, ki bi nam dajala potrebne podatke.¹⁶² Od izbranih primerjav v sledeči tabeli tako SSKJ ni funkcijskozvrstno uravnovežen, druga dva vzorca iz korpusa Nova beseda pa približno spadata v publicistiko (časnik Delo; različne vrste besedil) in leposlovje (ZD Ivana Cankarja; poleg treh tradicionalnih leposlovnih zvrsti še pisma, predavanja itd.).

Pred nami je seznam, kot sem ga obrazložil v poglavju I.4.3.1.1, po prvi izmed 7 primerjav, napovedanih v poglavju I.4.3.3:

Primerjava 1: podkorpus prostega govora in pisni (pod)korpusi													
[39 različnic v 4 vzorcih]													
korpus	2. gradivo: časnik Delo 1998–2000 vir: Jakopin 2001: 471–474			3. gradivo: Zbrana dela Ivana Cankarja vir: Jakopin 2001: 471–474			4. gradivo: ponazarjalno gradivo SSKJ vir: Jakopin 1995: 358–359			1. gradivo: besedila StAnId, ki imajo vrednost podatka [30] prostogovornost: <u>prosto</u> ali <u>prosto z navedki</u> ; <i>besedila</i> ¹⁶³			
	P= 47.219.000 (SSKJ)			P= 1.991.000 (Delo)			R= 222.176	P= 1.616.200	R/P= 0,137	R= 7.026	P= 42.750	R/P= 0,164	
zap. št.	različnica	delež [%]	razlika	različnic a	delež [%]	razlik a	različnica	delež [%]	razlika	različnic a	delež [%]	razlik a	
1	je	3,33	-1,18	je	6,19	1,69	je	5,02	0,52	pa	3,91	2,46	
2	v	2,86	0,78	in	3,96	1,86	se	2,13	0,22	je	3,48	-1,03	
3	in	2,48	0,38	se	2,82	0,91	v	2,11	0,03	ne	2,98	1,67	
4	na	1,61	0,31	v	1,96	-0,12	so	1,35	0,27	ja	2,75	NPP	
5	za	1,39	0,52	da	1,74	0,41	in	1,34	-0,77	da	1,89	0,56	
6	se	1,28	-0,64	na	1,31	0,00	na	1,29	-0,02	se	1,43	-0,49	
7	da	1,24	-0,09	ne	1,21	-0,10	za	0,88	0,01	v	1,4	-0,68	
8	so	1,19	0,11	so	1,11	0,03	z	0,82	0,21	*e	1,15	NPP	
9	ki	1,08	0,60	bi	1,09	0,34	s	0,66	NPP	bi	1,04	0,29	
10	pa	1	-0,45	sem	0,99	NPP	ga	0,56	NPP	na	1,01	-0,30	
11	z	0,73	0,12	ni	0,71	0,19	po	0,55	NPP	tak'	0,99	NPP	
12	tudi	0,7	NPP*	pa	0,69	-0,76	ne	0,48	-0,83	a	0,93	NPP	
13	bi	0,67	-0,09	kakor	0,68	NPP	mu	0,48	NPP	to	0,9	NPP	
14	po	0,61	NPP	ki	0,64	0,16	da	0,46	-0,87	jaz	0,83	NPP	

¹⁶² Pogostostni seznam najpogostejših 20 različnic, s podatkom že izračunanega deleža ali z absolutno pogostostjo iz podatkom o številu pojavnic. Idealno bi bilo tudi, da bi bil tudi primerjani seznam izdelan iz gradiva enakega velikostnega razreda, kot je naš (42.750 pojavnic). Ob v tabeli citiranih virih sem poznal še sezname Neweklowskega (1984 in 1988), ki pa so iz gradiva velikosti 10.000 besed.

¹⁶³ *Besedila podkorpusa prostega govora*: 1ned: 02, 04, 06, 09–11, 14, 16–19, 20, 22, 24–29, 34 –38, 40–41, 43, 46–48, 50, 52, 54–55, 57–58, 60, 62, 68, 70–72;

6pet: 01–05, 07, 09, 11, 13–19, 21–22, 24–25, 27–28, 30, 32, 34, 36–38, 40, 42–43, 47–49, 51–52, 54–55, 58, 62, 65, 67, 69, 71, 73–78, 80–81.

15	s	0,6	NPP	bil	0,63	NPP	iz	0,42	NPP	*s'n	0,79	NPP
16	ne	0,58	-0,73	z	0,61	-0,01	ni	0,39	-0,13	še	0,76	0,23
17	bo	0,53	NPP	še	0,59	0,06	to	0,35	NPP	če	0,74	NPP
18	še	0,51	-0,02	mi	0,56	NPP	si	0,31	NPP	si	0,68	NPP
18	ko	0,5	NPP	za	0,54	-0,33	že	0,29	NPP	za	0,67	-0,20
20	ni	0,4	-0,12	bilo	0,5	NPP	pri	0,28	NPP	so	0,66	-0,42

*NPP = ni podatkov o primerjavi – v virih za 2. in 3. gradivo so podatki le za najpogostejših 20 besed, za redkejša različnica pa nimamo deležev, zato izračun povprečnega deleža in razlike do njega ni mogoč.

Najprej ugotovimo, da so najbolj odstopajoče besede tudi med najpogostejšimi – razlaga tega je kvantitativna: najpogostejše besede imajo največje deleže, zato so tudi razlike posameznih primerjav do povprečja največje. Pri naslednjih besedah vendarle ni opaziti večje razlike med vsemi štirimi seznamami, v vseh so med prvimi sedmimi: <je|se|v>. Ostale besede, ki se pojavijo v vseh seznamih, pa so <na|ne|so|za>.

Naslednje večje razlike izhajajo iz nasprotja med gradivi pisnih korpusov in našim govornim, 1. gradivom. <pa> ima med vsemi 39 različnicami v primerjavi največji odstop v 1. gradivu, kjer prehitu v pisnih gradivih splošno najpogostejši <je> in izrine veznik <in> – le-tega med 20 najpogostejšimi različnicami 1. gradiva sploh ni, v ostalih treh pa je med prvimi petimi. V 1. gradivu je precej pogostejši kot v ostalih tudi <ne> (verjetno zaradi pogoste vloge vprašalne členice v govoru), precej redkejši kot v ostalih je <so>. Sploh pa se v govornem gradivu med 20 najpogostejšimi od takih besed, ki so v vseh ostalih treh seznamih, ne pojavlja nobeden od predlogov <z|s>, tudi ne različnici <da|ni>. Nasprotno se samo v seznamu najpogostejših različnic tega gradiva pojavljajo <a|če|ja|jaz> ter tudi 3 govorne posebnosti, ki jih v ostalih treh gradivih verjetno ne bi našli nikjer: <tak'|*e|*s'n>.

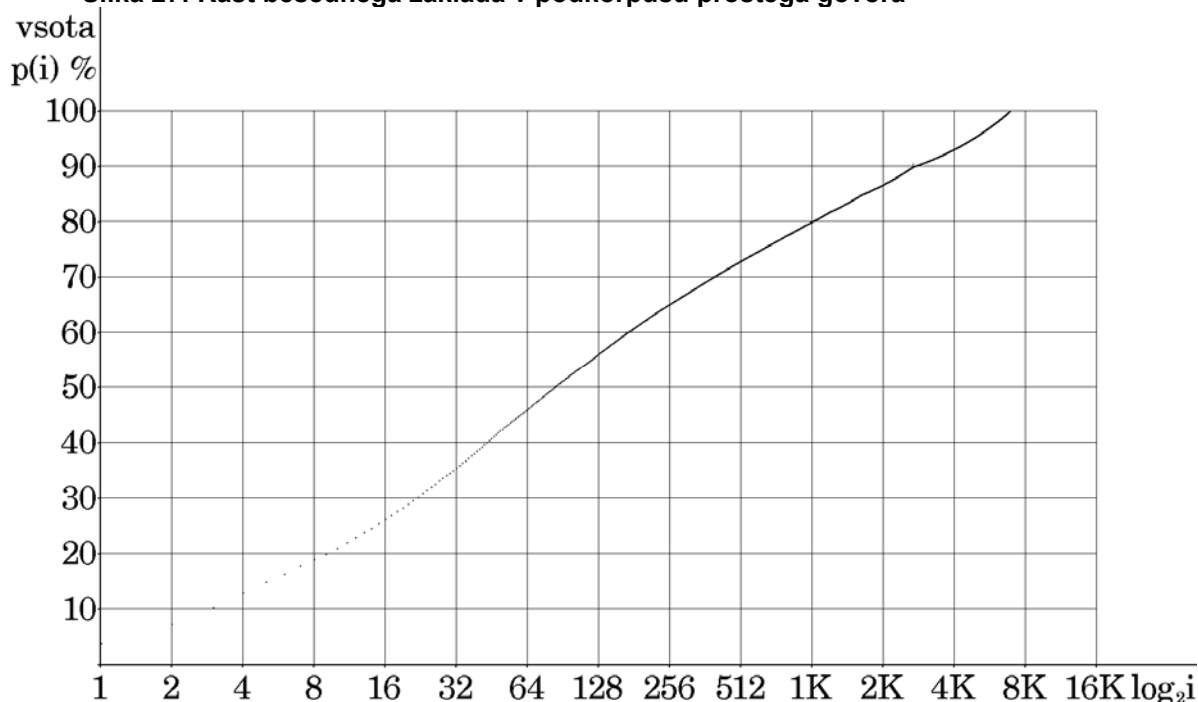
Poleg prvega, govornega gradiva, glede na vse ostale po specifičnih najpogostejših različnicah nekoliko izstopa še drugo gradivo, SSKJ; to se vidi tudi po številu različnic z oznako NPP – teh je v gradivih SSKJ in prostega govora po 9, v preostalih dveh pa 5.

Gradiva se izrazito razlikujejo po velikosti (gl. vrednost P; edino 3. in 4. gradivo sta morda primerljiva), kar se odraža tudi v pogostostih različnic ter razlikah do povprečne pogostosti. Večje kot je gradivo, počasneje se polni njegov besedni zaklad različnic (manjši so deleži najpogostejših različnic), manjše oz. večinoma negativne so razlike do povprečne pogostosti različnic, manjši bi bil verjetno tudi R/P, ki ga nismo mogli izračunati v vseh primerih. Toliko **o relativnosti naših kvantitativnih podatkov glede na velikost gradiva** – v preostalih stilometričnih primerjavah, izjema bo le še Primerjava 5, bodo primerjana gradiva precej bolj podobne velikosti.

Omenjena kvantitativna različnost se odraža tudi na grafu rasti besednega zaklada (različnic; Slika 27 spodaj). Gre za kumulacijo deležev različnic v podkorpusu prostega govora, začeni pri najpogostejših na levi abcise, ki je logaritmizirana.¹⁶⁴ V tem majhnem podkorpusu, ki ima tudi malo različnic, se krivulja zelo strmo vzpenja (v primerjavi z npr. milijonskimi korpusi): prvih 9 različnic pokrije petino podkorpusa, prvih 15 četrtno, prvih 28 tretjino, prvih 83 polovico; velik delež imajo enkratnice – 9,9 % (zadnji del krivulje, od preloma krivulje z navpično črtico dalje), pa dvakratnice 5,0 %, trikratnice 3,1 % itd. Za seznam 1212 najpogostejših različnic s podatki o deležih v tem podkorpusu gl. Priloga 4.

¹⁶⁴ Graf, v katerem je i zaporedno mesto različnice po pogostosti v besedilnem gradivu; število vrednosti i je omejeno, enako številu različnic in v grafu urejeno od najvišje proti najnižji pogostosti v smeri od leve proti desni. Skala vrednosti za i je logaritemska. Vrednost p(i) je vsota pogostosti različnice (v deležu zaklada besednih oblik) s pogostostmi vseh pogostejših različnic. Več o tem grafu gl. Jakopin 2002: poglavje 5.2.1, ali McEnery/Wilson 2001: 156–161.

Slika 27: Rast besednega zaklada v podkorpusu prostega govora



15. OKOLIŠČINE SPORAZUMEVANJA MED SEBOJ

Pregled okoliščin sporazumevanja dajem po oznakah, ki so pripisane posameznim besedilom. Za podroben pregled za vseh 153 polnih in praznih enot korpusa gl. Priloga 1, bolj zgoščen pregled „*polnih*“ besedil korpusa pa je v treh tabelah dve strani naprej (Tabela 44–Tabela 46). V njih je delež besedil, ki ustrezajo posamezni okoliščini, podan kot delež besedišča v korpusu (v %), ponekod pa je v oglatem oklepaju dodano še število besedil (npr. podatek [29] vgnezdeno).

Okoliščine so večinoma podane v obliki opisa, kot je bila določena v poglavju II.8.5, pri teh so vrednosti iz omejenih seznamov v tabelah tako kot drugje v diplomu podčrtane (npr. podatek [26] povod zaključka). Pregled podatkov, katerih potencialne vrednosti niso bile predpisane pred določevanjem, je podan bodisi v številskih razponih (številске vrednosti, npr. podatek [14] trajanje) bodisi po povzermalnih kategorijah (taki podatki so v tabelah posivljeni, npr. podatek [28] oblika sporazumevanja).

Pregled služi tudi kot izbor podkorpusov, ki jih uporabljam za stilometrično analizo. Zanj izberem podkorpuse oz. gradiva homogenih okoliščin, ki bodo čim večja ali pa toliko zmanjšana, da bodo podobne količine kot ostala gradiva v primerjavi. Izbrani podkorpusi so podrobneje opisani v tabeli posamezne analize, gl. Primerjava 1 zgoraj do Primerjava 7 spodaj.

Iz tabel povzemimo **okoliščine večine besedišča**:

- oznake in meje besedil: povod meja besedil je menjavanje navzočnosti soudeležencev (zaradi nosilca mikrofona ali drugih ssporazumevalcev), po obliki sporazumevanja (dolgi) pogovori z elementi pripovedi, vgnezdena besedila so redka;

- neodvisne okoliščine: besedila trajajo več kot 5 minut, čez dan je največ govora dopoldan (nedelja) oz. po večerji (oba dneva), v zasebnih prostorih, kjer pa se družijo več ljudi (presenetljivo dosti je posnetkov iz avtomobila), besedilo je posneto ob najslabši možni kvaliteti oz. ob največji kompresiji posnetka;

- sporazumevalne okoliščine: govor v osebni (slušni in vidni) stiku, brez drugih medijev;

- soudeleženci: 3 govorniki ali več pomeni veliko govora, ki pa pride v glavnem do 2–10 poslušalcev, ne eni ne drugi pa se ne zavedajo snemanja; socialnozvrstno je jezik narečni ali pokrajinski pogovorni, namen sporazumevanja je zabava, prosti čas ali gojenje osebnih odnosov, govorni položaj je zasebni, funkcijska zvrst jezika praktičnosporazumevalna, sporazumevanje pa poteka v obeh smereh oz. navzkrižno med govorniki.

15.1. Slogi sporazumevanja v pogostejših okoliščinah

Slog govora po okoliščinah (lahko bi rekli: slog različnih sociolektov ali diskurzov) bomo obdelali v dveh primerjavah, najprej besedila različnih in nato še besedila istih okoliščin (isti so v 3. primerjavi predvsem govorniki).

Med gradiva v primerjavi različnih okoliščin (gl. 2 strani naprej, Primerjava 2) gre v dveh primerih za dve največji besedili v korpusu (1. gradivo, sledi mu besedilo 1ned13Vg iz 3. gradiva, besedilo 1ned15Dm ima le 3 besede), ostala pa so sestavljena iz več besedil. Ker je primerjava najširša med vsemi (5 gradiv), je največje tudi število vseh različnic, ki se pojavijo znotraj posamezne primerjave seznamov 20 najpogostejših – 48. Gradiva imajo kar primerljivo velikost (med največjim in najmanjšim je 1140 besed razlike), niso pa posebno velika – 20. različnica iz 5., najmanjšega gradiva, ima absolutno pogostost 23 pojavitev. Iz izračunov razmerij R/P poskusimo oceniti bogatost besednih zakladov: najbogatejšega ima absolutno in relativno glede na svojo velikost 1. gradivo, saj ima kljub temu, da je največje med gradivi, šele drugi največji R/P; najrevnejši besedni zaklad ima absolutno gledano najmanjše gradivo, verjetno pa tudi v relativnem pogledu po R/P, saj kljub temu, da je najmanjše, nima največjega R/P.

(Razlaga se nadaljuje na str. 159 pod tabelo.)

Tabela 44: Korpus po prostih oznakah in mejah besedil

<i>meje besedil</i>	<i>26–27) povod začetka/zaključka</i>		<i>prostor</i>	43,7	<i>navzočnost</i>	39,9	<i>dejavnost</i>	14,6	<i>kontinuiteta</i>	0,01	<i>razločnost</i>	0,2	<i>vir</i>	1,6
<i>metaoznake, sobesedilnost</i>	<i>pozdrav/slovo /poziv</i>	0,5	<i>delovni pomenek</i>	2,7	<i>telefonski pomenek</i>	6,0	<i>neurejen večgovor</i>	2,8	<i>formalni pomenki</i>	1,0	<i>reklama RTV</i>	0,5	<i>glasba – besedila</i>	1,1
<i>28) oblika sporazumevanja</i>	<i>samogovor</i>	0,8	<i>klepet</i>	2,1	<i>pogovor /pripoved</i>	60,2	<i>urejen večgovor</i>	16,0	<i>javno obvestilo</i>	0,6	<i>oddaja RTV</i>	6,0	<i>mešano, razno</i>	0,02
<i>29) vgnezdjena (besedila)</i>	<i>brez</i>	53,1 [95]	<i>nabožna besedila</i>	21,0 [5]	<i>publicistika</i>	10,2 [3]	<i>glasba – pesmi</i>	4,9 [4]	<i>javna obvestila</i>	5,5 [2]	<i>strokovni podatki</i>	0,5 [1]	<i>mešano</i>	4,8 [5]

Za 30) prostogovornost gl. Tabela 39 zgoraj.

Tabela 45: Korpus po dejavnostih neodvisnih okoliščin in sporazumevalnih sredstev

<i>neodvisne okoliščine</i>	<i>14) trajanje</i>	<i>≤30''</i>	0,5	<i>31''–5'</i>	10,7	<i>301''–10'</i>	12,3	<i>601'–30'</i>	44,6	<i>30'<</i>	31,9		
	<i>16) čas konca</i>	<i>do 9:00</i>	0,02 [9 besedil]	<i>9.–12.</i>	31,3	<i>12.–16.</i>	19,6	<i>16.–19.</i>	5,4	<i>po 19.00</i>	43,7		
	<i>17) kraj</i>	<i>LJ</i>	24,6	<i>MB</i>	9,2	<i>Lovrenc</i>	53,0	<i>MB–LJ</i>	6,5	<i>MB–Lovrenc</i>	6,7		
	<i>18) prostor</i>	<i>zasebna soba</i>	14	<i>dnevni zasebni prostori</i>	34,8	<i>zasebni prehodi</i>	0,5	<i>odprte javne površine</i>	7,1	<i>zaprte javne površine</i>	19,8	<i>avto</i>	17,4
												<i>javni prevoz</i>	6,5
<i>19) kvaliteta</i>	<i>SP</i>	0	<i>LP2</i>	28,1	<i>LP4</i>	71,9							
<i>sporazumevalna sredstva</i>	<i>20) medij¹⁶⁵</i>	<i>sam zase</i>	0,9	<i>osebni stik</i>	88,4	<i>telefon /mobitel</i>	6	<i>TV/radio</i>	4,2	<i>zvočnik</i>	0,6		

18) prostor. zasebna soba (osebni prostori, kjer je posameznik običajno sam: soba, spalnica, kopalnica), *dnevni zasebni prostori* (kjer se skupnost zadržuje dalj časa: dnevna soba, kuhinja, jedilnica), *zasebni prehodi* (kjer se skupnost kratko srečuje: stopnišče, hodnik v stanovanju, hiši, bloku, manjšem podjetju), *odprte javne površine* (ulica, postaja, prostor pred vhodom), *zaprte javne površine* (trgovine, avle uradov, uradi, hodniki institucij, cerkev ipd.), *avto* (zasebno vozilo), *javni prevoz* (vlak, avtobus).

Tabela 46: Korpus po dejavnostih soudeležencev

<i>soudeleženci – aktualno + 21), 22)</i>	<i>23) št. govorcev</i>	<i>1</i>	3,2	<i>2</i>	20,7	<i>3</i>	28,9	<i>4–5</i>	16,1	<i>5<</i>	31,3
	<i>24) št. soudeležencev</i>	<i>1</i>	1,3	<i>2</i>	27,6	<i>3–10</i>	52,9	<i>11–99</i>	6,8	<i>100≤</i>	11,3
	<i>25) zavest snemanja</i>	<i>da</i>	7,1	<i>ne</i>	92,9					<i>neznano</i>	0,016
<i>soudeleženci – stalno</i>	<i>31) socialna zvrst</i>	<i>nar.</i>	44,1	<i>pok. pog.</i>	42,6	<i>knj.-pog.</i>	3,5	<i>zbor. gov.</i>	9,9		
	<i>32) namen sporazumevanja</i>	<i>INFO</i>	4,2	<i>JINST</i>	7,7	<i>POSL</i>	0,8	<i>ZPČO</i>	87,3		
	<i>33) govorni položaj</i>	<i>javni</i>	16,6	<i>zasebni</i>	83,4						
	<i>34) funkcijska zvrst</i>	<i>UM*</i>	9,1	<i>ST</i>	0,5	<i>PS</i>	87,3	<i>PU</i>	3,1		
	<i>35) smer sporazumevanja</i>	<i>brez</i>	0,9	<i>enosmerno</i>	6,0	<i>1,5-smerno</i>	1,3	<i>dvosmerno</i>	91,9		

Za 21) tema, 22) dejavnost gl. preglednico besedil, Priloga 1. *Pod umetnostni jezik je bil šteta tudi nedeljska maša.

¹⁶⁵ Besedila so razdelana po večinskem mediju v njih; za natančnejšo določitev bi pa bilo za vsakega posameznega govorce potrebno določiti medije, ki jih v posameznem besedilu uporablja: tako kot govorec in tako kot poslušalec.

okoliščine	Primerjava 2: različne okoliščine – podkorpusi/besedila														
	[48 različnic v 5 vzorcih]														
zap. št.	1. gradivo: nedeljsko družinsko kosilo doma, mati, oče in sin <i>besedila</i> : 1ned25Vg			2. gradivo: pevske vaje in avdicija mladinskega cerkvenega PZ, >10 govorcev <i>besedila</i> : 1ned: 09Vg, 18Vg			3. gradivo: nedeljska maša v podeželski župnijski cerkvi <i>besedila</i> : 1ned: 13Vg, 15Dm			4. gradivo: pogovor 2 študentov sostanovalcev ob srečanju po počitnicah, nedelja zvečer <i>besedila</i> : 1ned: 70Vg, 71Dg			5. gradivo: vožnja 6 priložnostnih znancev v avtu po MB, petek pozno zvečer <i>besedila</i> : 6pet: 74Vg, 75Vg		
	R= 1352	P= 4164	R/P= 0,325	R= 1030	P= 3984	R/P= 0,258	R= 1317	P= 3759	R/P= 0,350	R= 1045	P= 3519	R/P= 0,297	R= 877	P= 3024	R/P= 0,290
	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika
1	pa	4,47	1,37	pa	4,09	0,99	in	4,57	3,18	ne	4,23	1,61	je	4,43	1,09
2	je	3,75	0,41	ne	3,21	0,59	je	2,66	-0,68	je	3,69	0,35	ja	4,23	2,26
3	ne	2,16	-0,46	je	2,16	-1,18	nas	2,1	1,55	ja	2,9	0,93	pa	3,74	0,64
4	da	1,92	0,30	aleluja	2,06	pp	za	1,97	1,05	pa	2,73	-0,37	ne	2,94	0,32
5	tak'	1,61	pp	*e	1,71	0,76	v	1,89	0,67	se	1,96	0,30	se	1,52	-0,14
6	bi	1,46	0,69	da	1,66	0,04	se	1,73	0,07	da	1,9	0,28	a	1,29	0,50
7	ja	1,42	-0,55	se	1,66	0,00	da	1,33	-0,29	to	1,56	0,71	da	1,29	-0,33
8	se	1,42	-0,24	a	1,38	0,59	prosi	1,33	BP	na	1,28	0,36	tam	1,19	0,71
9	na	1,25	0,33	*s'n	1,2	pp	ki	1,22	pp	*e	1,22	0,27	to	1,19	0,34
10	če	1,08	0,40	ja	1,2	-0,77	s	1,2	0,71	kaj	1,22	0,69	v	1,16	-0,06
11	si	0,94	0,31	ti	1,18	0,44	Gospod	1,17	pp	v	1,14	-0,08	so	1,06	0,48
12	*s'n	0,91	pp	tak'	1,1	pp	vse	0,93	0,54	vem	0,99	0,55	*e	1,03	0,08
13	za	0,91	-0,01	v	1,05	-0,17	na	0,85	-0,07	in	0,94	-0,45	jaz	0,93	0,22
14	in	0,89	-0,50	jaz	1	0,29	o	0,77	pp	*s'n	0,88	pp	saj	0,89	0,46
15	v	0,84	-0,38	če	0,95	0,27	te	0,74	0,42	ti	0,85	0,11	ni	0,86	0,22
16	ni	0,82	0,18	še	0,85	0,21	z	0,74	0,35	al'	0,8	pp	*hehe	0,83	pp
17	z'aj	0,79	pp	tri	0,85	pp	aleluja	0,64	pp	še	0,77	0,13	*pol	0,83	pp
18	bo	0,77	0,30	okej	0,8	pp	si	0,64	0,01	saj	0,74	0,31	bi	0,83	0,06
18	jaz	0,77	0,06	bi	0,78	0,01	ti	0,64	-0,10	ni	0,71	0,07	tud'	0,79	pp
20	*e	0,74	-0,21	*mihoho	0,78	BP	po	0,61	0,27	so	0,71	0,13	al'	0,76	pp

Pregled seznamov najpogostejših različnic nam da nekaj slike o konkretnih okoliščinah gradiv. V vseh seznamih nastopajo samo 4 različnice: <da|je|v|se>. Najbolj samosvoj, izstopajoč seznam ima 3. gradivo: zelo nadpovprečno pogosta (tudi med vsemi primerjavami v diplomii) je različnica <in>, nasprotno pa med 20 pogostimi ni različnice <pa>, ki jo je najti med 4 najpogostejšimi v ostalih 4 seznamih. Samo v seznamu 3. gradiva se pojavi kar 10 različnic: <prosi|nas|Gospod|vse|s|z|te|po|ki|o>; 2 od njih skupaj z nadpovprečno pogostim <za> sestavljata frazo <prosti za nas> iz litanij, značilna različnica za 3. gradivo (maša) pa je tudi <aleluja>, ki se sicer pojavi še v 2. gradivu. Ostala gradiva imajo skupne še 4 različnice, ki jih ne najdemo v seznamu 3. gradiva <*e|ja|ne|pa> – 3 od njih se pojavljajo med 4 najpogostejšimi v seznamih teh gradiv, kar pomeni res veliko razliko med 3. gradivom in ostalimi. Razlog je v tem, da je v 3. gradivu prosti govor zelo redek, in še to večinoma iz ust enega govorca (maševalca).

Med ostalimi gradivi stojita 4. in 5. gradivo, v katerih prevladuje pokrajinski pogovorni jezik (govorci so iz različnih narečnih skupin), manj izrazito nasproti 1. in 2. gradivu z narečnim jezikom (koroške narečne skupine, mešajoče se s štajerskimi značilnostmi). 4. in 5. gradivo tako v seznamih nimata <tak'|*s'n>, imata pa druge posebne pogovorne ali tematsko značilne različnice. Samo v seznamu 4. gradiva se pojavljajo različnice <kaj|vem|to|al'>, samo v seznamu 5. gradiva pa <*hehe|*pol|tam> (očitno je situacija v nagnetenem avtu šaljiva, prislov kaže na „navigacijo“ med vožnjo), nadpovprečno pogost pa je v 5. gradivu tudi <ja>. V 2. gradivu se med 4 različnicami, značilnimi samo za njegov seznam, pojavljajo elementi pesmi <aleluja>, opevanja <*mihoho> in vodenja pevskih vaj <okej|tri> – slednja različnica iz fraze „tri štir!“; ki jo najdemo med najpogostejšimi povedmi. Najmanj lastnih različnic pa ima presenetljivo največje gradivo, prvo, le dve, in še ti malo povesta o tematiki sporazumevanja v okoliščinah tega gradiva: <z'aj|bo>.

Primerjava 3: besedila istih okoliščin (istih kot v primerjavah #4 in #7: govorec 001Mz z vsaj enim od govorcev 002Žz, 003Mz – sin, mati, oče; v osebni stiku, v Lovrencu ali v avtu na poti Lovrenc–MB) [31 različnic v 4 vzorcih]												
okoliščine	1. gradivo: družina pri nedeljskem kosilu <i>besedila</i> : 1ned: 24Dm, 25Vg			2. gradivo: družina ob večerji konec tedna <i>besedila</i> : 6pet: 78Vg, 80Vr			3. gradivo: družinski pogovori v nedeljo dopoldan in popoldan <i>besedila</i> : 1ned: 22Vg, 26Dr, 27Vr, 28Dm, 34Vg, 36Vg, 40Dg, 41Vr			4. gradivo: družina pri nedeljskem zajtrku, pred njim in po njem <i>besedila</i> : 1ned: 02Dm, 04Vg, 06Dm		
	zap. št.	R=	P=	R/P=	R=	P=	R/P=	R=	P=	R/P=	R=	P=
	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika
1	pa	4,46	0,04	pa	4,47	0,05	pa	4,2	-0,22	pa	4,53	0,12
2	je	3,74	0,19	je	4,21	0,66	je	2,76	-0,80	je	3,51	-0,05
3	ne	2,16	-0,17	bi	2,3	0,61	ne	2,73	0,40	ne	2,37	0,04
4	da	1,92	-0,22	da	2,25	0,11	da	2,27	0,13	da	2,12	-0,02
5	tak'	1,61	0,26	ne	2,05	-0,28	ja	2,09	0,64	v	1,84	0,30
6	bi	1,46	-0,24	v	1,91	0,37	bi	1,59	-0,11	tak'	1,51	0,16
7	ja	1,44	-0,01	se	1,57	0,13	v	1,56	0,02	bi	1,43	-0,27
8	se	1,42	-0,02	če	1,32	0,39	se	1,41	-0,03	se	1,35	-0,09
9	na	1,25	0,24	*s'n	1,12	0,19	tak'	1,26	-0,10	ja	1,26	-0,19
10	če	1,08	0,15	tak'	1,04	-0,32	še	1,2	0,39	si	1,18	0,39
11	si	0,94	0,15	ja	1,01	-0,44	na	0,98	-0,03	na	1,02	0,01
12	*s'n	0,91	-0,02	a	0,81	0,10	so	0,95	0,30	*s'n	0,94	0,01
13	za	0,91	0,11	*te	0,79	0,26	al'	0,92	0,14	še	0,82	0,01
14	in	0,89	0,22	jaz	0,79	0,08	za	0,89	0,09	al'	0,77	-0,01
15	v	0,84	-0,70	na	0,79	-0,22	a	0,8	0,09	to	0,73	0,07
16	ni	0,82	0,23	za	0,79	-0,01	kaj	0,8	0,29	smo	0,69	0,30
17	z'aj	0,79	0,22	al'	0,76	-0,02	to	0,8	0,14	sem	0,65	pp
18	bo	0,77	0,13	bo	0,73	0,09	če	0,77	-0,17	*te	0,61	0,08
18	jaz	0,77	0,06	in	0,73	0,05	*s'n	0,74	-0,19	jaz	0,61	-0,10
20	*e	0,74	0,13	*e	0,7	0,09	*e	0,71	0,10	za	0,61	-0,19

Gradiva sporazumevanj v istih okoliščinah srečamo v zgornji in še v 2 kasnejših primerjavah (govorcev med seboj ter posameznih vlog istega govorca med seboj). Zdaj primerjamo torej 4 podkorpuse, med katerimi je najmanjši velik skoraj za komaj polovico največjega; vseh različnic primerjanih seznamov je 31, skoraj pol manj kot v prejšnji primerjavi in tudi manj kot v prvi primerjavi, kjer smo tudi imeli 4 gradiva, a 39 različnic. Primerjana gradiva so torej zelo podobna; tudi iz izračunov R/P bi težko ocenili najrevnejše in najbogatejše gradivo po besednem zakladu. Najrevnejši je verjetno besedni zaklad 2. besedila, pri katerem je tudi R/P

najmanjši; to bi lahko potrdili s kvalitativnimi podatki iz drugih virov: 58 % govora v njem pripada govorcu 003Mz, ki pripoveduje v dolgih povedih o eni in isti temi (kopanju kanalizacije) – eno njegovo poved iz tega gradiva smo navedli tudi med najdaljšimi povedmi v korpusu, prim. Priloga 7.

Tudi v primerjavi seznamov najpogostejših besed ni velikih razlik: ne srečujemo več velikih odstopanj od povprečnega deleža različnice (le velikosti do 1 %), tudi le ena različnica v seznamih je taka, da je ni v katerem od seznamov (<sem> v 4. gradivu), obe najpogostejši različnici <pa|je> sta trdno na svojem mestu mesto v vseh seznamih, v vseh seznamih je skupnih kar 12 različnic, naštejmo jih od povprečno najpogostejše proti najredkejši: <pa|je|ne|da|bi|v|ja|se|tak'|na|*s'n|za>. Izstopajočega gradiva ali skupine gradiv ne moremo jasno oceniti. Naštejmo lahko različnice, ki nastopajo le v posameznih seznamih: <z'aj|ni> v 1., <so|kaj> v 3. in <sem|smo> v 4. Nadpovprečno pogoste različnice, ki jih je po več v posameznem gradivu, so <ja|ne|bi> v 3. seznamu in <je|bi|če|v> v 2. seznamu.

16. GOVORCI MED SEBOJ

Govorci so v korpusu sicer pojmovani kot okoliščine, a jih pregledujemo posebej, ker imajo svoj ritem variiranja: bodisi se z istimi lastnostmi ponavljajo v korpusu pri več besedilih bodisi variirajo v vlogah od besedila do besedila. Večina govorcev ni pomembna, ima le manjši prispevek v korpusu: 70 govorcev nastopi le v enem besedilu, 70 jih spregovori le do 40 besed. Zato je bolj kot število govorcev, ki ustrezajo posameznemu podatku o govorcih, reprezentativen podatek o tem, kolikšen delež besedišča korpusa (v %) spregovori ustrezen govorec. Zapis podatkov o govorcih je bil predstavljen v poglavjih II.8.3 in II.8.6, konkretni podatki za korpus pa so podani v naslednji tabeli. S sivo je označen podatek, katerega vrednosti (kategorije) niso bile vnaprej predvidene, ampak so bile povzemale sestavljene po koncu dopolnjevanja seznama govorcev. Podrobni podatki o posameznih govorcih so predstavljeni še na koncu (Priloga 2), v nadaljevanju pa povzemam **sliko tipičnega govorca** oz. govorca z najbolj verjetnimi lastnostmi v korpusu:

- je moškega spola, imamo zabeleženo njegovo ime ali javni medij pojavljanja, star je 20–39 let, iz koroške narečne skupine, končano ima srednjo šolo ali fakulteto ter je iz kroga družine izbranega govorca – snemalca;
- v sporazumevanju običajno nastopa zaznamovan predvsem s svojo vlogo v družini, v posameznem besedilu spregovori vsaj 11 izjav oz. vsaj 100 besed.

V bistvu smo opisali nosilca snemanja, „izbranega“ govorca s šifro 001Mz, za katerega bomo opravili tudi še stilometrično analizo. V korpusu spregovori 13.500 besed, naslednji „najzgovornejši“ govorec je njegov oče 003Mz s 6.800 besedami, tretja njegova mati 002Žz s 6.000 besedami, četrta njegova sestra 086Žz s 3.900 besedami in peti njegov sostanovalec študent 058Mz z 2.700 besedami. Vsi ti so nastopali so gradivih prejšnjih dveh primerjav, prve tri pa bomo primerjali še v naslednji primerjavi za tabelo.

Tabela 47: Pregled podatkov govorcev v korpusu in posameznih besedilih

raven govorca	podatek	vrednost ali obseg vrednosti		delež korpusa iz ust govorca, ki ustreza vrednosti [št. govorcev, ki ustrezajo vrednosti] ¹⁶⁶											
govorec – stalno	V) spol+število	<i>moški</i>	61,6 [49]	<i>ženska</i>	34,3 [48]	<i>neznano</i>	2 besedi [1]	<i>več (skupaj)</i>	4,1 [5]						
	VI) poznanost	<i>znan</i>	91,4 [39]	<i>javno znan</i>	3,7 [21]	<i>avtomat</i>	1,2 [4]	<i>neznano</i>	3,7 [39]			<i>40–46</i>	4,5 [1]		
	VII) starost ¹⁶⁷	<i>7–15</i>	0,8 [5]	<i>16–19</i>	0,9 [5]	<i>20–26</i>	39,0 [13]	<i>27–39</i>	16,0 [10]	<i>50–65</i>	26,3 [2]	<i>več</i>	4,1 [5]	<i>neznano</i>	8,3 [62]
	VII) govorno poreklo ¹⁶⁸	<i>kor.</i>	58,5 [24]	<i>štaj.</i>	19,7 [9]	<i>dol.</i>	0,9 [2]	<i>gor.</i>	3,0 [2]	<i>prim.</i>	5,6 [2]	<i>več</i>	4,1 [5]	<i>neznano</i>	7,2 [56]
	IX) izobrazba	<i>(še) brez</i>	0,8 [5]	<i>OŠ</i>	0,9 [5]	<i>SŠ</i>	37,8 [14]	<i>VŠUNI</i>	49,5 [13]	<i>PODIPL</i>	0 [0]	<i>več</i>	4,1 [5]	<i>neznano</i>	6,9 [61]
	X) razmerje do 001Mz	<i>isti</i>	27,9 [1]	<i>družina</i>	35,7 [6]	<i>prijatelj</i>	10,3 [5]	<i>znanec</i>	17,5 [27]	<i>javno spoznan</i>	4,8 [22]	<i>neznaneč</i>	3,8 [42]		
govorec v korpusu	XIO) prisoten v besedilih	<i>1</i>	18,3 [70]	<i>2</i>	5,1 [15]	<i>3</i>	1,1 [6]	<i>4–5</i>	18,6 [7]	<i>6–10</i>	2,7 [2]	<i>10<</i>	52,4 [3]		
	10O) izjave	<i>1–5</i>	4,1 [60]	<i>6–10</i>	0,5 [8]	<i>11–50</i>	9,8 [21]	<i>51–100</i>	4,4 [4]	<i>100<</i>	81,1 [10]				
	12O) besede	<i>1–10</i>	0,3 [30]	<i>11–100</i>	3,4 [40]	<i>101–500</i>	8,9 [18]	<i>501–1000</i>	9,8 [7]	<i>1000<</i>	77,6 [8]				
govorec v posameznem besedilu ¹⁶⁶	XOB) vloga	<i>sam</i>	0,9 [1]	<i>sostanovalec študent</i>	12,0 [6]	<i>priložnost ni znanec</i>	2,6 [6]	<i>civilnodružben a funkcija</i>	6,7 [8]	<i>novinar RTV</i>	2,6 [7]	<i>bralec RTV reklame</i>	0,7 [5]	<i>glasbenik</i>	13,3 [20]
		<i>družinsk i član</i>	42,1 [9]	<i>znanec /prijatelj</i>	16,1 [11]	<i>neznaneč</i>	0,1 [13]	<i>uradni/ poslovni odnos</i>	0,9 [16]	<i>vir novinarja</i>	1,4 [12]	<i>javno obveščanje</i>	0,6 [7]		
	10OB) izjave	<i>1</i>	1,1 [47]	<i>2–3</i>	3,0 [34]	<i>4–10</i>	4,9 [28]	<i>11–50</i>	27,9 [25]	<i>51–100</i>	26,9 [9]	<i>100<</i>	36,2 [8]		
	12OB) besede	<i>1–10</i>	1,0 [53]	<i>11–100</i>	7,7 [49]	<i>101–500</i>	35,5 [25]	<i>501–1000</i>	18,7 [11]	<i>1000<</i>	37,1 [7]				

¹⁶⁶ Na tretji ravni – govorec v posameznem besedilu – je v ogletem oklepaju zapisano število različnih govorcev, ki ustrezajo temu pogoju, ne posam. vloge, zato vsota govorcev za vse vrednosti posameznega podatka ni 103 (kolikor je sicer govorcev v korpusu)!

¹⁶⁷ Primerov za vrednosti <6 | 66= ni.

¹⁶⁸ Primerov za vrednosti pan. | rovt. | MB ni.

16.1. Slogi govorcev

Gradivo za naslednjo primerjavo smo že spoznali, a drugače organizirano – v podkorpusih, v katerih so bila cela besedila s sorodnimi okoliščinami. Posamezno gradivo v spodnji primerjavi pa ni sestavljeno iz celih besedil, marveč iz vseh besed, ki jih v naštetih besedilih spregovori določen govorec (ali z vidika partiture transkripcije: gradivo sestavljajo vsi verbalni – [v]-traki določenega govorca v naštetih besedilih). V naštetih 15 besedilih ni drugih govorcev kot omenjeni trije: v 9 nastopajo vsi trije, v 3 samo sin in mati, v 3 samo oče in sin. Poleg naštetih obstajajo še druga besedila v korpusu, v katerih nastopajo ti govorcev, tudi v medsebojni kombinaciji, a v drugačnih okoliščinah in z drugimi govorcev: 001Mz še v 75 drugih besedilih, 002Žz še v 3 drugih, 003Mz še v 8 drugih.

govorci	Primerjava 4: govorcev v istih okoliščinah								
	(istih kot primerjava #3 in #7: govorec 001Mz z vsaj enim od govorcev 002Žz, 003Mz – sin, mati, oče; v osebnem stiku, na domu v Lovrencu ali v avtu na poti Lovrenc–MB) [28 različnih v spodnji primerjavi]								
	besedila: (vsa besedila iz primerjave #3, razdeljena po govorcev) 1ned: 02Dm, 04Vg, 06Dm, 22Vg, 24Dm, 25Vg, 26Dr, 27Vr, 28Dm, 34Vg, 36Vg, 40Dg, 41Vr; 6pet: 78Vg, 80Vr								
zap. št.	2. gradivo: 002Žz			3. gradivo: 003Mz			1. gradivo: 001Mz		
	R= 1611	P= 5418	R/P= 0,297	R= 1620	P= 4985	R/P= 0,325	R= 1121	P= 3040	R/P= 0,369
	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika
1	pa	4,36	-0,04	pa	4,55	0,15	pa	4,28	-0,12
2	je	3,32	-0,21	je	4,15	0,62	je	3,13	-0,40
3	ne	2,42	0,02	da	2,31	0,22	ne	3,03	0,63
4	da	2,14	0,05	bi	1,83	0,13	ja	2,5	0,91
5	bi	1,64	-0,06	ne	1,75	-0,65	da	1,81	-0,28
6	tak'	1,46	0,14	v	1,69	0,23	bi	1,64	-0,06
7	v	1,4	-0,06	se	1,42	-0,05	se	1,64	0,17
8	se	1,35	-0,12	tak'	1,38	0,06	mhm	1,38	0,71
9	ja	1,27	-0,32	*s'n	1,1	0,14	v	1,28	-0,18
10	na	1,02	0,00	na	1,04	0,02	če	1,15	0,15
11	si	1	0,21	če	1	0,00	*s'n	1,12	0,16
12	za	0,9	0,08	ja	1	-0,59	tak'	1,12	-0,20
13	še	0,89	0,09	a	0,84	0,15	*e	1,02	0,33
14	to	0,87	0,24	in	0,8	0,13	na	0,99	-0,03
15	če	0,85	-0,15	*te	0,78	0,30	al'	0,95	0,15
16	so	0,81	0,20	al'	0,74	-0,06	si	0,92	0,13
17	kaj	0,79	0,30	jaz	0,74	0,02	še	0,92	0,12
18	a	0,74	0,05	za	0,7	-0,12	za	0,86	0,04
19	ko	0,74	0,25	so	0,64	0,03	bo	0,76	0,09
20	z'aj	0,74	0,18	*e	0,62	-0,07	jaz	0,72	0,00

Kaj lahko rečemo o slogu govorcev na podlagi zgornjih treh seznamov in izračunov? 001Mz je najbolj redkobeseden (čeprav nastopa v vseh 15 besedilih, ostala dva pa v le po 12), 002Žz najbolj gostobesedna. 003Mz ima besednjak nedvomno bogatejši od 002Žz, govorca 001Mz pa z njima ne moremo gotovo primerjati po R/P.

Primerjava je nasploh skromnejša od predhodnih treh primerjav (primerjamo samo 3 gradiva), zato je tudi število vseh različnic v 3 seznamih skupaj manjše – 28, kaka različnic manj je tudi zaradi sorodnosti okoliščin in sobesedilnosti gradiv, ki so si drug drugemu notranji kontekst. Tudi število različnic, ki jih najdemo v vseh treh seznamih, je veliko (12), pravzaprav gre za 12 povprečno najpogostejših različnic. Naštejmo jih po padajoči povprečni pogostosti:

<pa|je|ne|da|bi|ja|se|v|tak'|na|če|za>. Med posivljenimi različnicami si pri nadaljnji interpretaciji ne moremo pomagati z različnico, ki je v vseh seznamih na istem zaporednem pogostostnem mestu, <je>, zanimiva pa je izpostavljenost <ja> v vseh treh seznamih: pri govorcju 001Mz ima tolikšen delež, kot ga ima pri ostalih dveh govorcjih skupaj.¹⁶⁹

Najbolj izstopa govor 002Žz: njen seznam ima največ specifičnih različnic <kaj|ko|to|z'aj>, manjkajo pa ji tudi kar 4 različnice, ki jih najdemo v obeh ostalih seznamih <*e|*s'n|a'l|jaz>. Na podlagi teh njenih posebnosti lahko rečemo, da ima med vsemi tremi govorcji najmanj narečen govor oz. da njen govor vsebuje več značilnosti pokrajinskega pogovornega ali celo knjižnega pogovornega jezika.

Pri govorcju 001Mz sta izrazito nadpovprečno pogosti različnici <ja|mhm> (največja pozitivna odstopa v vseh treh seznamih), nekoliko tudi <*e>. Samo njegov seznam ima različnici <bo|mhm>, nima pa različnic <a|so>, ki sta prisotni v drugih dveh seznamih. Na podlagi izstopajočih različnic seznama govorca 001Mz lahko rečemo, da ima v govoru nekoliko več mašil (mhm|ja|*e) oz. da postavlja manj vprašanj in več pritrjuje.

Specifični različnici govorca 003Mz sta <*te|in>, prva pri njem tudi izrazito nadpovprečno izstopa; morda bi ju lahko interpretirali kot element pripovednega načina ubesedovanja, ki pri njem izstopa. Samo on nima različnic <si|še>, ki ju najdemo v drugih dveh seznamih. Govorec 003Mz naj bi skupaj z 001Mz bil bolj podoben in imel več značilnosti narečne socialne zvrsti jezika (zaradi izstopanja 002Žz v nasprotno smer), pa si tudi ta dva govorca pri eni različnici diametralno nasprotujeta – kolikor je delež <ne> pri 003Mz podpovprečen, je pri 001Mz nadpovprečen.

16.2. Vloge izbranega govorca proti sobesedilu

Zakaj je važno slogovna primerjava izbranega govorca s sobesedilom? V **sobesedilu** po eni strani gre za vse, kar govorec sliši, a je izvorno drugega sloga kot njegov govor – in tega sloga se govorec tudi navzema; glede na ta argument bi lahko vzeli za sobesedilo kar ves preostali del korpusa, ki ne pripada posameznemu govorcju. Po drugi strani gre pri sobesedilu za besedišče, ki šele kompatibilno z govorcjevimi sestavlja zaključeno besedilo ali zbirko besedil – pričakovati je, da se besedišče pri govorcju in v sobesedilu prekriva, ponavlja, saj oba skupaj obdelujeta isto temo pogovorov; po tem argumentu pa raje gradivo za sobesedilo *vzamemo le iz besedil, v katerih govori tudi izbrani govorec*. Še bolj natančno bi bilo za sobesedilo vzeti le tiste sogovorke, ki dvosmerno komunicirajo z izbranim govorcjem (ob strani pa pustiti enosmerna javna obvestila, predvajano glasbo ipd.) oz. se govorec nanje odziva, a tako daleč nismo šli.

V samem korpusu izbrani govorec 001Mz sodeluje v 90 besedilih (v 22 od teh je govorec le on sam, v njih spregovori 867 besed), ki pokrijejo 46.426 besed ali 96 % korpusa. V primerjavi, gl. spodaj, je razmerje govorca proti sobesedilu iz teh 90 besedil 1 : 2,4.

Glede na to, da primerjamo le 2 gradivi, bi bilo po izkušnjah prejšnjih primerjav pričakovati izredno podobnost obeh gradiv. Dejansko se v obeh seznamih skupaj pojavi le 23 različnic, 17 pa jih je skupnih; in tudi tisti pari od 17, ki so posivljeni <je|ne|ja|*e>, imajo v obeh seznamih seveda razlike deležev nasprotnega predznaka, saj je povprečen delež izračunan le iz njihovih dveh vrednosti. Le v sobesedilu se pojavijo 3 različnice – <in|so|to>, le pri govorcju 001Mz pa prav tako 3 – <če|mhm|ti>.

¹⁶⁹ Seveda gre samo za deleže, ki so relativni; ker je gradivo za 001Mz najmanjše, je delež <ja> pri njem v tem preračunavanju „na oko“ pretiran. Po absolutnem številu ima 001Mz 76x <ja>, 002Žz 69x, 003Mz pa 50x.

zap. št.	Primerjava 5: govorec proti sobesedilu v podkorpusu vseh besedil, kjer se pojavi (različne okoliščine) [23 različnic v spodnji primerjavi]					
	besedila ¹⁷⁰					
	1. gradivo: sobesedilo (brez 001Mz)			2. gradivo: samo govor 001Mz		
	R=6.265	P= 32.822	R/P=0,191	R=3.173	P= 13.485	R/P= 0,235
	različnica	delež [%]	razlika	različnic a	delež [%]	razlika
1	je	3,66	0,38	pa	3,71	0,05
2	pa	3,6	-0,06	ne	3,21	0,30
3	ne	2,61	-0,30	ja	2,97	0,35
4	ja	2,27	-0,35	je	2,91	-0,38
5	da	1,96	0,21	*e	1,7	0,46
6	se	1,54	0,09	da	1,54	-0,21
7	v	1,46	0,00	v	1,46	0,00
8	in	1,25	0,46	se	1,37	-0,09
9	bi	1,02	0,06	na	1,13	0,10
10	to	0,95	0,18	tak'	0,97	0,04
11	na	0,94	-0,10	*s'n	0,93	0,14
12	a	0,92	0,11	bi	0,9	-0,06
13	tak'	0,89	-0,04	še	0,89	0,10
14	za	0,81	0,04	če	0,82	0,09
15	*e	0,78	-0,46	ti	0,82	0,13
16	jaz	0,77	-0,01	jaz	0,78	0,01
17	še	0,69	-0,10	si	0,76	0,05
18	si	0,66	-0,05	mhm	0,74	0,23
19	so	0,66	0,07	za	0,73	-0,04
20	*s'n	0,65	-0,14	a	0,7	-0,11

Kljub majhnemu številu primerjav pa lahko vseeno s pomočjo razlik v deležih posameznih različnic in s pomočjo pravkar naštetih specifičnih različnic vsakega seznama razločimo sloga gradiv. Govor 001Mz je toliko bolj *prostogovoren*, da imata v njem mesto pogostejše različnice <ne> kot vprašalna členica ter <ja> kot oporni signal. Sobesedilo pa je toliko manj prostogovorno, da se v njem <je> prebije na mesto najpogostejše različnice (čeprav mu je <pa> takoj za petami), <in>, ki ga v drugem gradivu ni, pa ima v tem seznamu najvišjo pozitivno razliko v deležu različnice. Tudi nadpovprečni deleži drugih značilno prostogovornih (ter narečnih in pogovornih) različnic pri 001Mz, ki smo jih kot take spoznali *že v prejšnjih primerjavah*, pričajo o tem, da je jezik 001Mz v korpusu nasploh prostogovoren: <pa|*e|tak'|*s'n|mhm>.

Najvažnejši od teh so: mašilo <*e> kot različnica z največjo pozitivno razliko v deležu, oporni signal <mhm> kot specifična različnica tega seznama, <in>-izrinjajoči <pa> kot najpogostejša različnica.

16.3. Vloge izbranega govorca med seboj

Slog idiolekt smo že določili v primerjavi z drugimi idiolekti in sociolektom, sedaj se spuščamo še na nižjo raven primerjave sporazumevanja: variiranje idiolekt posameznika v različnih tipiziranih okoliščinah (*diskurzih*) – prva sledeča primerjava – ter med besedili istih tipiziranih okoliščin (besedilo kot v jezikoslovju bolj uveljavljena enota sporazumevanja, ki ima zaokrožen sklop sporazumevalnih okoliščin) – druga sledeča oz. zadnja primerjava.

Gradivo našega korpusa omogoča, da govor izbranega govorca spremljamo z vidika spreminjanja sporazumevalnih okoliščin neprekinjeno za dva dneva, za ta čas bi lahko izrisali celo nekakšen **sociogram** ali graf sporazumevalnih okoliščin in spreminjanja govora. Naj kot takšen služi pregled besedil (Priloga 1), v katerem so besedila razporejena po vrsti, kot so si dejansko sledila. Lahko pa z vidika našega korpusa ovrednotimo trditev, da **človek dnevno spregovori** od 10.000 do 18.000 besed (Suhadolnik/Janežič 1962: 48–49) – govorec 001MZ jih v nedeljo spregovori 8.015, v petek pa 5.470. Težje ovrednotimo na istem mestu zapisano trditev, da je v tej

¹⁷⁰ Besedila za obe gradivi: 1ned: 02, 04–06, 08–11, 13–19, 20, 22, 24–30, 32, 33–37, 40–41, 43, 45, 47–48, 52, 55, 57–58, 60, 62, 70–72; 6pet: 01–03, 05, 07, 09, 11, 13, 15, 17–19, 21–22, 24–25, 27–28, 32, 34, 36–38, 40, 42–43, 47, 51–52, 54–55, 58, 61, 65, 67, 69, 71, 73–78, 80–81.

dnevni količini govora človeka 1.000 leksemov; govorec 001Mz v nedeljo spregovori 2.190, v petek pa 1.583 različnic, leksemov pa je v vsakem posameznem dnevu verjetno manj kot tisoč.

Primerjava 6 spodaj je med vsemi v diplomu, ki primerjajo 4 gradiva (še 1., 3. in 7. primerjava), enaka najbogatejši po številu skupnih različnic – prvi, ki jih ima prav tako 39. To dejstvo je povezljivo s tem, da primerjamo govor v različnih okoliščinah, kot smo v prvi primerjavi primerjali podkorpuse različnih prenosnikov in funkcijskih zvrsti. Torej tudi najpogostejše različnice istega govorca (in idiolekta) močno variirajo (sic!) – tega znotraj idiolekta morda ne bi pričakovali.

Pomemben razlog raznovrstnosti 6. primerjave bi znal biti velikost oz. majhnost njegovih gradiv – v manjših gradivih imajo posamezne različnice večji delež, seznam najpogostejših pa je manj predvidljiv. V četrtem, najmanjšem gradivu, je absolutna pogostost 20. različnice <jaz> le 6 pojavitev (20. različnica najmanjšega gradiva 1. primerjave, prostega govora, <so>, ima 238 pojavitev). Ali drug izračun, ki pove, da razlika v velikosti gradiv 1. in 6. primerjave vendarle ni tako zelo velika: vseh najpogostejših 20 različnic najmanjšega gradiva 1. primerjave pokrije 29 % tega gradiva, pri najmanjšem gradivu 6. primerjave pa 33 % gradiva. 6. primerjava torej vseeno priča, da idiolekt v različnih okoliščinah močno variira, čeprav manj kot gradiva v 1. primerjavi.

Gradiva 6. primerjave so velikostno kar primerljiva, še najbolj prvi največji dve in najmanjši dve med seboj v paru; med prvima dvema gradivoma ima najraznovrstnejši zaklad besednih oblik 1., med drugima dvema pa 4. gradivo. Če upoštevamo ugotovitev, da je R/P z večanjem gradiv običajno pada, lahko ocenimo, da je tudi relativno najraznovrstnejši zaklad besednih oblik v vsej 6. primerjavi v 1. gradivu.

Primerjava 6: govorec 001Mz v različnih okoliščinah (4 besedila različnih okoliščin z absolutno največjim prispevkom govorca) [39 različnih v spodnji primerjavi]														
vloga	1. gradivo: pogovor z znanko študentko med vožnjo z vlakom <i>besedilo:</i> 6pet54Vg			2. gradivo: vodenje avdicije v pevskem zboru <i>besedilo:</i> 1ned18Vg			3. gradivo: klepet s študentom sostanovalcem pred spanjem <i>besedilo:</i> 1ned71Dg			4. gradivo: družina pri nedeljskem kosilu <i>besedilo:</i> 1ned25Vg				
	zap. št.	R=	P=	R/P=	R=	P=	R/P=	R=	P=	R/P=	R=	P=	R/P=	
	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika		
1	ne	5,41	1,71	pa	5,31	0,79	je	4,41	1,25	pa		5	0,48	
2	pa	4,59	0,07	*e	2,74	1,21	ne	4	0,30	je		3,4	0,24	
3	ja	3,77	1,49	ne	2,74	-0,96	pa	3,18	-1,34	ne		2,66	-1,04	
4	je	3,03	-0,14	da	1,82	0,19	ja	2,46	0,18	ja		2,23	-0,05	
5	vem	1,56	0,70	je	1,82	-1,35	da	2,16	0,53	se		1,7	0,21	
6	na	1,56	0,18	tri	1,74	1,19	*s'n	2,05	1,03	na		1,6	0,22	
7	mhm	1,48	0,62	tak'	1,74	0,55	na	1,85	0,47	mhm		1,6	0,74	
8	v	1,23	0,09	ti	1,66	0,82	al'	1,64	0,70	bi		1,6	0,73	
9	se	1,23	-0,26	se	1,58	0,09	vem	1,44	0,58	ni		1,49	0,91	
10	tak'	1,15	-0,04	*mm	1,33	pp	se	1,44	-0,05	tak'		1,38	0,19	
11	da	1,15	-0,48	v	1,33	0,19	*e	1,44	-0,10	si		1,38	0,54	
12	še	1,07	0,13	ste	1,24	0,81	pač	1,33	0,64	da		1,38	-0,25	
13	so	0,9	0,37	še	1,24	0,30	bil	1,13	pp	če		1,28	0,51	
14	samo	0,9	0,29	okej	1,08	BP	kaj	1,03	0,48	in		1,06	0,52	
15	*e	0,9	-0,64	če	1,08	0,31	a	1,03	0,32	*e		1,06	-0,48	
16	za	0,82	0,24	bi	1,08	0,21	v	1,03	-0,11	v		0,96	-0,18	
17	pač	0,82	0,13	to		1	0,31	to	0,92	0,23	še		0,96	0,01
18	pr'	0,74	pp	a		1	0,29	samo	0,82	0,21	bo		0,96	0,38
19	bo	0,74	0,16	lahk'		0,91	pp	b'lo	0,72	pp	*s'n		0,96	-0,06
20	bom	0,66	0,29	si		0,91	0,07	jaz	0,72	0,05	jaz		0,64	-0,03

V vseh seznamih se pojavlja 7 različnic, prve tri našteje so redno med 5 najpogostejšimi: <pa|ne|je|da|*e|se|v>. Glede na prejšnje primerjave lahko povzamemo **stalnice izbranega govorca**: nasploh prostogovorne značilnosti (<pa> običajno pogostejši kot <je>, redke <in>, fonetično reducirane različnice in mašila <*e|*s'n|tak'> idr.), specifično pogost pri njem je <ne> (vedno med najpogostejšimi tremi), med bolj pogostimi različnicami pa morda še mrmranje <mhm|*mm>, čeprav nobene teh različnic ni v 3. gradivu zgornje primerjave.

Razlike med seznama gradiv kar nekaj povedo o temi posameznega gradiva. Najbolj izstopa **2. gradivo** – avdicija v pevskem zboru, ki jo vodi prav izbrani govorec. Samo v tem gradivu se med vsemi seznama pojavi kar 6 različnic <*mm|lahk'|okej|ste|ti|tri>, v primerjavi z vsemi ostalimi tremi seznama pa samo v njem umanjata različnici <ja|na>. Velika pozitivna razlika v deležu <*e> ali specifična različnica <*mm|okej> so primeri mašil v govoru, neverbalnih in verbalnega – očitno je v besedilu mnogo negotovosti pri dogovarjanju, odločanju ali čem podobnem, o tem govorita tudi pogostejši <pa> in specifična različnica <lahk'>. Mnogo je v gradivu nagovarjanja v 2. osebi – nadpovprečno pogosti sta tudi različnici <ti|ste>. <tri> je značilen element vodenja upevanja („tri štir'!“), ki je izstopal že v 2. gradivu 2. primerjave (ki je poleg avdicije zajemalo še pevske vaje); tudi umanjkanje <ja>, ki je pri ostalih 3 seznama zelo pogost, pa kaže na to, da 001Mz bolj vodi sporazumevanje kot v njem odgovarja.

Ostala gradiva imajo manj posebnosti, prvo med njimi je **3. gradivo**, v katerem si sestanovalec ob srečanju pred začetkom študijskega leta pripovedujeta, kaj sta doživela v preteklem poletju. Samo v njegovem seznamu se pojavijo različnice <al'|bil|b'lo|kaj>, v primerjavi z vsemi ostalimi tremi seznama pa samo v njem umanjata različnici <še|tak'>. Dve specifični različnici <bil|b'lo> lahko povežemo z največjima odstopoma deležev različnic v pogostost <je|*s'n> v npr. <je+bil>: skupaj povesta, da je za gradivo značilno pripovedovanje v pretekliku. Podobno bi se dalo povezati različnici <ne+vem>, ki v tem gradivu sicer nista med najbolj izrazitimi, a imata pozitivno razliko do povprečnega deleža različnice. Presenetljivo je v 3. gradivu redke <pa>, česar ne znam razložiti (veznik <in> nasproti tej različnici ni toliko pogost, da bi ga našli med prvimi 20). Zanimiva je tudi izstopajoča pozitivna pogostost različnic <al'|pač> – morda gre za vpliv sogovorca, utrujenost v pozni uri pogovora ali za podobno zadržanost kot pri <ne+vem>.

Samo v seznamu **1. gradiva** so 4 različnice, <bom|pr'|so|za>, ki jih srečamo v drugi polovici seznama. Povsem pri vrhu seznama pa najdemo <ne|ja>, dve različnici z največjima nadpovprečnima pozitivnima odstopanjema deležev različnic v vseh 4 seznama, in to kljub temu, da gre za največje gradivo (kjer je zaradi običajnih velikostnih razmerij pričakovati nasploh manjše deleže različnic). Obe različnici se v prostem govoru pogosto pojavljata kot oporna signala, <ne> še kot vprašalna členica, še bolj pa je verjetna njuna raba pri odgovarjanju na t. i. odločevalna vprašanja. Za to rabo govori tudi različnica s tretjo najbolj izstopajočo pozitivno razliko v tem seznamu <vem>, ki jo lahko kot pri 3. gradivu povežemo s pogostim <ne> v <ne+vem>.

Najmanj izstopajoče **gradivo je 4.**: različnici, ki se pojavljata samo v njegovem seznamu, sta le 2, <in|ni>, seznam pa je tudi brez „pp“ in „BP“. Je pa 4. gradivo (po R/P) vseeno relativno drugo najbogatejše po besednem zakladu vseh različnic glede na velikost v pojavnihah.

Število različnic naslednje, 7. in zadnje primerjave (30), je najmanjše med vsemi 4 primerjavami, ki primerjajo 4 gradiva, za eno različnico pa manjše od 3. primerjave, ki primerja gradiva v istih okoliščinah kot 7. primerjava. Če ob tem še upoštevamo, da bi zaradi majhnosti gradiv v primerjavi lahko pričakovali raznovrstnejše različnice (manj predvidljive v posameznem seznamu), lahko že iz najosnovnejših podatkov potrdimo, da se *slogi gradiv ne razlikujejo dosti*. Največje gradivo, 1., niti ni dosti manjše od gradiv v prejšnji primerjavi, najmanjše gradivo, 4., pa je pol manjše kot najmanjše gradivo v prejšnji primerjavi – 20. najpogostejša različnica v njem, <pač>, ima le 4 absolutne pojavitve. Pri samih deležih vidimo, da se ponavljajo za različne

besede – te različnice imajo enako absolutno pogostost, nekaj takih, ki imajo enako pogostost kot 20. najpogostejša v seznamu, pa sem tudi odrezal, kar našo primerjavo oz. sklepanje slabi; različnicam od 15. mesta navzdol pri interpretaciji raje ne damo važnejšega pomena.

Če pogledamo R/P glede na velikost primerjanih gradiv (spet sta paroma primerljivi le največji dve in najmanjši dve gradivi med seboj), ugotovimo, da iz gradiv ne moremo z gotovostjo izločiti relativno najbolj bogatega ali revnega glede na bogatost zaklada različnic.

V vseh seznamih primerjanih gradiv se ponavlja 12 različnic, večinoma v zgornjem delu seznama (naštete so padajoč po povprečni pogostosti): <pa|je|ne|ja|da|se|bi|mhm|v|če|*s'n|al'>. To veliko število pomeni veliko podobnost gradiv v primerjavi, o tem pa govori tudi dejstvo, da se samo pri eni različnici (od 30 v primerjavi) pojavi „pp“. Glede na specifične različnice, ki nastopijo samo v enem seznamu, je najbolj izstopajoče **3. gradivo** – samo v njegovem seznamu so 4 različnice <jaz|ka'|samo|si>, samo pri njem nista <bo|tak'>, ki ju najdemo v vseh ostalih treh seznamih. Druga gradiva so manj izstopajoča po teh dveh pokazateljih: samo v seznamu 4. gradiva sta <ko'|vem> (samo tam presenetljivo ne <*e>), samo v seznamu 1. gradiva je <in> (samo tam pa ne <za>), samo v seznamu 2. gradiva je <to> (ne pozna pa različnic, ki bi jih našli v vseh ostalih treh seznamih, pri njem pa ne).

Primerjava 7: govorec 001Mz v istih okoliščinah (istih kot v primerjavah #3 in #4: govorec 001Mz z vsaj enim od govorcev 002Žz, 003Mz – sin, mati, oče; v osebni stiku, v Lovrencu ali v avtu na poti Lovrenc–MB) [30 različnih v spodnji primerjavi]												
vloge	3. gradivo: družinski pogovori v nedeljo dopoldan in popoldan <i>besedila:</i> 1ned: 22Vg, 26Dr, 27Vr, 28Dm, 34Vg, 36Vg, 40Dg, 41Vr			1. gradivo: družina pri nedeljskem kosilu <i>besedila:</i> 1ned: 24Dm, 25Vg			2. gradivo: družina ob večerji konec tedna <i>besedila:</i> 6pet: 78Vg, 80Vr			4. gradivo: družina pri nedeljskem zajtrku, pred njim in po njem <i>besedila:</i> 1ned: 02Dm, 04Vg, 06Dm		
	zap. št.	R=	P=	R/P=	R=	P=	R/P=	R=	P=	R/P=	R=	P=
	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika	različnica	delež [%]	razlika
1	pa	4,63	0,60	pa	4,99	0,96	je	4,1	0,73	je	3,94	0,57
2	ja	3,43	1,14	je	3,4	0,03	pa	3,17	-0,86	ne	3,94	0,90
3	ne	3,33	0,29	ne	2,65	-0,39	se	2,61	0,93	pa	3,32	-0,71
4	da	2,13	0,33	ja	2,34	0,05	ne	2,24	-0,80	ja	2,28	-0,01
5	je	2,04	-1,33	se	1,7	0,02	da	2,05	0,25	bi	1,66	0,08
6	bi	1,94	0,36	bi	1,59	0,01	*s'n	1,87	0,70	da	1,66	-0,15
7	se	1,39	-0,30	mhm	1,59	0,16	v	1,87	0,47	v	1,66	0,26
8	*e	1,2	0,25	na	1,59	0,72	mhm	1,68	0,25	mhm	1,45	0,01
9	al'	1,2	0,25	ni	1,49	0,79	če	1,49	0,31	si	1,24	0,24
10	še	1,11	0,25	da	1,38	-0,43	tak'	1,49	0,30	tak'	1,24	0,05
11	v	1,11	-0,29	si	1,38	0,38	za	1,31	0,39	vem	1,24	pp
12	*s'n	1,02	-0,15	tak'	1,38	0,19	*e	1,12	0,17	če	1,04	-0,14
13	jaz	1,02	0,36	če	1,27	0,09	al'	1,12	0,17	se	1,04	-0,65
14	mhm	1,02	-0,42	*e	1,06	0,11	bi	1,12	-0,46	za	1,04	0,12
15	bom	0,93	0,39	in	1,06	0,63	bo	1,12	0,30	*s'n	0,83	-0,34
16	če	0,93	-0,25	*s'n	0,96	-0,21	ja	1,12	-1,17	al'	0,83	-0,12
17	ka'	0,93	0,43	bo	0,96	0,14	si	1,12	0,12	bo	0,83	0,01
18	na	0,93	0,06	še	0,96	0,10	bom	0,93	0,39	ko'	0,83	0,53
19	samo	0,93	0,27	v	0,96	-0,44	pač	0,93	0,34	ni	0,83	0,13
20	za	0,93	0,01	al'	0,64	-0,31	to	0,93	0,38	pač	0,83	0,24

Pri opazovanju razlik glede na povprečen delež (posivljene različnice) odkrijemo nekaj diametralnih nasprotij pri najbolj pogostih različnicah: <pa> v vseh 4 seznamih, <je> izstopa v 3. gradivu nasproti ostalim; paroma si deleža <ja> stojita nasproti v 3. in 2. gradivu, deleža <ne>

pa v 2. in 4. gradivu. Menjavanja teh 4 različnic <pa|je|ne|ja> na prvih 4 mestih pogostostnega seznama smo navajeni že iz prejšnjih primerjav (oz. iz gradiv tipično prostega govora), zato nam posamič v tej primerjavi ne povedo dosti, zlasti zato ne, ker se pri najpogostejših različnicah velika razlika v deležu hitro najde.

Zanimivi pa so pogostostni odstopi po ene ali dveh različnic v vsakem gradivu. V 1. gradivu je tak <in>, ki smo ga v primerjavah običajno interpretirali kot pokazatelja višje socialne zvrsti jezika, odstopa <na|ni> pa sta verjetno pogojena s tematiko sporazumevanja. V 2. gradivu je opazen odstop <se> – njegov pomen odkrije pogled v besedilo (konkordance v tem besedilu) – pogosto nastopa v trpnih oblikah glagola. V 3. gradivu (največjem, sic!) imamo največji odstop od povprečne pogostosti pri <je> – govorec v pogovoru sodeluje z izrazito kratkimi, okrnjenimi povedmi (ali kar z odvisnimi stavki, ki dopolnjujejo izjave drugih dveh govorcev), tudi ne pripoveduje dosti, zato <je> prehitijo oporni signali <ja|ne>, veznik <da> in ne mu približajo drugi glagolski prosti morfemi <bi|se>. Pri 4. gradivu lahko podpovprečni odstop <se> zanemarimo v odnosu do nadpovprečnega <se> v 2. gradivu in se obrnemo k <ne>. Odstopanje tega lahko povežemo s specifično različnico tega seznama, <ne>; odgovor <ne+vem> je v tem gradivu očitno pogost.

Za zadnjo primerjavo lahko torej rečemo, da *se idiolekt v isti okoliščinah malo razlikuje* in da nobeno od gradiv ne izstopa izrazito. Glede na druge primerjave zanjo še posebej velja, da izstopanja posameznih različnic sicer lahko hitro interpretiramo v zvezi s tematiko sporazumevanja gradiva, da pa so gradiva tako majhna, da na naš kvantitativni model pomembneje vpliva (pre)več dejavnikov in so zato interpretacije na njegovi podlagi vedno manj zanesljive.

17. OVREDNOTENJE KVOCIENTA R/P

V 7 primerjavah se je zvrstilo 26 primerjanih gradiv in za vsakega od njih smo poskušali ovrednotiti bogatost besednega zaklada (različnic) na podlagi kvocienta R/P (tega smo predstavili v poglavju I.4.3.2). Kvocient pa ni zanesljiva metoda zaradi jezikovnega gradiva, v katerem R ne narašča premosorazmerno z P. Skušajmo ovrednotiti gibanje vseh teh spremenljivk (P, R, R/P) na podlagi naših gradiv – pregled za vseh 26 gradiv je v naslednji tabeli, urejen po naraščanju P, graf pa je narejen za 20 najmanjših gradiv, ki jih je bilo še lahko povezati z linearnimi preračunavanji oz. funkcijo prve stopnje (Slika 28 spodaj).

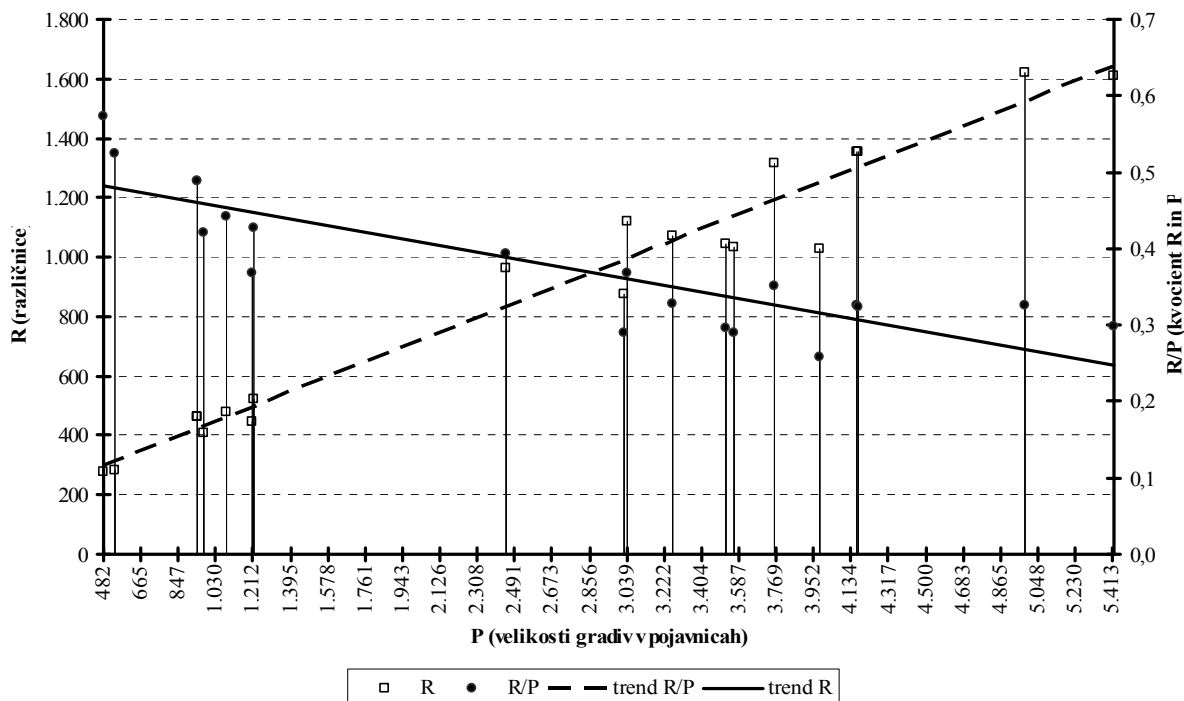
Tabela 48: Spremenljivke P, R, R/P v gradivih slogovnih primerjav (naraščanje po P)

primerjava in gradivo	P	R	R/P	zap. št.	primerjava in gradivo	P	R	R/P
7p4g	482	276	0,573	1/14	3p2g	3.560	1.032	0,290
7p2g	536	281	0,524	2/15	2p3g	3.759	1.317	0,350
6p4g	940	460	0,489	3/16	2p2g	3.984	1.030	0,259
7p1g	942	461	0,489	4/17	2p1g	4.164	1.352	0,325
6p3g	974	409	0,420	5/18	3p1g	4.168	1.353	0,325
7p3g	1.080	478	0,443	6/19	4p3g	4.985	1.620	0,325
6p2g	1.206	444	0,368	7/20	4p2g	5.418	1.611	0,297
6p1g	1.220	522	0,428	8/21	5p2g	13.485	3.173	0,235
3p4g	2.452	962	0,392	9/22	5p1g	32.822	6.265	0,191
2p5g	3.024	877	0,290	10/23	1p1g	42.750	7.026	0,164
4p1g	3.040	1.121	0,369	11/24	1p4g	1.616.200	222.176	0,137
3p3g	3.261	1.069	0,328	12/25	1p3g	1.991.000	NPP	NPP
2p4g	3.519	1.045	0,297	13/26	1p2g	47.219.000	NPP	NPP

Opomba: Posivljene vrednosti iz tabele niso predstavljene na grafu. x-os grafa ima premosorazmerno oz. linearno skalo (skrajna leva vrednost je 482, skrajna desna pa 5.418), prikazane vrednosti P pa niso tiste iz tabele (tazen skrajne leve), marveč preračunane vmesne enakomerne stopnje.

Slika 28: Trendi gibanja spremenljivk P, R in R/P v gradivih slogovnih primerjav

Spreminjanje R in R/P glede na P



Primerjanje leve polovice grafa z desno pove, da je za manjše vrednosti P gibanje R in R/P bolj predvidljivo kot za velike vrednosti. Sicer so vrednosti R in R/P dokaj razpršene, ne ravno idealno skladne z linearno funkcijo. Za ugotavljanje in preverjanje le-te bi bilo potrebnih več gradiv.

V. ZAKLJUČKI

Zaključki so podani po 7 tematskih sklopih opravljenega dela, vsak sklop ima 3 dele: GL. POGLAVJA (reference na poglavja s to tematiko v diplomu), PREDNOSTI in POMANJKLJIVOSTI (načina dela; hkrati s prednostmi povzetek tega, kar je bilo v diplomu opravljeno, in najzanimivejših spoznanj, zato so prednosti nekoliko obsežnejše).

1)) ZAJEM GRADIVA, SNEMANJE

GL. POGLAVJA: I.5.2.1, III.10

PREDNOSTI:

+ zanimiva sociometrična zasnova, zaokroženost gradiva: neprekinjeno snemanje od jutra do večera v dveh dneh (izbrana sta delovnik in prosti dan, petek in nedelja), osredotočenost na enega govorca, nosilca mikrofona

+ posneto je realno gradivo, sogovorniki se snemanja ne zavedajo

+ kljub preprostem pristopu je zajet širok spekter situacij, oseb in nadaljnja obdelava tega gradiva daje izkušnje za gradnjo t. i. referenčnega govornega korpusa

+ po avtorjevem vedenju je ustvarjen zaenkrat največji govorni korpus (transkripcij, povezanih s posnetki) realnega prostega govora v slovenščini, velik 48.500 besed¹⁷¹

POMANJKLJIVOSTI:

– snemanje ni (formalno) avtorizirano, čeprav je za snemanje v zasebnih situacijah pridobljeno ustno soglasje sogovorcev (daleč vnaprej pred snemanjem ali po njem); podatki govorcev v transkripcijah so šifrirani (zakriti), na samih posnetkih pa zaenkrat še ne

– izbrani (preučevani) govorec je hkrati tudi snemalec in raziskovalec

– korpus je majhen (od 7 posnetih dni sta prepisana le 2, od načrtovane količine 100. besed le slaba polovica): le za majhen delež situacij je dovolj gradiva za kvantitativne raziskave

– nastopajoči govorniki niso natančneje (p)opisani, npr. z anketo, ampak le z poznavanjem snemalca, ki vodi tudi dnevnik snemanja

2)) TRANSKRIBIRANJE GOVORA IN ZAPIS OKOLIŠČIN SPORAZUMEVANJA

GL. POGLAVJA: I.2, I.3, II,

PREDNOSTI:

± nečrkovna pisna znamenja so v primerjavi s črkami semiotično reflektirana in jezikovnoravninsko razporejena, predvidena kot pisni paraverbalni ekvivalent govornemu paraverbalnemu izražanju: artikulaciji, prozodiji, stavčni fonetiki

+ izdelana je transkripcijska konvencija, ki podrobno dokumentira pravila zapisa s posebnim pretresom problematičnih (težavnih) primerov (npr. leksika brez uveljavljenega zapisa, neverbalni oz. paraverbalni govorni znaki)

+ primerjava transkripcijske konvencije z drugimi zgledi (BNC, Turdis, UKGS)

+ modifikacija ortografske transkripcije (načeli fonografskega prepisa in primerjave s standardno pisno obliko) odkriva posebnosti (prostega) govora in daje možnosti primerjave s pisnim jezikom

+ odločitev za partiturni format transkripcije, ki je novost v slovenski analizi govornega diskurza

+ transkripcije (transkripcijske datoteke) so povezane z zvočnim posnetkom, kar omogoča primerjavo transkripcije z (nezapisano) prozodijo in stavčno fonetiko

¹⁷¹ No, to avtorjevo vedenje je zelo omejeno; v kolikor so trenutno prepisani vsi posnetki korpusa Turdis (vseh 5 ur, ne le Turdis-1), bi ta znal biti za nekaj sto besed večji.

+ transkripcijska konvencija je oblikovana tako, da omogoča enostavno strojno štetje segmentov jezika (povedi, izjav; s programom EXMARaLDA)

+ opravljeno je kontroliranje transkribiranja, predstavljene so najpogostejše napake pri zapisu

+ definicija besedila glede na sporazumevalne dejavnike in uporaba te definicije konkretno na primeru rezanja materinskih transkripcij v hčerinska „besedila“ po kriterijih, določenih na podlagi opazovanja sporazumevalnih dejavnikov

+ pripravljen je model sporazumevalnih dejavnikov, ki je izpeljan v konkretno kot osnova za natančen opis okoliščin posameznih besedil (povezava s tradicionalnimi klasifikacijami besedil in jezikovnih zvrsti)

+ po pomenu v sporazumevanju in za semiotični vidik jezikovnega znaka so zanimivo in dodelano razločena t. i. „sporazumevalna sredstva“: medij/posrednik vs. prenosnik/kanal vs. kod/jezik (socialno zamejen na visokem nivoju – nacionalnem)

POMANJKLJIVOSTI:

– relativno počasna transkripcija, kljub izpustitvi natančnejših časovnih referenc; hitrejši bi bil ortografski pristop (z več ločili) po zgledu korpusov BNC ali Cobuild

– samosvoja, izvorna transkripcijska konvencija, primeri zanjo so žal samo v tem korpusu; primernejša bi bila navezava na utečeno transkripcijsko konvencijo, npr. HIAT, ali običajni pravopisni zapis

– kljub pregledovanjem transkripcij se v korpusu še najdejo kakšne napake; njegovo velikost lahko zato najbolj natančno ocenimo le na 48.450 besed ± 80 besed

– transkripcijska konvencija le malo nakazuje možnosti zapisa z drugih jezikoslovnih vidikov (fonetični, morfološka členitev, lematizacija, obliko- in pomenskoskladenjske določitve), saj te možnosti niso bile med cilji diplome ali sem se jih zaradi količine dodatnega dela odpovedal

3)) PROCEDURA STROJNEGA DELA IN RAČUNALNIŠKA ORGANIZACIJA KORPUSA (PODATKOV)

GL. POGLAVJA: I.5, II.6, III

PREDNOSTI:

+ postopek dela in urejenost podatkov sta dognana, premišljena in opisana, npr. tudi z vidika časovne zamudnosti

+ uporabljeni so prosto dostopni računalniški programi: Audacity, Elan, EXMARaLDA Partitur Editor, Eva; izjema je MS Access (urejevalnik zbirk podatkov)

+ s pomočjo teh programov je odprta možnost enostavne pretvorbe besedil (transkripcijskih datotek) v standarde jezika markup (XML, TEI), v različne urejevalnike transkripcij (EXMARaLDA, Elan, TASX itd.) in v različne možnosti predstavitve (brez ali s pomočjo XSLT v TXT, HTML, RTF, SVG itd.)

+ datotečna struktura in struktura znotraj datotek sta takšni, da zavzameta malo prostora (zvočne datoteke v formatu MP3), da so priročne za kvantitativne analize besedila (besedilne datoteke formata EVA z možnostjo pretvorbe v TXT) in sporazumevalnih okoliščin (zbirka podatkov s tabelami, poizvedbami in enostavnimi obrazci, ki omogočajo tudi preverjan in objektivni vnos podatkov); možne so tudi konkordance

+ prečiščen korpus je pripravljen na razpolago drugim raziskovalcem

POMANJKLJIVOSTI:

– organizacija samega korpusa ni standardna, podatki o okoliščinah besedil so ločeni od samih transkripcijskih datotek

– zapleteno „šifriranje podatkov“ o sporazumevalnih okoliščinah besedil in njihovih statistikah

– smelo samostojno zasnovan pristop kljub relativno majhni podkovanosti v računalništvu; uporabljena programska oprema in struktura datotek korpusa nista bili primarno razviti za namene te diplome

4) PROSTOGOVORNI JEZIK V KVANTITATIVNEM POGLEDU

GL. POGLAVJA: I.2.4, IV

PREDNOSTI:

+ reflektirano je pojmovanje osnovnih kvantitativnih enot: pisnih znamenj, besed pojavnic in (t. i. „grafofonemskih“) različnic, povedi; pa tudi gradiv analiz: besedil in prispevkov posameznih govorcev v besedilih

+ izbrane so preproste kvantitativne metode: pogostost različnic (z izpeljavami: pogostostni seznam različnic – najpogostejše besedne oblike v besedilu, primerjava deležev iste različnice v več gradivih), kvocient različnic in pojavnic idr.

+ primerjave vključujejo gradiva takšnih različnih velikosti, da to omogoča interpretacijo kvantitativnih rezultatov z relativiziranjem velikosti gradiva

+ prosti govorni jezik je primerjan s (pod)korpusi pisnega jezika (za interne primerjave prostega govora gl. naslednji razdelek zaključka): najočitnejše posebnosti prostega govora so med različnicami velika pogostost <pa|ne|da|*e|tak'|jaz|*s'n/sem> (ki odrinejo z mesta najpogostejše različnice <je>, z redkejših mest pa npr. <z/s|in>), med povedmi pa pogostejše kratke povedi (npr. oporni signali)

+ metoda iskanja najpogostejših n-kratnic besed v govoru ($n > 1$) ter večkratnic daljših povedi se izkaže kot učinkovita za odkrivanje neprostega govora (npr. pesmi, javnih obvestil, pozdravov, recitacijskih obrazcev – molitev)

POMANJKLJIVOSTI:

– majhna refleksija tujih in domačih izkušenj v kvantitativni lingvistiki
– odsotnost bolj poglobljenih in izdelanih kvantitativnih metod (npr. več primerov za rast besednega zaklada, hi-kvadratnega testa primerjav gradiv in težjih izpeljav)

– gradivo je relativno (pre)majhno za možnost kvantitativnega posploševanja spoznanj na razmere splošnega slovenskega sporazumevanja; predvsem gre res za idiolekt

5) SLOGOVNE ANALIZE GOVORA IN IDIOLEKT

GL. POGLAVJA: I.4, IV

PREDNOSTI:

+ razločena je popularna predstava „pogovornega jezika“ od govora kot prenosniško pojmovanega diasistema (zvrsti) jezika; pri opisovanju besedil korpusa je uveden kriterij „prostogovornost“, ki po 5-stopenjski lestvici razlikuje prosti govor od navajanja (branja, petja po predlogi, recitiranja)

+ odločitev za sociometriko ter raziskave idiolekta kot novo področje v slovenistiki

+ razvit je strukturalistično dosleden model opisovanja sloga, ki je uporaben za katerokoli jezikovno ravnino: 3 ravni (enačenja, razlikovanja in primerjanja) ter 2 pola (izraz–konkretno v. vsebina–abstraktno)

+ izbran je zanimiv razpon primerjav, ki omogoča dobre predstave o variiranju govora in prenašanje spoznanj z opisov sociolektov na opis idiolekta: pisni jezik vs. prosti govor, različne in enake okoliščine govora (diskurzi) med seboj, izbrani govorec vs. sogovorci (enake okoliščine) in vs. heterogeni kontekst (različne okoliščine) ter same vloge izbranega govorca v različnih in istih okoliščinah govora (variiranje samega idiolekta)

+ glede variiranja govora z vidika najpogostejših različnic je bilo ugotovljeno, da splet različnih okoliščin sporazumevanja (predvsem prenosnik, medij, več soudeležencev, funkcija sporazumevanja) bolj določa jezik (njegov diasistem, zvrst, notranjo diferenciacijo) kot konceptualna navezava diasistema jezika na posamezen sporazumevalni dejavnik – tudi na

posameznika (idiolekt); z drugimi besedami: besedišče posameznika *v različnih situacijah nič manj (nepredvidljivo) ne variira* kot besedišče več ljudi (sociolekta)

+ glede variiranja sociolekta in idiolekta v istih okoliščinah (več besedil med seboj) pa je gradivo vendarle pokazalo, da besedišče idiolekta *v istih okoliščinah manj variira* kot skupno besedišče več ljudi – čeprav nismo ugotovili nič novega, smo to vsaj pokazali z metodo primerjanja najpogostejših različnic

+ za najpogostejše različnice izbranega govorca so bile na podlagi korpusa vendarle ugotovljene nekatere značilnosti: je redkobeseden, pogosto uporablja mašila oz. pritrjevalne oporne signale <*e|mhm|ja|ne>, njegov jezik je v situacijah korpusa dokaj prostogovoren (lahko tudi: narečen, vsakdanji, pogovoren) – <pa> je njegova najpogostejša različnica, pogosta sta tudi <tak'|*s'n>

POMANJKLJIVOSTI:

– minimalna sporočilnost izbrane sociometrične metode o vsebini govora (jezika) – razen z različnicami se ne ukvarjamo z drugimi ali drugače pojmovanimi jezikovnimi sredstvi (npr. leksemi, samostalniki, glagoli, besedni red, dolžine povedi)

– interpretacija vsebine najpogostejših različnic – enakoizraznic, čeprav so enostavne (kot opornih signalov, vprašalnih členic, veznikov, mašil, medmetov itd.), ni povsem utemeljena oziroma nima znane verjetnosti

– za širša spoznanja o variiranju besedišča so potrebne še raziskave najpogostejših leksemov in še širše primerjave (več gradiva za posameznega govorca), raznolikejša gradivo (več govorcev, sporazumevalnih situacij)

6)) NADALJNJE MOŽNOSTI RAZISKOVANJA

GL. POGLAVJA: I.5, II.7.4, II.8.7, I

PREDNOSTI:

+ diploma ima raznovrstne priloge in nakazane nadaljnje možnosti dodelave gradiva, pripravljen je prečiščen korpus v elektronski obliki za druge raziskovalce

+ partiturno oz. XML-strukturo transkripcijskih datotek je lahko širiti in dodajati vzporedne informacije, npr. fonetični zapis ali metaoznake besed

+ model raziskave idiolekta je uporaben pri gradivu večjega korpusa, ki je podobno demografsko zasnovan (iz več gradiva za posameznega govorca, taka praksa je izpričana za korpus BNC)

+ korpus je kot statistični model ali kot (učno) gradivo, ki pa ga je potrebno dodelati v veliki meri, lahko uporaben pri strojnih postopkih analize prostega govora, kjer taka gradiva šele nastajajo: v lematizatorjih in določevalnikih (POS-taggerjih), v fazah razpoznavanja in tvorjenja sporočil (v prevajalnikih speech-to-speech)

+ druge možne raziskave: fonetične in raziskave kompleksnejših jezikovnih ravnin, sociolingvistične, dialektološke

POMANJKLJIVOSTI:

– v diplomu sami je bil sprva cilj pripraviti tudi besedno določitev in lematizacijo, toda zaradi upočasnitve dela pri transkribiranju sem se odločil diplomu skrajšati za to dodatno delo

– za nadaljnje raziskave je potrebno ali več gradiva ali dodatne označitve – v korpusu so neoznačene transkripcije (nelematizirano, niti besedne vrste niso označene, kaj šele natančnejše oblikoskladenjske določitve)

7)) REFLEKSIJA RAZISKOVALNEGA PRISTOPA

teoretična utemeljenost in zasnova, aktualnost dela, osebni pogled

GL. POGLAVJA: I.1, uvod v poglavje III in pričujoče poglavje V – zaključki

PREDNOSTI:

+ aktualnost raziskave govornega jezika in relativno velika obsežnost korpusa

- + izpeljava dela od ustvarjanja korpusa do primera njegove uporabe – analize in slogovne interpretacije (dolga in zato zahtevna paleta raziskovalnih vprašanj)
- + teoretična utemeljitev vseh stopenj obdelave
- + preglednost diplome zaradi razčlenjenosti, raznovrstnost prikaza (tabele, slike in skice, primeri)

POMANJKLJIVOSTI:

- kljub relativno malo pregledovanja domače in tuje literature so izvedena široka sklepanja
- krčenje ciljev zaradi obsežnosti dela, časovna razvlečenost ustvarjanja diplome na 3 leta
- nepovezanost z drugimi raziskovalci na tem področju govornih korpusov, preveč individualizirano delo
- dolga diploma, nekateri postopki so (pre)podrobno opisani
- pogled te naloge je „pogojen z družbeno zgodovino jezika“ (Skubic 2005: 53) – z določenim trenutkom družbenega razumevanja jezika, ki mu pripadam tudi avtor te naloge; konkretno gre za popularnost računalniškega, kvantitativnega, korpusnega in sociolingvističnega jezikoslovja

VI. SEZNAMI

18. KRAJŠAVE, POJASNILO OBLIKOVANJA TISKA

<i>krajšava</i>	<i>pomen</i>	<i>krajšava</i>	<i>pomen</i>
ASCII	American Standard Code for Information Interchange	mp3	format zvočne datoteke, MPEG (Moving Picture Experts Group)-1, Audio Layer 3
b. k., b. l.	brez kraja, brez letnice (pri literaturi)	MSD	angl. MorphoSyntactic Description (oblikoskladenjska določitev)
BNC	British National Corpus	NPP	ni podatkov o primerjavi (v seznamu različnic za potrebe slogovnih primerjav v tej diplomi)
BP	brez primerjave (v seznamu različnic za potrebe slogovnih primerjav v tej diplomi)	OLA	Общеславянский лингвистический атлас
BSJ	Besedišče slovenskega jezika	pp	premalo primerjav (v seznamu različnic za potrebe slogovnih primerjav v tej diplomi)
CES	Corpus Encoding Specification (for XML → XCES)	POS-tagging	angl. Part of Speech (besedna vrsta), označevanje oz. določevanje besednih vrst
CHAT	transkripcijska konvencija projekta CHILDES – Child Language Data Exchange System	R/P	kvocient R-azličnic in P-ojavnic
DIDA	Diskursdatenbank (na inštitutu za nemški jezik, Mannheim), transkripcija	SAMPA	Speech Assessment Methods Phonetic Alphabet (fonetični zapis po IPA, ki uporablja črkovni nabor ASCII)
.eaf	Elan annotation format – format transkripcijske datoteke programa Elan	SGML	Standard Generalised Markup Language
EAGLES	Expert Advisory Group on Language Engineering Standards	SP ²⁰⁰¹	Slovenski pravopis 2001
ESJ	Jože Toporišič: Enciklopedija slovenskega jezika 1992	SSKJ	Slovar slovenskega knjižnega jezika
Eva	urejevalnik besedil in format njegovih datotek	StAnId	korpus ali transkripcijska konvencija Stilometrična analiza idiolekta
EXM.	projekt EXMARaLDA, njegov program Partitur Editor za urejanje transkripcij	TEI	Text Encoding Initiative
GAT	Gesprächsanalytisches Transkriptionssystem	Turdis-1	korpus telefonskih pogovorov v turizmu
GT	govorne tehnologije (veja jezikovnih tehnologij)	UKGS	Učni korpus govornjene slovenščine
HIAT	Halbinterpretative Arbeitstranskriptionen	UTF-16,-8	Unicode Transformation Format, 16- in 8-bitni nabor pisnih znamenj
IPA	International Phonetic Alphabet	XML	eXtended Markup Language; tudi format datotek projekta EXM.
mio.	milijon		

POMEN OBLIKOVANJA BESED

v okvirjih, pomanjšani

podčrtane

ležeče
odebeljeni

<>

[]

//

{}

|

so navedeni t. i. podatki o sporazumevalnih okoliščinah besedil, snemanja, o statistiki besedilnih enot v korpusu;

so t. i. vrednosti teh podatkov, tj. konkreten vpis, ki ga ima besedilo za posamezen podatek, npr. za podatek [17] kraj ima besedilo 6pet09Dv vrednost oz. vpis Ljubljana; podčrtane in ležeče so tudi šestere stopnje obdelave gradiva, prva je npr. StAnId.1.Snemanje;

so besede, ki želijo biti pomensko poudarjene;

so ključni strokovni izrazi, ko se prvič pojavijo v poglavju;

lomljena oklepaja obdajata pisna znamenja ali besede, transkripcije in tako izpostavljata njihov pisni izraz kot objekt razprave;

oglata oklepaja zaznamujeta ali fonetični zapis ali v partituri transkripcije imena tipa trakov (npr. verbalni trak [v]) ali v seznamu literature pripise pisca diplome;

poševna oklepaja zaznamujeta ali foneme ali vrivke v navedke;

valovita oklepaja zaznamujeta morfeme;

pokončna črta ločuje primere izpiskov iz transkribiranega gradiva (besede, povedi, govorne dogodke).

19. KAZALA TABEL, SLIK IN PRIMEROV

TABELE

Tabela 1: Umestitev jezikovnega sredstva z vidika prenosnika in koda (pomenskorazločevalna raven)	18
Tabela 2: Zmožnosti izraznih enot (sredstev) za razlikovanje in strukturiranje pomena	24
Tabela 3: Transkribcije govora med možnostma ortografskost in fonetičnost.....	25
Tabela 4: Pregled kvantitativnih jezikoslovnih razprav o slovenščini.....	33
Tabela 5: Ohlapna in aktualna plat dejavnikov sporazumevanja	44
Tabela 6: Kriteriji za ločevanje govora in spremljajočih okoliščin na enote gradiva – „besedila“	46
Tabela 7: Ocena razmerij med besedišči: SSKJ vs. BSJ in nevtralno vs. zaznamovano	57
Tabela 8: Tipi dokumentov transkripcijskih datotek po programih Elan in EXMARaLDA	76
Tabela 9: Pregled tipov datotek v korpusu in podatkov v njih.....	77
Tabela 10: Podatki v korpusu, stopnje njihove obdelave in pravila zapisa	81
Tabela 11: Možna mesta zapisa transkripcijskih podatkov.....	86
Tabela 12: Izbrana ločila v transkripcijski konvenciji StAnId4.....	89
Tabela 13: Skupen pregled znamenj transkribcije po jezikovnih ravneh in pomenu v transkribciji.....	93
Tabela 14: Izbrana skladenjska ločila v transkripcijski konvenciji StAnId4.....	105
Tabela 15: Pregled oznak na traku [pv] v partituri.....	107
Tabela 16: Tipi opisov oznake nvg: s traku [pv] v partituri transkribcije	107
Tabela 17: Primeri gradiva za posamezne opise ob oznaki nvg:.....	108
Tabela 18: Primerjava transkripcijskih rešitev za govor: StAnID, Turdis, UKGS, BNC	109
Tabela 19: Pregled oznak in primeri opisov na traku [Ozadje] v partituri transkribcije	111
Tabela 20: Podatki o transkripcijskih enotah na stopnji .3 <i>Transkribcije</i>	112
Tabela 21: Podatki o govornih na ravni celega korpusa, stopnja .2 <i>Okoliščine</i>	113
Tabela 22: Poimenovanje (šifriranje) „besedil“ v korpusu, stopnja .4 <i>Besedila</i>	114
Tabela 23: Podatki o rezanju in časovni razsežnosti „besedil“, stopnja .4 <i>Besedila</i>	114
Tabela 24: Datotečni podatki „besedil“, stopnja .4 <i>Besedila</i>	115
Tabela 25: Podatki o neodvisnih okoliščinah „besedil“, stopnja .4 <i>Besedila</i>	116
Tabela 26: Podatki o soudeležencih kot okoliščinah „besedil“, stopnja .4 <i>Besedila</i>	116
Tabela 27: Podatki o sporazumevalnih sredstvih „besedil“, stopnja .4 <i>Besedila</i>	117
Tabela 28: Podatki o statistikah jezikovnih enot v „besedilih“, stopnja .4 <i>Besedila</i>	117
Tabela 29: Jezikoslovne metaoznake „besedil“ in sobesedila, stopnja .4 <i>Besedila</i>	118
Tabela 30: Podatki o govornih v „besedilih“, stopnja .4 <i>Besedila</i>	118
Tabela 31: Primerjava transkripcijskih rešitev za opis sporazumevalnih okoliščin besedil: StAnID, UKGS, BNC.....	120
Tabela 32: Podatki o obdelavi in strukturi korpusa, stopnja .6 <i>Informacije</i>	122
Tabela 33: Primeri površnega transkribiranja	123
Tabela 34: Zamudnost transkribiranja v korpusu StAnId	126
Tabela 35: Zamudnost obdelave gradiva po stopnjah obdelave.....	127
Tabela 36: Pregled posnetkov po dneh.....	129
Tabela 37: Podatki o rezanju na besedila v obrazcu 03_razrez in viri teh podatkov.....	136
Tabela 38: Mapna in datotečna struktura korpusa, interno in javno.....	146
Tabela 39: Pregled korpusa po št. govorcev, dnevih in deležu prostega govora.....	147
Tabela 40: Pregled črk in pisnih znamenj v jezikovnem gradivu korpusa StAnId	149
Tabela 41: Pregled različnic korpusa po oblikoskladenjskih ločilih in pomožnih znamenjih	150
Tabela 42: Variiranje števila različnic glede na upoštevanje posebnih ločil transkribcije	151
Tabela 43: Korpus po velikosti besedil v jezikovnih enotah	153
Tabela 44: Korpus po prostih oznakah in mejah besedil.....	158
Tabela 45: Korpus po dejavnih neodvisnih okoliščin in sporazumevalnih sredstev	158
Tabela 46: Korpus po dejavnih soudeležencev	158
Tabela 47: Pregled podatkov govorcev v korpusu in posameznih besedilih	162
Tabela 48: Spremenljivke P, R, R/P v gradivih slogovnih primerjav (naraščanje po P).....	169

SLIKE

Slika 1: Strukturiranje izrazno-pomenskih (z)možnosti idealne besede.....	29
Slika 2: Razmerje stilometrije idiolekta do drugih ved s funkcionalističnim pogledom na jezik	37

Slika 3: Jakobsonov model konstitutivnih dejavnikov sporazumevanja.....	38
Slika 4: Model sporazumevalnih dejavnikov	39
Slika 5: Strukturalni model učinkovanja sloga.....	58
Slika 6: Pogostostni seznam različnic v slogovnem modelu.....	61
Slika 7: Formati transkripcije v programu Transcriber.....	83
Slika 8: Razdelana in hierarhizirana partitura med označevanjem besednih vrst, program Elan	84
Slika 9: Relacije med tabelami metapodatkov besedil v programu Access	87
Slika 10: Nadbesedna segmentacija v transkripciji	104
Slika 11: Izsek podatkov iz tabele 08 BESEDILA, Access	116
Slika 12: Stopnje konkretnosti/abstraktnosti podatkov v korpusu.....	125
Slika 13: Začetek transkribiranja, EXMARaLDA Partitur Editor.....	130
Slika 14: Urejanje trakov idr. v partituri, EXMARaLDA Partitur Editor	131
Slika 15: Obrazec za vpis transkripcij, Access	133
Slika 16: Obrazec za vpis govorcev, Access.....	133
Slika 17: Partitura transkripcije z označenim besedilom za izrez, EXMARaLDA	135
Slika 18: Obrazec za vpisovanje in odbiranje razreza (primer hčerinskega besedila označen), Access.....	136
Slika 19: Spektrogram posnetka transkripcije z označenim intervalom za izrez besedila, Audacity	137
Slika 20: Preglednica govorcev v besedilih (označen primer besedila), Access	140
Slika 21: Obrazec štetja segmentov besedila, EXMARaLDA.....	140
Slika 22: Obrazec za vpis metaoznak besedil pred ročnim vnosom preostalih oznak, Access.....	141
Slika 23: Obrazec za vpis metaoznak besedil, izpolnjen, Access	142
Slika 24: Začetek skupne korpusne datoteke, program Eva.....	143
Slika 25: Iskanje v konkordančniku EXAKT	145
Slika 26: Primerjava dolžin povedi v korpusu: številčnost in pokritje besednega zaklada korpusa.....	153
Slika 27: Rast besednega zaklada v podkorpusu prostega govora	156
Slika 28: Trendi gibanja spremenljivk P, R in R/P v gradivih slogovnih primerjav	170

PRIMERI

Primer 1: Mešanje vplivov pisnega in govornega izraznega sistema ob prevzemanju besed.....	16
Primer 2: Omejitve sistema alofonov in alografov pri transkribiranju	23
Primer 3: Oblikoglasnost in razlikovalni sistem oblik	27
Primer 4: Pravopisne posebnosti	27
Primer 5: Kohezija elementov besede.....	28
Primer 6: Katere slovnične kategorije želimo razlikovati z grafofonemskimi različnicami	30
Primer 7: Odvisnost seznama različnic od natančnosti transkripcije.....	31
Primer 8: Kaj pravi SSKJ o slogu oz. stilu.....	47
Primer 9: Slog nad ravnino skladnje – pragmatika.....	52
Primer 10: Okoliščine dojemanja denotata, zajete v konotativnem pomenu	54
Primer 11: Označevanje sloga s kvalifikatorskimi pojasnili	55
Primer 12: Enote, ki jim vrednotimo zaznamovanost: njihove jezikovne ravnine in sistemi primerjav	56
Primer 13: Besedni primeri za strukturalni model učinkovanja sloga.....	60
Primer 14: Pogostostni seznam različnic za potrebe stilometrične analize.....	63
Primer 15: Velikosti zvočnih datotek	74
Primer 16: Primera uporabe oklepajev v modificirani ortografski transkripciji	91
Primer 17: Zapis v dnevnik snemanja.....	129

20. SEZNAM NAVEDENE LITERATURE

Vse spletne strani so bile nazadnje preverjene avgusta 2007, za nekatere pa so podani tudi natančnejši podatki (nastanka ali zadnje spremembe), kjer so bili na voljo.

BARRAS, Claude idr. (2006): *Transcriber: a tool for segmenting, labeling and transcribing speech*. [Računalniški program. B. k.] [Http://www.etca.fr/CTA/gip/Projets/Transcriber/](http://www.etca.fr/CTA/gip/Projets/Transcriber/) (uporabljeno 17. 7. 2006).....78

BARTHES, Roland (1990): *Retorika starih. Elementi semiologije*. Ljubljana: ŠKUC, FF (Studia humanitatis, V. letnik).....42, 59

- BGSPEECH ([2004]): *сайт за българската разговорна реч*. [Http://www.slav.uni-sofia.bg/bgspeech/sites/index.htm](http://www.slav.uni-sofia.bg/bgspeech/sites/index.htm). 69
- BOERSMA, Paul, WEENINK, David (2007): *Praat: doing phonetics by computer*. [Računalniški program.] Amsterdam: University of Amsterdam, Institute of Phonetic Sciences. [Http://www.praat.org/](http://www.praat.org/) (avgust 2007). 78
- BREZNIK, Anton (1924): *Slovenska slovnica za srednje šole: Tretja izdaja*. Prevalje: Družba sv. Mohorja. 15
- BRINKER, Klaus, SAGER, Sven F. (2001³): *Linguistische Gesprächsanalyse: Eine Einführung: 3., durchgesehene und ergänzte Auflage*. Berlin: Erich Schmidt (Grundlagen der Germanistik, 30). 9, 25, 84, 91, 108, 125
- BSJ (1999): *Besedišče slovenskega jezika z oblikoslovnimi podatki: A – Ž: po gradivu za slovar sodobnega knjižnega jezika zbrane besede, ki niso bile sprejete v Slovar slovenskega knjižnega jezika*. Ljubljana: ZRC SAZU, Založba ZRC (Slovarji). 33, 34, 57, 92, 102, 149, 150, 176
- CROWDY, Steve (1995): 19 The BNC spoken corpus. *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 224–234. 71, 110, 119
- DOBRIŠEK, Simon, VESNIČER, Boštjan, MIHELČ, France, PAVEŠIČ, Nikola (2006): Biometrični sistem za razpoznavanje govorcev. *SloFon1: 1. slovenska mednarodna fonetična konferenca: zbornik povzetkov*. [Ljubljana]: Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU. 22. 70, 130
- DOBROVOLJC, Helena (2005): *Slovenska teorija jezikovne naravnosti s slovenskim (obliko)skladenjskim gradivom*. Ljubljana: Založba ZRC, ZRC SAZU (Linguistica et philologica, 12). 57, 59
- DUKAT, Zdeslav (1985): [gesli:] stil, stilistika. *Rečnik književnih termina*. Beograd: Nolit. 759–766. 49
- EAGLES (1996): [Expert Advisory Group on Language Engineering Standards: Spoken Language Working Group:] *EAGLES preliminary recommendations on Spoken Texts, 1996*. [Http://www.ilc.cnr.it/EAGLES96/spokentx/spokentx.html](http://www.ilc.cnr.it/EAGLES96/spokentx/spokentx.html) 69, 72, 108, 109
- EDWARDS, Jane A. (1995): 1 Principles and alternative systems in the transcription, coding and mark-up of spoken discourse. *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 19–34. 68, 69
- ELAN (2007): [Eudico Linguistic Annotator.] [Nijmegen: Max Planck Institute for Psycholinguistics.] [Http://www.lat-mpi.eu/tools/elan/](http://www.lat-mpi.eu/tools/elan/) (nazadnje spremenjeno 1. 3. 2007). 78
- ERI-BIRK, Marijana (1976): Pogostnost in skladenjska vloga sklonov v besedilu Šeligovega Triptiha Agate Schwarzkobler. *SR XXIV*. 119–126. 33
- ERJAVEC, Tomaž (1997): Računalniške zbirke besedil. *JiS XLII/2–3*. 81–96. 78
- ERJAVEC, Tomaž (2003): Označevanje korpusov. *JiS XLVIII/3–4*. [61]–76. 75, 77
- Erjavec, Tomaž (2004): Multext-East Resources, Version 3. (7. 2. 2004.). [Http://nl.ijs.si/ME/V3/](http://nl.ijs.si/ME/V3/). 70
- FREEBORN, Dennis (1996): *Style: text analysis and linguistic criticism*. Basingstoke, London: Macmillan (Studies in English language). 48, 51
- GARSDIE, Roger (1995): 13 Grammatical tagging of the spoken part of the British National Corpus: a progress report. *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 161–167. 71
- GORJANC, Vojko (2005): *Uvod v korpusno jezikoslovje*. Domžale: Izolit (Zrenja). ... 30, 33, 34, 35, 71, 119, 120, 126
- GRGIČ, Matejka (2006): Historična perspektiva preučevanja jezikoslovnih ved: Nekateri vidiki antičnega razumevanja jezika. *JiS LI/2*. 3–18. 15
- GRZYBEK, Peter (2000): Pogostnostna analiza besed iz elektronskega korpusa slovenskih besedil. *SR XLVIII/2*. 141–157. 31, 34
- HAJNŠEK – HOLZ, Milena, NOVAK, France (1970): Kvalifikatorji. [V sklopu članka Slovar slovenskega knjižnega jezika, str. 137–141.] *Naši razgledi* (6. 3. 1970). 140. 55
- HASLERUD, Vibecke, STENSTRÖM, Anna-Brita (1995): 20 The Bergen Corpus of London Teenager Language (COLT). *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 224–234. 70, 108
- HELLWIG, Birgit, UYTVANCK, Dieter Van (2006): EUDICO Linguistic Annotator (ELAN) version 2.6: Manual. [Nijmegen: Max Planck Institute for Psycholinguistics.] [Http://www.mpi.nl/tools](http://www.mpi.nl/tools). 76, 144
- HLADNIK, Miran (1995): Količinske in empirične analize v literarni vedi. *SR XLIII/3*. 319–340. [Tudi na <http://www.ijs.si/lit/empir1.html-12>.] 68
- HOLMES, David I. (1994): Authorship Attribution. *Computers and the Humanities* 28. 28–106. 49, 64, 66
- JAKOBSON, Roman (1989): Lingvistika in poetika. *Lingvistični in drugi spisi*. Ljubljana: ŠKUC, ZIFF (Studia humanitatis, letnik 4). 147–190. 38

- JAKOPIN, Primož (1995): Nekaj števil iz Slovarja slovenskega knjižnega jezika. *SR XLIII/3*. 341–375...32, 34, 149, 154
- JAKOPIN, Primož (1996): Ali so rojstna imena krajša od drugih samostalnikov? *SR XLIV/2*. 193–200.....34
- JAKOPIN, Primož (2000): *EVA – urejevalnik besedil in podatkovnih zbirk*.
[Http://www.laze.org/eva/eva_index_si.html](http://www.laze.org/eva/eva_index_si.html) (4. 2. 2000).79
- JAKOPIN, Primož (2001): O nekaterih kvantitativnih kazalcih v slovenskem jeziku. *Slovenski knjižni jezik – aktualna vprašanja in zgodovinske izkušnje*. Ljubljana: FF, Oddelek za slovanske jezike in književnosti (Obdobja: Metode in zvrsti, 20). 471–482..... 34, 154
- JAKOPIN, Primož (2002): *Entropija v slovenskih leposlovnih besedilih*. Ljubljana: ZRC SAZU (Linguistica et philologica). [Tudi na <http://www.ff.uni-lj.si/hp/pj/disertacija/naloga.html>..... 27, 32, 33, 34, 149, 155
- JAKOPIN, Primož (2003): Nekaj zanimivosti iz besedilnega korpusa Nova beseda. *Jezikoslovni zapiski IX/2*. 145–152..... 32, 34, 152
- Jakopin, Primož, Bizjak, Aleksandra (1997): O strojno podprtem oblikoslovnem označevanju slovenskega besedila. *SR XLV/3–4*. 513–532. [Tudi v angleščini na http://bos.zrc-sazu.si/bibliografija/POS_tagging.html.] 30, 96, 125
- JURANČIČ, Janko, LEGIŠA, Lino (1970): Izbor besed. [V sklopu članka Slovar slovenskega knjižnega jezika, str. 137–141.] *Naši razgledi* (6. 3. 1970). 13757
- KAČIČ, Zdravko (2002): Pomen združevanja raziskovalnih potencialov pri preseganju jezikovnih pregrad v okviru jezikovnih tehnologij naslednjih generacij. *Jezikovne tehnologije: Zbornik B 5. mednarodne multikonference Informacijska družba IS 2002*. Ljubljana: Institut Jožef Stefan. [Tudi na <http://nl.ijs.si/isjt02/zbornik/sdjt02-19kacic.pdf>..... 69
- KLOPČIČ, Mile, SUHADOLNIK, Stane (1970): Glavne značilnosti slovarja. [V sklopu članka Slovar slovenskega knjižnega jezika, str. 137–141.] *Naši razgledi* (6. 3. 1970). 137.....54
- KMECL, Matjaž (1995): *Mala literarna teorija*. Ljubljana: Mihelač in Nešović..... 50, 52
- KÖHLER, Reinhard (1995): *Bibliography of Quantitative Linguistics* [BQL]. Amsterdam/ Philadelphia: John Benjamins.32
- KOPŘIVOVÁ, Marie, WACLAWIČOVÁ, Martina (2005): Construction of Spoken Corpus Based on the Material from the Language Area of Bohemia. *Computer treatment of Slavic and East European languages: Third International Seminar, Bratislava, Slovakia, 10–12 November 2005: proceedings*. Bratislava: Veda. 137–140.....70
- KOROŠEC, Tomo (1986): Pretvorbeni modeli (k teoriji pretvorb pisnost-govornost, govornost-pisnost). *SR XXXIV/4*. 393–402..... 13, 51
- KOROŠEC, Tomo (1998): *Stilistika slovenskega poročevalstva*. Ljubljana: Kmečki glas.51
- KOZLEVČAR, Ivanka (1970): O kvalifikatorskih pojasnilih. [V sklopu članka Slovar slovenskega knjižnega jezika, str. 137–141.] *Naši razgledi* (6. 3. 1970). 139.....55
- KRAJNC, Mira (2004): Besedilnoskladenjske značilnosti javne govorne besede (na gradivu mariborščine). *SR LII/4*. [475]–498.....25
- KRANJC, Simona (1997): Govorjeni diskurz. *JiS XLII/7*. 307–319.....70
- KRANJC, Simona (1998): Govorjena besedila in korpus slovenskega jezika. *Jezikovne tehnologije za slovenski jezik: zbornik konference*. Ljubljana: Institut Jožef Stefan. [Tudi na: <http://nl.ijs.si/isjt98/zbornik/sdjt98-Kranjc.pdf>.]..... 12, 13
- KRANJC, Simona (2004): Jezikovna zvrstnost v sodobnih medijih. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*. Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22). 395–405.....41
- LEECH, Geoffrey N., MYERS, Greg, THOMAS, Jenny (ur.1995): *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 9, 68, 69
- MARTINČIČ IPŠIČ, Sanda, MATEŠIČ, Mihaela, IPŠIČ, Ivo (2004): Korpus hrvatskoga govora. *Govor XXI/2*. 135–150.....70
- MATTHEWS, Peter H. (1997): *The Concise Oxford Dictionary of Linguistics*. 1997: New York: Oxford University Press. 20, 28, 48
- MAZZONI, Dominic idr. (2004): Audacity: A Free Digital Audio Editor. [Http://audacity.sourceforge.net](http://audacity.sourceforge.net). [Verzija programa 25. 8. 2004.]74
- MCCARTHY, Michael (1998): *Spoken language and applied linguistics*. Cambridge: Cambridge University Press.69
- MCENERY, Tony, WILSON, Andrew (2001): *Corpus Linguistics*. Edinburgh: Edinburgh University Press. 69, 155

- NEWEKLOWSKY, Gerhard (1984): Die Häufigkeit grammatikalischer Kategorien in slowenischen Zeitungstexten. *Didaktische Reflexion und Wissenschaft in einer sich ändernden Welt*. Celovec. 383–400. [V slovenščini v skrajšani obliki objavljeno kot: NEWEKLOWSKY, Gerhard (1983/84): Pogostnost gramatičnih kategorij v slovenskih časopisnih besedilih. *JiS* XXIX. 50–62.]..... 29, 31, 33, 62, 64, 154
- NEWEKLOWSKY, Gerhard (1984): *Trubarjev Katekizem 1550: Konkordanca, indeks besed, pogostostni spiskii*. Ljubljana: DZS.....33
- NEWEKLOWSKY, Gerhard (1986): Zur Paradigmatik in Trubars Katechismus 1550. *16. stoletje v slovenskem jeziku, književnosti in kulturi: mednarodni simpozij v Ljubljani od 27. do 29. junija 1984*. Ljubljana: Filozofska fakulteta (Obdobja 6). 307–317. [Podobna razprava v nemščini: Trubars Katechismus von 1550 – eine Konkordanz. *Protestantismus bei den Slowenen/Protestantizem pri Slovencih*. Dunaj 1984. 133–152 (WSA, Sonderband 13).].....33
- Neweklowsky, Gerhard (1988): Zur häufigkeit morphologischer Kategorien in slowenischen Prosatexten. Sodobni slovenski jezik, književnost in kultura: mednarodni simpozij, v Ljubljani od 1. do 3. julija 1986. Ljubljana: Filozofska fakulteta (Obdobja 8). 337–349. 30, 33, 154
- Nova beseda 2000. [Ljubljana:] ZRC SAZU, Inštitut za slovenski jezik Frana Ramovša. http://bos.zrc-sazu.si/a_about_si.html in http://bos.zrc-sazu.si/s_beseda.html (oboje 2. 5. 2000 in kasnejše dopolnitve).25
- NOVAK, Nina (2006): Prvine spontanega govora v zapisniku. *SloFon1: 1. slovenska mednarodna fonetična konferenca: zbornik povzetkov*. [Ljubljana]: Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU. 52.....25
- OCVIRK, Andrej (1978): *Literarna teorija*. Ljubljana: DZS (Literarni leksikon, 1).50
- OOSTDIJK, N. (2006): *Corpus Gesproken Nederlands, CGN; versie 1.0*. http://lands.let.kun.nl/cgn/doc_English/topics/project/pro_info.htm (1. 2. 2006).69
- PAYNE, Jonathan (1995): 16 The COBUILD spoken corpus: transcription conventions. *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman. 203–207..... 108
- PEČNIK, Daniela (2002): *Versuch einer computerunterstützten quantitativen Textanalyse am Beispiel des slowenischen Romans Namesto koga roža cveti: Diplomarbeit*. Dunaj. [Članek na podlagi te diplome pa tudi: Računalnik pri jezikovni obdelavi – možnosti in dimenzije na osnovi literarnih besedil. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*. Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22.). 487–494.]33
- PETERLIN, Primož, KOŠIR, Aleš, ERJAVEC, Tomaž (1998): Digitalni zapis slovenskih znakov. *Jezikovne tehnologije za slovenski jezik: zbornik konference*. Ljubljana: Institut Jožef Stefan. [Tudi na <http://nl.ijs.si/isjt98/zbornik/sdjt98-Peterlin.pdf>.]20
- POGAČNIK Jože (1967): Poglavja iz splošne stilistike. *Jezikovni pogovori II*. Ljubljana: CZ: 173–240. 42, 51
- Pogorelec, Breda (1965): Vprašanja govorenega jezika. *Jezikovni pogovori* [1]. Ljubljana: CZ, 132–156. .14
- POGORELEC, Breda (2003): Teorija zvrstnosti in slovensko jezikovno normiranje. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti: povzetki predavanj*. Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22.). 26. [V samem zborniku razprave Pogorelčeve ni!]14
- PONIŽ, Denis (1974): *Slovenski jezik, literatura, računalniki: numerično-statistično raziskovanje konstantnih in spremenljivih količin v slovenskem jeziku, prozi in poeziji*. Maribor: Obzorja (Znamenja, 45–46). 9, 34, 68
- REHBEIN, Jochen, SCHMIDT, Thomas, MEYER, Bernd, WATZKE, Franziska, HERKENRATH, Annette (2004): *Handbuch für das computergestützte Transkribieren nach HLAT*. Hamburg: SFB 538, Universität Hamburg (Arbeiten zur Mehrsprachigkeit, 56). [Tostopno tudi po spletu v elektronski obliki.]..... 108
- ROBINSON, Andrew (1999): *The Story of Writing: Alphabets, Hieroglyphs & Pictograms*. London: Thames & Hudson.....15
- ROHLFING, KATARINA idr. (2006): Comparison of multimodal annotation tools: workshop report. *Gesprächforschung – Online-Zeitschrift zur verbalen Interaktion* 7. 99–123.....79
- ROMIH, Miro (1998): Direktorijška struktura korpusa FIDA. *Uporabno jezikoslovje: Tematska številka Jezikovne tehnologije* 6. 70–84.78
- RUPEL, Mirko (1946): *Slovensko pravorečje: navodila za zborni ali knjižno izreko*. Ljubljana: DZS..... 19, 20
- SCHMIDT, Thomas ([2006; b. l.): EXMARALDA Partitur-Editor: Handbuch: Version 1.3.2. [b. k.] <http://www1.uni-hamburg.de/exmaralda/Daten/2D-Download/Partitur-Editor/Handbuch-aktuell.pdf>. 76, 104
- SCHMIDT, Thomas (2005): *Computergestützte Transkription: Modellierung und Visualisierung gesprochener Sprache mit texttechnologischen Mitteln*. Frankfurt a. M. idr.: Peter Lang (Sprache, Sprechen und Computer, 7).79
- SCHMIDT, Thomas, WÖRNER, Kai (2007): EXMARALDA. Hamburg: Universität Hamburg: Sonderforschungsbereich Mehrsprachigkeit (SFB 538): Teilprojekt Mehrsprachige Datenbank. <http://www1.uni-hamburg.de/exmaralda/index.html> (b. l.; preverjeno avgust 2007). 78, 79, 144

SCHNEIDER, Stefan (2007): Banca dati dell'italiano parlato – BADIP. Gradec: Karl-Franzens-Universität Graz, Institut für Romanistik. http://languageserver.uni-graz.at/badip/ (zadnja sprememba 12. 4. 2007).....	70
SIM, Tze Jan, KAZI, Sabbir Ahmed, HONG, Huaqing (2005): A Survey of Transcription, Annotation and Query Tools for Development of a Classroom Discourse Corpus. <i>Redisigning Pedagogy: Research, Policy, Practice: 30 May - 1 June 2005</i> . Singapore: Centre for Research in Pedagogy and Practices [CRPP], National Institute of Education, Nanyang Technological University. http://conference.nie.edu.sg/paper/Converted%20Pdf/ab00405.pdf	79
SINCLAIR, John (1991): <i>Corpus, concordance and Collocation</i> . Oxford: Oxford University.....	77
SINCLAIR, John (1995): 9 From theory to practice. <i>Spoken English on Computer: Transcription, mark-up, and application</i> . New York: Longman, 99–109.	34, 68, 69, 81, 96
Skubic, Andrej E. (2005): <i>Obrazi jezika</i> . Ljubljana: Študentska založba (Posebne izdaje/Študentska založba).	14, 81, 175
SMOLE, Vera (2000): <i>Uvod v slovensko dialektologijo: Del vsebine za predmet ZSD II</i> . [Skripta.] Ljubljana: FF Univerze v Ljubljani, Odd. za slovanske jezike in književnosti, Katedra za zgodovino slovenskega jezika in dialektologijo.....	22
SMOLEJ, Mojca (2004): Načini tvorjenja govornega diskurza – paradigmatska in sintagmatska os. <i>Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti</i> . Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22). 423–436.....	10
SP (Slovenski pravopis, 2001). Ur. Jože Toporišič idr. Ljubljana: Založba ZRC SAZU.....	15, 16, 17, 19, 20, 24, 27, 30, 31, 60, 89, 90, 91, 100, 101, 176
SPERBERG-MCQUEEN, C. M., BURNARD, Lou (2001): <i>TEI P4: Guidelines for Electronic Text Encoding and Interchange: XML-compatible edition</i> . TEI Consortium. http://www.tei-c.org/	12
SSKJ: BAJEC, Anton idr. (ur.): <i>Slovar slovenskega knjižnega jezika</i> . Ljubljana: DZS (Slovarji DZS). [Uporabljena knjižna izdaja iz 2000. Tudi na http://bos.zrc-sazu.si/] 14, 15, 16, 23, 33, 47, 54, 55, 56, 57, 60, 98, 99, 100, 102, 149, 150, 154, 155, 176	
STABEJ, Marko (2003): Jezikovne tehnologije in jezikovno načrtovanje. <i>JiS XLVIII/3–4, Jezikovne tehnologije za slovenščino</i> . [5]–18.....	19
STABEJ, Marko, VITEZ, Primož (2000): KGB (korpus govornih besedil) v slovenščini. <i>Jezikovne tehnologije: zbornik konference</i> . Ljubljana: Institut Jožef Stefan. [Tudi na http://nl.ijs.si/isjt00/zbornik/sdjt00-Stabej11.pdf].....	13, 14
STARC, Branko (2006): Vokografija in njena uporaba v fonetiki. <i>SloFon1: 1. slovenska mednarodna fonetična konferenca: zbornik povzetkov</i> . [Ljubljana]: Inštitut za slovenski jezik Frana Ramovša, ZRC SAZU. 68.....	25
SUHADOLNIK, Stane (1974): Frekvenčna analiza besedišča v Tavčarjevi povesti Cvetje v jeseni. <i>SR XXII</i> . 31–40.	31, 33
SUHADOLNIK, Stane, JANEŽIČ, Marija (1962): Plasti in pogostnost leksike. <i>JiS VIII</i> , št. 1–2, 45–49, in št. 3, 73–78.	27, 31, 56, 62, 66, 67, 165
ŠKOFIC, Jožica (1991): <i>Problemi slovenskega pogovornega jezika: magistrska naloga</i> . Srednja Dobrava: [samozal.].	11, 16, 70
ŠKOFIC, Jožica (2006): Prevajanje govornega narečnega besedila v pisani knjižni jezik. <i>Diabronija in sinbronija v dialektoloških raziskavah</i> . Maribor: Slavistično društvo (Zora, 41). 174–182.	25, 26, 35
TIVADAR, Hotimir, JURGEC, Peter (2003): Podoba govornega slovenskega knjižnega jezika v Slovenskem pravopisu 2001. <i>SR LI/2</i> . [203]–220.	26
Toporišič [SS] (2000 ⁴): Slovenska slovnica. Maribor: Obzorja.	11, 22, 38, 52
TOPORIŠIČ, Jože (1962): Razgledi po stilistiki in stilu. <i>JiS VIII/1–2</i> . 34–44.	49, 52, 53
TOPORIŠIČ, Jože (1965): Pogovori o stilu. <i>Jezikovni pogovori [1]</i> . Ljubljana: CZ. 157–181.	53, 57
TOPORIŠIČ, Jože (1967): Nekaj strani iz slovenske slovnice: X. Stavčna intonacija. <i>Jezikovni pogovori II</i> . Ljubljana: CZ. 167–172.	21
TOPORIŠIČ, Jože (1992): <i>Enciklopedija slovenskega jezika</i> . Ljubljana: CZ (Leksikoni Cankarjeve založbe, Sopotnik).....	11, 12, 18, 19, 20, 22, 27, 52, 104, 176
TOPORIŠIČ, Jože, GJURIN, Velemir (1981): <i>Slovenska zvrstna besedila</i> . Ljubljana: Filozofska fakulteta.	26
TULDAVA, Johan (2005): Stylistics, author identification. <i>Quantitative Linguistik</i> . Berlin: de Gruyter (HSK – Handbücher zur Sprach- und Kommunikationswissenschaft, 27). 368–386.	49
TWEEDIE, Fiona J. (2005): Statistical Models in Stylistics and Forensic Linguistics. <i>Quantitative Linguistik</i> . Berlin: de Gruyter (HSK – Handbücher zur Sprach- und Kommunikationswissenschaft, 27). 387–396.	32, 49

- VERDONIK, Darinka (2006): *Analiza diskurza kot podpora sistemom strojnega simultanege prevajanja govora: doktorska disertacija*. Maribor: Pedagoška fakulteta. [Tudi na http://www.dsplab.uni-mb.si/Pdf_files/424a89810f2a1JiS_Darinka_Verdonik_zadnjaverzija.pdf (pregledano 10. 5. 2006).] ..73, 78, 110
- VERDONIK, Darinka, ROJC, Matej, KAČIČ, Zdravko (2003): Analiza jezikovnih vprašanj, nastalih pri gradnji SImflexa - oblikoslovnega in glasoslovnega slovarja za slovenski knjižni jezik. *Raziskovalno delo podiplomskih študentov v Sloveniji - ena znanost : e-zbornik*. Ljubljana: Društvo mladih raziskovalcev Slovenije. 434–443. [Tudi na http://www.dsplab.uni-mb.si/Pdf_files/41c2ecdf8cfdeClanek_SImflex_za_sln_jezvidik_net.pdf.]29
- VIDOVIČ MUHA, Ada (1971/72): Oris dveh osnovnih pojavnih oblik sistema knjižnega jezika. *JiS* XVII/6. 178–186..... 56, 57
- VIDOVIČ MUHA, Ada (2003): Kaj je novega v knjižnem jeziku? – Ob izidu Slovenskega pravopisa. *SR* LI/2. 117–122.....25
- VIDOVIČ-MUHA, Ada (2000): *Slovensko leksikalno pomenoslovje: Govorica slovarja*. Ljubljana: Znanstveni inštitut Filozofske fakultete (Razprave Filozofske fakultete)..... 31, 53, 54, 56
- VITEZ, Primož (1999): Od idealnih jezikovnih struktur k strategiji realnega govora. *SR* XLVII/1. 23–48. 11, 14, 18
- VITEZ, Primož, ZWITTER VITEZ, Ana (2004): Problem prozodične analize spontanega govora. *JiS* XLIX/6. 3–24.....22
- VLAJ, Damjan, KAČIČ, Zdravko, HORVAT, Bogomir (2000): Detektiranje pogrešanih značilk v šumnem okolju. *Jezikovne tehnologije: zbornik konference*. Ljubljana: Institut Jožef Stefan. [Tudi na <http://nl.ijs.si/isjt00/zbornik/sdjt00-Vlaj14.pdf>.]96
- VOGEL, Jerica (2004): Nekateri vidiki zvrstnosti govornega besedila s stališča poslušalca. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*. Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22). 453–466.41
- WAGENER, Peter (2004): *Deutsches Spracharchiv (DSAv) und Datenbank Gesprochenes Deutsch (DGD): Öffentliche Version*. Mannheim: Institut für Deutsche Sprache (IDS). [Http://dsav-oeff.ids-mannheim.de/DSAv/DSAVINFO.HTM](http://dsav-oeff.ids-mannheim.de/DSAv/DSAVINFO.HTM) (zadnja sprememba 12. 2. 2004).70
- WALES, Katie (1994³): *A dictionary of stylistics*. London, New York: Longman (Studies in language and linguistics).48, 49, 51
- Wikipedia: The Free Encyclopedia* (2007): .[gesla:] MP3; MPEG-1; WAV; Digital audio; Audio data compression; Speech encoding. [Http://en.wikipedia.org/wiki](http://en.wikipedia.org/wiki) (uporabljeno 21. 2. 2007).....74
- ZEMLJARIČ MIKLAVČIČ, Jana (2004): Taksonomija besedilnih tipov za gradnjo govornega korpusa. *Aktualizacija jezikovnozvrstne teorije na Slovenskem: Členitev jezikovne resničnosti*. Ljubljana: CSDTJ (Obdobja: Metode in zvrsti, 22). 503–522..... 11, 119
- Zemljarič Miklavčič, Jana (2005): Učni korpus govornega slovenskega jezika. [Računalniške prosojnice predstavitev.] Jezikovnotehnološki abonma – JOTA (15. 3. 2005). Ljubljana. [Http://www2.arnes.si/~svinta/zemljariic.ppt](http://www2.arnes.si/~svinta/zemljariic.ppt)..... 78, 119
- ZEMLJARIČ MIKLAVČIČ, Jana (2006): Korpus govornega slovenskega jezika. *Jezikovne tehnologije: zbornik 9. mednarodne multikonference Informacijska družba IS 2006, 9. do 10. oktober 2006*. Ljubljana: Institut..... 73, 110, 119
- ZEMLJARIČ MIKLAVČIČ, Jana, STABEJ, Marko (2005): Building a Pilot Spoken Corpus. *Computer treatment of Slavic and East European languages: Third International Seminar, Bratislava, Slovakia, 10–12 November 2005: proceedings*. Bratislava: Veda. 229–240..... 73, 110, 119, 120, 126
- ZEMLJARIČ MIKLAVČIČ, Jana, STABEJ, Marko (2006): *Zapisati nezapisljivo: Transkribiranje za referenčni govorni korpus*. [Računalniške prosojnice predstavitev na konferenci SloFon I, posredovano 14. 6. 2006.] (22. 4. 2006)..... 23, 91, 151
- ZUPAN, Jure (1994): Lahko računalnik prebira pesnike in ugotavlja njihov slog? *Delo – Književni listi* (21. julij 1994). 13–15. 34, 60
- ŽGANEC GROS, Jerneja, MIHELIČ, France, DOBRIŠEK, Simon (2003): Govorne tehnologije: pridobivanje in pregled govornih zbirk za slovenski jezik. *JiS* XLVIII/3–4, *Jezikovne tehnologije za slovensko*. [47]–59.70
- ŽGANK, Andrej, ROTOVNIK, Tomaž, VERDONIK, Darinka, KAČIČ, Zdravko (2004): Baza Broadcast News za slovenski jezik (BNSI) in sistem za razpoznavanje tekočega govora. *Jezikovne tehnologije: Zbornik 7. mednarodne multikonference Informacijska družba IS 2004: Zvezek B*. Ljubljana: Institut Jožef Stefan. [Tudi na <http://nl.ijs.si/isjt04/zbornik/sdjt04-17zgank.pdf>.] 70, 110

- ŽIBERT, Janez, MIHELIC, France (2004): Development, Evaluation and Automatic Segmentation of Slovenian Broadcast News Speech Database. *Jezikovne tehnologije: Zbornik 7. mednarodne multikonference Informacijska družba IS 2004: Zvezek B*. Ljubljana: Institut Jožef Stefan. [Tudi na <http://nl.ijs.si/isjt04/zbornik/sdjt04-14zibert.pdf>.]70

20.1. Druga zanimiva literatura – transkribiranje

- CAZINKIČ, Robert (2001): O vsakdanji govorici Ljubljane in o zapisu nezbornih besedil. *Bela Ljubljana: zgodbe iz slovenske prestolnice*. Ur. Stanonik, Marija idr. Ljubljana: Kmečki glas (Glasovi, 23). 27–32.
- GJURIN, Velemir (1990): Beseda avtorju. Gradišnik, Brane: *Nekdo drug*. Ljubljana: samozaložba. 163–171.
- Goody, Jack (1993): *Med pisnim in ustnim: študije o pisanosti, družini, kulturi in državi*. Ljubljana: Škuc, ZIFF (Studia humanitatis).
- KOLTER, Franz (1994): *Kvantitativna fonološka analiza slovenščine: Kosmačev roman Pomladni dan: Diplomarbeit*. [Mentor Gerhard Neweklowsky.] Celovec: [samozal.].
- LEECH, Geoffrey N., MYERS, Greg, THOMAS, Jenny (ur., 1995): *Spoken English on Computer: Transcription, mark-up, and application*. New York: Longman.
- SMOLE, Vera (1994): Folklorist med prevajanjem in zapisovanjem. *Traditiones*. 23. 143–154.
- STANONIK, Marija (2004): Folkloristično o razmerju med govornim in zapisanim. [Poglavje v:] *Slovsvena folkloristika*. Ljubljana. 174–203.
- TOPORIŠIČ, Jože (1973): Stilna vrednost glasovnih, prozodijskih, (pravo)pisnih, morfemskih in naglasnih variant slovenskega knjižnega jezika. *SR XXI/2*. [217]-263.

21. ZANIMIVE SPLETNE POVEZAVE

21.1. Seznamei govornih korpusov

- ARCHIV FÜR GESPROCHENES DEUTSCH (2007): *Links auf externe Archive für gesprochene Sprache*. Mannheim: IDS. [Http://agd.ids-mannheim.de/links.html](http://agd.ids-mannheim.de/links.html) (zadnja sprememba 25. 6. 2007).
- BREXER, Yvonne (2006): *Gateway to Corpus Linguistics on the Internet: Spoken Corpora*. New South Wales: Macquarie University, European Languages, German Division. http://www.corpus-linguistics.de/html/nav/corp/corp_spoken.html (b. l., zadnja sprememba [2006]).
- LLISTERI, Joaquim (2007): *Speech and Spoken Language Resources – Bibliography*. Barcelona: Universitat Autònoma de Barcelona. http://liceu.uab.es/~joaquim/language_resources/spoken_res/biblio_corpus_orals.html (zadnja sprememba 24. 5. 2007).
- SCHLICHTING, Frank ([2001]): *SFB 441: Projekt B5: Phonologie und Semantik von Fokus und verwandten Phänomenen (beendet 2001): Korpora gesprochener Sprache*. [Http://www.sfb441.uni-tuebingen.de/b5/spoken_corpora.html](http://www.sfb441.uni-tuebingen.de/b5/spoken_corpora.html) (zadnja sprememba 12. 4. [2001], b. l.).
- SCHNEIDER, Stefan (2007): *Banca dati dell'italiano parlato – BADIP: lista di corpora*. [Seznam italijanskih korpusov.] Gradec: Karl-Franzens-Universität Graz, Institut für Romanistik. [Http://languageserver.uni-graz.at/badip/](http://languageserver.uni-graz.at/badip/) (zadnja sprememba 12. 4. 2007).

21.2. Seznamei orodij za urejanje transkripcij

- DYBKJÆR, Laila idr. (2001): *Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data*. ISLE Natural Interactivity and Multimodality Working Group Deliverable D11.1. Januar 2001. [Http://isle.nis.sdu.dk/reports/wp11/](http://isle.nis.sdu.dk/reports/wp11/).
- LINGUISTIC ANNOTATION WIKI (2007). [Http://www.exmaralda.org/annotation/](http://www.exmaralda.org/annotation/) (zadnja sprememba 13. 8. 2007).
- SCHMIDT, Thomas, WÖRNER, Kai (2007): *EXMARALDA: Linksammlung*. Hamburg: Universität Hamburg: Sonderforschungsbereich Mehrsprachigkeit (SFB 538): Teilprojekt Mehrsprachige Datenbank. [Http://www1.uni-hamburg.de/exmaralda/Deutsch/5-Links/linksammlung.html](http://www1.uni-hamburg.de/exmaralda/Deutsch/5-Links/linksammlung.html) (b. l., zadnja sprememba [2007]).

VII. PRILOGE

PRILOGA 1: PREGLED BESEDIL V KORPUSU

Besedila so podana v dveh sklopih, za vsak dan posebej. Predstavljeni so izbrani podatki, vrednosti so za potrebe strnjene prikaza pri nekaterih podatkih prikazane z okrajšavami, a še vedno tako, da se da prepoznati oznako, kot je opisana v poglavju II.8. Druge oznake so: <-> pomeni brez, <nn> pomeni neznano; <+> pomeni da, pri podatku [29] vgnezdno pa to, da so vgnezdna besedila prisotna, niso pa v tabeli naštet.

NEDELJINA BESEDILA (pred ID besedila manjka <1ned>)

01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgnezdno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govortci
															21) tema	22) dejavnost					
01P0	0:23	8:40	Lov	osebna soba	0	0,00	-	-	-	-	-	-	-	-	-	jutranje oblačenje	0	1	+	-	
02Dm	0:05	8:40	Lov	kuhinja /jedilnica	9	1,80	pozdrav	-	P	osebni	nar	zpčo	Z	ps	pozdrav	jutranji pozdrav	2	2	-	2	001Mz, 002Žz
03P0	3:08	8:41	Lov	kopalnica	0	0,00	-	-	-	-	-	-	-	-	-	jutranja higiena	0	1	+	-	
04Vg	26:10	8:44	Lov	kuhinja /jedilnica	2440	1,55	pogovor ob zajtrku, branje novic	+	PzN	osebni	nar	zpčo	Z	ps	¹⁷²	zajtrk	3	3-10	-	2	001Mz, 002Žz, 003Mz

¹⁷² spanje, telesno počutje, pitje pijač, znanci na trgatvi, načrti za večer (vožnja z vlakom, zborovski koncert), brki, nogometna tekma, branje teletekska, povojni poboji

12P0	11Vr	10Dg	09Vg	08Mr	07P0	06Dm	05Mm	01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgruzdeno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorniki	
																							21) tema	22) dejavnost						
-	-	-	+	+	-	-	+				Lov	osebna soba	11	0,03	popevanje, glasno razmišljanje	+	P+N	sam	ppj	zpčo	Z	um	popevanje, ravnanje obleke	urejanje pred odhodom od doma	1	1	+	-	001Mz	
-	-	-	+	NzP	-	P	-				Lov	kuhinja /jedilnica	4	0,50	slovo	-	P	osebn	ppj	zpčo	Z	ps	slovo	slovo pred odhodom od doma	2	2	-	2	001Mz, 003Mz	
-	-	-	+	+	-	-	-				Lov	hoja po lokalni cesti, poljski poti	0	0,00	-	-	-	-	-	-	-	-	-	oblačenje, odhod od doma, hoja po cesti in poljski poti proti cerkvi		0	1	+	-	
-	-	-	+	+	-	-	+				Lov	pevska soba pri župnišču	25	0,07	popevanje	+	NzP	sam	kpj	zpčo	Z	um	pesmi za mašo (izbiranje)	priprave na pevske vaje v pevski sobi	1	1	+	-	001Mz	
-	-	-	+	PzN	-	P	+				Lov	pevska soba pri župnišču	2217	1,48	pevske vaje	+	PzN	osebn	ppj	zpčo	J	ps	izbiranje in vaja pesmi za mašo, razno	pevske vaje pred mašo, razni dogovori in klepet	16	11-99	-	2	173	
-	-	-	+	+	-	P	-				Lov	hoja od pevske sobe v cerkev	120	2,50	pomenek	-	P	osebn	nar	zpčo	Z	ps	preteklo počutje govorke	pripoved sogovornice med potjo iz mežnarije v cerkev	2	2	-	2	001Mz, 008Žz	
-	-	-	+	+	-	P	-				Lov	kor v cerkvi	72	0,43	navodila, klepet	-	P	osebn	nar	zpčo	Z	ps	navodila pevcem, verouk prejšnji dan	navodila pevcem na koru, klepet	3	3-10	-	2	001Mz, 018Mz, 019Žz	
-	-	-	+	+	-	-	-				Lov	kor v cerkvi	0	0,00	-	-	-	-	-	-	-	-	-	tišina pred začetkom maše		0	11-99	-	-	

¹⁷³ 001Mz, 004Žz, 005Žz, 006Žz, 007Žz, 008Žz, 009Žz, 010Žz, 011Mz, 012Žz, 013Žz, 014Mz, 015Vz, 016Vz, 017Žz, 018Mz

27)r	26)r	25)Vg	24)Dm	23)P0	22)Vg	21)P0	20)Dg	01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgrazdeno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorci	
																							21) tema	22) dejavnost						
2:07	8:37	50:38	1:17	24:00	9:30	1:24	1:57		1:57	11:14	Lov	hoja po cesti proti domu	124	1,06		-	P	osebno	nar	zpčo	Z	ps	dogovor za pevske vaje, šola	pogovor ob hoji skupaj	2	2	-	2	001Mz, 009Žz	
12:51	12:43	11:52	11:51	11:27	11:17	11:16	11:14		11:16		Lov	veža v domači hiši	0	0,00		-	-	-	-	-	-	-		oblačenje in slačenje ob prihodu domov		0	1	+	-	
											Lov	kuhinja /jedilnica	949	1,66		-	P	osebno	nar	zpčo	Z	ps	krst v družini znancev, nova družina v sosesčini, načrt dostave reči sorodnikom	pogovor pred kosilom o novicah in dogovarjanje o načrtih	3	3-10	-	2	001Mz, 002Žz, 003Mz	
											Lov	osebna soba	0	0,00		-	-	-	-	-	-	-		zapiski za snemanje, sam zase		0	1	+	-	
											Lov	osebna soba	5	0,06		-	P	osebno	nar	zpčo	Z	ps	poziv h kosilu	poziv sogovorca h kosilu, nato zaključevanje dela v tišini	2	2	-	2	001Mz, 003Mz	
											Lov	kuhinja /jedilnica	4165	1,37		+	PzN	osebno	nar	zpčo	Z	ps	¹⁷⁷	družinsko kosilo		3	3-10	-	2	001Mz, 002Žz, 003Mz
											Lov	kuhinja /jedilnica	516	1,00		-	P	osebno	nar	zpčo	Z	ps	prehrana, pospravljanja kuhinjske posode	pospravljanje posode	2	2	-	2	001Mz, 003Mz	
											Lov	kuhinja /jedilnica	104	0,82		-	P	osebno	nar	zpčo	Z	ps	ogrevanje hiše	dogovarjanje	3	3-10	-	2	001Mz, 002Žz, 003Mz	

¹⁷⁷ molitev pred jedjo, hrana, izobrazba znanke, spored martinovanja, vreme, pevski zbor, gojenje rož, poraba papirja za printanje, popoldanski načrti

35Dg	34Vg	33Mr	32Dg	31P0	30Mr	29Mm	28Dm	01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgrazdno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorci	
																							21) tema	22) dejavnost						
											Lov	kopalnica, kuhinja	8	0,03		-	P	osebni	nar	zpčo	Z	ps	popivanje, neznan	osebna higiena po kosilu	2				001Mz, 002Žz	
											Lov	osebna soba	4	0,00		-	P	sami	nar	zpčo	Z	ps	mrmljanje, popivanje	zapiski in urejanje zgoščenek	1	1	+	-	001Mz	
											Lov	osebna soba	31	0,01		+	NzP	sami	kpj	zpčo	Z	um	popivke	poslušanje zgoščenke in urejanje zgoščenek	1	1	+	-	001Mz	
											Lov	osebna soba	0	0,00		-	-	-	-	-	-	-	-	-	zapiski za urejanje zgoščenek		1	+	-	
											Lov	osebna soba	526	0,53		+	N	RA	zgj	zpčo	Z	um	balade rock/pop	poslušanje zgoščenke in urejanje zgoščenek	2	2	+	1	001Mz, 029Va	
											Lov	osebna soba in delovni kabinet z računalnikom	66	0,01		+	P+N	sami	kpj	zpčo	Z	um	popivanje bivših pesmi, urejanje na računalniku	urejanje ovitkov za zgoščenke, tudi s pomočjo računalnika	1	1	+	-	001Mz	
											Lov	osebna soba in kuhinja	233	0,11		-	P	osebni	nar	zpčo	Z	ps	pakiranje študenta: zapiski, hrana, obleka	pakiranje prtljage, pregledovanje urnika	3	3-10	-	2	001Mz, 002Žz, 003Mz	
											Lov	osebna soba	302	1,39		-	P	mobi	ppj	zpčo	Z	ps	čas in lokacija prihodnjega snidenja	pogovor po mobitelu	2			2	001Mz, 030Mz	

52Dm	0:02	19:28	Lj-MB	regionalni potniški vlak	4	2,00	vprašanje in odgovor	-	P	osebno	ppj	zpčo	J	ps	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorniki
															21) tema	22) dejavnost					
52Dm	0:02	19:28	Lj-MB	regionalni potniški vlak	4	2,00	vprašanje in odgovor	-	P	osebno	ppj	zpčo	J	ps	povpraševanje po prostem sedežu	posedanje sopotnice	2		2	001Mz, 037Žn	
51P0	27:21	19:02	Lj-MB	regionalni potniški vlak	0	0,00	-	-	-	-	-	-	-	-	vožnja z vlakom ce-zidani most-hrastnik	vožnja z vlakom mb-ce	0	3-10	-		
50Vg	0:12	19:01	Lj-MB	regionalni potniški vlak	17	1,42	pomenek	-	P	osebno	ppj	zpčo	Z	ps	vsebina in odložitve prtljage	pogovor v skupini sopotnikov	4	3-10	-	033Mn, 034Mn, 035Žn, 036Žn	
49P0	52:42	18:08	Lj-MB	regionalni potniški vlak	0	0,00	-	-	-	-	-	-	-	-	vožnja z vlakom mb-ce	vožnja z vlakom mb-ce	0	1	+		
48Dm	0:05	18:07	Lj-MB	regionalni potniški vlak	9	1,80	kontrola	-	P	osebno	kpj	posl	J	ps	kontrola vozovnice	žigosanje vozovnice	2	2	-	001Mz, 032Mn	
47Mm	12:27	17:55	Lj-MB	regionalni potniški vlak	7	0,01	samogovor	-	P	sam	ppj	zpčo	Z	ps	čudenje praznemu vlaku	čakanje na odhod vlaka in naposled odhod vlaka	1	1	+	001Mz	
46Mr	5:48	17:49	MB	železniška postaja	25	0,07	(javna) napoved	-	P	zvočni	zgj	info	J	ps	čas odhoda vlaka	napovedovanje odhoda vlaka, vmes premor	1	11-99	-	031Žj	
45Mm	12:10	17:37	MB	peron in vlak na žel. p.	5	0,01	popevanje sebi	-	P+N	sam	kpj	zpčo	Z	um	popevanje, vlak	čakanje na odhod vlaka	1	1	+	001Mz	
01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgrazdeno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst							

61P0	19:06	21:15	Lj	kuhinja/skupni prostor	0	0,00	-	-	-	-	-	-	-	-	-	-	-	-	-	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorci
																				21) tema	22) dejavnost					
61P0	19:06	21:15	Lj	kuhinja/skupni prostor	0	0,00	-	-	-	-	-	-	-	-	-	-	-	-	-	vožnja z vlakom hrastnik-lj	0	11-99	-	-		
60Mm	0:05	21:14	Lj	osebna soba	3	0,60	samogovor	-	P	sam	nar	zpčo	Z	ps	kašljanje in voščilo samemu sebi	pospravljanje sobe	1	1	+		1	+	-	001Mz		
59P0	33:47	20:41	Lj	osebna soba	0	0,00	-	-	-	-	-	-	-	-	namestitve v stanovanju: higiena, razpakiranje, hoja po prostorih		0	1	+		0	+	-			
58Mm	2:13	20:39	Lj	stopnišče večstanovajske hiše	4	0,03	samogovor	-	P	sam	nar	zpčo	Z	ps	prazno stanovanje	vstopanje v hišo in dalje v stanovanje, odklepanje	1	1	+		1	+	-	001Mz		
57Dg	7:21	20:31	Lj	avto, vožnja Bežigrad-Vodmat	986	2,24	pogovor	-	P	osebni	ppj	zpčo	Z	ps	smer vožnje, izmenjava reči, brada, zdravje, snemanje, slovo	pogovor med vožnjo	2	2	+		2	+	2	001Mz, 030Mz		
56Mr	0:19	20:30	Lj	Avtobusna postaja Lj	10	0,53	obvestila	-	N	zvočni	zgj	info	J	ps	obvestila potnikom	čakanje pred postajo na ulici na prevoz	1	100+	-		1	1	040Ma			
55Mg	0:32	20:30	Lj	avla žel. p.	22	0,69	pogovor po mobitelu	-	P	mobi	ppj	zpčo	Z	ps	snidenje	čakanje na prevoz	1	2	-		1,5	001Mz				
54Dm	1:54	20:28	Lj	peron na žel. p.	9	0,08	vzkliki v gneči	-	P	osebni	ppj	zpčo	Z	ps	dežnik, prihod avtobusa	hoja iz perona skozi avlo železniške postaje	2	100+	-		1	038Mn, 039Žn				
53P0	58:58	19:29	Lj-MB	regionalni potniški vlak	0	0,00	-	-	-	-	-	-	-	-	-	-	-	-	-	vožnja z vlakom hrastnik-lj	0	11-99	-	-		

69Vg	13:28 21:53	Lj	kuhinja/skupni prostor	928 1,15	TV-dokumentarec	+	NzP	TV	kpj	info	J	pu	življenje slikarja Oskarja Pistorja	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorniki
														21) tema	22) dejavnost					
62Mm	12:20 21:34	Lj	kuhinja/skupni prostor	4 0,01	4 0,01	-	P	TV	sam	nar	zpčo	Z	ps	popevanje	pripravljanje večerje: čaja, hrane, razgledovanje po kuhinji in skozi okno	1	1	+	-	001Mz
63Mm	0:26 21:46	Lj	kuhinja/skupni prostor	6 0,23	6 0,23	-	N	TV	zgj	info	J	pu	napoved TV-oddaje	večerjanje, tv: napovednik oddaje	1	100+	+	1	041Žj	
64Mg	0:50 21:46	Lj	kuhinja/skupni prostor	16 0,32	16 0,32	-	N	TV	zgj	info	J	pu	napovednik TV-oddaje	večerjanje, gledanje tv	1	100+	+	1	042Mj	
65Mg	1:44 21:47	Lj	kuhinja/skupni prostor	187 1,80	187 1,80	-	N	TV	zgj	info	J	pu	reklama za PVC-pohištvo	večerjanje, gledanje tv	1	100+	+	1	043Žj	
66Mg	0:16 21:49	Lj	kuhinja/skupni prostor	25 1,56	25 1,56	-	N	TV	zgj	info	J	pu	reklama za revijo	večerjanje, gledanje tv	1	100+	+	1	044Žj	
67Mg	1:15 21:49	Lj	kuhinja/skupni prostor	16 0,21	16 0,21	-	N	TV	zgj	info	J	pu	reklama za kurilno olje	večerjanje, gledanje tv	1	100+	+	1	043Žj	
68Vg	3:21 21:50	Lj	kuhinja/skupni prostor	315 1,57	315 1,57	-	PzN	TV	kpj	info	J	pu	reprezentančna nogometna tekma	večerjanje, gledanje tv	7	100+	+	1	045Mj, 046Mj, 047Mj, 048Mj, 049Mj, 050Mj, 051Mj	
69Vg	13:28 21:53	Lj	kuhinja/skupni prostor	928 1,15	928 1,15	+	NzP	TV	kpj	info	J	pu	življenje slikarja Oskarja Pistorja	večerjanje, gledanje tv	6	100+	+	1	052Mj, 053Žj, 054Mj, 055Mj, 056Vn, 057Žj	

72Mm	71Dg	70Vg	01) ID besedila		18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgnezdno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	NEDELJA pred ID besedila manjka <1ned>		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govortci
			21) tema	22) dejavnost																		
2:47 22:56 LJ	34:54 22:21 LJ	13:51 22:07 LJ	2	0,01	kopalnica	2	0,01	samogovor	-	P	sam	nar	zpčo	Z	ps	milo	večerna higiena pred spanjem	1	1	+	-	001Mz
			2671	1,28	osebna soba	2671	1,28	pogovor pred spanjem	-	P	osebn	ppj	zpčo	Z	ps	¹⁷⁸	pogovor o preteklih počitnicah in prihajajočem študiju, pripravljanje na spanje	2	2	-	2	001Mz, 058Mz
			849	1,02	kuhinja/skupni prostor	849	1,02	pogovor ob TV	+	P	osebno	ppj	zpčo	Z	ps	snidenje, vsebina TV-dokumentarca, nogomet, rokomet, formula 1, pomivanje posode	pogovor sstanovalcev ob snidenju po dolgem času; ob tv, pomivanje posode	3	3-10	-	2	001Mz, 053Žj, 058Mz

PETKOVA BESEDILA (pred ID besedila manjka <6pet>)

02Dr	01Dg	01) ID besedila		18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgnezdno	30) prostogovom.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govortci	
		21) tema	22) dejavnost																			
5:51 9:32 LJ	11:50 9:20 LJ	11:50 9:20 LJ	547	1,56	kuhinja/skupni prostor	547	1,56	pogovor po mobitelu in telefonu	-	P	mobi	ppj	zpčo	Z	ps	delitev hrane za brezdomce, čakajoči klic po telefonu	Pogovor po mobitelu, nato še čakajoč klic po stacionarnem telefonu.	2	3-10	-	1	001Mz, 059Mz
			641	0,90	kuhinja/skupni prostor	641	0,90	pogovor ob zajtrku	-	P	osebno	ppj	zpčo	Z	ps	vstajanje, zajtrk, sanje, kostanjev piknik, nočni nemir, verzi, delitev hrane brezdomcem	Jutranji pogovor ob kavi, pospravljanje posode in pripravljanje zajtrka.	2	2	-	2	001Mz, 059Mz

¹⁷⁸ vodovod, začetek predavanj, brada, sstanovalci, absolventski izleti, počitnice, volitve, študentski sejem, lahko noč

10P0	09:35 11:49	09Dm	08Pn	07Mr	06Pn	05Dr	04Mr	03Dr	01) ID besedila 14) trajanje [mm:ss] 15) čas začetka 17) kraj	18) prostor	12) besede 13) besede/s	28) oblika sporaz.	29) vgnezdno 30) prostogovorn.	20) medij	31) soc. zvrst 32) namen sprz. 33) gov. položaj 34) funkc. zvrst	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zaveš snem.	35) smer sprz.	XI) prisotni govorci			
																21) tema	22) dejavnost								
										kuhinja/skupni prostor	30 0,58	pomenek	- P	osebn ppj	zpčo Z	ps	neznani zlicatelj po telefonu	Pogovor - ugibanje o klicatelju po telefonu, ki je prekinil linijo.	2	2	-	2	001Mz, 059Mz		
										kuhinja/skupni prostor	61 0,97	pogovor po telefonu	- P	telef ppj	zpčo Z	ps	pot domov	Pogovor po telefonu.	1	3-10	-	1,5	059Mz		
										kuhinja/skupni prostor	41 0,21	pomenek	- P	osebn ppj	zpčo Z	ps	neznani klicatelj, pospravljanje kuhinje	Pospravljanje po zajtrku.	2	2	-	2	001Mz, 059Mz		
										kuhinja/skupni prostor, kopalnica, osebna soba	0 0,00	nn	+ nn	nn	nn	nn	nn	nn	Pomivanje posode, umivanje zob, urejanje študijskih stvari in MID v sobi, sostanovalčevo slovo.		0	2	-	nn	
										osebna soba	245 0,06	polglasno (samo)učenje	+ PzN	sam	kpj	info	Z	st	predmet Zgodovina institucij	Samoučenje arhivistike iz zvezka, pregled skripta.	1	1	+	-	001Mz
										stanovanje, hoja po Vodmatu	0 0,00	nn	+ nn	nn	nn	nn	nn	nn	nn	179	0	1	-	nn	
										ulica na Vodmatu	11 0,92	spraševanje o času	- P	osebn	ppj	info	Z	ps	informacija o času	Srečanje na ulici.	2	3-10	-	2	001Mz, 060Žn
										hoja Vodmat-kolodvor	0 0,00	-	-	-	-	-	-	-	Hoja skozi mesto.		0	+	-		

¹⁷⁹ Preoblacanje, stranišče, pripravljanje nahrbtnika, obuvanje, odhod iz hiše, zaklepanje, hoja po ulici (z Vodmata proti žel. p.).

28Dr	1:02 14:10 Lj	pult kuhinje v menzi	26 0,42	7 0,09	pozdrav	-	P	osebno	ppj	posl	J	ps	prevzem hrane	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zaveš. snem.	35) smer sprz.	XI) prisotni govorci
														21) tema	22) dejavnost					
28Dr	1:02 14:10 Lj	pult kuhinje v menzi	26 0,42	7 0,09	pozdrav	-	P	osebno	ppj	posl	J	ps	prevzem hrane	21) tema	22) dejavnost	2	2	-	2	001Mz, 075Mn
27Dm	1:15 13:08 Lj	vhod v dijaški dom	7 0,09	7 0,09	pozdrav	-	P	osebno	kpj	zpčo	J	ps	pozdrav	21) tema	22) dejavnost	2	2	-	2	001Mz, 074Mn
26P0	9:12 12:59 Lj	hoja po Poljanah	0 0,00	0 0,00	-	-	-	-	-	-	-	-	-	21) tema	22) dejavnost	0	nn	+	-	
25Vg	1:11 12:58 Lj	mladinska pisarna	116 1,63	116 1,63	pomenek ob predaji	-	P	osebno	ppj	posl	Z	ps	predaja zgoščanje, prevzem tiskovin	21) tema	22) dejavnost	3	3-10	-	2	001Mz, 072Žz, 073Mn
24Dm	0:34 12:57 Lj	vhod v urad	7 0,21	7 0,21	pogovor po domofonu	-	P	telef	ppj	zpčo	J	ps	odpiranje vrat	21) tema	22) dejavnost	2	2	-	2	001Mz, 072Žz
23P0	5:27 12:52 Lj	hoja po Poljanah	0 0,00	0 0,00	-	-	-	-	-	-	-	-	-	21) tema	22) dejavnost	0	nn	+	-	
22Dr	1:23 12:51 Lj	prodajalna pekarnice na tržnici	32 0,39	32 0,39	popivanje, pogovor ob nakupu	-	P	osebno	ppj	posl	J	ps	popivanje, nakup pekovskih izdelkov	21) tema	22) dejavnost	2	2	-	2	001Mz, 071Žn
21Dg	5:58 12:45 Lj	tržnica	912 2,55	912 2,55	pomenek	-	P	osebno	ppj	zpčo	Z	ps	aktualne poti govorcev, skupni znanci, skavtske aktivnosti, dvigalo na tržnici, malica	21) tema	22) dejavnost	2	2	-	2	001Mz, 070Mz
20P0	4:44 12:40 Lj	hoja skozi Center	0 0,00	0 0,00	-	-	-	-	-	-	-	-	-	21) tema	22) dejavnost	0	nn	+	-	

37Mm	6:07	14:02	Lj	kopalnica	3	0,01	samogovor	-	P	osebni	ppj	zpčo	Z	ps	slab zrak, spomin na sostanovalca	Olajšanje in umivanje v kopalnici, planiranje popoldneva v sobi.	1	1	+	-	-	001Mz																					
																							36Dm	1:11	14:01	Lj	kuhinja/skupni prostor	15	1,36	pozdrav, vprašanje	-	P	osebni	ppj	zpčo	Z	ps	pozdrav, prisotni sostanovalci	2	2	-	2	001Mz, 058Mz
35P0	1:52	13:59	Lj	osebna soba	0	0,00	-	-	-	-	-	-	-	-	-	Razpakiranje doma.	0	1	+	-	-																						
34Vg	1:27	13:58	Lj	stopnišče	179	2,06	seznanjanje, pomenek	-	P	osebni	ppj	zpčo	Z	ps	seznanjanje, kosanje piknik	Predstavljanje, dogovor za kosanje piknik.	4	3-10	-	2		001Mz, 076Mz, 077Mz, 078Mz																					
33P0	15:00	13:42	Lj	hoja po Poljanah in Vodmatu	0	0,00	-	-	-	-	-	-	-	-	-	Hoja skozi mesto.	0	nn	+	-	-																						
32Dm	0:08	13:42	Lj	vhod v dijaški dom	3	0,38	slovo	-	P	osebni	ppj	zpčo	J	ps	slovo	Slovo od vratarja pri odhodu iz dijaškega doma.	2	2	-	2		001Mz, 074Mn																					
31P0	19:51	13:22	Lj	jedilnica dijaškega doma	0	0,00	-	-	-	-	-	-	-	-	-	Kosilo, zraven tiho branje revije, konec kosila, pospravljanje pribora, odpravljanje.	0	11-99	-	-	-																						
30Mm	0:05	13:22	Lj	jedilnica dijaškega doma	8	1,60	poziv po zvočniku	-	P	zvočni	zgj	zpčo	J	ps	napoved obiska gojenke doma	Napoved obiska za dijakinjo v domu; po zvočniku.	1	nn	-	1		074Mn																					
29P0	11:46	13:11	Lj	jedilnica dijaškega doma	0	0,00	-	-	-	-	-	-	-	-	-	Izbor solate, posedanje v jedilnici, kosilo, zraven tiho branje revije.	0	3-10	-	-	-																						
01) ID besedila		14) trajanje [mm:ss]		15) čas začetka		17) kraj		18) prostor		12) besede		13) besede/s		28) oblika sporaz.		29) vgnedeno		30) prostogovorn.		20) medij		31) soc. zvrst		32) namen sprz.		33) gov. položaj		34) funkc. zvrst		21) tema		22) dejavnost		23) št. govorcev		24) št. soudelež.		25) zavest snem.		35) smer sprz.		XI) prisotni govorci	

53Mr	0:44 18:57	Lj-MB	peron žel. p.	25 0,57	javno obvestilo	-	P+N	zvočni	zgj	info	J	ps	lokacija vlakov	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zavezi snem.	35) smer sprz.	XI) prisotni govornici
														21) tema	22) dejavnost					
52Vr	10:17 18:47	Lj-MB	lokalni potniški vlak	17 0,03	napoved, obvestilo, ugotovitev	-	PzN	zvočni	kpj	info	J	ps	nasledna postaja, prestop z vlaka	Vožnja z vlakom, pripravlanje na prestop.	1 100+	+	1	001Mz, 089Ža, 091Mn		
51Mr	2:00 18:45	Lj-MB	lokalni potniški vlak	138 1,15	pogovor po mobitelu	-	P	mobi	nar	zpčo	Z	ps	dogovor za prevoz domov	Pogovor po mobitelu.	1 3-10	-	1,5	001Mz		
50Pn	36:00 18:09	Lj-MB	lokalni potniški vlak	0 0,00	nn	+	nn	nn	nn	nn	nn	nn	nn	Vožnja z vlakom na relacija LJ-Zidani Most, nosilec spi.	0 11-99	-	nn			
49Dr	9:42 17:59	Lj-MB	lokalni potniški vlak	29 0,05	napoved, pogovor ob kontroli	-	P	osebno	ppj	posl	J	ps	napoved postaje, kontrola vozovnic	Vožnja z vlakom, napovedi postaj, sprevodnik kontrolira vozna karte.	2 3-10	-	1	089Ža, 091Mn		
48Mr	1:29 17:58	Lj-MB	lokalni potniški vlak	92 1,03	dogovarjanje po mobitelu	-	P	mobi	ppj	zpčo	Z	ps	snidenje	Pogovor sopotnika po mobitelu.	1 3-10	-	1,5	090Mn		
47Dr	11:28 17:47	Lj-MB	lokalni potniški vlak	10 0,01	1) vprašanje, 2) napoved	-	P	osebno	ppj	zpčo	J	ps	1) zasedanje prostora, 2) naslednja postaja	Namestitev v vlaku, začetek vožnje.	2 2	-	1	001Mz, 089Ža		
46Mg	1:40 17:45	Lj	peron žel. p.	147 1,47	javno obvestilo	+	P+N	zvočni	zgj	info	J	ps	zamuda vlaka, odhod vlaka	Na žel. p., obvestila po zvočniku.	1 100+	-	1	088Žj		

61) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgnezdno	30) prostogovorn.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govornici	
															21) tema	22) dejavnost						
54Vg	23:45	18:58	Lj-MB	lokalni potniški vlak	2517	1,77	pogovor	+	P	osebno	ppj	zpčo	Z	ps	180	Vožnja z vlakom, pogovor med znancema.	5	2	-	2	001Mz, 089Ža, 093Žz, 094Mn, 095Žn	
55Vm	1:19	19:22	Lj-MB	lokalni potniški vlak	7	0,09	vprašanje	-	P	osebno	ppj	zpčo	J	ps		zasedanje sedeža, kontrola vozovnic	3	3-10	-	2	001Mz, 094Mn, 096Nn	
56Mm	7:14	19:23	Lj-MB	lokalni potniški vlak	3	0,01	napoved-obvestilo	-	N	zvočni	zgj	info	J	ps		napoved postaje	1	100+	+	1	089Ža	
57Pn	10:00	19:30	Lj-MB	lokalni potniški vlak	0	0,00	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	Vožnja z vlakom, branje knjige.	0	11-99	-	nn	
58Mr	3:14	19:40	Lj-MB	lokalni potniški vlak	177	0,91	dogovarjanje po mobitelu	-	P	mobi	nar	zpčo	Z	ps		snidenje, prevoz, čakanje na to	1	3-10	-	1,5	001Mz	
59Mm	3:35	19:43	Lj-MB	lokalni potniški vlak	7	0,03	napoved-obvestilo	-	N	zvočni	zgj	info	J	ps		napoved postaj	1	100+	+	1	089Ža	
60Pn	43:00	19:47	Lj-MB	lokalni potniški vlak	0	0,00	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	Vožnja z vlakom.	0	11-99	-	nn	
61Dm	3:48	20:30	Lj-MB	lokalni potniški vlak	4	0,02	napoved, slovo	-	N	zvočni	zgj	info	J	ps		napoved postaje	2	100+	+	1	001Mz, 089Ža	

¹⁸⁰ vozovnice, študij (didaktični nastopi, diploma, zaposlitev), pristočasne dejavnosti (zborovodstvo, skavti, animiranje), skupni znanci, stan

70Pn	17:00	21:13	MB	Glavni trg v MB	0	0,00	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	PETEK (pred ID besedila manjka <6pet>)		23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.	XI) prisotni govorniki
																			21) tema	22) dejavnost					
70Pn	17:00	21:13	MB	Glavni trg v MB	0	0,00	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	nn	
69Dg	1:27	21:11	MB	Glavni trg v MB	228	2,62	pogovor po mobitelu	-	P	mobi	nar	zpčo	Z	ps	poti družinskih članov	Pogovor po mobitelu.	2	2	-	2	001Mz, 002Žz				
68P0	18:24	20:53	MB	hoja po Centru MB	0	0,00	-	-	-	-	-	-	-	-	-	Hoja po mestu.	0	1	+	-	-				
67Dg	0:49	20:52	MB	na ulici, Center MB	95	1,94	pogovor po mobitelu	-	P	mobi	nar	zpčo	Z	ps	termin in kraj snidenja	Pogovor po mobitelu.	2	2	-	2	001Mz, 003Mz				
66P0	6:33	20:45	MB	hoja po Centru MB	0	0,00	-	-	-	-	-	-	-	-	-	Hoja po mestu.	0	1	+	-	-				
65Mr	1:04	20:44	MB	hoja po Centru MB	79	1,23	pogovor po mobitelu	-	P	mobi	ppj	zpčo	Z	ps	možnost srečanja	Pogovor po mobitelu.	1	2	-	1,5	001Mz				
64P0	6:37	20:37	MB	hoja po Centru MB	0	0,00	-	-	-	-	-	-	-	-	-	Hoja po mestu.	0	1	+	-	-				
63Mr	0:54	20:36	MB	peron že postaje	21	0,39	obvestilo	-	P+N	zvočn	zgj	info	J	ps	zamuda vlaka	Napoved vlaka na žel. p.	1	100+	-	1	099Mj				
62Dr	2:52	20:34	Lj-	lokalni potniški vlak	99	0,58	pomenek v živo in pomenek po mobitelu	-	P	mobi	nar	zpčo	Z	ps	nn	Pogovor sopotnikov na vlaku, eden od njiju dobi pe klic po mobitelu.	2	3-10	-	2	097Žn, 098Mn				
01) ID besedila	14) trajanje [mm:ss]	15) čas začetka	17) kraj	18) prostor	12) besede	13) besede/s	28) oblika sporaz.	29) vgnedeno	30) prostogovorn.	20) medij	31) soc. zvrst	32) namen sprz.	33) gov. položaj	34) funkc. zvrst			23) št. govorcev	24) št. soudelež.	25) zavest snem.	35) smer sprz.					

PRILOGA 2: BAZA GOVORCEV

Podani so izbrani podatki o govornih, ki so obrazloženi v poglavju II.8.3. Opis govornca v podatku [0) ostalo] je v glaven prisoten pri govornih, ki nastopajo le v 1 ali 2 besedilih, in zato lahko opisuje njihovo značilno vlogo.

I) ID govornca	12O) besede	XIO) prisoten v besedilih	VII) starost	VIII) govorno poreklo	IX) izobrazba	X) razmerje do 001Mz	0) ostalo
001Mz	13.537	90	20-26	kor.	SŠ	isti	Izbrani govorec, nosilec mikrofona.
002Žz	5.988	15	50-65	štaj.	VŠUNI	družina	
003Mz	6.782	20	50-65	kor.	VŠUNI	družina	
004Žz	202	4	16-19	kor.	OŠ	znanec	
005Žz	22	2	7-15	kor.	(še) brez	znanec	
006Žz	119	3	16-19	kor.	OŠ	znanec	
007Žz	344	4	20-26	kor.	SŠ	znanec	
008Žz	1.244	6	27-39	kor.	SŠ	prijatelj	
009Žz	97	3	7-15	kor.	(še) brez	znanec	
010Žz	162	3	20-26	kor.	VŠUNI	prijatelj	
011Mz	3	2	20-26	kor.	SŠ	znanec	
012Žz	84	2	27-39	kor.	VŠUNI	družina	
013Žz	39	2	7-15	kor.	(še) brez	znanec	
014Mz	64	2	27-39	kor.	VŠUNI	družina	
015Vz	877	2	več	več	več	znanec	Petje zbora; osnovni glas, kjer poje večina pevcev isto besedilo. Gre izključno za recitirano/brano besedilo.
016Vz	18	3	več	več	več	znanec	Petje zbora; spremljevalni glas, ki ga zgolj občasno poje manjšina zbora. Gre izključno za recitirano/brano besedilo.
017Žz	8	1	16-19	kor.	OŠ	znanec	
018Mz	17	2	16-19	kor.	OŠ	znanec	
019Žz	165	5	7-15	kor.	(še) brez	znanec	
020Mz	2.182	1	40-49	štaj.	VŠUNI	znanec	Duhovnik, v vlogi maševalca; mikrofona - zvočnik.
021Vn	577	1	več	več	več	neznanec	Občestvo/množica pri sveti maši, kadar recitira/poje skupaj; mašni deli, molitveni obrazci, odpevi in odgovori pri litanijah, prošnjah ...
022Žz	193	1	nn	kor.	SŠ	znanec	Bralka pri maši: mikrofona - zvočnik.
023Mj	110	1	27-39	štaj.	VŠUNI	javno spoznan	Bralec pri maši: zvočnik - mikrofona.
024Žz	86	1	7-15	kor.	(še) brez	znanec	Bralka pri maši: mikrofona - zvočnik.
025Mz	90	2	20-26	kor.	SŠ	prijatelj	
026Mz	2	1	20-26	kor.	SŠ	znanec	Delivec obhajila.
027Žz	232	2	20-26	kor.	SŠ	znanec	
028Žz	111	3	16-19	kor.	OŠ	znanec	
029Va	509	1	več	več	več	javno spoznan	Glasbena zgoščenka slovenskega „rokerja“.
030Mz	484	2	27-39	LJ	VŠUNI	družina	
031Žj	25	1	nn	nn	nn	javno spoznan	Napovedovalka prometa vlakov na železniški postaji Maribor.
032Mn	7	1	nn	nn	nn	neznanec	Sprevodnik na vlaku.

I) ID govorca	12O) besede	XIO) prisoten v besedilih	VII) starost	VIII) govorno poreklo	IX) izobrazba	X) razmerje do 001Mz	0) ostalo
033Mn	4	1	nn	štaj.	nn	neznane	Sopotnik v istem razdelku na vlaku.
034Mn	5	1	nn	štaj.	nn	neznane	Sopotnik v istem razdelku na vlaku.
035Žn	5	1	nn	štaj.	nn	neznane	Sopotnica v istem razdelku na vlaku.
036Žn	3	1	nn	štaj.	nn	neznane	Sopotnica v istem razdelku na vlaku.
037Žn	3	1	nn	nn	nn	neznane	Sopotnica v istem kupeju na vlaku.
038Mn	3	1	nn	štaj.	nn	neznane	Sopotnik na vlaku.
039Žn	6	1	nn	nn	nn	neznane	Dekle v gneči na postaji.
040Ma	10	1	nn	nn	nn	neznane	Postajni napovednik (stalna obvestila za potnike) na Avtobusni postaji Ljubljana.
041Žj	6	1	nn	nn	nn	javno spoznan	Napovedovalka oddaje na TVSlo.
042Mj	16	1	nn	LJ	nn	javno spoznan	Napovedovalec svoje oddaje na TVSlo.
043Žj	203	2	nn	nn	nn	javno spoznan	Bralka reklamnega sporočila na TV.
044Žj	25	1	nn	nn	nn	javno spoznan	Bralka reklamnega sporočila na TV.
045Mj	28	1	nn	prim.	nn	javno spoznan	Športnik, član nogometne reprezentance, dajalec komentarja o nogometni tekmi.
046Mj	24	1	nn	nn	nn	javno spoznan	Športnik, član nogometne reprezentance, dajalec komentarja o nogometni tekmi.
047Mj	57	1	nn	nn	nn	javno spoznan	Športnik, član nogometne reprezentance, dajalec komentarja o nogometni tekmi.
048Mj	62	1	nn	nn	nn	javno spoznan	Športni novinar, avtor prispevka o nogometni tekmi za POPTV.
049Mj	13	1	nn	nn	nn	javno spoznan	Športnik, član nogometne reprezentance, dajalec komentarja v studiu za športno oddajo na POPTV.
050Mj	88	1	nn	nn	nn	javno spoznan	Športni novinar, voditelj športne oddaje na POPTV.
051Mj	43	1	nn	nn	nn	javno spoznan	Športnik, član nogometne reprezentance, dajalec komentarja v studiu za športno oddajo na POPTV.
052Mj	438	1	nn	nn	nn	javno spoznan	Bralec spremnega besedila v dokumentarcu na TV.
053Žj	250	2	nn	nn	nn	javno spoznan	Dajalka informacij v dokumentarcu, verjetno umetnostna zgodovinarica.
054Mj	55	1	nn	nn	nn	javno spoznan	Dajalec informacij v dokumentarcu na TV.
055Mj	29	1	nn	nn	nn	javno spoznan	Dajalec informacij v dokumentarcu na TV, župnik, igralec krajšega inserta.
056Vn	14	1	več	več	več	neznane	Petje ljudskih pevk v cerkvi.
057Žj	149	1	nn	nn	nn	javno spoznan	Bralka pisem, prevedenih iz nemščine, v dokumentarcu na TV.
058Mz	2.703	4	20-26	prim.	VŠUNI	prijatelj	Sostanovalec v sobi v študentskem stanovanju.
059Mz	819	5	27-39	kor.	SŠ	prijatelj	Sostanovalec v študentskem stanovanju.
060Žn	6	1	nn	nn	nn	neznane	Mimoidoča na ulici, ki vpraša za uro.
061Žn	30	1	nn	nn	nn	neznane	Prodajalka v trgovini pekarnice Vrhnik na železniški postaji Ljubljana.
062Mn	11	1	nn	nn	nn	neznane	Naključni kupec v trgovini pekarnice Vrhnik na

I) ID govorca	12O) besede	XIO) prisoten v besedilih	VII) starost	VIII) govorno poreklo	IX) izobrazba	X) razmerje do 001Mz	0) ostalo
							železniški postaji Ljubljana.
063Žn	9	1	nn	nn	nn	neznane	Prodajalka vozniških kart na železniški postaji Ljubljana.
064Mn	7	1	nn	nn	nn	neznane	Mimoidoči na Miklošičevi ulici v LJ, ki se pogovarja po telefonu.
065Žn	2	1	nn	nn	nn	neznane	Prodajalka v knjigarni pod hotelom Kompas v Ljubljani.
066Mn	6	1	nn	nn	nn	neznane	Prodajalec v trgovini Metalka na Ajdovščini v LJ.
067Žn	10	1	nn	nn	nn	neznane	Prodajalka v trgovini Nama v LJ.
068Žn	2	1	nn	nn	nn	neznane	Sopotnica v dvigalu v Nami v LJ.
069Mn	2	1	nn	nn	nn	neznane	Sopotnik v dvigalu v Nami v LJ.
070Mz	515	1	20-26	gor.	SŠ	znane	
071Žn	10	1	nn	nn	nn	neznane	Prodajalka pekarnice na osrednji tržnici v LJ.
072Žz	59	2	27-39	nn	VŠUNI	znane	Uradnica v mladinski pisarni.
073Mn	4	1	nn	nn	nn	neznane	Obiskovalec v mladinski pisarni.
074Mn	11	3	nn	nn	nn	neznane	Vratar v dijaškem domu; v 2 besedilih, enkrat v živo in enkrat po zvočniku.
075Mn	11	1	nn	nn	nn	neznane	Kuhar v menzi dijaškega doma.
076Mz	1	1	20-26	nn	SŠ	znane	Sostanovalec študent v isti hiši.
077Mz	49	1	20-26	dol.	SŠ	znane	Sostanovalec študent v isti hiši; spoznan šele med snemanjem.
078Mz	14	1	20-26	nn	SŠ	znane	Sostanovalec študent v isti hiši.
079Mn	76	1	nn	nn	nn	neznane	Bralec reklame na radiu Antena.
080Žn	20	1	nn	nn	nn	neznane	Bralka reklame na radiu Antena.
081Mn	9	1	nn	nn	nn	neznane	Bralec reklame na radiu Antena, glas je spačen.
082Mj	340	1	nn	nn	nn	neznane	Napovedovalec na radiu Antena.
083Žj	297	1	nn	nn	nn	neznane	Napovedovalka na radiu Antena.
084Žj	3	1	nn	nn	nn	neznane	Sodelujoča v studiju na radiu Antena.
085Žn	20	1	nn	nn	nn	neznane	Sodelujoča v oddaji na radiu Antena, vzdevek „Ta mala“.
086Žz	3.933	5	27-39	kor.	VŠUNI	družina	
087Ža	7	1	nn	nn	nn	neznane	Avtomatski odzivnik na telefonski centrali.
088Žj	147	1	nn	nn	nn	javno spoznan	Napovednik vlakov (po zvočniku) na žel. p. LJ.
089Ža	42	7	nn	nn	nn	neznane	Posnetki napovedi postaj na vlaku Siemens. ¹⁸²
090Mn	92	1	nn	nn	nn	neznane	Sopotnik na vlaku, ki se pogovarja po telefonu.
091Mn	28	2	nn	nn	nn	neznane	Sprevodnik na vlaku LJ-Senovo.
092Mj	25	1	nn	nn	nn	javno spoznan	Napovednik vlakov (po zvočniku) na žel. p. Zidani Most.
093Žz	1.273	1	20-26	štaj.	SŠ	znane	
094Mn	11	2	nn	nn	nn	neznane	Sprevodnik na vlaku Zidani Most-MB.
095Žn	5	1	nn	nn	nn	neznane	Sopotnica na vlaku, ki se jo med postankom na postaji Rimske Toplice.
096Nn	2	1	nn	nn	nn	neznane	Sopotnik, ki prisede na žel. p. CE.

¹⁸² Ta govorec (avtomat) ima v vseh besedilih 18) prostor: lokalni potniški vlak, 21) tema: naslednja postaja, 28) oblika sporazumevanja: napoved, 20) medij: zvočnik, 24) soudeležencev (naslovnikov): 100+, 25) zavest snemanja: da, 30) prostogovornost: navajanje, 31) socialno zvrst: zbor. gov., 32) namen sporazumevanja: INFO, 33) govorni položaj: javni, 34) funkcijska zvrst: PS.

I) ID govorca	12) O) besede	XI) O) prisoten v besedilih	VII) starost	VIII) govorno poreklo	IX) izobrazba	X) razmerje do 001Mz	0) ostalo
097Žn	10	1	nn	nn	nn	neznanec	Sopotnica na vlaku.
098Mn	89	1	nn	nn	nn	neznanec	Sopotnik na vlaku.
099Mj	21	1	nn	nn	nn	javno spoznan	Napovednik vlakov (po zvočniku) na žel. p. MB.
100Žz	896	4	nn	gor.	VŠUNI	znanec	Spoznana med snemanjem.
101Žz	553	1	27-39	nn	VŠUNI	znanec	Spoznana med snemanjem.
102Žz	390	1	27-39	dol.	nn	znanec	Spoznana med snemanjem.
103Žn	21	1	nn	gor.	nn	neznanec	

PRILOGA 3: ABECEDNI SEZNAM RAZLIČNIC KORPUSA

Naslednji abecedni sezname so narejeni iz korpusa, v katerem podatki govorcev v besedilih niso preimenovani; besedne zveze s podčrtajem niso razlomljene na posamezne besede in vsaka različnica razen imen in podatkov govorcev ima tudi pogostost pojavitve v korpusu. Za pomen ločil, ki so uporabljena poleg črk, gl. Tabela 13, str. 93, in interpretacijo v poglavju IV.14.1, Tabela 40. Sezname so po vrstnem redu, kot so navedeni, izbrani iz enotnega seznama različnic korpusa – nobena različnica ne nastopa v dveh seznamih. V drugi tabeli je navedeno, koliko različnic in pojavnic je v posameznem seznamu.

Posebej oz. drugače kot v transkripciji so v spodnjem seznamu označene besede, ki se v transkripciji nahajajo v oklepaju, kar pomeni, da je njihov zapis negotov: posamezni besedi sledi zaklepaj. Takih besednih je skupaj 342 pojavnic oz. 236 različnic.

KRATICE

*A ·2	*DESUS ·1	*ena_A ·2	*GEOSS ·1	*MIK ·5	*U ·1	*ZeeSKaeSe
*A_štiri ·1	*DESUS-a ·2	*eSDeeS ·2	*GEOSS-a ·2	*O ·1	*UeSBe-	S ·1
*B ·1	*De_dva ·2	*eSDeeS-	*IVe ·2	*OeMVe ·1	priključek ·	*ZeeSKaeSe
*Ce ·1	*DR ·3	ovca ·1	*JLA ·1	*PeVeCe-	2	S-a ·1
*CeDe ·2	*eL ·1	*eSeMeS ·6	*K ·1	okna ·1	*VD ·3	
*CeDe-jev ·1	*eLDeeS ·2	*eSeMeS-a ·1	*Ka ·4	*PeVeCe-	*VeCe ·1	
*CeDe-ju ·1	*eMPe_tri-	*eSeMeS-e ·1	*KR ·1	pohištva ·1	*VeCe-je ·1	
*CSOD-ju ·1	je ·1	*eSHaeS ·1	*M ·2	*TIMP ·3	*VeCe-ju ·1	

GOVORCI

Andreja	Barbaro	Franc	Lučka	Matejina	Rok	Vando
Andrej	Bečan	Furlan	Lučko	Mirijam	Simona	Dita
Anžėja	Borut	Katja	Marcela	Mojca	Simoni	*Leja
Anžejem	Breda	Katjo	Marcel	Neža	Simono	
Anžejovo	Edita	Lara	Marjeta	Nino	Tevža	
Anžej	Edita	Laro	Marjeta	Perneta	Tevž	
Barbara	Edita	Laura	Marko	Pero	Tina	
Barbari	Franca	Lovrenca_na	Mateja	Roka	Vanda	

DRUGA IMENA

Afriki ·2	Bela_krajina ·1	Cerkvi ·1	Dunaj ·2	Glavni_trg ·2
Afriko ·1	Belgije ·1	Cesar ·4	Dunajčan ·1	Gorazd ·4
Akademijo_za_glas	Beljaka ·3	Cesti_proletarskih_	Dunaju ·4	Gorenjci ·1
bo ·1	Benjamino ·2	brigad ·1	Dušanom ·1	Gorica ·7
Aleš ·6	Beograda ·1	Čeha ·1	Dva_bela_galeba ·2	Gorice ·5
Aleša ·1	Betnavska ·2	Čehi ·1	Elizabeta_Jakoš ·1	Gorici ·1
Aljaž ·1	Betnavski ·1	Černetič ·3	Elizej ·1	Goričkem ·1
Amerika ·1	Blagovici ·1	Černetiča ·1	Elizeja ·1	Gospod ·51
Ana ·2	Bog ·24	Črnega_morja ·1	Ema ·3	Gospoda ·5
Andraž ·3	Boga ·9	Črnomlja) ·2	Emo ·2	Gospodarskem_raz
Ane ·1	Boga_Očeta ·1	Danica ·1	Ervin_Šiljak ·1	stavišču ·1
Anine ·1	Bogu ·11	Dani) ·1	Evi ·1	Gospodov ·2
Ano ·3	Bolfenk ·1	Danka ·1	Evrope ·1	Gospodovem ·1
Antona ·1	Bolgarijo ·1	Danko ·1	Evropi ·3	Gospodovo ·1
Antonijo ·1	Boris ·4	Dan ·1	Evroval ·1	Gospodu ·9
Anton_Martin_Slo	Bosan'c ·2	Darje ·1	Fakultete_za_farma	Gradišniki ·1
mšek ·1	Boštjan ·3	Darujemo_Gospod ·	cijo ·1	Grčiji ·2
Anžėja ·1	Boštjana_Cesarja ·2	2	Ferenc ·1	Grdadolnik ·3
Arena ·3	Boštjan_Cesar ·2	Davidov ·1	Feri_Horvat ·3	Grdadolniki ·3
Arene ·1	Brane_Oblak ·1	David ·1	Filip ·3	Grosuplju ·1
Areni ·2	Braziliji ·1	Devica_Marija ·1	Filozofske ·2	Habakuk ·6
Arhivistika ·1	Bresternica ·1	Device ·2	Franc ·3	Habakuka ·3
Ars ·1	Brezočniki ·1	Devico ·4	Franca ·2	Habakukom ·2
Avstrija ·1	Brezočnikovega ·1	Dnevnik ·1	Francoški_simboliz	Habakuku ·4
Avstro-Ogrska ·1	Brezovica ·1	Dolenja_Trebuša ·1	em ·1	Hočem_s_teboj ·1
A_si_ti_tud'_noter	Brezovico ·1	Dolga_Gora ·1	Frucov ·1	Hozjan ·2
_padel ·1	Brezov'ca ·1	Dragice ·1	Fruška_gora ·1	Hozjana ·1
Bahon ·3	Brežice ·1	Drago ·1	Galilejo ·1	Hrastnik ·1
Bambi ·1	Celja ·2	Dravčami ·1	Gašperja_Terška ·1	Hrvaškem ·1
Barbare ·2	Celje ·10	Dravi ·3	Geršaka ·1	Istrabenz ·1
Barbari ·1	Celju ·5	Dravska_dolina ·1	Gerškovi ·1	Italija ·1
Barbarini ·1	Celovca ·3	Druge_knjige_kralje	Glavnega_trga ·2	Italijani ·3
Barbi ·1	Cerkev ·1	v ·1	Glavnem_trgu ·5	Italije ·1
Bastijan) ·1	Cerkve ·3	Duhu ·2	Glavnem_*trgi ·3	Italiji ·2

Italijo 2	Kreta 7	Materjo 2	O_Marija_prelepo_	Ruski_realizem 1
Ivana 2	Krete 1	Mati 1	bitje 1	Rušah 1
Ivanka 4	Kreto 2	Matjaž 1	Pahor 1	Ruše 2
Izlak 1	Krf 2	Matjažev 1	Palačinko 1	Ruto 3
Izlakah 1	Krfa 1	Matjažu 1	Pangerl 2	Samarijan 2
Izlakami 1	Krfu 7	Metliko 1	Partizanska 1	Samarijo 1
Izlake 4	Kristina 3	Meto_Rovškovo 1	Patmosa 1	Samo 2
Izraelove 1	Kristina_Kovačič 1	Miha 2	Pavla 2	Sarajeva 1
Izraelu 1	Kristus 4	Milan 2	Peci 2	Sašo 1
Jagnjetovo 1	Kristusa 1	Milana_Bizanta 1	Pekersko_gorco 1	Sašo_Popijal 1
Jagnje_božje 4	Kristusa_Jezusa 1	Mile_Ačimović 1	Pekrska_gorca 1	Saturnus 1
Jakca 1	Kristusom 2	Mimara 2	Peterlinovi 1	Savo 1
Jalovcu 1	Kristusov 1	Mirijam 1	Peter_Kovačič_Perš	Seničice 2
Janeza 4	Kristusova 1	Mitjem 1	in 1	Senovo 2
Janeza_Menarta 1	Kristusove 1	Mitrakoviča 1	Petro 1	Senožče 1
Jani 1	Kristusovega 2	Mlada_lipa 1	Pevski_zbor_in_zb	Silvana 1
Janja 3	Kristusovem 2	Mladinski_čeh 3	orovodstvo 1	Simona 4
Janjo 2	Kristusovo 1	Mobitel 1	Pistor 3	Simoni 1
Jankovič 4	Kristusu 5	Mojca 1	Pistorja 1	Sin 1
Jelovec 2	Krstnika 1	Mojca 1	Pistorje 1	Sina 8
Jeruzalem 2	Laško 2	Moj_ata_socialističn	Pistorjeva 1	Sinom 2
Jesenic 3	Lauri 1	i_kulak 1	Pistorjevih 1	Sinu 3
Jezus 9	Laze 1	Moli_in_delaj 1	Pistorjevim 1	Sirijec 1
Jezusa 2	Laznico 1	Monika 1	Pod_žarometom 1	Skavtič 1
Jezusa_Kristusa 7	Lehna 3	Monika_Kovačič 1	Pohorci 1	Skavtiči 1
Jezusi 1	Leo 1	Moniko_Kovačič 1	Pohorje 3	Slak 1
Jezusu 1	Lepa_in_zdrava 2	Moravč 1	Polja 1	Slavije 3
Jezusu_Kristusu 2	Lipici 2	Moravčam 1	Poljaki 1	Slomška 1
Jezus_Kristus 3	Lipnika 1	Moravčami 1	Poljane 1	Slomškova 1
Jordanu 1	Litije 1	Moravče 1	Polje 1	Slomškovem_trgu 1
Jožef 1	Litijo 2	Most 1	Polona_Rozman 1	Slomškov_trg 2
Jožefu 1	Litijo 1	Mravljaka 1	Poncijem_Pilatom	Slovence 1
Jožetu 2	Ljubljana 2	Muri 4	1	Slovencev 3
Jožica 2	Ljubljane 5	Murske 1	Poročilih 1	Slovenci 6
Jožici 1	Ljubljani 11	Murski_Soboti 4	Posavja 1	Slovenija 7
Jožico 2	Ljubljano 7	Mursko 1	Postojnčanka 1	Sloveniji 3
Jugoslavija 2	Ljubljan 1	Naaman 2	Postojne 1	Slovenske_novice 1
Jugoslavijo 1	Lovrenc 2	Nastje_Čeha 1	Pota 1	Slovenskih_novicah
Juču 2	Lovrenca 1	Nataša 2	Pozor_v_eter 1	1
Jurček 5	Lovrenca_na_Poho	Nedelo 2	Pragerskega 1	Smrekarjevo_Marjet
Jurčka 1	rju 1	Nedelu 1	Pragersko 3	ko 1
Jure 1	Lovrencu 6	Nemci 1	Prekmurja 3	Snežki 1
Jurija_Vege 1	Luč 1	Nemčija 1	Prekmurje 1	Sobote 1
Kardeljeva 3	Luči 1	Nemška_romantika	Preske 2	Soboto 1
Kardeljevi 2	Luka 4	1	Primor'c 1	Sonja 2
Karmen 1	Lukačeva 2	Nič_te_ne_moti 4	Primož 4	Stane 1
Karting_center_*Be	Luku 1	Nikolaja 1	Primoža 1	Stari_zavezi 1
TeCe 1	L'ubljan 1	Njegoseva 1	Proletarskih 1	Stični 2
Katr'ca 1	Madžari 1	Nogometne_zveze	Pušcave 1	Stično 1
Kidričevem_trgu 1	Majda 1	1	Pušnik've 1	Strgar 1
Kidričev_trg 1	Maks 3	Nova_vas 2	Pušnik'vo 1	Strgarju 1
Kitajci 1	Maksa 2	Nove 1	Radgoni 1	Struna 4
Knausa 1	Maksu 1	Novega_mesta 2	Radovljice 1	Struni 1
Knjižni_sejem 1	Maribor 11	Novi_Gorici 1	Radvanje 1	Strunjan 2
Kocbeka 1	Maribora 11	Novi_svet 1	Razdevšekovi 4	Strunjanu 1
Kogelniku 1	Mariborčani 2	Novi_zavezi 1	Razdevšek'vi 2	Struno 1
Koglera 1	Mariborske_koče 2	Novo_mesto 2	Rimske_Toplice 1	Stvarnikova 1
Kolenc 1	Mariboru 8	Nov'ga_mesta 1	Rimske_(Toplice) 1	Svetega_Duha 6
Kolodvorska 1	Marija 43	Obali 2	Ritem_Duha 1	Svetem_Duhu 3
Kolpi 1	Marija_ti_mati_si_	Oblak 2	Robič 1	Sveti_Duh 5
Koper 1	mo- 1	Obrul- 1	Robinsona 1	Svetozarevsko 1
Korana_Kovačič 1	Marije 2	Oče 16	Rojšek 2	Šveto_pismo 1
Koroška 1	Mariji 1	Očeta 8	Rokom 1	Šauperl 1
Koroško_cesto 1	Marijin 1	Očetom 2	Rolando 1	Šentjanža 1
Kožljeku 1	Marijine 1	Očetovi 1	Romi 1	Šentjernej 1
Kožljek- 1	Marijino 1	Očetu 2	Rop 2	Škotsko 1
Ko_rano_zjutraj 2	Marijo 3	Odrešenika 1	Rosande 1	Štajerc 1
Ko_so_pastirji 1	Mariša 1	Odrešenikova 1	Rosando 1	Štajerci 1
Krajncu 1	Marjetka 1	Oskar 8	Rotovška_klet 3	Štajersko 1
Kraljevina_Jugoslav	Marjetki 1	Oskarja 1	Rotovski 1	Štajer'c 1
ija 1	Markom 1	Oskarju 1	Rotovski_kleti 1	Štefan 1
Kraljica_venca_rož	Marmeljada 1	Oskar_Pistor 2	Rotovski_trg 1	Štihovi 1
nega 1	Marmeljade 1	Ostrožno 1	Rotovško_klet 1	Štore 1
Kraljica_venca_rož	Maša_Vesel 3	O_Marija 2	Rotovžu 1	Švedske 1
ne- 1	Matej 1	O_Marija_o_Marija	Rozi 1	Tabor 4
Kranjčani 1	Mateja 1	2	Rudolf_*Maister 1	Taboru 2
Krasu 2	Matej_Hozjan 1		Rudolfa_*Maistra 1	Tanja 4

Ta_mala 4	Urška 2	Zidanem_Mostu 1	*Hotižo 1	*Romija 1
Ta_malo 1	Urško 1	Zidani 1	*Intercity_Slovenija 1	*Sacrum_romanu m_*imperium_*nationis_*germaniae 1
Teharjah 1	Vačam 1	Zidani_Most 8	*Interspar_Vič 1	*Simon_Robinson 1
Tiger 6	Vače 3	Zinka 1	*Jolka_*Tatič 1	*Simon_*Robinson 1
Timotej 1	Valerija 3	Zojo 2	*Jožeja 1	*Sonijeve 1
Timoteju 1	Valerije 1	Zupančičevi 1	*Jožo 1	*Stankeja 1
Titov_most 4	Valerijo 1	Žiga 1	*Juršaja 1	*Studiju_*City 1
Titov_m_ 1	Vanka 1	Žiga_Nastran 1	*Juršaja 1	*Švetiga_Duha 3
Tomaju 1	Vanko 2	Žontar 1	*Lepe_Njive 1	*Šmelcervo 1
Tomaž_Repnikov 1	Vasko 1	*An 1	*Lihtegerja 1	*Spara 1
Tomaž_(Nemec) 1	Večera 2	*Animation_*Team 1	*Lovrenci 1	*Tatič 1
Tone 1	Vesna 1	*Baden_*Powell 1	*Lublano 1	*Tatič_*Jolka 1
Toneta 2	Vida 1	*Barbarove 1	*Mastru 1	*Televiziji 1
Tonija 2	Vilharjeva 1	*Božeja 1	*Maksi 1	*Tomasa 1
Topniško 1	Vojkovo 1	*Božica 1	*Marcello_*Lippi 2	*Trošter 1
Toti_Rotovž 1	Vrhonkovih 1	*Brazill 1	*Maribori 9	*Vanko 1
Transverzala 2	Vuzenica 2	*Brezočniki 1	*Milerja 1	*Volfruberva 1
Trbovljami 1	Vuzenice 1	*British_*Council 1	*Minhen 1	*Zagorica 1
Trbovlje 1	Vuzenico 3	*Bufonovo 1	*Mitrakoviča 1	*Zambrotta 1
Trebnjega 1	Vzpenjača 2	*Danka 1	*Münchna 2	*Zavrhi 1
Trgu_Borisa_Kidriča 1	Vzpenjačo 1	*Dankota 1	*Nečak 1	*Zimrajh 1
Trojan 1	Zagorice 1	*Dragejom 1	*Nov_mu_mestu 1	*Zimrajhovega_*Drageja 1
Trojanah 1	Zagorje 1	*D_R_ago 1	*Obruljove 1	*Zoja 1
Trojica 1	Zagreb 3	*Eko 1	*Obrul'ova 1	*Zojko 1
Tuhinjsko 1	Zalog 1	*Eurocity 2	*Rajkeju 1	*Z 1
Tuhinjsko_dolino 1	Zaloške 1	*Fernand_*Bernot 1	*Reichstag 1	
Tuhinj- 1	Zasavja 1	*Garneži 1	*Riata 1	
Tuš 1	Združena_lista 1	*Gianluca_*Zambrotta 1	*Riotu 1	
Uha 2	Združenja 1		*Robin 1	
Ulice 1	Združenje 1		*Roll_bar 1	
Umek 4	Zidanega_Mosta 3		*Roll_baru 1	

VEČBESEDNE ZVEZE

dvanajst_deset 1	sto_petdeset 2	tisoč_devetsto_deve	tisoč_osemsto_*šec	tristo_šestindvajset 1
dva_tisoč 2	sto_petindvajset 2	tindvajset 1	set 1	tristo_šestindvajset 1
dva_tisoč_dva 1	sto_tisoč 1	tisoč_devetsto_ose	tisoč_petsto 1	tristo_šestindvajset 1
enajsti_deseti 1	štirindvajset_ur_pika_*kom 1	mnajst 1	tisoč_sto_dvainšest	tristo_trideset 1
petsto_devetnajst 1	štiristo_trideset_(de	tisoč_osemsto_šest 1	deset 1	*tavžent_dvesto 1
petsto_petnajst 1	vetindvajset)_dev	tisoč_osemsto_*ena	tisoč_šestindvajset 1	-
petsto_šestnajst 2	etindvajset 1	inšeset 1	trinajst_petdeset 1	najststo_*os'nšt' r'deset_*šeset 1
sto_osemdeset 1			tristo_devetdeset 1	

OKRNJENE BESEDE

absolvent- 1	dvan- 1	la- 2	oble- 1	pre- 1	št- 1	vz- 2	*mhmhmh
absol- 1	dva_npe- 1	le- 1	ob- 3	pr ne- 1	š- 2	v- 9	m- 2
ab- 1	d- 7	l- 6	od- 1	pust- 1	tak- 2	zač- 1	*mhmhm-
aj- 1	edin- 1	mal- 1	omejit- 1	p- 14	ta- 5	zad- 1	1
akli- 1	enako- 1	mas- 1	omeji- 1	račun- 1	teb- 1	zau- 1	*mh- 1
alelu- 2	en- 4	me- 4	oni- 2	razum- 1	tempa- 1	za- 3	*mihoh- 1
anglešč- 1	e- 1	mhm- 1	on- 2	ra- 2	te- 1	zb- 1	*miho- 1
avtob- 1	fo- 1	misl- 1	opet- 1	rd- 1	toč- 1	zdr- 1	*mih- 1
av- 1	frančiškan- 1	mis- 1	op- 1	rek- 1	to- 3	zd- 2	*mi- 1
a- 5	gled- 1	mi- 1	osem- 1	re- 1	tr- 1	zg- 1	*m- 1
bom- 1	gobl- 1	modro- 1	osn- 1	roč- 1	tuk- 1	zmen- 1	*nananana
bo- 3	g- 1	bele 1	os- 1	r- 2	tu- 1	z- 9	na- 1
budil- 1	haba- 1	mo- 3	ozirom- 1	sez- 1	tvo- 1	ži- 1	*ratata- 1
b- 8	hor- 1	m- 18	ozir- 1	sk- 1	t- 15	'ma- 2	*ssve- 1
ci- 1	ho- 1	nast- 1	o- 4	sl- 1	ultraskle- 1	'me- 1	*stuč- 1
čak- 1	ja- 8	natip- 1	pa- 3	slu- 1	umazan- 1	*čeretc- 1	*taratata- 1
če- 1	je- 2	naum- 1	poča- 1	sm- 2	ume- 1	*ejo- 1	*ta- 1
či- 2	ju- 1	na- 5	podrobn- 1	sobot- 1	u- 1	*fejs- 1	*tritimtirir
čut- 1	ju- 1	nedostop- 1	pokazal- 1	sob- 3	vaš- 1	*ff- 1	i- 1
č- 1	j- 8	neizraelsk- 1	pol- 1	soš- 1	vč- 1	*lalalalalala	*tot- 1
dec- 1	kak- 1	1	posku- 1	spodr- 1	veličast- 1	lala- 2	*t- 1
dej- 1	kam- 1	neprim- 1	posk- 1	sred- 1	ve- 3	*lalalalalala	*zaintiran-
de- 1	ka- 3	nim- 1	pozad- 1	sre- 1	vid- 1	- 3	1
dne- 1	ke- 2	ni- 1	pozn- 1	str- 1	vin- 1	*lala- 3	*zunajšnj-
dob- 1	kopa- 1	nje- 1	po- 3	st- 3	vo- 4	*la- 1	1
dos- 1	ko- 2	nn- 1	preb- 1	svetlo- 1	vsemogoč- 1	*ma- 1	*-bro 1
do- 2	kraja- 1	nova- 1	pred- 1	š- 26	1	*mel- 1	-buja 1
dru- 1	kup- 1	no- 1	pres- 1	študi- 1	vska- 1	*mhmhmh	*-če 1
dvanajst- 1	k- 7	n- 20	preve- 1	št' r- 1	vznemi- 1	mhm- 1	-et 1

*magari 2	*mmhmmm 1	*niši 1	*pečet 2	*properties 1	*stariga 1	*text 1
*mah 1	*mmhm 11	*ni'sn 1	*peeset 1	*provocirat 1	*starsem 1	*te) 1
*maja 1	*mmh 1	*ni 2	*pejd 2	1	*stinstal 1	*te 167
*maji 1	*mmmha 1	*nja 3	*pej 1	*prov 1	*stolčt 1	*tharam 1
*majl 1	*mmmhm m 1	*njiju 1	*per 1	*prvo 1	*storžeke 1	*that 1
*majn 14	*mmmhmh	*nonstop 1	*pesa 1	*pravle 1	*stroji 1	*the 2
*majster 1	*mmm 1	*novi 2	*petkret 1	*pr'su 3	*stuckali 1	*this 1
*malokakšna 1	*mm 27	*nucala 1	*photograph 1	*pšpšpš 2	*sunce 1	*tič 1
*mal' 1	*mobitela 1	*nucalo 1	*piksno 1	*pš 1	*system 1	*tirajoče 1
*mantrali 1	*mobitel 3	*nuca 1	*piskerči 1	*pubecov 1	*s'ndeset 1	*tiririm 1
*man 1	*modliš 1	*obadva 1	*piskerčom 1	*pucal 2	*s'n 338	*tiririti 1
*marem 4	*mogel 14	*obenoma 1	*piš 1	*pucat 1	*s 1	*tirititum 1
*mare 8	*mogla 12	*običnih 1	*pje 3	*puf 1	*šalčke 1	*tismu 1
*marmo 1	*mogle 1	*obično 1	*placa 6	*puno 2	*seeset 2	*tistiga 1
*marzlo 1	*mogli 12	*obidveh 1	*placu 1	*radijator 3	*šeri 1	*tjararam 1
*mar'jo 1	*moglo 1	*obodvojno 1	*plac 2	*radiju 1	*šetiga 1	*toke 1
*mar'mo 1	*mogo 4	1	*plačuvat 1	*rajnglo 1	*šetundvajš't i) 1	*tokiga 1
*mar'm 3	*molo 2	*očese 3	*please 1	*rečt 6		*tolarjov 1
*mar's 1	*moramo 1	*očes 1	*pleh 1	*reč mo 1	*šlambre 1	*tolčt 1
*mar'ta 1	*morem 1	*odkapal 2	*pobojnih 1	*redi 69	*šnicle 1	*tole 2
*masa 2	*more 1	*odparla 1	*pocukraš 1	*rek 3	*šnicli 1	*tota 37
*masovke 3	*mor'ma 2	*odrabljene 1	*počajte 1	*restat 1	*šniclov 1	*toteğa) 1
*mater 1	*mor'mo 1	1	*počaj 2	*re'i 4	*šnicl 1	*toteğa 10
*maturanc 1	*mor's 2	*ofnaj 3	*podni 2	*rihtat 1	*špad 1	*totem 3
*maunt 3	*mosta 1	*ofnl 1	*podstrešji 1	*rrahm 1	*špajs 1	*tote 17
*ma'em 1	*mo 7	*of 1	*pogledajmo 1	*rufni 1	*šparat 1	*totiga 1
*ma 1	*mrkhm 1	*oglas'ma 1	1	*rukzak 1	*špilaj 1	*totih 3
*mein 1	*muti 1	*ohtat 1	*pojstlo 2	*ru'zake 1	*špilala 1	*totile 1
*melšpajs 1	*my 1	*oh 2	*pojst'li 2	*saj 1	*špilali 1	*totim 2
*mesari 1	*m 15	*ojeja 1	*pojte 1	*salote 1	*špilal 1	*toti 50
*mesenc 3	*nadriblal) 1	*ojej 1	*poj 5	*saloto 1	*špilam 1	*totole 2
*mesidž 1	*naenkret 2	*okej 1	*pokrampam 1	*sam 6	*špuknilo 1	*toto 83
*mesti 2	*nagle 1	*okna 1	1	*sed'nsto 1	*ššš 2	*to'ih 1
*meš'cov 1	*nahecali 1	*okne 2	*polji 1	*sed'n 6	*ššt 4	*to i 2
*meša 1	*nahm) 1	*okni 1	*pol) 1	*see 3	*šš 1	*to'o 2
*meše 1	*nahm 3	*okno 1	*pol 179	*sej 3	*štikel 2	*to 6
*meši 4	*naja 1	*oniga 2	*pomaj 1	*selavi 1	*štirji 1	*train 1
*mešna 1	*najboljši 2	*on 1	*pometnega 1	*se 1	*štosno 1	*trgi 2
*mešno 2	*najdel 2	*oo 1	1	*se 2	*studijom 1	*trijimi 1
*mešo 4	*najdli 1	*opasno 1	*ponorjenih) 1	*sfalil 1	*štuklala 1	*triji 2
*me 1	*najdl 1	*opeto 1	1	*sfolgala 1	*št'r npe'ese t 1	*trof'la 1
*mhkrghm 2	*najj 6	*opucali 1	*ponucali 1	*sigurno 3		*trola 1
*mhmmhmm	*najmanjši 1	*orehemi 1	*popka 1	*sila 1	*švoh 3	*tss 1
hmmhmm	*naklačeno 1	*orehi 1	*poplozaš 1	*skapal 1	*tacajt 1	*tš 1
hmm 4	*nam 2	*organiziram a 1	*popu'li 1	*skapon 1	*tac'ga 1	*tule 1
*mhmmhmm	*nana 1	*osen 5	*porajtala 1	*skavtuj 1	*taizejska 2	*tutu 7
hmmhmm	*napakira 1	*os 2	*porihtal 1	*skop 1	*taizejsko 2	*tvaja 1
2	*napihkala 1	*os najstega 2	*poslal 1	*skoro 2	*taj 1	*tvaji 1
*mhmmhmm	*naraste 1	*os'ndeset 1	*potelefonira mo 1	*skor 9	*takiga 2	*type 2
hm 3	*naredo 1	*os'n 5		*skoz 2	*taki 8	*t 1
*mhmmhmm	*narete 1	*otročke 1	*povedeti 1	*skup 4	*takou 1	*uah 1
6	*nareto 2	*ova 8	*povedet 1	*sma) 1	*tak 1	*ubitačno 1
*mhmmmm 2	*naret 8	*ovi 1	*požike 1	*sma 38	*tananata 1	*uf 6
	*naret) 1	*o 1	*po 1	*smotane 1	*taplem 1	*uhuhu 1
*mhmm 8	*naret 2	*pajem 1	*prejšn 1	*smo 2	*taplo 1	*ultraskleroz a 1
*mhmm 2	*nateknila 1	*pajes 2	*prejšn 3	*smrkaš 1	*tararam 1	*una 1
*mhmm 1	*natokal 1	*pajeta 1	*prekel 1	*some 1	m 1	*und 2
*mhm 12	*navadne 1	*paje 2	*previsaka 1	*sori 2	*targat 1	*unem 2
*mičkeni 1	*navzgur 1	*pajša 1	*previsako 1	*sosedi 1	*tatatatatahm 1	*unc 2
*mičken 2	*nazodno 1	*pajštva 3	*pridema 1	*so 2	1	*unih 1
*midve 1	*na'a 1	*pakel 1	*pridet 1	*space 1	*tatatatatata 1	*unimi 2
*mihoho 31	*necem 1	*pale 1	*primeri 1	*spodi 1	*tatatatata 2	*uni 2
*mijadva 3	*neki 4	*pal 69	*primlo 1	*spomeniki 1	*tatatatata 2	*un' 3
*mijav 1	*nemo 2	*papapapapa papapapa papapapa pa 1	*prim'lo 1	*spotekneš 1		*un 1
*mija 1	*nene 3		*printer 1	*spredi 1	*taum 1	*upss 1
*mimimi 1	*nes 17		*print 1	*sprobali 1	*tavžent 5	*ups 2
*minhensko 1	*ne'a 7		*prislušimo 1	*sprobat 1	*ta 4	*uštimal 1
	*ne'e 23			*sprobavlem 1	*tc 30	*uuuu 1
*minidisk 1	*ne 1	*paper 2	*probajmo 4	1	*teh 1	*uuu 2
*miri 1	*nice 1	*parajt 2	*probajte 1	*spucat 2	*tej 5	*uuvovo 1
*misl'la 1	*nih 1	*parkplac 1	*probaj 1	*spučem 1	*teko 15	*uu 4
*mladi 1	*nisen 30	*parvi 1	*probal 1	*spumpala 1	*tek 4	*uvav 1
*mmhmmhmm	*nisma 3	*pasa 1	*probamo 1	*spumpal 1	*tenka 1	*u 4
mmhmm 1	*nis'n) 1	*pavre 1	*probam 1	*spumpa 1	*termoflis 1	*vade 1
	*nis'n 20	*pa 1	*probata 1	*ss 3		

*vado ·1	*velnes ·1	*vračaval ·1	*vun ·31	*zaglihano ·1	*zmarzjene ·1	*zun ·3
*vaji ·6	*vel'ki ·1	*vrate ·2	*vzdržovat' ·1	*zagrob'lo ·1	*zmarzнем ·1	*zutra ·9
*valda ·4	*verjemem ·1	*vremeni ·1	1	*zajtrki ·1	1	*z'ej ·1
*varhka ·1	*verjeme ·1	*vrhi ·1	*vzemeš ·2	*zapaješ ·1	*zmerom ·2	*žeganje ·1
*varhniga ·1	*ves ·1	*vrhka ·1	*vzemetē ·1	*zapel'et ·1	*zmer ·6	*žganke ·1
*vau ·1	*ve'te ·1	*vrti ·1	*vzeme ·1	*zastopfalo ·1	*zminka ·3	*žgečo ·3
*vav ·1	*vic ·1	*vsaki ·3	*wellness ·1	*zbori ·2	*zmodlili ·2	*župi ·1
*včer ·1	*vijadva ·6	*vsak'kret ·1	*yes ·1	*zdej ·2	*zmučkano ·1	*žvancat' ·1
*včosih ·1	*vijadve ·1	*vsok'kret ·2	*you ·3	*zdevli ·1	*znajdel ·1	*ž ·1
*večerko ·1	*vinogradi ·2	*vstola ·1	*začetki ·1	*zgira ·3	*zobeke ·1	*'ajde ·1
*večer ·1	*virhank ·1	*vstran ·1	*zadi ·2	*zgir ·3	*zoo ·1	*'lih ·10
*večki ·1	*visako ·1	*všeno ·1	*zadnjokret ·1	*zglihali ·1	*zos ·3	*'mama ·1
*večkret ·1	*viš ·2	*vtarga ·2	1	*zgužvani ·1	*zramal ·1	*'ni ·4
*veda ·2	*vjutro ·11	*vtek ·jen ·1	*zadn ·1	*ziher ·9	*zrihtali ·1	*'no ·1
*vejo ·3	*vkup ·1	*vtihni ·1	*zad ·1	*zluštalo ·1	*zunajšnje ·1	
*veliki ·1	*vlaki ·2	*vunta ·3	*zaglihali ·1	*zmantran ·1	*zunam ·1	

PREOSTALE BESEDE

absolutizem ·1	arhivih ·1	balonu ·1	bitki ·1	boljš' ·4	brizat ·1	copati ·1
absolutnih ·1	arhivistika ·3	bal ·2	bit' ·28	bolj ·83	britofa ·1	cuker ·1
absolventa ·2	arhivistiko ·2	banano ·1	bit ·1	bolnika ·1	brke ·5	cvetela ·3
absolvente ·2	arhivom ·2	banka ·1	bivšega ·1	bolnikov ·1	brki ·3	cvetelo ·1
absolventske ·m ·2	arhivskega ·1	barve ·1	bivši ·2	bolno ·1	brošure ·1	cvetijo ·1
absolventske ·1	arhiv ·1	barvne ·2	bivšo ·2	bombeto ·1	bršljana ·1	cveti ·1
absolventske ·3	aristokratska ·1	barvni ·1	bij ·5	bomo ·107	bršljan ·2	cvetje ·1
absolvent ·1	1	barvno ·1	bi ·463	bom ·192	brundal ·1	cvetju ·2
absurd ·1	armirana ·1	barvo ·1	blagajničarka ·1	bonbon ·2	bundo ·1	cvetove ·2
adijo ·43	asfaltu ·1	bar ·1	blagajni ·1	borovnicami ·1	burek ·3	cvetovi ·1
aerobike ·1	asfalt ·1	basala ·1	blagajno ·1	1	burjo ·1	cvetov ·1
aha ·113	asistenta ·1	basi ·1	blagor ·2	borovnice ·1	buteljko ·1	cvet ·1
ah) ·1	asistent ·1	bas ·1	blagoslavljam ·1	bosta ·3	b'la) ·4	čaja ·1
ah ·18	astro ·2	baterije ·2	blagoslovljen ·1	bošte ·5	b'la ·98	čaj ·6
aja ·118	ata ·21	beatifikacije ·1	blagoslovi ·2	boš ·83	b'le ·11	čakaj ·9
ajda ·1	ati ·1	1	blagoslovljen ·1	boter ·1	b'li) ·1	čakala ·2
ajde ·1	avgusta ·4	bedela ·1	blagoslovom ·1	botrom ·1	b'li ·31	čakali ·2
ajdove ·1	avstrijski ·1	bedet' ·1	1	bova ·5	b'lo) ·1	čakalnici ·1
ajdov ·1	avstrijsko ·2	bedne ·1	blagovest ·1	božic ·3	b'lo ·153	čakalo ·1
akademija ·1	avta ·1	bela ·1	blazno ·3	božja ·5	b' ·5	čakal ·2
akademiji ·2	avte ·1	belega ·1	blaži ·1	božjega ·3	cela ·4	čakam ·3
akademijo ·2	avti ·1	bele ·2	bleščijo ·3	božjemu ·1	celega ·1	čakate ·2
akademijski ·1	avtobusa ·1	beli ·3	blešči ·1	božje ·2	celem ·1	čakat' ·5
akcija ·1	avtobusi ·3	belo) ·1	blizu ·10	božji ·3	celih ·1	čakat ·4
aktualno ·1	avtobusna ·2	belo ·2	blizu ·10	božjo ·6	celino ·1	čaka ·3
albumu ·1	avtobusne ·1	bencina ·2	bliz ·1	bo'ita ·1	celi ·6	čao ·1
aleluja ·106	avtobusni ·2	berilo ·4	bliža ·1	bo'ta ·12	celotnega ·1	čari ·1
aleluje ·1	avtobusom ·4	berilu ·1	bližini ·3	bo'te ·12	celo ·43	čas ·10
alelujo ·5	avtobusov ·1	besedah ·2	bližje ·2	bo) ·2	celuloza ·1	časih ·1
ali ·14	avtobusu ·1	beseda ·6	blodimo ·1	bo ·262	cel' ·3	časov ·1
alta ·1	avtobusu ·15	besede ·1	bloki ·4	brada ·1	cel ·12	častitljiva ·1
alte ·1	avtocesta ·1	besedic ·1	bluzijo ·1	bradica ·1	cenjene ·1	časti ·1
alti ·4	avtoceste ·1	besedila ·2	boban ·1	brado ·5	cenjeni ·1	čast ·2
altov ·1	avtocesti ·1	besedilo ·1	bodimo ·1	brala ·1	center ·3	času ·3
alt ·10	avtomatsko ·1	besedil ·1	bodisi ·1	bral ·3	centimetrov ·1	čas ·2
al) ·1	1	besedi ·3	bodi ·7	brate ·1	1	čase ·1
al' ·268	avtomobile ·1	besedne ·1	bodoče ·1	bratje ·7	centru ·2	čedalje ·1
amaterje ·1	avtom ·2	besednih ·1	bodočnost ·1	brat' ·1	cepn'lo ·1	čedno ·1
amen ·24	avtom ·13	besedo ·2	bodo ·10	brat ·6	cerkev ·3	čepet' ·1
ampak) ·2	avtu) ·1	besed ·2	bođ'te ·1	bravo ·7	cerkve ·3	čepprav ·20
analize ·1	a) ·2	betnavskega ·1	bođ' ·1	brega ·2	cerkvi ·4	čeri ·1
analiziral ·1	a ·397	1	bođ ·1	brezmadežna ·1	cesarski ·1	česa ·3
angele ·2	baba ·1	betnavskem ·1	bogat ·1	1	cesarstvo ·3	čestitamo ·1
angeli ·4	babica ·3	1	bogoslov'c ·2	brezmadežno ·1	cesar ·1	češčena ·1
angelov ·2	babice ·2	beton ·3	bogovom ·1	1	cestah ·2	češ ·2
angelski ·1	babici ·2	bicikli ·1	bogovem ·1	brezplačne ·1	cesta) ·1	četi ·1
angleščino ·2	bagrom ·1	bife ·3	bogveka' ·1	brezveze ·1	cesta ·2	četrt ·1
animatorja ·2	bajadere ·1	bila ·15	bojum ·2	brez ·12	ceste ·2	četrt ·10
apartmaji ·2	baje ·4	bili ·8	boju ·1	bridkosti ·1	cesti ·5	četrt ·3
apostola ·1	bajtam ·1	bilo ·12	boju ·1	bridkost ·2	cesto ·9	četrti ·2
apostolom ·1	bajta ·2	bil) ·1	boleča ·1	bridko ·1	cevi ·14	četrtka ·1
apostolov ·1	bajte ·3	bil ·159	bolečine ·1	briga ·1	cev ·20	četrtkih ·2
apostolsko ·1	bajti ·1	biser ·1	bolezen ·1	brijem ·2	cicibanček ·5	četrt ·1
	bajto ·1	bistva ·1	bolezen ·1	brjesh ·2	čigani ·3	četrt ·5
	baj' ·3	bistveno ·1	bolje ·3	bril ·1	čilji ·1	čevapčiči ·2
	balkonu ·1	bistvu ·17	boljše ·5	brinjevec ·2	cingljalo ·1	čez ·67
	balone ·1	biti ·5	boljši ·1	brisače ·1	cisterna ·4	če) ·1
					cisterno ·3	če ·325

čik ·1	darove ·2	desnico ·1	dobrega ·4	dosti ·20	dr'gje ·1	edinosti ·1
čim ·5	darovi ·1	desni ·1	dobre ·2	dost'krat' ·1	duhom ·4	edino ·6
čista ·2	darujemo ·4	desno ·8	dobrih ·1	dost' ·13	duhovne ·1	edin' ·1
čistega ·1	dar ·1	destinacija ·1	dobrina ·1	dotiki ·1	duhovnih ·2	eh) ·1
čistim ·2	daš ·9	deščicami ·1	dobrine ·2	dovod ·2	duhovnike ·1	eh ·15
čisti ·2	datume ·1	dete ·2	dobri ·8	dovoli ·1	duhovnikom ·1	ej ·27
čisto ·11	datum ·1	det'la ·1	dobrodelneg ·1	dovolj ·5	1	ekipa ·2
čist' ·48	dat' ·4	devetdeseteg ·1	a ·1	dovoz ·1	duhovnik ·1	ekipe ·1
čist ·4	dava ·1	a ·1	dobrote ·1	dovrš ·1	duhovni ·1	ekipi ·1
članke ·1	da'ta ·3	devetdeset ·2	dobroto ·1	doživi ·1	duhu ·1	ekipo ·1
človeka ·5	da'te ·1	devetih ·18	dobro) ·2	do) ·1	dunajska ·1	ekonomijo ·1
človeku ·1	da) ·3	devetinpetde ·1	dobro ·59	do ·93	dunajskimi ·1	ekonomsko ·1
človek ·9	da ·876	seti ·1	dobr' ·5	draga ·1	dunajski ·3	1
človeške ·1	debela ·1	devetintrides ·1	dob'la ·1	drage ·1	duša ·1	ekskurzija ·1
človeško ·2	debelega ·1	et ·1	dob'li ·5	dragi ·3	duše ·1	eksplozija ·1
čohala ·1	debelina ·1	deveti ·2	dob'r' ·2	dragocena ·1	duši ·3	elegantno ·1
čohat' ·1	debeli ·3	devetletke ·1	dob'r ·1	dragonec ·1	dušo ·5	elektrarna ·1
čokolado ·2	debelo ·2	devetnajstega ·1	docent ·1	drag ·1	dvaindvajseti ·1	elektriko ·1
čorba ·1	deci ·1	1	dodatne ·1	dramili ·1	h ·1	elementi ·1
črnilom ·1	dečki ·1	deveto ·3	dodatni ·1	drekarija ·2	dvainosemde ·1	enaindvajset ·1
črno ·1	dedek ·3	devetsto ·1	dodatno ·4	drekarije ·2	seti ·2	1
črpala ·1	dediščino ·1	devet ·12	dodelat' ·1	drekarijo ·1	dvainosemde ·1	enaintrideset ·1
črta ·1	ded ·1	devica ·6	dogaja ·2	drek ·3	set ·3	1
črte ·1	definitivno ·1	devic ·2	dogorijo ·9	drobne ·1	dvainsedem ·1	enajstih ·1
črtil ·1	dejanju ·2	deviška ·1	dojenček ·1	drobno ·1	eseti ·1	enajst ·2
črva ·1	dejansko ·3	dežela ·1	dokler ·6	drugačen ·1	dvainšestdese ·1	enake ·2
čuden ·1	dejavnosti ·2	dežele ·1	dokončno ·2	drugače) ·1	t ·1	enakomerno ·1
čudna ·1	dekan ·1	deželi ·1	dokumentar' ·1	drugače ·3	dvainštirides ·1	2
čudno ·12	deklamirat ·1	deževnem ·1	c ·1	drugačne ·1	t ·3	enakovreden ·1
čudn' ·1	dekle ·1	dežja ·3	dokumente ·1	drugačno ·2	dvajsetem ·1	1
čudovita ·2	dekllice ·1	dežju ·1	dokupil' ·1	drugač' ·7	dvajseti ·1	enako ·3
čujejo ·1	deklino ·1	dežnik ·2	dolenjski ·1	drugam ·4	dvajset ·2	enak ·2
čujem ·3	delajo ·4	dež ·10	dolge ·4	druga ·6	dvajs' t ·9	ena ·45
čuješ ·2	delajte ·1	didaktikah ·1	dolgi ·2	drugega ·13	dvakrat ·3	enega ·5
čuje ·5	delaj ·1	didaktiki ·1	dolgoročno ·1	drugemu ·2	dvanajstih ·4	enemu ·1
čujta ·1	delala ·2	digitalca ·1	1	drugem ·6	dvanajsto ·1	enem ·3
čujte ·1	delali ·10	dijaško ·1	dolgo) ·1	druge ·7	dvanajst' ga ·1	energije ·1
čuj ·16	delal ·7	diktatura ·1	dolgo ·21	drugič ·7	dvanajst ·8	energijo ·1
čukov ·2	delamo ·1	dilema ·2	dolg ·2	drugih ·1	dva' ndvajste ·1	ene) ·1
čula ·5	delam ·1	dimenzije ·1	določ ·1	drugim ·3	ga ·1	ene ·15
čuli ·2	delaš ·3	diploma ·1	določ ·1	drugi ·36	dva' nosemde ·1	enih ·12
čulo ·1	delat' ·9	diplome ·1	določ'la ·1	drugie ·5	seta ·1	enim ·6
čul ·5	delat ·2	diplomiral ·1	določ'li ·1	drugo ·19	dva' nosemde ·1	eni ·18
čutil' ·1	delavnica ·1	diplo ·1	dolžan ·1	drug'mu ·1	sti ·1	enkrat ·27
čutim ·4	delavno ·1	direkten ·1	dolžine ·1	drug' ·1	dva ·58	enk'at ·2
čutiš ·1	dela ·17	direktni ·1	dolžnikom ·4	drug ·8	dveh ·1	enodnevno ·2
čuti ·1	deležni ·3	direktno ·7	dolž ·2	druščini ·1	dvema ·1	enoglasna ·1
daje ·3	delež ·2	direkt ·7	določ ·1	društvo ·2	dvesto ·1	enostavno ·7
dajmo ·1	dele ·2	dirigente ·1	dol) ·1	družabnih ·1	dve) ·1	enoto ·1
dajo ·2	delijo ·1	dirigentu ·1	dol) ·59	družbe ·1	dve ·36	eno ·67
dajte ·3	delila ·1	dirigent ·1	domače ·2	družinah ·1	dvigalo ·1	en'ga) ·1
daj ·31	delili ·1	dirigiranje ·1	domačije ·1	družina ·4	dvigneš ·1	en'ga ·14
dala) ·2	delil ·1	dirigirat' ·2	domačijo ·1	družine ·2	dvignil ·1	en'mu ·6
dala ·13	deliš ·1	diši ·1	domačinko ·1	družini ·1	dvig) ·1	en' ·3
daleč ·6	delno ·1	divjaki ·1	domačin ·1	družino ·1	dvojčka ·6	en) ·1
dali ·10	delom ·1	dlaka ·2	domačo ·1	družinske ·1	dvojčke ·2	en ·92
dalje ·8	delovalo ·1	dlje ·1	domač' ·1	družin ·1	dvoje ·1	eter ·1
dalo ·9	delovnih ·1	dneh ·3	doma) ·1	drva ·2	dvojko ·1	evangelija ·2
dal'č ·2	delovnik ·1	dneva ·1	doma) ·52	držat' ·1	dvojna ·1	evangeliju ·1
dal' ·3	delo ·8	dneve ·1	domov ·19	državah ·1	dvojno ·1	evangelij ·1
dal ·26	delu ·4	dnevi ·1	domo' ·8	država ·1	dvome ·1	evangeljskem ·1
damo ·3	del ·11	dnevni ·3	domo ·1	države ·1	dvorane ·1	1
dam ·16	denarja ·2	dnevu ·2	domu ·1	državicah ·1	dvorani ·1	evo ·9
današnja ·1	denarn'co ·3	dni ·44	dom ·2	državljanstev ·1	dvorano ·2	evropski ·2
današnje ·2	denar ·3	dober ·44	dopade ·2	državna ·1	dvorišče ·1	faksu ·1
današnji ·1	desetih ·12	dobijo ·3	dopisovala ·1	državnega ·1	dvorišču ·1	fakture ·2
danes ·22	desetimi ·1	dobila ·8	dopoldan ·4	državne ·2	2	fakulteta ·1
danica ·1	deseti ·1	dobili ·7	dopolodne ·4	državnih ·1	2	fakultete ·1
dan's ·27	desetko ·1	dobil' ·1	dopolnitvi ·1	državni ·1	džungli) ·1	fanta ·1
dan ·57	desetletja ·1	dobil ·10	dopolni ·1	držav ·1	d'narja ·1	fantje ·1
darilo ·3	desetminutn ·1	dobimo ·2	dopovedoval ·1	držav ·1	d'narje ·1	fantov ·2
daritev ·3	o ·3	dobim ·2	a ·1	držav ·1	d'nar ·1	farmacije ·1
daritve ·4	deseto ·2	dobiš ·1	doseči ·2	držiš ·1	eden ·17	farmacijo ·1
daritvi ·1	deset ·25	dobite ·1	dosegla ·1	drži ·2	edina ·1	farmacijo ·1
daroval ·1	desko ·1	dobit' ·3	dosegli ·1	dr'gačen ·1	edini ·6	februarja ·2
darovanje ·1	desnica ·1	dobi ·2	dosegljiva ·1	dr'gač ·23	edinoxrojeneg ·1	februarski ·1
darovano ·1	desnici ·1	dobra ·4	doslej ·1	dr'gam ·2	2	fiksnega ·1
						film ·1

fino 2	glede 2	greje 1	hozana 13	in 495	izvoli 2	jurjev 2
finte 1	glejte 1	gremo 31	hrana 1	iskali 2	izvoljenemu 1	jutranja 1
firma 1	glej 6	grem 33	hranil 1	iskal 4	1	jutranjo 1
fis 1	gle'al 4	grenko 1	hrastovi 1	iskanje 1	izvoljenih 1	jutri 21
fižola 1	gle'as 3	grešen 1	hrbet 2	iskat' 1	izvoljeni 1	jutro 5
fižoli 1	gle'at 4	grešila 1	hrepnenja 1	iskat 2	izvol' te 2	jutr' 4
fižol 2	globoka 1	grešil 1	hrib 1	iskreno 3	izvora 1	južne 1
folkloro 3	gmota 1	grešniki 1	hripava 1	ista 2	izvrstno 1	južnih 1
folkoro 1	gnetlo 1	grešnikov 1	hripav 2	istem) 1	izza 1	južni 1
formularje 1	gnojšnica 1	grešnost 2	huda 2	istem 1	iz 105	južno 2
formulo 1	gnojšnico 1	greš 25	hudega 6	isti) 1	jabolk 1	ju) 2
fotka 1	gobavca 1	greva 4	hudi 1	isti 1	jagal 1	ju 2
fotoaparat 1	gobavcev 2	greznica 8	hudodelec 1	isto 8	jah 1	kačo 1
fotografije 1	gobavce 1	greznice 7	hudo 4	iščejo 1	ajca 1	kadarkoli 1
fotografijo 1	gobavih 1	greznici 1	hujša 1	iščes 1	fajce 3	kadar 14
fotokopiranje 3	gobavosti 1	greznico) 1	hujši 1	itak 12	jajc 1	kadi 1
3	gobavost 1	greznico 11	huligani 1	italijanski 3	jakna 2	kafe 1
frajerja 1	gobe 3	grezn'ca 3	hvala) 2	iti' 1	jakno 2	kajti 7
francoski 2	gobice 3	grezn'ce 1	hvala 96	iti 6	jami 1	kajzarice 1
francoščine 1	gobic 1	grezn'co 2	hvaležen 1	it' 18	januarja 1	kaj) 3
franciškani 2	gobovo 1	gre'ta 1	hvaležna 1	izberite 2	jarek 2	kaj 254
franciškanom 2	gokarta 1	gre'te 1	hvaležne 1	izbiraš 1	jasno 2	kakava 2
2	gola 1	gre 65	hvaležni 2	izbire 1	jašek 12	jakav 3
franciškanski 1	golimi 1	grize 6	hvaležnosti 2	izbrala 3	jaška 2	kaka) 1
1	golobi 1	grizlja 1	hvaležnost 2	izbrali 1	jaški 1	kaka 12
funkcije 1	golu 1	grob 1	hvale 1	izbrane 2	javna 1	kakega 5
furmani 1	gol 5	groza 3	hvalo 2	izdali 1	javnih 1	kakemu 1
gala 2	gondola 1	grozdje 1	h'tela 2	izdal 1	jaza 1	kakem 3
garali 1	gondolo 1	grozd 2	h'teli 1	izginejo 3	jaz) 2	kake 10
garaži 1	gorice 2	grozilo 1	h'telo 1	izgled 1	jaz 363	kakih 9
garažnih 1	gorici 2	grozno 4	h'tel 7	izgovarjaš 1	ja) 5	kakimi 1
garažo 1	gori 4	gumbek 1	h 1	izgradnjo 1	ja 1175	kaki 8
gardi 1	goro 2	gumb 1	idealno 1	izgubili 3	jedel 2	kakorkoli 2
gaz'li 1	gor 33	hali 1	ideja 1	izgubi 1	jedi 2	kakorkol' 2
ga) 1	gospa 2	halo 8	ideje 1	izgubljen 1	jedla 7	kakor 18
ga 94	gospodoval 1	hamburger 1	identitete 1	izhaja 1	jedli 2	kako 53
generacije 1	gospod 1	ha 1	iger 1	izhlapi 1	jed 4	kakšen 4
generalnem 1	gostega 1	ha 1	ignacijanske 1	izhoda 1	jejte 1	kakšna 1
gladek 1	hecal 1	hčerko 2	1	izhodiščna 1	jel' 2	kakšnega 1
gladino 3	gostilne 1	hecam 1	igrajo 1	izhodu 1	jemljejo 1	kakšne 2
glasba 1	gostilni 1	hecna 1	igralcev 2	izhod 10	jemlji 1	kakšni 1
glasbeniki 1	gostim 1	hecna 1	igralce 2	izjave 2	jemo 1	kak) 1
glasbeni 1	gostov 1	hec 1	igralci 2	izjemoma 1	jem 4	kak' 100
glasbe 1	gosto 6	hiša 6	igrali 2	izkazal 1	jeruzalemske 1	kak 17
glasbi 1	gost 1	hiše 10	igral 1	izkaznico 1	1	kamenje 1
glase 1	gotovo 2	hiši 3	igra 2	izkoristili 1	jeseni 1	kamen 1
glasilce 1	govorijo 1	hišo 10	igre 1	izkoristil 2	jesenskega 1	kame'je 1
glasilke 2	govorila 1	hiš 1	igrišče 2	izkušena 1	jesenske 1	kamne 2
glasno 2	govorili 1	hitih 1	igro 5	izmeček 1	jesti 2	kamnokreč 2
glasom 2	govoril' 1	hitro 12	igr'ca 1	izmed 1	jest' 8	kam 9
glasovih 1	govoril 6	hja 1	igr'co 1	izmeni 1	jest 1	kanalizacija 2
glasovi 1	govoriš 1	hkrati 7	igr'c 1	izmenjalo 1	ješ 4	kanalizacije 1
glasov 1	govorite 1	hlače 1	imajo 4	izmišljotina 1	jetnikom 1	kanalizacijo 5
glas 12	govori 3	hladilnik 2	imamo 4	izneverimo 1	jezen 2	kanalu 1
glava 4	govorjenega 1	hladilno 1	imaš 5	izneveriti 1	jeze 1	kanal 2
glave 2	1	hladno 3	ima 2	izognit' 1	jezika 1	kandidirali 1
glavi 1	govorjenje 1	hm 10	imela 5	izpise 1	jeziki 1	kante 1
glavna 1	govorjenju 1	hočemo 1	imel 7	izpite 4	jeziku 1	kanto 1
glavnem) 1	govorne 1	hočem 18	imenoval 1	izpit 2	jezuitski 1	kapetan 1
glavnem 11	govornih 1	hočeš 4	imenu 8	izpljune 1	jezuit 1	kaplan 1
glavni 2	govor' t' 1	hoče 3	imen) 1	izpolnil 1	jez 1	kapljice 1
glavno 2	govor 1	hoč'jo 1	imeti 1	izpolni 1	je) 16	karieri 1
glavo 7	gozdu 1	hodijo 3	imet' 1	izrednega 1	je 1656	karkoli 2
glazutah 1	grabi 1	hodil 8	ime 10	izročil 1	jih 63	karkol' 1
gledala 3	gradiva 1	hodim 2	informacije 1	izstopa 1	jim 17	karte 2
gledalci 1	gradu 1	hodit' 3	instant 1	izstopi 2	ji 19	kartica 1
gledališčem 1	grah 1	hodit 1	institucije 1	izstop 1	joj 13	kartice 2
gledališču 2	grajali 1	hodi 7	institucij 1	izsuši 1	jopič 2	kartico 1
gledali 4	grame 1	hod'jo 1	inšpektor 1	izšel 1	jo' 3	kar 97
gledalke 1	gramski 2	hod'la 6	inštrumenti 1	izšla 1	jo) 1	kasneje 2
gledal 8	granit 1	hod'li 4	intenzivne 1	iztekel 1	jo 52	kasnejša 1
gledam 1	grdo 1	hoje 1	intenzivnih 1	izven 1	ju) 2	kasnej' 1
gledaš 6	greha 5	hojla) 1	interesantno 1	izvidniške 1	juho 2	kaši 1
gledat' 2	grehe 6	hojo 1	1	izviralo 1	juh'ca 1	kašljamo 1
gledat) 1	grehov 2	hotelov 1	interneta 1	izvira 1	julij 6	kataloge 1
gledat 1	grehu 1	hotelu 1	internetu 1	izvirnega 1	julij 1	katedre 1
gleda 1	greh 1	hotel 10	in) 2	izvir 1	jurja 4	

katerega ·1	knjige ·3	korenčkovi ·1	kuh'nji ·1	lepo ·30	lopov ·3	maša ·5
katerem ·3	knjigi ·1	korenčkov ·2	kuh'n'i ·2	lepše ·1	lovimo ·2	maše ·1
kateri ·2	književnost ·2	korenje ·1	kulturnem ·2	lepš' ·1	lovrenškega ·2	maše ·6
katerih ·3	knjiž'ce ·1	korenju ·2	kulturne ·2	lep ·7	lovrenški ·3	mašo ·2
katerim ·1	kod' ·1	koristno ·1	kulturni ·1	lesarji ·1	luči ·2	materialov ·1
kateri ·2	kod ·5	korita ·1	kulturo ·1	leseno ·1	lučka ·1	materini ·1
katero ·2	koga ·3	koroško ·1	kumar'ce ·1	lesom ·2	lučke ·1	mati ·13
katoliško ·1	koledar ·2	kosa ·1	kupeju ·1	lestvico ·1	lučko ·1	matraš ·1
kavč ·2	kolega ·5	kosila ·1	kupil ·4	les ·2	luč ·5	maturantka ·1
kavo ·2	kolegice ·3	kosilo ·9	kupimo ·2	lešnikovo ·1	luknj ·4	maturantske ·1
kazalo ·1	kolegij ·1	kosilu ·1	kupim ·1	letake ·2	luknje ·7	maturantski ·2
kaže ·3	kolegi ·4	kosil ·1	kupit' ·4	letaki ·1	luknjo) ·1	ma ·15
ka'sen ·3	kolegom ·1	kostanjev ·2	kupit' ·4	letak ·1	luknjo ·6	medenih ·2
ka'sna ·2	kolegu ·1	kostanje ·6	kupi ·1	leta ·11	luksuz ·1	med ·26
ka'sne ·1	kolesih ·1	kostanj ·1	kupovala ·1	letih ·4	luštna ·1	megla ·1
ka'sni ·2	kolesi ·1	kostan'e ·1	kupu ·1	letijo ·1	luštni ·1	megle ·3
ka'sno ·1	kolesom ·1	kosta'je ·2	kup'la ·2	leti ·8	luštno ·2	meĝlo ·1
ka' ·165	kolikokrat ·1	koš ·3	kup'li ·1	letnih ·1	l' ·1	meĝniki ·1
kdaj ·31	koliko ·1	košarke ·1	kup ·7	letniku ·5	madeža ·1	mejo ·1
kda' ·12	kol'kor ·1	košati ·1	kura ·1	letnik ·3	magari ·1	melodija ·3
kdor ·1	kol'ko ·11	koščene ·1	kurba ·1	letom ·2	mah ·1	menda) ·3
kdo) ·2	kol'k'r ·2	košta ·1	kurilno ·1	letos ·13	maja ·2	menda ·26
kdo ·29	kol'k') ·1	koš ·1	kuriš ·1	letošnje ·1	majhna ·2	mene ·15
kegljanja ·1	kol'k' ·22	kotu ·2	kurit' ·1	letovala ·1	majhnega ·1	menil ·1
kegljanje ·2	kol'k ·2	kot ·45	kurjave ·1	letoviščarji ·1	majhnih ·1	menite ·1
kelih ·2	komadov ·1	kotarce ·3	kuža ·8	letoviščarski ·1	majhni ·1	meni ·17
kelnarji ·1	komaj ·3	ko' ·54	kvalifikacijah ·1	leto ·20	maj ·1	menjal ·2
ker) ·2	koma' ·1	ko) ·3	kvalifikacije ·1	letu ·1	maj ·1	menjavajo ·1
ker ·170	komentatorji ·1	ko ·189	kvalitete ·1	let ·11	maksimimum ·1	menjavo ·1
kili) ·1	komentira ·1	kraji ·1	kvalitetna ·1	levo ·8	mala ·3	menoj ·3
kilogramov ·1	komisije ·1	krajše ·1	kvaliteto ·1	leže ·4	malega ·3	meno' ·2
kilometrov ·2	komisiji ·1	kraj ·7	kvečjemu ·1	leži ·1	malemu ·1	men'la ·3
kil ·2	komot' ·1	kraljestvo ·7	kvišku ·1	le ·24	malem ·1	men'li ·3
kiosku ·2	kompatkne ·1	kraljestvu ·2	k'ero ·1	lica ·1	malenkost ·1	meri ·2
kiosk' ·1	komplifikacij ·1	kraljica ·13	k'r ·14	licih ·1	male ·5	mero ·1
kip ·2	kompost ·2	kraljuješ ·2	k'ra ·14	liniji ·1	malih ·1	mesa ·2
kište ·1	kompozicija ·1	kraljuje ·2	k'ra ·1	lisaste ·1	malimi ·3	meseca ·1
kitara ·1	komu ·5	kralj ·1	k'terega ·1	listat' ·1	mali ·8	mesec ·8
kitaro ·3	komu' ·5	krasni ·1	k'teregi ·1	listek ·1	malokdaj ·1	mesecini ·2
kitica ·2	konca ·6	kratek ·2	k' ·5	liste ·9	malokje ·1	meso ·4
kitico ·3	koncertirala ·1	krati ·1	k ·40	lističa ·1	malokje ·1	mesta ·2
ki) ·1	koncert ·1	kratka ·2	labilno ·1	listi ·1	mal'ci ·2	mestna ·1
ki ·115	koncev ·2	kratko ·2	lačen ·4	listke ·1	mal'co ·3	mestne ·2
kjerkoli ·1	konce ·1	kregat' ·1	lačna ·4	listki ·1	mal'krat ·1	mestni ·1
kjer ·9	koncu ·6	krijejo ·1	ladjah ·1	listkov ·1	mal) ·3	mestno ·2
kje' ·1	koncu' ·1	kriko ·1	ladja ·4	listov ·1	mal' ·32	mestni ·1
kje ·78	končala ·2	kristjanov ·1	ladjo ·2	listu ·1	mama ·15	mestno ·2
klancih ·1	končala ·2	kritika ·1	lahke ·1	litanije ·3	mamica ·1	mestno ·2
klanci ·1	končal) ·1	krivo ·1	lahko ·117	literarna ·2	mamino ·1	mesto ·11
klasična ·2	končal) ·1	kriza ·2	lahk) ·1	literarno ·1	mamin ·1	mestu ·2
klasično ·1	končal) ·2	križan ·1	lahk' ·94	literaturo ·1	mami ·1	mest ·1
klasiki ·1	končani ·1	križišče ·3	lahk' ·94	litrov ·1	mandate ·1	meš'ca ·1
klavirja ·1	konča' ·1	križ ·6	lahk' ·94	liturgična ·1	mandati ·1	mešal ·1
klavne ·1	končno ·2	kri ·9	lahk' ·94	ljubeče ·1	mandatov ·1	mešano ·1
klepetoš ·1	kondenz ·1	kroj ·2	lahk' ·94	ljubezen ·13	mandatu ·1	meša ·1
klet ·2	kondicijo ·1	kroji ·2	lahk' ·94	ljubezniva ·1	manjka ·1	meter ·6
klicala ·7	kondominij ·2	krompir ·2	lahk' ·94	ljubezni ·8	manjkrat ·1	metre ·1
klicali ·1	konec ·1	krsti ·1	lahk' ·94	ljubica ·2	manjši ·1	metrov ·4
klicali ·16	konec ·1	krst ·2	lahk' ·94	ljubila ·2	manjši ·1	mezga ·1
klicari ·1	konec ·1	kršćeni ·2	lahk' ·94	ljubila ·2	manj ·6	mežikala ·1
klicat' ·1	kongresom ·1	kruha ·5	lahk' ·94	ljubili ·2	mano ·3	mežika ·2
kličemo ·2	kontrola ·1	kruhovec ·1	lahk' ·94	ljubim ·2	mape ·1	mežiš' ·1
kličeva ·1	kontroliran ·1	kruh ·15	lahk' ·94	ljubi ·2	mapo ·2	me'jat' ·1
kličice ·1	kon'c ·9	kruta ·1	lahk' ·94	ljub ·1	marele ·1	me) ·2
kličimo ·1	kopali ·3	krvi ·1	lahk' ·94	ljudi ·7	marelo ·4	me ·60
kličimo ·1	kopalnici ·2	krvjo ·2	lahk' ·94	ljudje ·4	mariborskih ·1	mhm ·194
ključ ·3	kopat' ·2	kubika ·1	lahk' ·94	ljudmi ·1	1	midva ·7
kljuka) ·1	kopirala ·2	kubik ·1	lahk' ·94	ljudska ·1	marmelado ·1	midve) ·1
kljuko ·2	kopiral ·1	kugo ·1	lahk' ·94	ljudskih ·1	marmor ·1	midve ·6
klobase ·1	kopiramo ·1	kuhala ·1	lahk' ·94	ljudski ·1	maršikoga ·1	miga ·4
klobaso ·1	kopirat ·1	kuhali ·1	lahk' ·94	ljudsko ·1	martinovanje ·2	mika ·1
kmečki ·1	kopira ·1	kuhal ·2	lahk' ·94	ljudstvo ·1	2	milijonov ·4
kmečko ·1	korajžni ·1	kuhamo ·1	lahk' ·94	ljudstvu ·1	mar ·2	milosti ·2
kmeti ·1	korajžo ·1	kuhana ·1	lahk' ·94	logiko ·1	masa ·1	milostjiva ·1
knezi ·1	korenčkova ·1	kuhat ·1	lahk' ·94	lokalni ·4	maso ·1	milostljiva ·1
knezov ·1	1		lahk' ·94	lokomotiva ·1	masne ·1	milostno ·2
knjigarna ·1			lahk' ·94	lonaj ·3		

nimate 1	nosila 10	bojestranski 2	odpirajo 2	oktobra 3	or'ginalno 1	oženit' 1
nima 11	nosil 6	o 2	odpira 1	oktobrška 1	osamelec 1	oživlja 1
nisem 28	nosim 1	oboje 1	odpočijte) 1	oktobru 1	osebe 1	ožji 1
nisi 23	nosit' 1	obračat' 1	odpovedal 1	okupatorski 1	osebno 1	o 73
nismo 16	nostalgija 1	obračunu 1	odpovem 1	1	osemdesetih 1	pač) 1
niso 36	nos'la 1	obrani 1	odpravi 5	okus 5	1	pač 110
nista 4	nos'li 1	obrat 1	odpremo 1	okvirje 1	osemdeset 3	padavine 1
niste 8	noter) 2	obraz 2	odprem 1	okvir 1	osemindevet 1	padavin 1
nisva 2	noter 45	obresti 1	odpret' 1	olajšalo 1	deset 1	padec 1
nis' 2	note 4	obribalo 1	odpre 1	olimpijskih 1	osemindvajset 1	padel 1
nit 9	notranja 1	obril 1	odprla 1	olje 1	t 1	padel 1
nivoju 2	not) 1	obrišeš 1	odprli 2	oltar 1	oseminšestdeset 1	padla 1
nivo 1	not 43	obrneš 1	odprl 2	omadeževan 1	set 1	padli 3
nizka 2	not(nova) 1	obrnil 1	odprt 1	o 1	osemnajsti 2	paket 2
nizki 2	nova 6	obrnit' 2	odpusti 5	omako 1	osemsto 15	pakiral 1
nizko 1	novega 5	obrokov 3	odpuščamo 4	omari 1	osem 6	palučinko 1
nižja 2	novjši 1	obsedeno 1	odpuščanja 1	omaro 1	osišče 1	pametne 1
nižje 4	novembra 4	obstajata 1	odpuščanje 2	omejenost 1	osmega 1	pametno 2
niž'e 1	novem 2	obstaja 2	odraščal 1	omejitev 1	osmih 18	pamet 3
ni'amo 1	nove 4	obvestila 1	odrašal 1	omogočajo 1	osmi 3	panike 1
ni'am 4	novih 3	obvestilo 1	odrešenja 3	onadva 1	osnovi 1	panonski 1
ni) 3	novinarji 1	obvestit' 2	odrezal 1	ona 32	osnovni 3	papeža 1
ni 259	novi 5	obveznosti 1	odstotke 1	onega 4	osnovnošolci 1	papirja 11
njega 7	novogradnja 1	obvezno 1	odstotkov 2	onemu 2	1	papirjem 1
njegova 2	1	obvoza 1	odstranili 1	onem 2	ostajal 1	papirjev 1
njegovega 2	novoletna 1	obvozih 1	odtehtalo 1	one 9	ostajam 1	papirje 3
njegovemu 2	novoletni 1	obzov 1	odtehalo 2	onidve 1	ostala 8	papirja 1
njegovem 1	novomeški 1	obžalujem 1	odtoka 3	onih 6	ostale 4	papirnica 1
njegove 1	novost 2	obžalujmo 1	odtok 7	onimi 1	ostalih 1	papir 16
njegovo 5	novo 12	ob) 1	odveč 1	onim 4	ostali 1	paradajz 1
njegov 4	nov'ga 1	ob 120	odvijanje) 1	oni 45	ostali 5	paralelka 1
njej 1	nov' 1	očakov 1	odvisni 1	ono 25	ostal 5	paralelki 1
njemu 6	nov 2	očeh 1	odvisno 3	on' 1	ostaneš 2	pardon 4
njem 3	no) 2	očem 1	odvrni 1	on 26	ostanete 1	parka 1
njena 2	no 170	očetovski 1	odzadaj 1	on 26	ostane 6	parkirali 1
njene 1	nujno 1	oče 6	od 118	opazil 2	ostane 6	parkiral' 1
njenih 1	nula 4	očičenih 1	oglarji 1	opazimo 1	ostani 2	parkiral 1
njeni 2	n'č 2	očičeni 1	oglasila 1	opazi 1	ostati 4	parkirani 1
njeno 3	n'č 2	očičen 1	oglasil 1	opazovalne 1	ostat' 1	parkiraš 1
njen'mu 1	obadva 3	očitek 1	oglasno 1	opaz'la 2	osvetliš 1	parkirate 1
njen 8	obarva 1	očitno 6	ogledalu 1	opazam 2	osvetljeno 1	parkrat 1
nje 2	oba 5	oči 1	oglje 1	opereto 1	osvetljen' 1	park 1
njihova 2	občestvo 1	odbor 1	ognji 9	opere 1	oteti 1	paro 1
njihovi 1	občestvu 2	odcepa 1	ogoljufani 1	opete 2	otoka 1	par 8
njih 6	občudoval 1	odcepa 1	ogorke 2	opet 2	otok 2	pasalo 1
niju 1	občudujem 1	odcep 1	ogrejte 1	opevamo 1	otoci 8	paše 7
njima 1	občutka 1	oddaje 1	ogrodja 1	opevat' 1	otroke 1	paštete 1
njimi 4	občutljiva 2	oddaji 1	ogromnega 1	opira 1	otroke 3	pašteto 1
njim 6	obdobje 3	oddajo 1	ogromno 4	opogumil 1	otroki 1	patent 1
njivo 2	obeh 3	odgovarjat' 1	ogrski 2	opoješ 1	otrok 8	pater 1
njo 5	obema 2	odgovarja 1	ohranila 1	opoldne 1	otroške 1	pavza 1
nobeden 3	obesiš 1	odgovoril 1	oh) 1	opozarjal' 1	otroškimi 1	pavzo 1
nobel 1	obe 1	oddidemo 1	oh 4	opozorilo 1	otroški 2	pazijo 1
nobena 1	obhajali 2	odigrali 1	ojej 2	opravičujem 1	otroško 2	pazili 1
nobenega 2	obhajilo 2	odjmlješ 3	okej) 1	o 2	otvoritev 1	pazite 1
nobene 2	običajno 4	odjemlje 1	okej 97	opravičujem 1	ovčarja 1	pazit' 1
nobenim 1	obira 1	odklonil 1	okence 1	1	oven'la 1	pazi 8
noben'ga 3	obisk 3	odkopljem 2	okensko 1	opravilo 1	ovratnik 1	pa) 6
noben 3	obiščete 1	odkrh'jen 1	oklicev 1	opravljeno 1	ozdravil 1	pa 1691
nocoj 1	obiščite 1	odkritju 1	okna 7	oprostite 1	ozdravi 1	pa 6
nočem) 1	objavlja 1	odlična 1	okni 1	oprost' 1	ozdravljena 3	pecivo 3
noče 2	objemal 1	odlični 1	okno 2	opustila 1	ozdravljeneg 1	pečene 2
noči 2	objemi 3	odlično 2	okno) 1	opustil 2	a 1	pečenje 1
nočna 1	objeta 1	odlikuje 1	okno 5	orehe 1	ozdravljenih 3	pečenko 2
noč) 1	oblasti 1	odločili 1	oknu 1	1	3	pekli 4
noč 21	oblast 1	odločil 4	okolici 1	organizatorji 1	ozdravljeni 2	peklo 1
noge 1	oblečena 1	odločitev 1	okolico 1	1	ozdravljen 1	pelargonije 2
nogometna 1	oblekel 1	odločite 1	okol' 7	1	ozirali 1	pela 8
nogometne 1	oblekla 1	odločit' 3	okorajž'la 1	orglala 1	ozziroma 19	pele 3
nogomet 6	obleko 1	odlomku 1	okovjem 1	orglale 1	ozki 1	peli 13
nog'ce 1	obletnica 1	odložen 1	okraj 2	orglam 1	oznanjajo 1	peljala 1
nohte 1	obletnice 1	odmrznili 1	okrajih 1	orglat' 1	oznanjal 2	peljale) 1
noht 1	oblaki 1	odnese 1	okraju 1	orientacijo 1	oznanjamo 1	peljale 1
nonstop 2	oblikovat' 1	odpadne 1	okrog 5	1	1	peljali 5
normalen 1	obnašal 1	odpeljal) 1	oktavo 1	original 2	ozri 2	peljal) 1
normalno) 1	obnovil 1	odpelje 4	oktober 4	orto 1	ožarjena 1	peljal) 9
normalno 5		odpel'ala 1				peljat' 1

povemo ·1	praznimi ·1	prekidal ·1	preusmerjene ·1	priključka ·2	pri ·39	prvotno ·1
povem ·3	prazni ·2	prekine ·1	1	priključ ·1	probaj ·2	prvo ·14
poveš ·1	prazno ·3	prekinitvah ·1	preveč ·24	priklopa ·1	probala ·1	pr'bližno ·4
povezано ·1	praženec ·1	prekleta ·1	preverim ·1	priklopili ·1	probali ·1	pr'čakujem ·1
povezavo ·1	pražnje ·1	preko ·5	preverjanje ·1	priklop ·1	probali ·4	pr'jat'1 a ·1
pove'al ·2	prebavo ·2	prekratka ·1	previdnost ·1	priklop ·6	probal ·3	pr'javil' ·1
pove'te ·1	preberemo ·1	prekratko ·1	previ's'ko ·1	prilagodit' ·1	probam ·1	pr'javil' ·1
pove ·2	preberem ·1	prekriva ·1	prevladuje ·1	prilika ·1	probate ·1	pr'javljen' ·1
povoha ·2	preberi ·2	prek' ·3	prevozom ·1	priliki ·1	probat' ·1	pr'jazen ·1
povsod ·2	prebiva ·2	prek ·5	prevoz ·1	priložnosti ·2	probat' ·1	pr'jet' ·1
povšeči ·1	prebiva ·2	prelivat' ·1	prevoz ·1	priložnost ·1	problema ·4	pr'letel ·1
povzdignjen ·1	preblašeno ·1	preliva ·2	prevoznik ·1	primarno ·3	probleme ·2	pr'leti ·1
pozabil' ·1	prebrala ·1	premagala ·1	prevoznik ·1	primerjalna ·2	problemov ·1	pr'nesejo ·1
pozabil ·6	prebrali ·2	premagane ·1	prevoznik ·1	primerjalno ·1	problem ·16	pr'nesel ·6
pozabim ·2	prebrat' ·1	premala ·1	prevoznik ·1	primerna ·1	procentov ·2	pr'nesem ·1
pozabiš ·1	prebudila ·1	premalo ·1	prevoznik ·1	primerno ·2	prodekan ·1	pr'neseš ·1
pozabljeni ·1	prebudim ·3	premalo ·3	prevoznik ·1	primeru ·3	profesionalci ·1	pr'nese ·1
pozab'la ·10	prebujam ·4	premika ·1	prevoznik ·1	primer ·3	1	pr'nesla ·2
pozanimal ·1	prebujam ·3	premislili ·1	prevoznik ·1	prinašam ·1	profesorja ·1	pr'nesli ·1
pozdrave ·2	precej ·2	premislit' ·2	prevoznik ·1	prinesejo ·1	proga ·2	pr'nesti ·1
pozdravi ·4	prečista ·1	premisli ·1	prevoznik ·1	prinesel ·3	progo ·1	pr'nest' ·4
pozdravlja ·5	prečka ·2	premražena ·1	prevoznik ·1	prinese ·1	programe ·1	pr'nest ·1
pozdravljena ·1	prečko ·1	1	prevoznik ·1	prinesi ·1	program ·3	pr'peljala ·1
pozdravljeni ·2	predajva ·1	prenašam ·2	prevoznik ·1	prinesli ·1	proizvajalce ·1	pr'peljali ·3
pozdravljen ·5	predal' ·1	prenašas' ·1	prevoznik ·1	printala ·1	proizvodnja ·1	pr'peljal' ·1
pozdrav'la ·1	predal' ·1	prenavlja ·1	prevoznik ·1	printanje ·1	prometa ·1	pr'pel'ala ·1
pozdrav'te ·1	predavanja ·10	prenavljaj ·1	prevoznik ·1	printerju ·1	promet ·1	pr'pel'ali ·3
pozdrav' ·2	predavanje ·4	prenovljeni ·1	prevoznik ·1	printer ·5	proniciljiv ·1	pr'pel'al ·1
pozdrav ·1	predavanj ·3	prepevajte ·1	prevoznik ·1	prpada ·2	prosil ·2	pr'pognem ·1
pozen ·2	predbeli ·2	prepiha ·1	prevoznik ·1	prpeljala ·1	prosimo) ·1	pr'pravil' ·1
pozicije ·1	predel ·1	prepih ·1	prevoznik ·1	prpeljali ·1	prosimo ·14	pr'pravljajo ·2
pozimi ·4	predel ·1	prepozno ·1	prevoznik ·1	prpeljalo ·1	prosim ·50	pr'pravljat' ·1
pozirala ·1	predenj ·1	preprečujo ·1	prevoznik ·1	prpeljal ·4	prosite ·2	pr'pravljen ·1
pozirala ·1	predenj ·1	1	prevoznik ·1	prpomb ·1	prosi ·50	pr'prav'la ·2
poznalo ·1	predlagam ·1	prepričana ·2	prevoznik ·1	prporočajo ·1	proslavil ·1	pr'prav'li ·1
poznal ·1	predlagat' ·1	prepričan ·1	prevoznik ·1	prporočamo ·3	proslavimo ·2	pr'šel ·3
poznam ·19	predlog ·3	preprost ·1	prevoznik ·1	prporočilo ·1	prostega ·1	pr'šla ·2
poznaš ·10	predmeta ·1	prepustilo ·1	prevoznik ·1	prpovedova ·1	prosti ·1	pr'šle ·2
pozna ·2	predmete ·1	prerežem ·1	prevoznik ·1	nja ·1	prostora ·5	pr'šli ·1
pozneje ·1	predmetni ·1	prerivali ·1	prevoznik ·1	prprave ·3	prostorov ·2	pr'šlo ·1
pozno ·8	predniku ·1	prerodim ·1	prevoznik ·1	prpravljeni ·2	prostor ·6	pr'tisneš ·1
pozorni ·1	predno ·1	prerokih ·2	prevoznik ·1	prpravljenost ·1	prostost ·1	pr'tisnilo ·1
pozorno ·1	predpisih ·1	prerokov ·1	prevoznik ·1	1	prosto ·2	pr'vošč'la ·2
požegnaj ·2	predprejšnji ·1	presadila ·2	prevoznik ·1	prprirošnje ·1	prosveti ·1	pr'žgala ·1
poženeš ·1	predrli ·1	presadili ·1	prevoznik ·1	prpreditev ·1	prošnje ·1	pr'žgan ·1
požene ·1	predrli ·1	presadil ·1	prevoznik ·1	prrisije ·1	prošnje ·1	pr) ·2
požirek ·1	predsednik ·4	presadiš ·1	prevoznik ·1	prrisiljen ·1	prošnje ·1	pr' ·61
požulil ·1	predstavitev ·1	preselili ·1	prevoznik ·1	prsluhnim ·1	prošnje ·1	psa ·2
po ·142	predstavitev ·1	presenečen ·1	prevoznik ·1	prsmrdelo) ·1	prošnje ·1	pučes' ·1
prakse ·2	predstavljaj ·1	presenetil ·1	prevoznik ·1	prisotni ·1	prošnje ·1	pukel ·1
prakso ·1	predstavljala ·1	presenet'lo ·1	prevoznik ·1	prisotnosti ·1	prošnje ·1	puli ·1
praktično ·3	predstavljali ·1	presnemaval ·1	prevoznik ·1	prispel ·1	prošnje ·1	pulover ·2
prat' ·2	predstavljaj ·1	1	prevoznik ·1	prisrčno ·2	prošnje ·1	pumpat' ·1
prava ·1	predstojnik ·1	prespat' ·1	prevoznik ·1	pristane ·1	prošnje ·1	punce ·1
pravega ·2	predvidevas' ·1	prestal ·1	prevoznik ·1	prišel ·27	prošnje ·1	punc ·6
pravice ·1	predvid ·1	prestari ·1	prevoznik ·1	prišla ·11	prošnje ·1	punčko ·1
pravičnost ·1	predvsem ·5	prestavljali ·1	prevoznik ·1	prišle) ·1	prošnje ·1	puranje ·1
pravično ·2	predzadnja ·1	prestavljaj ·1	prevoznik ·1	prišle ·2	prošnje ·1	puranji ·1
pravijo ·1	pred ·47	prestavljali ·1	prevoznik ·1	prišli ·20	prošnje ·1	pustila ·1
pravil ·8	pregledal ·1	prestop ·1	prevoznik ·1	prišlo ·2	prošnje ·1	pustili ·3
pravim ·5	pregledne ·1	prestop ·1	prevoznik ·1	pritegnila ·1	prošnje ·1	pustil ·5
praviš ·2	prehodno ·1	presvetega ·1	prevoznik ·1	pritegnilo ·1	prošnje ·1	pustit' ·2
pravit' ·1	prehod ·2	presveto ·1	prevoznik ·1	pritiska ·1	prošnje ·1	pusti ·5
pravi ·18	prehuda ·1	preštudirat' ·1	prevoznik ·1	prit' ·5	prošnje ·1	pust' ·1
pravljica ·1	prejela ·1	1	prevoznik ·1	privlačno ·2	prošnje ·1	puščal ·1
pravokotno ·3	prejšnjem ·2	prešvicali ·1	prevoznik ·1	privlekli ·1	prošnje ·1	pr ·3
pravo ·4	prejšnjih ·1	preteklih ·3	prevoznik ·1	prizadeta ·1	prošnje ·1	rabijo ·1
pravzaprav ·5	prejšnji ·7	preteklosti ·1	prevoznik ·1	prizadetost ·1	prošnje ·1	rabil ·1
prav' ·2	prejšnji ·3	pretežko ·1	prevoznik ·1	prizadeval ·1	prošnje ·1	rabimo ·3
prav) ·2	prej ·53	pretipkal ·1	prevoznik ·1	prizajmo ·1	prošnje ·1	rabim ·3
prav ·47	prekata ·12	pretoč'la ·1	prevoznik ·1	priznam ·3	prošnje ·1	rabiš ·3
prazen ·1	prekatoma ·4	pretok ·2	prevoznik ·1	priznavam ·1	prošnje ·1	rab'la ·1
prazna ·3	prekatom ·2	preučuješ ·1	prevoznik ·1	prizna ·1	prošnje ·1	rab'li ·1
	prekatov ·1	preudarkom ·1	prevoznik ·1	pr' ·em ·2	prošnje ·1	računaj ·1
	prekat ·9	preudarno ·1	prevoznik ·1	pr' ·i ·1	prošnje ·1	

računalnika 2	razmišlja 1	resnično 1	samo 190	siguren 1	skup) 1	smeri 5
računalniko	razodel 1	resno 1	samuraja 1	sigurno 4	skup 30	smerokaz 2
m 1	razpisov 1	restavracija 1	samuraji 2	sijajem 1	skušal 2	smer 2
računalnik 3	razpis 2	restavraciji 1	sam' 43	sije 1	skušnjavo 4	smešna 1
računalniško	razpoči 1	res) 1	sam 20	silil 1	slabega 2	smešno 1
2	razporedi 1	res 102	sanjalo 1	sili 1	slabe 2	smeš 1
računal 1	razporedu 1	rešene 1	sarajevski 1	simpatičen 1	slabosti 1	sme 2
računamo 1	razporeja 1	rešeni 1	savske 1	sine 1	slabo 6	smisla 2
računam 1	razredi 1	rešenje 6	sa' 11	sinoči 1	slabše 2	smotano 1
računaš 1	razredni 1	rešila 2	sčohat' 1	sinova 1	slab 3	smo 6
računa 1	razredno 1	rešitev 1	sebe 3	sin 2	sladke 1	sme 204
račun 5	razredu 1	reši 5	sebi 5	sirov 1	sladko 1	smrada 1
rada 4	razred 7	rešnjim 1	seboj 2	sirup 6	slasti 1	smrdelo 8
radenska 2	razrezali 1	rešuje 1	sebo' 14	sistematično 1	slava 11	smrdi 4
rade 1	razstavljajo 1	režež 1	sedela 1	sistemi 1	slave 2	smrti 2
radio 3	razumela 1	revije 1	sedeli 2	sistemom 1	slave 2	smrtne 1
radiram 1	razumemo) 1	reviji 1	sedel 2	sistem 1	slavijo 3	smrt 4
radi 1	razumeš 1	revmo 1	sedemkrat 1	sita 1	slavili 1	smučarje 1
radosti 1	rezervanje 2	rezervo 1	sedemnajsti 5	sitnosti 1	slavil 1	smučat 1
radost 1	razumljivo 1	rigalo 1	sedem 3	situacija 1	slavimo 6	smučišče 1
radujte 1	razveselil 1	rinil 1	sedet' 1	sit 2	slavi 3	sna 1
rad 3	razviti 1	ritem 1	sedet 2	si) 4	slavni 1	snega 1
rahle 2	razvit 4	ritjo 1	sedežni 1	si 318	slavo 2	snež 4
rahlo 1	rdeča 3	ritke 1	sedež 2	skala 1	sledimo 1	snemana 1
rah'o 1	rdečelistno 1	rit 1	sede 4	skalnjak 1	sledi 1	snemaš 1
rajnega 1	rdeče 4	rjave 2	sediš 2	skapiral 1	slednjč 1	soba 1
rajne 1	rdečih 1	rjavih 1	sedl 2	skavte 8	sled 1	soba 4
rajni 2	rdeči 1	rjavi 1	sedli 1	skavtih) 1	slekla 1	sobotah 5
rajši 6	rdečo 7	rjavo 1	sedmega 2	skavtih 1	slepa 1	sobotnega 1
rajš) 1	recimo 16	robu 1	sedmih 12	skavti 4	slikah 4	sobotnem 2
rajš 1	reci 1	rodijo 2	segala 1	skavtom 1	slikal 1	sobotno 2
ramo 3	rečejo 1	rodil 1	sej m 1	skavtov 2	slikami 1	soboto 6
rampa 1	rečemo 1	rodu 1	sekač) 1	skavtska 2	slikarja 1	sobot 1
ranjeno 1	rečem 6	rogliček 1	selektor 3	skavtske 1	slikarjev 2	sobo) 1
rano 7	rečeno 1	roglič 1	selil 1	skavt 2	slikarji 1	sobo 2
rastejo 1	rečen 1	rojen 3	semafor 1	sklada 1	slikarstva 2	sodeč 1
raste 1	rečeš 2	rojstni 1	semester 1	sklad 2	slikarstvu 1	sodelavko 1
rastlina 1	rečete 2	rokah 2	semestru 1	sklanjaš 1	slikar 3	sodel'vali 1
ratali 1	reče 15	rokami 1	seminarskega 1	skleda 2	slika 2	sodit 1
ratalo 1	reči 3	roke 1	1	skledi 1	slike 2	sodi 1
ravnokar 1	reč' 1	roki 2	seminarsko 1	skledo 1	slikice 1	sodstvo 1
ravno 17	reč 3	rokomet 1	seminar 1	sklenil 1	sliki 1	soka 3
razbije 1	redke 4	rokomet) 1	semlaj 1	sklonjen 1	sliko 2	sok 7
razbila 1	redki 1	rokomet 6	semle 1	skočiš 1	slik 2	solato 2
razbit' 1	redkokdaj 1	roko 2	sem) 2	scombinali 1	slišala 3	sonca 3
razblinimo 1	redko 11	rok 1	sem 166	skopali 3	slišali 1	sonce 4
razcvetela 1	rednim 4	romal 1	senat 1	skopat 1	slišal 2	soncu 1
razdalja 2	redno 1	rožah 1	senčne 1	skorajda 1	slišim 1	sončni 1
razdalj 1	redu 41	roža 5	sendviče 3	skoraj 2	slišiš 1	sončno 1
razdelili 4	red 3	rože 9	septembra 5	skorja 1	slišiva 1	song 9
razdelil' 1	regionalni 3	roži 1	septembru 1	skorjo 1	sliš' mo 1	sopara 2
razdelimo 1	registracijo 1	rožnega 4	serbus 4	skozi 6	sliš'va 1	soprani 15
razdelit' 1	rekel) 1	rožni 4	seri 1	skoz' 27	slonokosteni 1	sopranov 1
razdiral 1	rekel 72	rožo 7	serviete 1	skoz 1	slovel 1	sopran 13
razdret' 2	reke 1	rubriko 1	sestavljali 1	skrajšane 1	slovenska 1	soprogo 1
razen 8	reki 1	ruski 1	sestavljat 1	skrbet' 1	slovenske 1	sorte 1
raziskovalno 1	reklama 1	ruščine 1	sesti 1	skrbi 3	slovenski 1	sosednja 1
razkomotil 1	reklamo 2	rutka 1	sestrama 1	skrinja 1	slovenščina 1	sosednji) 1
razkopana 1	rekla 58	rutko 2	sestra 1	skrinjo 1	slovenščine 2	sosedom 1
razkošno 1	rekli 15	sabo 3	sestre 8	skripti 1	slovenščini 2	sošolci 1
razkril 1	rekl' 1	sadja 1	sešila 1	skrivaj 3	slovo 2	sošolka 3
razlagala 1	reko 1	sadje 1	seveda) 1	skriva 1	slučajno 5	sošolko 1
razlagal 3	reku 1	saj) 2	seveda 18	skrivnosti 1	slušala 1	sošol'c 2
razlaga 1	rek 1 1	saj 192	severni 2	skrivnostna 1	slutil 2	sovražnega 1
različne 1	remi 1	salonitna 1	severno 2	1	služabnika 2	so 306
različno 1	repa 1	sama) 1	severu 1	skrivnost 2	služabnike 1	spada 1
razlije 1	repkom 4	sama 16	seznamom 1	skuhali 3	služabniku 1	spakira 1
razločno 1	reprezentanca 1	samem 1	seznam 4	skuhal 3	služabnik 1	spala 3
razloga 1	1	same 3	se'endeset 1	skuhano 1	služba 1	spal 6
razlog 2	reprezentance 1	samimi 1	se) 5	skuhan 2	službi 2	spat 14
razlomil) 1	1	sami 8	se 716	skupaj 2	službo 1	specialistične) 1
razložil 1	reprezentanci 1	samohvale 1	shajala 1	skupini 1	služijo 2	1
razloži 1	1	samopostrežni 1	shladilo 1	skupino 3	smehljaj 1	specializacije 1
razmerja 1	reprezentanco 1	1	sicer 18	skupne 1	smeli 1	specialko 1
razmigal 1	1	samopostrežn 1	sifonih 1	skupnimi 1	smel 1	spečem 1
razmišljala 2	resnice 1	o 1	sifoni 1	skupno 1	smemo 1	spečene 1
razmišljaj 5	resnici 2	samoti 1	sifon 4	skup' 3	smem 2	speljat' 1

spet) 1	sprehod 1	stane 2	storila 1	svetovljansko 1	šofer) 1	tačas 1
spet 40	sprejeli 1	stanje 2	storil 2	1	šofer 1	tačko 6
spijem 2	sprijel 1	stanovanja 1	stožec 1	svetovne 1	šola 2	tažajska 1
spije 1	sprejema 1	stanovanje 1	sto 7	svetovnih 1	šole 5	taka 29
spijo 1	sprejeta 1	stanovanjski 2	strah 1	svetovni 1	šoli 8	takega) 1
spila 1	sprejmejo 1	stanova je 1	straneh 2	svetovno 4	šolo 8	takega 15
spili 1	sprejme 3	stanujejo 3	stranišča 1	sveto 5	šolskem 1	takem 4
spim 1	sprejmi 3	stanuje 1	stranišče 1	svetu 6	šolsko 3	take 30
spisek 3	spremenijo 2	stan 1	strani 28	svet 7	šolstvo 1	takih 6
spiska 1	spremenile 1	stara 5	stran 1	svidenje 8	šopek 2	takile 1
spiš 1	spremenili 1	starejša 4	stran 11	svinčnike 1	šopke 1	takimi 1
spite 2	spremenilo 1	starejših 1	strasti 3	svinčnik 1	šop 1	takim 1
spit 2	spremenil 1	starejši 3	strašna 2	svinjsko 3	šotora 1	taki 15
spi 1	spremenil 1	starem 2	strašno 5	svobodi 1	šotor 1	takoj 19
splačalo 1	spreminjali 1	stare 3	streljen 1	svoja 1	šparat 1	takole 1
splača 1	spreminja 1	starinska 1	stregel 1	svojega 5	špilala 2	tako 54
splanirali 1	spremljal 1	stari 7	strela 1	svojemu 1	špilat 1	takratne 1
splavih 1	spremlja 1	starost 2	streljal 1	svojem 4	športu 1	takrat 21
splavi 1	spremnstvom 1	staro 1	streljaš 1	svoje 19	špork 2	taksiji 1
spletni 1	sprintal 1	staršem 3	strelu 1	svojih 1	šprica 1	tak at 1
sploh) 1	sprintano 1	staršev 2	strgalo 1	svojimi 3	šprintat) 1	tak le 11
sploh 48	sprintaš 1	starše 1	strica 2	svojim 7	štajerska 1	tak ne 1
spočeta 1	sprobali 1	starših 1	strip 1	svoji 3	štajerski 1	tak) 3
spočijes 2	sprobal 1	starši 3	strmih 1	svojo) 1	šteje 3	tak 420
spočit 1	sprostiti 1	star 12	strmo 1	svojo 9	štel 1	tak 7
spodaj 5	sprožit 1	stat 2	stroju 1	svoj ga 2	številka 4	talal 1
spoda 3	sprva 1	stavka 1	strokovnega 1	svoj 6	številke 1	talas 2
spodnjem 1	spučal 1	stavkov 1	stropa 1	s noč 3	številko 3	tale 5
spodnjim 1	spučem 1	sta 74	stroši 1	s 2	številko 3	tali 1
spodnji 2	spumpala 2	stekla 2	stroške 2	s 162	štima 1	talna 1
spodobi 4	spustili 2	steklena 1	stroški 1	šajkah 1	štirih 2	tamle 5
spodrslijaj 2	sraka 2	steklo 5	stroškov 2	šajke 1	štirikrat 1	tam 3
spod 11	srali 1	stena 2	struna 1	šal ca 1	štirinajstih 1	tam 203
spomenik 1	sram 1	stene 2	stržena 1	šal ce 2	štirinajst 1	tančica 1
spomini 1	sra je 1	steno 3	stvarem 1	šal co 1	štiri 27	tankala 1
spominjamo 1	srca 3	stestiral 1	stvari 20	šans 1	štirje 5	tankat 1
spominjam 1	srcem 2	ste) 1	stvarnika 2	šef 1	štir krat 1	taoist 1
spomin 3	srce 3	ste 51	stvarstva 1	še 16	štir 18	tat 1
spomladi 2	srčno 1	stiki 1	stvarstvom 1	šel 56	štir 1	ta 1
spomnili 1	srebrna 1	stiku 1	stvarstvo 2	šepečem 5	štokali 1	ta) 2
spomnil 1	srebrno 1	stik 2	stvar 8	šestdeset 4	štopali 1	ta 83
spomnil 4	srečala 4	stil 1	suha 2	šestega 1	štrajkal 1	tebe 17
spomnim 12	srečal 4	stiskah 1	suhe 2	šestih 23	štrli 2	tebi 29
spomniš 6	srečanjih 2	stiski 1	suho 6	šestindvajseti 1	študenta) 1	teboj 29
spomnite 1	sreča 1	stisnemo 1	suh 1	šestka 5	študente 1	tebo 1
spomni 2	sreče 1	stisneš 3	sukal 1	šestko 1	študenti 1	teb 2
spomn la 1	srečna 1	stisnil 1	sukat 1	šestnajstega 2	študentki 1	tečaj 1
sponsoriral 1	srečno 9	stisnit 1	suknjarna 1	šestnajst 2	študentko 1	teče 2
sporočam 1	srečujemo 1	stisjeno 1	sumim 1	šesto 1	študentov 3	tečna 1
sporočila 1	srečujes 1	stoje 2	super 22	šeststo 1	študentsko 2	tedaj 1
sporočilo 4	sređa 1	stojijo 2	sušilnica 1	šest 17	študent 3	teden) 1
sporočil 4	srede 2	stojim 1	sušit 1	še) 1	študijske 1	teden 53
sporočit 1	redinca 1	stojišča 1	sušit 2	še 352	študiju 3	tedna 2
sporoči 1	redini 2	stojiš 1	suš 2	šibam 1	študij 2	tedne 4
sposodil 2	rediščih 1	stoji 8	svaki 1	šipe 2	študirajo 1	tednih 2
spoznala 2	redišču 1	stojnico 1	svate 1	širok 1	študirala 2	tedni 1
spoznali 1	redil 1	stokrat 1	svatovna 1	širo 1	študiral 13	tednov 1
spoznal 1	rednja 1	stolčem 2	sva 22	šivilja 1	študiram 2	tednu 1
spoznam 1	rednješolci 1	stoletja 2	svočana 3	škarpa 1	študiraš 3	ted n 1
spoznati 7	rednješolsko 1	stoletju 2	svčanosti 1	škarpo 1	študirat 3	tega 21
spoznavalcev 1	rednje 2	stole 1	svčke 1	škatlico 2	štukom 1	tehnično 1
1	rednje 2	stolkel 2	svčo 2	škatli 1	št r deset 1	tehniko 1
spoznavas 1	rednji 1	stolnici 4	sveta 11	škoda 3	št r najstdnev	teh 4
sprali 1	redno 7	stolnico 1	svetega 4	škodi 2	no 1	tej 4
spraševala 1	red 2	stolni 3	svetemu 1	škod lo 1	št r najstem 1	tekaj 1
spraševali 1	rednost 1	stolp 2	svete 3	škofa 2	št r najstimi 1	tekem 1
spraševali 2	stadion 1	stol 3	svetila 1	škofe 1	št r najst 20	tekla 1
spraševal 1	stala 3	stol 3	sveti 8	škofje 1	švedski 1	tekli 1
sprašvala 1	stale 1	stopil 1	svetla 1	školkji 3	tabele 1	tekma 1
spravil 1	stalilo 1	stopinj 3	svetlobe 1	škrjančki 1	tabel 1	tekme 1
spravimo 1	stališča 1	stopnicah 1	svetlobi 2	šla 24	table 1	tekme 2
spravim 1	stali 3	stopnja 1	svetlobo 1	šle 2	tabli 1	tekmi 1
spraznili 1	stalo 2	stopnji 1	svetlo 2	šli 43	taborne 2	tekmovanje 1
spraznite 1	stal 1	stopnjo 2	svetnike 2	šlo 19	taboru 1	1
spraznit 1	standardiziran 1	stoprocentno 1	svetniki 1	šment 1	tabor) 1	
spredaj 4	o 1	1	svetnikov 1	šoder 1	tabo 1	
sprehod li 1	standard 1	stop 1	svetosti 1	šodra 2		

tekmo 2	tleh 5	trpel 4	ujeli 1	uspešen 1	večerja 4	verujem 1
tekstilstvo 1	tlom 1	trto 1	ukradel 1	uspeva 2	večerje 1	verze 1
tek 2	točka 1	truden 5	ukradli 1	ustalila 1	večerji 3	verzije 1
telefonske 1	točke 1	trudili 1	ulica 2	ustanovah 1	večerjo 6	verzi 1
telefonu 4	točno 1	trudim 1	ulice 1	ustanovitelj 1	večer 19	vesela 2
telefon 6	točno 14	tudi 1	ulici 3	ustavili 1	večglasen 1	veselica 1
telesna 1	točn 1	tudi 83	ulilo 2	ustavil 2	večglasnosti 1	veselja 2
telesno 1	toda 3	tud 2	umaknil 1	ustaviš 3	večina 1	veseljem 4
telesom 1	tod 4	tud 189	umazanost 1	ustavit 1	večinoma 1	veselje 3
televiziji 1	tolarjev 4	tuja 1	umazano 2	ustavlja 1	večja 3	vesel 3
tele 1	tolar 1	tujca 1	umetna 1	usta 1	večje 2	vesolja 2
telovadbe 1	tolažnica 1	tujem 1	umetnikov 1	ustnice 1	večjih 1	vesoljni 1
telovadbo 1	tole 14	tuji 1	umetnik 2	ustrašili 1	večji 3	vesoljno 1
telo 10	tol ko 1	tukaj 14	umetniški 1	ustrašil 1	večjo 3	vespo 1
temeljito 1	tol k 1	tule 1	umetnosti 1	ustraš la 1	večkje 1	ves 4
teme 2	tol k 19	tule 19	umetnostna 1	ustreza 1	večkrat 2	veš 1
temi 1	topel 1	tuširala 1	umetnost 1	ustrežljiv 1	večnamenski 1	veš 147
temne 1	tople 2	tu 3	umetno 1	ustvaril 1	1	vetra 1
temno 1	toplota 1	tu 103	umiril 1	ustvarjeno 1	večnem 1	vezala 1
temperatura 2	torbami 1	tvoja 13	umirjeno 1	utelesil 1	večne 2	veze 2
tempirano 1	torej 13	tvojeja 2	umreti 1	utici 1	večni 1	vezne 1
temu 4	torek 4	tvojemu 2	umrla 1	utrudi 1	večno 4	ve el 2
temveč 5	torkih 1	tvojem 12	umrl 3	utrgan 1	več 73	ve te 1
tem 34	torto 2	tvoje 20	umrl 1	utrjeni 1	vedeli 1	ve te 3
tenorji 3	tor k 4	tvojih 1	uničil 1	utrujen 1	vedel 12	ve 10
tenor 1	tor k 4	tvojimi 1	univerzi 1	užaljena 1	vedet 3	vhoda 2
teorija 1	tor k 4	tvojim 6	uokvirjanje 1	uživaj 1	vedl 1	vhodom 1
terenu 1	tor k 4	tvoji 1	upajmo 1	vabilo 2	vedno 30	videla 1
teren 1	to 4	tvojo 6	upali 1	vabi 2	vedve 1	videli 2
terjali 1	to 406	tvoja 1	upal 2	vabljeni 1	ved la 7	videl 1
termine 1	tradicije 1	tvoj 17	upamo 1	vadit 1	ved li 4	videl 23
ter 2	trajala 1	tvorbe 1	upam 13	vadit 1	veke 2	videokaseti 1
tesnilo 2	trajalo 1	t a 1	upanja 1	vagon 1	veki 1	videt 4
tesno 1	traja 1	t a 4	upaš 1	vajah 4	vekomaj 5	vidijo 1
teto 2	travo 2	t kole 2	uporabil 1	vajami 2	vekov 2	vidimo 8
teža 1	trdi 2	t ko 1	uporabiva 1	vajeni 1	veku 1	vidim 9
težek 3	treba 38	t ko 54	uporabljali 1	vajen 1	veličastvu 1	vidiš 9
težil 1	treh 1	t lele 3	uporabljal 2	vaje 11	velika 2	vidi 4
težje 3	treh 2	t le 18	uporabljat 1	vaj 5	velikem 1	vidnih 1
težkega 1	tremi 1	t l ko 1	uporabno 1	vama 3	velikim 2	vidva 1
težko 16	trenerja 1	t sta 1	uprava 1	vami 7	veliko 5	vidve 7
te 1	treningu 2	t stem 1	upravno 2	vam 29	velik 6	vid la 14
te 107	trenirali 2	t ste 2	upret 1	varianta 1	velik 2	vid le 1
tiče 1	trenira 1	t sti 2	uprla 1	variante 1	veljala 1	vid li 5
tičijo 2	trenutek 2	t sto 1	uradnih 1	varna 2	veljalo 1	vid lo 3
tiho 9	trenutka 1	t ubogim 1	uradnika 1	varni 1	velja 5	vid l 2
tikat 1	trenutkov 1	t ubogi 1	uradno 2	varnosti 1	vel ka 5	vijolica 1
tik 3	trenutku 1	t ubrat 1	ura 9	varnostna 1	vel kem 1	vijolice 1
tipična 1	trenutno 2	učencem 1	ureditev 1	varnostnim 1	vel ke 1	vijolico 3
tipkal 1	tretjem 3	učenik 1	ureditve 1	varnostni 1	vel ki 1	vikendi 1
tipologija 1	tretje 2	učila 1	urejene 1	varnostno 1	vel ki 4	vikend 3
tip 4	tretji 2	učili 2	uresničimo 1	varnost 2	vel ko 12	vila 1
tira 4	trtjo 1	učil 3	ure 13	varovali 1	vemo 1	vil ce 1
tiru 2	tre a 1	učiš 1	urice 1	varstva 1	vem 2	vina 8
tir 2	trgala 1	učitelj 1	uri 18	varuh 1	vem 206	vinca 1
tiskajte 1	trgatev 2	učitelj 2	urnik 1	varuje 2	venca 3	vinom 1
tiskalnik 1	trgatvi 2	učit lji 1	urnik 4	varuj 1	vendarle 3	vino 3
tiskanje 2	trgat 1	učit 2	uro 22	var val 1	vendar 6	vinska 3
tiskat 1	trga 1	uči 2	ur 4	vase 1	venerini 1	vinske 1
tiska 1	trgovini 1	učne 2	usedeš 1	vase 1	ven 14	vinski 4
tisoč 18	trgu 1	uč te 1	usedeta 1	vas 39	vera 2	vinsko 1
tista 24	trg 1	udari 1	usedla 1	vaša 1	vere 2	vir 1
tistega 11	trideset 6	udarja 1	usedlo 1	vaščanov 1	verige 1	visi 1
tistem 4	trindvajset ga 1	udar lo 3	usest 1	vaše 4	verjamejo 1	visoka 1
tiste 22	1	udeležencev 1	uskladil 2	vaško 1	verjamem 1	visoke 2
tistih 4	trije 7	udobje 2	usklajevali 1	vašo 1	verjameš 1	visokih 1
tistimi 1	trikratne 1	ugasnil 2	usliši 6	vate 1	verjeli 1	visokimi 1
tistim 4	trikratni 1	ugled 2	usmeril 1	važen 1	verjelo 1	visoki 1
tisti 62	trikrat 2	ugotavljala 1	usmerjaj 1	važnega 2	verjel 1	visoko 5
tisto 83	trinajst ga 1	ugotavljal 1	usmer li 1	važno 3	verjetnost 1	višvah 6
tist ga 3	trijom 1	ugotovili 1	usmili 18	včasih 28	verjetno 2	višina 4
tist 7	tristo 2	ugotovil 1	usmiljenja 1	včasi 1	verjetno 28	višine 1
tišina 1	tri eset 1	ugotovil 3	usmiljeno 2	včerašnjega 1	verjet 1	višini 3
ti 1	trindvajst ti 2	ugotovimo 1	usodno 2	včera 18	verouk 3	višino 1
ti 296	1	ugotovit 1	uspela 1	vdrla 1	verovali 1	višja 2
tja 1	1	ugrabi 1	uspelo 3	večera 1	vero 4	višjem 1
tja 35	tri 60	uh 2	uspel 2		verujejo 1	višje 3

zgornjim ·1	zmenil' ·1	znašel ·1	zvestobo ·2	železo ·1	žlic ·1	'mam) ·2
zgradili ·2	zmenil' ·5	znašle ·1	zvest ·14	želim) ·1	žogo ·1	'mam ·66
zgradit' ·1	zmenimo ·1	znaš ·4	zvezan) ·1	želim ·4	žrebanje) ·1	'maš ·61
zgrešim ·2	zmeniš ·2	znata ·1	zvezde ·5	želi ·1	žreb ·1	'mata ·3
zgubila ·1	zmenit' ·5	znate ·4	zvezek ·2	želje ·2	žrtev ·1	'mate ·7
zgubljeni ·1	zmenjena ·1	znat' ·1	zveze ·3	željo ·1	žrtvam ·1	'mava ·2
zidajo ·1	zmen'la ·3	zna ·7	zvezi ·1	žemlje ·1	žulil ·1	'ma ·56
zidali ·2	zmen'li) ·1	znova ·4	zviti ·1	žemlji ·2	župa ·2	'mejte ·1
zidal ·1	zmen'li ·8	zob ·1	zvočne ·2	žena ·3	župi ·1	'mej ·1
zidovi ·1	zmeraj ·4	zoper ·1	zvočnike ·1	ženiš ·1	župnija ·1	'mela ·31
zima ·1	zmerit' ·1	zopet ·1	z'ajle ·2	ženi ·1	župniji ·1	'meli ·58
zimski ·3	zmerne ·1	zračunat' ·1	z'aj) ·1	ženo ·2	župnika ·1	'melo ·1
zimsko ·1	zmerom) ·1	zradiram ·1	z'aj) ·178	ženska ·3	župniku ·1	'mel' ·6
zjutraj ·14	zmerom ·6	zrahljat ·1	z'aj ·36	ženske ·4	župo ·2	'mel ·64
zlagat' ·1	zmešana ·1	zramal ·1	z ·174	žensko ·1	žurali ·1	'met' ·10
zlasti ·2	zmetalo ·1	zrastejo ·2	žagala ·1	žetoni ·1	žur ·6	'met ·3
zlata ·2	zmisljio ·1	zrastel ·2	žalostna ·1	že) ·3	žvižga ·1	'ne ·9
zleze ·2	zmisliš ·1	zraste ·1	žalostnih ·1	že ·219	žogi ·1	'nih ·5
zlijemo ·1	zmočil' ·1	zrastla ·1	žal ·8	žgalne ·1	žogo ·1	'ni ·1
zloglasna ·1	zmolili ·1	zraven ·4	žanr ·2	žgance ·1	žifnabahijo ·1	'no ·9
zloglasne ·1	zmolimo ·2	zrihtala ·1	žarela ·1	žiga ·1	1	'om ·4
zloma ·1	zmore ·1	zrihtali ·1	žarel ·1	živa ·3	1	'ota ·1
zlomijo ·1	zmotim ·1	zrihtal ·2	žarki ·1	živčen ·1	1	'ote ·1
zloženke ·1	zmot ·1	zrihtal ·2	žegen ·1	živela ·2	1	'o'ta ·1
zloženo ·1	znajo ·1	zrno ·1	žegnaj ·1	živeli ·2	1	'saj ·1
zložiš ·1	znake ·1	zunaj ·7	žegna' ·1	živeti ·1	1	'sedete ·1
zlušta ·1	znak ·1	zunanje ·1	žegna' ·1	žive ·1	1	'sedi ·1
zmagala ·1	znali ·1	zunanji ·2	žejcn ·2	živijo ·4	1	'večer ·9
zmagali ·1	znal ·2	zunanjo ·1	žejna ·1	živimo ·1	1	'zgledalo ·1
zmagal ·1	znamenitega ·1	zuna' ·3	žejni ·1	živiš ·2	1	'zgledaš ·1
zmage ·1	1	zvadiš ·1	želela ·1	živi ·4	1	'zgleda ·7
zmagovalni ·1	znamenitih ·1	zvalali ·1	želeli ·1	živjo ·4	1	'zgle'a ·1
zmagovalno ·1	znamenje ·1	zvedel ·13	želel ·3	življenja ·3	1	'zvedel ·1
zmago ·6	znamo ·2	zvedele ·1	železnice ·2	življenje ·7	1	'z ·14
zmanjkalo ·1	znam ·2	zvedel ·1	železnic ·1	življenju ·2	1	
zmanjkat' ·1	znana ·1	zved'li ·2	železniška ·1	živl'e'je ·1	1	
zmanjka ·1	znane ·1	zveličanja ·1	železniške ·1	živo ·2	1	
zmanjšaj ·1	znani ·1	zveličavno ·1	železniški ·1	živ'jo ·9	1	
zmanjšate ·1	znano ·2	zveni ·2	železniško ·1	žlahti ·2	1	
zmatrana ·2	znan) ·1	zvestobe ·1	železobeton ·1	žleht ·1	1	
zmatra ·1	znan ·1	zvestobi ·1	1	žlico ·1	1	

PRILOGA 4: 1210 NAJPOGOSTEJŠIH RAZLIČNIC V PROSTEM GOVORU KORPUSA

V seznamu so vse različnice, ki se pojavijo v podkorpusu prostega govora (gl. Primerjava 2, str. 159) vsaj štirikrat, razvrščene po najprej po pogostosti, tiste z isto pogostostjo pa po abecedi. Različnici za piko sledi absolutna pogostost. Najpogostejšim 99 je za piko dodan še njihov delež v % (tj. absolutna pogostost, deljena s številom pojavnic v podkorpusu). Za graf rasti besednega zaklada različnic gl. Slika 27, str. 156.

pa ·1673 ·3.93	aha ·112 ·0.26	tri ·59	pred ·37	strani ·27	med ·19	stvari ·16
je ·1487 ·3.49	po ·112 ·0.26	´meli ·58	*tota ·37	štiri ·27	m- ·19	takega ·16
ne ·1273 ·2.99	pač ·111 ·0.26	ki ·57	drugi ·36	*mm ·27	petek ·19	takrat ·16
ja ·1176 ·2.76	ob ·110 ·0.26	rekla ·57	dve ·36	nimam ·26	poznam ·19	tebi ·16
da ·809 ·1.90	lahko ·109 ·0.26	dva ·56	mal´ ·35	s- ·26	sva ·19	težko ·16
se ·612 ·1.44	*ahm ·106 ·0.25	tisti ·56	tja ·35	greš ·25	šlo ·19	*ki ·16
v ·600 ·1.41	s ·105 ·0.25	´ma ·56	z´lo ·35	ono ·25	večer ·19	avtobus ·15
*e ·492 ·1.16	b´la ·102 ·0.24	t´ko ·55	kot ·34	potem ·25	zadnjič ·19	čepav ·15
bi ·446 ·1.05	tu ·102 ·0.24	*ful ·55	*kuj ·34	pride ·25	dr´gač´ ·18	dam ·15
na ·432 ·1.01	kak´ ·101 ·0.24	ko´ ·54	dni ·33	´majo ·25	eni ·18	eden ·15
tak´ ·423 ·0.99	od ·101 ·0.24	šel ·53	gor ·33	*hehehe ·25	it´ ·18	en´ga ·15
a ·397 ·0.93	mislim ·100 ·0.23	teden ·53	grem ·33	deset ·24	moti ·18	jim ·15
to ·386 ·0.91	23	doma ·52	vas ·33	on ·24	najbrž ·18	ma ·15
jaz ·353 ·0.83	res ·100 ·0.23	ste ·52	ona ·32	šla ·24	naredil ·18	mama ·15
*s´n ·338 ·0.79	lahk´ ·95 ·0.22	tudi ·52	spet ·32	kol´k´ ·23	nekje ·18	mene ·15
še ·327 ·0.77	okej ·94 ·0.22	*fajn ·52	b´li ·31	naj ·23	osmih ·18	mimo ·15
če ·317 ·0.74	hvala ·92 ·0.22	me ·51	prvi ·31	preveč ·23	oziroma ·18	misliš ·15
si ·292 ·0.69	malo ·92 ·0.22	dan ·50	skup ·31	prišel ·23	sicer ·18	morem ·15
za ·288 ·0.68	bomo ·91 ·0.21	prej ·50	zrav´n ·31	šestih ·23	šest ·18	papir ·15
so ·283 ·0.66	en ·91 ·0.21	*toti ·49	´mela ·31	tista ·23	štir´ ·18	reče ·15
in ·274 ·0.64	kar ·89 ·0.21	čist´ ·48	*mihoho ·31	včasih ·23	takoj ·18	seveda ·15
ti ·273 ·0.64	ga ·86 ·0.20	nekaj ·48	*vun ·31	*ne´e ·23	tega ·18	soprani ·15
al´ ·269 ·0.63	do ·84 ·0.20	noter ·48	gremo ·30	ata ·22	t´le ·18	šele ·15
ni ·250 ·0.59	tisto ·83 ·0.19	sploh ·48	kdam ·30	tem ·22	včera´ ·18	taki ·15
kaj ·247 ·0.58	aleluja ·82 ·0.19	´mamo ·48	kdo ·30	videl ·22	vsak ·18	točno ·15
bo ·246 ·0.58	bolj ·81 ·0.19	kako ·46	nam ·30	*keri ·22	*dolta ·18	t- ·15
že ·217 ·0.51	*toto ·81 ·0.19	nazaj ·46	pri ·30	dolgo ·21	*gir ·18	zjutraj ·15
vem ·205 ·0.48	nič ·80 ·0.19	zato ·46	take ·30	nisi ·21	dal ·17	*lalalalalalalal
tam ·204 ·0.48	te ·80 ·0.19	jo ·45	verjetno ·30	super ·21	devetih ·17	a ·15
smo ·195 ·0.46	boš ·79 ·0.19	oni ·45	´lej ·30	uro ·21	drugo ·17	*m ·15
mhm ·194 ·0.46	iz ·78 ·0.18	ena ·44	*nisen ·30	včera ·21	kak ·17	*teko ·15
mi ·193 ·0.45	kje ·77 ·0.18	not ·44	*tc ·30	vedno ·21	le ·17	=086Žz=ime ·
saj ·191 ·0.45	pol ·77 ·0.18	zdravo ·44	daj ·29	*he ·21	lepa ·17	15
tud´ ·191 ·0.45	ta ·77 ·0.18	*bojo ·44	k ·29	*nis´n ·21	mislil ·17	cevi ·14
samo ·189 ·0.44	*hehe ·72 ·0.17	dober ·43	Marija ·29	ah ·20	proti ·17	dela ·14
bom ·184 ·0.43	*redi ·69 ·0.16	prav ·43	menda ·29	cev ·20	*nes ·17	k´r ·14
*pol ·180 ·0.42	*pal ·68 ·0.16	sam´ ·43	niso ·29	jutri ·20	*tote ·17	mogoč´ ·14
z´aj ·178 ·0.42	sta ·67 ·0.16	adijo ·42	*ha ·29	minut ·20	bistvu ·16	pravi ·14
ko ·176 ·0.41	čez ·66 ·0.16	mogoče ·42	bit´ ·28	n- ·20	čuj ·16	p- ·14
no ·172 ·0.40	´mam ·66 ·0.16	o ·42	lepo ·28	prišli ·20	eh ·16	sebo´ ·14
*te ·168 ·0.39	rekel ·65 ·0.15	prosim ·42	mu ·28	št´r´najst ·20	ene ·16	spat ·14
ka´ ·165 ·0.39	gre ·64 ·0.15	tako ·42	neka´ ·28	tisoč ·20	klical ·16	sto ·14
ker ·165 ·0.39	´mel ·64 ·0.15	zdi ·42	pet ·28	tiste ·20	mojem ·16	tebe ·14
b´lo ·154 ·0.36	pr´ ·63 ·0.15	=001Mz=ime ·	taka ·28	tol´k´ ·20	moram ·16	tole ·14
sem ·150 ·0.35	eno ·62 ·0.15	42	*gljh ·28	tule ·20	najbolj ·16	ven ·14
veš ·146 ·0.34	nas ·62 ·0.15	celo ·41	dan´s ·27	vam ·20	najprej ·16	vid´la ·14
bil ·141 ·0.33	več ·62 ·0.15	redu ·41	ej ·27	vsaj ·20	nismo ·16	vmes ·14
zdaj ·140 ·0.33	´maš´ ·61 ·0.14	šli ·41	enkrat ·27	*hm ·20	noč ·16	´z ·14
vse ·129 ·0.30	ampak ·59	men´ ·40	mor´š ·27	domov ·19	problem ·16	*kera ·14
z ·129 ·0.30	dobro ·59	vsi ·39	naprej ·27	dosti ·19	ravno ·16	*majn ·14
*ko ·128 ·0.30	dol ·59	*sma ·39	nisem ·27	ji ·19	recimo ·16	*mogel ·14
aja ·117 ·0.27	jih ·59	treba ·38	sko ·27	leto ·19	sama ·16	avto ·13

dala ·13	nared ·11	daš ·9	jest ·8	klicala ·7	ali ·6	priklop ·6
danes ·13	nima ·11	delat ·9	j- ·8	Kreta ·7	bila ·6	pr ·nesel ·6
dost ·13	novo ·11	devet ·9	kaki ·8	Krfu ·7	bodi ·6	punc ·6
joj ·13	papirja ·11	dvajs ·t ·9	kuža ·8	kup ·7	bodo ·6	rajši ·6
kaka ·13	peš ·11	dvanajst ·9	laja ·8	k- ·7	brat ·6	razred ·6
letos ·13	redko ·11	evo ·9	let ·8	leti ·7	celi ·6	rečem ·6
mesto ·13	sa ·11	hiše ·9	levo ·8	Ljubljano ·7	Cesar ·6	sami ·6
nekak ·13	spod ·11	ime ·9	mali ·8	luknje ·7	čaj ·6	sirup ·6
neki ·13	stran ·11	kakih ·9	Maribor ·8	luknjo ·7	delo ·6	skozi ·6
pelj ·13	tak ·le ·11	kam ·9	napisano ·8	midva ·7	dobili ·6	slabo ·6
petsto ·13	teklo ·11	kon ·c ·9	naš ·8	midve ·7	druga ·6	Slovenci ·6
pod ·13	tistega ·11	kosilo ·9	niti ·8	mobitel ·7	drugač ·6	Slovenija ·6
postaja ·13	tvoj ·11	lani ·9	ostalo ·8	moj ·7	druge ·6	soboto ·6
rekli ·13	vaje ·11	liste ·9	par ·8	mora ·7	drugem ·6	spal ·6
sopran ·13	vseeno ·11	most ·9	pazi ·8	nami ·7	enim ·6	spomniš ·6
študiral ·13	začne ·11	našem ·9	pela ·8	nikogar ·7	en ·mu ·6	stari ·6
upam ·13	zakaj ·11	neko ·9	petkih ·8	nimaš ·7	gledaš ·6	suho ·6
ure ·13	*ee ·11	one ·9	pozno ·8	odtok ·7	gosto ·6	tačko ·6
vlak ·13	*kao ·11	peljal ·9	pravil ·8	okol ·7	grize ·6	telefon ·6
*boma ·13	*mmhm ·11	petdeset ·9	rano ·8	ostala ·7	Habakuk ·6	Tiger ·6
*eh ·13	*totege ·11	poznaš ·9	skavte ·8	paše ·7	hod ·la ·6	torej ·6
*haha ·13	*vjutro ·11	prekat ·9	smrdelo ·8	pet ·7	julija ·6	tristo ·6
bo ·ta ·12	alt ·10	prišla ·9	srečno ·8	počasi ·7	klet ·6	tvoja ·6
bo ·te ·12	amen ·10	rože ·9	stoji ·8	pojedel ·7	kostanje ·6	ura ·6
cel ·12	Bog ·10	tvoje ·9	stvar ·8	prejšnji ·7	kraj ·6	večerjo ·6
čudno ·12	četrtak ·10	ve ·9	svidenje ·8	prvič ·7	lep ·6	velik ·6
enih ·12	dali ·10	vi ·9	šoli ·8	rokomet ·7	ljudi ·6	vreme ·6
glavnem ·12	delali ·10	vidiš ·9	šolo ·8	sok ·7	Lovrencu ·6	vsako ·6
greznico ·12	dež ·10	v- ·9	tukaj ·8	sredo ·7	l- ·6	všeč ·6
itak ·12	dobil ·10	zarad ·9	vidim ·8	tak ·7	manj ·6	vzel ·6
jašek ·12	glas ·10	zbor ·9	vi ·iš ·8	tiho ·7	Mariboru ·6	začela ·6
kda ·12	hišo ·10	zmen ·li ·9	vlakom ·8	tist ·7	mesec ·6	zadaj ·6
kruh ·12	hitro ·10	z- ·9	vzela ·8	trg ·7	meter ·6	zmago ·6
možno ·12	hm ·10	živ ·jo ·9	začeli ·8	trgu ·7	minuti ·6	žur ·6
nekam ·12	izhod ·10	ne ·9	zanimivo ·8	trideset ·7	mislila ·6	mel ·6
neke ·12	kake ·10	no ·9	zgodi ·8	trije ·7	mogel ·6	*ajerspajs ·6
nesel ·12	meni ·10	večer ·9	*co ·8	ved ·la ·7	mor ·jo ·6	*alelu ·6
prekata ·12	mogla ·10	*bek ·9	*čao ·8	vidimo ·7	mor ·m ·6	*domu ·6
prvo ·12	nad ·10	*cajta ·9	*mare ·8	vidve ·7	nama ·6	*eSeMeS ·6
sam ·12	naslednja ·10	*češ ·9	*naret ·8	vina ·7	naša ·6	*faks ·6
sedmih ·12	nedeljo ·10	*hahaha ·9	*ou ·8	začelo ·7	naše ·6	*fejst ·6
spomnim ·12	petih ·10	*heh ·9	*taki ·8	zaka ·7	nikol ·6	*grema ·6
uri ·12	petnajst ·10	*jast ·9	=010Žz=ime ·8	zapet ·7	niste ·6	*gužva ·6
vedel ·12	piše ·10	*Maribori ·9	=101Žz=ime ·8	zaradi ·7	njega ·6	*jah ·6
vel ·ko ·12	pogledal ·10	*skor ·9	bilo ·7	zdajle ·7	nogomet ·6	*jit ·6
zvečer ·12	pozab ·la ·10	*ziher ·9	bravo ·7	zmerom ·7	nosil ·6	*kej ·6
*hehem ·12	predavanja ·10	*zutra ·9	delal ·7	zunaj ·7	nova ·6	*naji ·6
*jel ·12	pridem ·10	brez ·8	direkt ·7	mate ·7	okno ·6	*placa ·6
*k ·12	prideš ·10	b- ·8	direktno ·7	zgleda ·7	osem ·6	*rečt ·6
*kene ·12	teboj ·10	cesto ·8	drug ·7	*au ·7	osemsto ·6	*sam ·6
*mhm ·12	vozi ·10	dalje ·8	drugič ·7	*črez ·7	ostali ·6	*sed ·n ·6
*mogla ·12	vzeli ·10	desno ·8	d- ·7	*dama ·7	pesem ·6	*tavžent ·6
*mogli ·12	zelo ·10	dobila ·8	enostavno ·7	*ehehe ·7	pijmo ·6	*uf ·6
=028Žz=ime ·12	met ·10	domo ·8	Gorica ·7	*kek ·7	poje ·6	*vaji ·6
	*cajt ·10	Glavnem ·8	greznice ·7	*keko ·7	pomeni ·6	*vijadva ·6
b ·le ·11	*enkret ·10	gledal ·8	hodi ·7	*mhmhm ·7	pon ·delj ·k ·6	*zmer ·6
del ·11	*lih ·10	Gospod ·8	hodil ·7	*mo ·7	poročila ·6	=006Žz=ime ·6
desetih ·11	blizu ·9	greznica ·8	h ·tel ·7	*ne ·a ·7	posebej ·6	=013Žz=ime ·6
drugega ·11	čakaj ·9	halo ·8	isto ·7	=003Mz=ime ·7	poslal ·6	=014Mz=ime ·6
kol ·ko ·11	časa ·9	hočem ·8	jedla ·7	=004Žz=ime ·7	pozabil ·6	a- ·5
Ljubljani ·11	čisto ·9	hotel ·8	kakor ·7	=030Mz=ime ·7	pridejo ·6	boljše ·5
Maribora ·11	dalo ·9	ja- ·8	kjer ·7	Aleš ·6	pridi ·6	Boštjan ·5

bova ·5	njo ·5	voziš ·5	Duha ·4	narobe ·4	repkom ·4	vseh ·4
brado ·5	normalno ·5	vprašanje ·5	dvanajstih ·4	našo ·4	roža ·4	začel ·4
brke ·5	novi ·5	vrata ·5	en- ·4	naučit´ ·4	rožni ·4	zadnji ·4
b´ ·5	O ·5	vsega ·5	gledali ·4	navodila ·4	sede ·4	zaenkrat ·4
Celju ·5	očitno ·5	zanima ·5	gle´al ·4	nekdo ·4	serbus ·4	zamenjal ·4
cesti ·5	odpravi ·5	zazidal ·5	gle´at ·4	Nič ·4	seznam ·4	zastopil ·4
cicibanček ·5	oh ·5	zdelo ·5	Gorazd ·4	nista ·4	sifon ·4	zdravje ·4
čakat´ ·5	okus ·5	zemlji ·5	greva ·4	niže ·4	sigurno ·4	Zidani ·4
četrt ·5	peljali ·5	zgle´a ·5	grozno ·4	ni´am ·4	Simona ·4	zmeraj ·4
čim ·5	pesmi ·5	zgodilo ·5	Habakuku ·4	note ·4	skavt ·4	znaš ·4
človek ·5	piknik ·5	zmenil ·5	hkрати ·4	novega ·4	skavti ·4	znete ·4
čul ·5	pila ·5	zmenit´ ·5	hočeš ·4	novembra ·4	slikah ·4	zraven ·4
čula ·5	pili ·5	zna ·5	hod´li ·4	nula ·4	smrdi ·4	zveze ·4
daleč ·5	ploščo ·5	zvest ·5	hudega ·4	običajno ·4	sobi ·4	žal ·4
dobri ·5	podporo ·5	želim ·5	hudo ·4	odločil ·4	Soboti ·4	ženske ·4
dobr´ ·5	poklicala ·5	´dijo ·5	Ivanka ·4	okrog ·4	sonce ·4	živijo ·4
dob´li ·5	poklicali ·5	´nih ·5	Izlake ·4	onega ·4	sporočil ·4	živjo ·4
drugje ·5	polnoči ·5	*čaj ·5	izpite ·4	onih ·4	sporočilo ·4	´om ·4
dvojčka ·5	ponikovalnico ·5	*den ·5	Jankovič ·4	onim ·4	spredaj ·4	*ah ·4
edino ·5	ponoči ·5	*du ·5	jed ·4	ostale ·4	srečal ·4	*aha ·4
glavo ·5	popoldne ·5	*ehe ·5	jem ·4	ostane ·4	srečala ·4	*ateja ·4
glej ·5	postaje ·5	*folk ·5	ješ ·4	otrok ·4	starejša ·4	*baš ·4
gol ·5	potniški ·5	*girta ·5	jurja ·4	o- ·4	steklo ·4	*do ·4
Gorice ·5	potrebno ·5	*itak ·5	jutr´ ·4	pardon ·4	stolnici ·4	*dvekret ·4
govoril ·5	pozdravlja ·5	*jurjov ·5	kolegi ·4	pekli ·4	Struna ·4	*ej ·4
hiša ·5	pozdravljen ·5	*kafe ·5	komu ·4	petindvajs´t ·4	svetovno ·4	*evo ·4
igro ·5	pravim ·5	*kaki ·5	konca ·4	počakal ·4	svoje ·4	*grejo ·4
imaš ·5	printer ·5	*mhmhmm ·5	Kristina ·4	počakat´ ·4	svojo ·4	*heheh ·4
imel ·5	prit´ ·5	*osen ·5	križ ·4	pojem ·4	šestdeset ·4	*I ·4
Jurček ·5	probali ·5	*os´n ·5	kruha ·4	pojemo ·4	Tabor ·4	*Ka ·4
jutro ·5	prostor ·5	*poj ·5	kupil ·4	pojest´ ·4	takem ·4	*kak ·4
kakega ·5	prostora ·5	*tej ·5	kupit ·4	poklical ·4	tale ·4	*keda´ ·4
kanalizacijo ·5	pusti ·5	*to ·5	kupit´ ·4	pokličem ·4	Tanja ·4	*kero ·4
kod ·5	pustil ·5	*trgi ·5	lačen ·4	poleti ·4	tedne ·4	*kod´ ·4
kolega ·5	račun ·5	=002Žz=ime ·5	ladja ·4	polno ·4	telefonu ·4	*kranc ·4
kom ·5	septembra ·5	=012Žz=ime ·5	lažje ·4	popoldan ·4	temu ·4	*marem ·4
koncu ·5	slučajno ·5	=025Mz=ime ·5	leže ·4	popoldansko ·4	tip ·4	*meši ·4
Kovačić ·5	sobotah ·5	=100Žz=ime ·5	Ljubljane ·4	postal ·4	tistem ·4	*mešo ·4
k´ ·5	spodaj ·5	alelujo ·4	luknja ·4	pot ·4	tistih ·4	*mhmhmmhmh
letniku ·5	stara ·5	alti ·4	mala ·4	povej ·4	tistim ·4	mhmhmmh
male ·5	strašno ·5	avtobusom ·4	marelo ·4	pozdravi ·4	tod ·4	mhm ·4
maši ·5	šestka ·5	baje ·4	maša ·4	pravo ·4	tolarjev ·4	*mogo ·4
minus ·5	šole ·5	bili ·4	metrov ·4	pravzaprav ·4	torek ·4	*nahm ·4
misl´la ·5	´Ta ·5	bloki ·4	me- ·4	prebujam ·4	tor´k ·4	*neki ·4
mizi ·5	takih ·5	boljš´ ·4	miga ·4	predavanje ·4	tvojem ·4	*probajmo ·4
močno ·5	tamle ·5	Boris ·4	milijonov ·4	predsednik ·4	Umek ·4	*re´i ·4
moje ·5	ta- ·5	Celje ·4	mirno ·4	prek ·4	vajah ·4	*skup ·4
moral ·5	Titov ·5	cista ·4	mislili ·4	prekatoma ·4	večerja ·4	*šeeset ·4
more ·5	tleh ·5	cisterna ·4	mogli ·4	prijateljici ·4	ved´li ·4	*ššt ·4
najboljš´ ·5	truden ·5	čakat ·4	moja ·4	Primož ·4	veliko ·4	*ta ·4
naredim ·5	t´a ·5	čist ·4	mor´mo ·4	problema ·4	venca ·4	*tek ·4
nared´la ·5	ur ·5	čuje ·4	Most ·4	pr´bližno ·4	videt´ ·4	*u ·4
naslednji ·5	urnik ·5	dat´ ·4	mošta ·4	pr´nest´ ·4	vidi ·4	*uu ·4
našega ·5	vaj ·5	delajo ·4	mo- ·4	rada ·4	vinski ·4	*valda ·4
našli ·5	velja ·5	dokler ·4	Muri ·4	razdelili ·4	višina ·4	*ni ·4
na- ·5	vel´ka ·5	dopoldan ·4	Murski ·4	Razdevšekovi ·4	vode ·4	=005Žz=ime ·4
nebesih ·5	vel´ki ·5	dopoldne ·4	namoč´lo ·4	rozen ·4	vozili ·4	=009Žz=ime ·4
nesreča ·5	vid´li ·5	dovolj ·4	naporno ·4	razmišljal ·4	vo- ·4	=017Žz=ime ·4
nimamo ·5	visoko ·5	drugače ·4	naravnost ·4	razvit ·4	vsaka ·4	-luja ·4
njemu ·5	voda ·5	drugam ·4	naredimo ·4	redke ·4	vsakem ·4	

PRILOGA 5: PRIMERI KONKORDANC DVEH RAZLIČNIC

Konkordance so narejene na podlagi vseh transkripcij prvega dne v korpusu (nedelje).

RAZLIČNICA <*te>

besedilo	govorec	časovno umeščena celica traku (angl. annotation), v kateri se pojavi različnica – časovne oznake so umaknjene
1ned04Vg	001Mz	samo mi včasih tak' ne vem ka' jaz vem pač pižama zleze izza *pasa a ne *te se pa zbudim *ko mi je hladno.
1ned04Vg	002Žz	zdaj pa eno mor'sh 'met tak' vedno v Ljubljani *te tak' sploh nimaš kaj ...
1ned04Vg	002Žz	di'gač' sem v *redi spala samo tisto ko včasih na *hr'ti spim *te se mi zdi da 'mam včasih čist' trdi hrbet *nazodno.
1ned04Vg	002Žz	kaj *te vem ...
1ned04Vg	002Žz	pa mena se bliža ka' *te veš ka' nas bo zagrab'lo še vse do mene ...
1ned04Vg	002Žz	*te pa namest' da bi vstal *vjutro *frišen pa spočit če osem al' devet ur spiš ... ponoči (). se mi zdi da () mogla vstat' ob pol sedmih.
1ned04Vg	002Žz	v *vinogradi *s'n pa samo enkrat *te nared'la od tebe *ni požirek.
1ned04Vg	002Žz	vmes se mi sploh ni zdelo potrebno če pa tak' je ... grozdje ješ p- ... pač pa soka pa niso *ekstra ponujali tak' da *s'n že *pol jaz rekla *ki je pa *te kak sok je?
1ned04Vg	002Žz	*sigurno jih povabijo na kosilo *te pa študirajo a lahko zun' *jejo al' ne mor'jo ne?
1ned04Vg	002Žz	smo *te direktno šli ().
1ned04Vg	002Žz	ker pač midva vedno k frančiškanom *hodima *te pa *toto *s'n si mislila da *grema k frančiškanom.
1ned04Vg	002Žz	kdo je pa *te dal gol?
1ned04Vg	003Mz	*te pa ... čuj (=001Mz_ime) *te pa še *totole *finto mi povej. kako prideš da stisneš točno nazaj kje si bil?
1ned04Vg	003Mz	*te *s'n pa študiral ka' te bi bila svečana *masa *pal *s'n spomnil svečana maša je a ne?
1ned09Vg	001Mz	() *te *pal mar ne?
1ned09Vg	001Mz	pa rečemo vedno: *ki je *te *Božica pa *Jožo ...
1ned09Vg	001Mz	nič ne laja nič ne grize ... okej *te pa tak'. še ... tri štir!
1ned09Vg	001Mz	*mmh *te pa gremo običajno *taizejsko aleluja a- ... 'sedi se malo bolj naprej da se *bojo čuli.
1ned09Vg	007Žz	tak' da če bi ob šestih 'meli ni problema *te lahko *pal. do *nih pol osmih al'?
1ned09Vg	007Žz	aja *te *pal že m- *muti ...
1ned09Vg	007Žz	čuj a veš *te pa () ... jaz 'mam en'ga strica *Dankota *ko 'ma teto Vanko. in jaz rečem zmerom: kje *te sta Danko pa Vanka? ne! *Danka pa *Vanko!
1ned09Vg	008Žz	() *te pa mi *toto ...
1ned09Vg	008Žz	zdaj pač smo ta stare *tacajt 'meli zdaj pa *s'n *te ...
1ned09Vg	012Žz	*te pa jaz () ...
1ned16Vg	001Mz	(=025Mz_ime) do naslednjč pr'nesel seznam *te bomo pa terjali.
1ned16Vg	001Mz	ka' *te vsi tak' ste to *ha?
1ned16Vg	027Žz	*te pa bom ... vseeno pa daj zaokroži pa tistih petsto gor al' pa dol () al' ka' jaz vem ka' ...
1ned16Vg	008Žz	ne 'večer! ka' *te jaz vem ka' bi ... dopoldan ne zato ker dopoldan jaz težko *pajem! no tak' ...
1ned16Vg	008Žz	ja *te bi pa z'aj ob šestih v petek bi še nekak' šlo no ...
1ned16Vg	008Žz	*te *nis'n 'mela in *s'n čist' pozab'la.
1ned16Vg	008Žz	ker *te *nis'n 'mela *pal pa pozabim pr'nest'.
1ned17Vg	004Žz	*te se samo () če še sploh *če ...
1ned17Vg	028Žz	ja bo pa *te na vsakih št'r'najst dni bolj dolgo ().
1ned18Vg	001Mz	*te pa j- to je pa potrebno zato da *pal se ti *ne'e glas pokvari če *paješ tisto ka' ti *ne'e paše ne.
1ned18Vg	001Mz	t- *te pa bo'te vid'le če lahk' tisto zapojete pa bom jaz po t'stem vedel al' ste prvi al' ste drugi sopran.
1ned18Vg	001Mz	a lahk' *te vidve *probata skup zapet'?
1ned18Vg	001Mz	*te pa z (=028Žz_ime) skup no? (=028Žz_ime) pa (=006Žz_ime)! tri štir'!
1ned18Vg	001Mz	če bi jaz z'aj govoril tak' bi *te takoj vid'lo da *s'n en bas.

<i>besedilo</i>	<i>govorec</i>	<i>časovno umeščena celica traku (angl. annotation), v kateri se pojavi različnica – časovne oznake so umaknjene</i>
1ned19Vg	001Mz	kak' *te lahk' žvižga?
1ned19Vg	008Žz	seveda *s'n čula *s'n pa *te tam zrav'n tebe stala v dnevni sobi.
1ned19Vg	008Žz	tak' je bil lep! tam *spod za ... čakaj kje smo *te b'li?
1ned19Vg	008Žz	'lej *jojil! ka' pa *te je vidva?
1ned19Vg	008Žz	tak' da mu je *mogel eden neka' *naret več ko *te je () ...
1ned19Vg	008Žz	vprašanje ka' mu je naredil ne. al' mu je vrgel kaj al' ga je mogoče ka' *te več ...
1ned19Vg	008Žz	ja se pa *te Uha tud' samo Uha je bil še debeli *te .
1ned20Dg	001Mz	v petek *te ko *s'n ti hotel za vaje povedat'.
1ned22Vg	001Mz	ne vem (). *te pa tud' ne'e pridem kaj dosti prej ob osmih pridem.
1ned22Vg	001Mz	a jo *te jaz pokličem? kda' je kda' sta rekli da bi se poklicali?
1ned22Vg	002Žz	vselili so se *te so pa 'meli otvoritev hiše pa ...
1ned25Vg	002Žz	*te pa *pol smo malo spraševali ...
1ned25Vg	002Žz	hkrati bi pa *te vzeli ().
1ned25Vg	002Žz	razen *te ko je sončno.
1ned25Vg	002Žz	lepo bi že b'lo je pa tak' *te bi *mogla bit' voda gor na *podstrešji. jaz že ne bom nos'la take kante vode gor ne ...
1ned25Vg	002Žz	kak pa *te printer z'aj? ni ()?
1ned25Vg	002Žz	da prerežem. vidim da včasih kako obvestilo kje kdo kaj napiše *te pa sta dva stavka pa je cela stran ne.
1ned25Vg	001Mz	kako bo pa *te tako bo z- vinski *zos?
1ned25Vg	001Mz	*te pa je dan's ni b'lo.
1ned25Vg	001Mz	jaz *s'n razmišljal če bi v Ljubljano lahk' nesel kako rožo. pa- no pač *te ko mi bo'te stvari peljali.
1ned25Vg	003Mz	ka' pa *te je v bistvu primerjalna književnost?
1ned25Vg	003Mz	kaj s' pa *te nadevala za 'no ()?
1ned25Vg	003Mz	ka' *te misliš!
1ned25Vg	003Mz	on je rekel da tud' *kene *nuca *te ko pa da () *te bomo pa mi kup'li *pol pa je.
1ned25Vg	003Mz	ka' *te nam niso ponud'li bolj poceni papir ()!
1ned26Dr	003Mz	tak' sine (). *zutra *sma pa *te dokončno sama.
1ned26Dr	003Mz	če jem fižol *te 'mam dobro prebavo me *kene *gir suši.
1ned26Dr	003Mz	kje *s'n pa *te jaz videl *nes 'no tako () debelo?
1ned26Dr	003Mz	aja tisto boš tak' ti nesel *te tak' v *redi.
1ned26Dr	003Mz	no ka' je pa *te tu noter b'lo?
1ned26Dr	003Mz	*te pa ne vem al' ga je v stroju obrivalo () ... al' ka' se je zgodilo.
1ned26Dr	003Mz	kam si pa *te one *šalčke za *kafe dal?
1ned27Vr	003Mz	ja ka' pa *te *pal z veseljem to naredim.
1ned34Vg	002Žz	*te pa vsaj *anorak vzemi.
1ned36Vg	001Mz	*te pa daj.
1ned41Vr	001Mz	v petek *te ko *s'n ti hotel za vaje povedat'.
1ned57Dg	001Mz	*ahm nisem niti kaj posebej razmišljal pa nimam kakih posebnih načrtov pač ... *probal sem kaj *te ne ...
1ned57Dg	001Mz	samo moram 'met' sto_tisoč besed *te pa ne vem kak' je ... a ta 'ma *e *e *UeSBe-priključek za na računalnik?
1ned70Vg	001Mz	pr' nas z'aj kmeti *e *gnojšn'co okol' po *polji vozijo *te pa mogoč' je kaj *bek prišlo *ehe ...
1ned70Vg	001Mz	*keremu je *te uspelo tak' zažgat' *toto not?
1ned71Dg	001Mz	a res kdaj *te to?
1ned71Dg	001Mz	*te pa *s'n *nis'n mislil vprašanje če je z'aj septembra bil ...

RAZLIČNICA <*ko>

besedilo	govorec	časovno umeščena celica traku (angl. annotation), v kateri se pojavi različnica – časovne oznake so umaknjene
1ned04Vg	001Mz	samo mi včasih tak' ne vem ka' jaz vem pač pižama zleze izza *pasa a ne *te se pa zbudim *ko mi je hladno.
1ned04Vg	001Mz	*ko je *M. tu stisneš pa se takoj gor pojavi *hold. in če hočeš dalje sprožit' pa spet *toti *M pr'tisneš.
1ned04Vg	001Mz	ja. ni tak' ko' računalnik *ko bi stisnil *back pa bi te vrglo tja nazaj.
1ned04Vg	002Žz	*ko nas bo še vse *zagrob'lo v *totih *letah ...
1ned04Vg	002Žz	ja! no pa to *ko sta včeraj () Danica pa Stane razmišljala ne?
1ned04Vg	003Mz	() *toke *ko 'majo take vel'ke dimenzije () se reče ...
1ned04Vg	003Mz	pa tam *ko je Franc *natokal *s'n *vsok'kret samo *molo vzela.
1ned04Vg	003Mz	sploh ni več t'ste ... tak' *s'n iskal ki je t'sta stara bajta *ko je Katr'ca noter b'la pa nisem več se znašel.
1ned04Vg	003Mz	a tam *ko stanujejo?
1ned04Vg	003Mz	ne tam *ko *s'n bil n- *ko *s'n na za- na *začetki ... n'č na sto ...
1ned04Vg	003Mz	je rekel t'sti *ko je *ee kip postavljaj je rekel da ga () tak' visoki kraj tak' da se vidi čez vse vrhove širom ().
1ned09Vg	001Mz	tak' ti je *hica da si () *kuj *du slekla ... ma *ko je mrzlo *ha?
1ned09Vg	001Mz	note ... najprej gledaš do *sam do do k- *ko piše ena ne? *pal pa greš od začetka pa *tole skočiš na dvojko.
1ned09Vg	004Žz	*ko maša bo *tot-?
1ned09Vg	007Žz	čuj a veš *te pa () ... jaz 'mam en'ga strica *Dankota *ko 'ma teto Vanko. in jaz rečem zmerom: kje *te sta Danko pa Vanka? ne! *Danka pa *Vanko!
1ned09Vg	007Žz	*ko *čaki z'aj *ko pa ...
1ned09Vg	008Žz	ja tisti *ko znajo besedilo na pamet!
1ned09Vg	012Žz	*toto *taizejsko *ko je *obično malo ponavljamo je tak' enoglasna žalostna!
1ned09Vg	015Mz	ta drugi del samo *ko je večglasen.
1ned10Dg	008Žz	čuj včera' *s'n *vstola taka da *gir *nis'n *odparla *očes! *ko *s'n v petek ... si dobil *eSeMeS?
1ned11Vr	001Mz	je prav pre- primerna pesem *ko je tak mrak.
1ned16Vg	027Žz	to je tisto dodatno meso pa pijača *ko smo nabavljali ...
1ned16Vg	025Mz	'lej to 'mam jaz jaz *s'n ono *ko *s'n jaz zbral *s'n deset mi jih je dalo ne a spisek a () spiska z'aj nimam zravn' () ...
1ned16Vg	025Mz	ne to je tak' pet *jurjov *ko je ostalo spisek ti bom pa drugič dal *ala?
1ned17Vg	001Mz	okej. vsi vse *ko nimate kaj z'aj to prosim *vun!
1ned18Vg	001Mz	mi se *mar'mo zdaj zmenit' *e ... tak' je v *zbori ne da vas je *ful dosti *ko *ko vas je zdaj pelo sopran ne.
1ned18Vg	001Mz	pa tud' tak' je da *e vas *ko ste zdaj na novo pr'sle ne: (=013Žz_ime) (=009Žz_ime) l- (=006Žz_ime) (=028Žz_ime) pa (=005Žz_ime) da vas nismo sprobali.
1ned18Vg	001Mz	*ahm a mi še lahk' neka' še *kuj tak' pove'te? *e a ve'te? k- *ko kak' glas se čuje tudi *kuj iz tega kak' govoriš ne.
1ned18Vg	001Mz	in vedve bi peli zravn' altov to je tam *ko *pajeta (=004Žz_ime) z'aj po novem pa (=027Žz_ime) že po starem pa ().
1ned18Vg	008Žz	(=001Mz_ime) a te lahk' zmotim ono k- kje so kje 'mamo seznam pesmi tistih *ko znamo?
1ned18Vg	008Žz	() *ko je napisano?
1ned19Vg	001Mz	*ovi *ko so se *toti teden z bagrom *kapoli *e v petek al' ka' je že?
1ned19Vg	008Žz	ja samo veš kak' je b'lo čist' or'ginalno tak' al' tak' samo pač *ko je malo višje. al'. veš kak' *frajersko!
1ned19Vg	008Žz	bom čist' z'aj rekla *froc'i *ko so malo bolj žleht ne ...
1ned19Vg	008Žz	Brezočniki () Rolando ne 'majo nemškega ovčarja pa en'ga psa *ko mu je ime Tiger.
1ned22Vg	001Mz	ena dvojčka pa še dva *ko nista dvojčka.
1ned22Vg	001Mz	*ko 'ma (=030Mz_ime)?
1ned22Vg	002Žz	Razdevšekovi od *totiga *ko 'ma dvojčke ne ... veš *keri 'majo dvojčke?
1ned22Vg	002Žz	Razdevšek'vi so tista hiša veš če greš k *čebelarjom na desno stran so najprej Gradišniki *ko hodijo k sosedom ta druga hiša so pa Razdevšekovi.
1ned22Vg	002Žz	*ko 'majo eni Mariborčani.
1ned22Vg	002Žz	brat od *tote Valerije *ko se piše Grdadolnik brat je pa ... Aleš.
1ned22Vg	002Žz	tista montažne *e *ko je ...

<i>besedilo</i>	<i>govorec</i>	<i>časovno umeščena celica traku (angl. annotation), v kateri se pojavi različnica – časovne oznake so umaknjene</i>
1ned22Vg	002Žz	*tote malo *pogledajmo *ko so t- suhe *tote lahko b- *bek damo.
1ned22Vg	002Žz	na njen ... pač na tisti mobitel *ko 'ma (=030Mz_ime).
1ned25Vg	002Žz	ste peli dan's *kero tisto novo *ko ste se v petek učili na vajah?
1ned25Vg	002Žz	a veš *ko vsake tol'ko obarva s- noter srce rdečo ...
1ned25Vg	002Žz	dr'gač' pa ona soba *ko je gor na vrhu ...
1ned25Vg	002Žz	*he *e ona roža *ko je gor na vrhu v sobi ...
1ned25Vg	002Žz	*tota *ko mi jo je menda (=086Žz_ime) kup'la za rojstni dan () gor pa sta rekla (=086Žz_ime) pa (=030Mz_ime): glej kak' bo to cvetelo!
1ned25Vg	002Žz	to je pet *kil *ko morš' nest' gor ne pet litrov vode ...
1ned25Vg	002Žz	in no *toti je menda z'aj bolj gladek prejšnji *ko smo ga pa 'meli pa *s'n si jaz kopirala za priprave liste in () kopiral ko piše s črnilom ne s penkalom.
1ned25Vg	002Žz	vi pa sploh *ko ste 'meli par sto strani včasih zanič pa sprintano pa ste stran vrgli.
1ned25Vg	001Mz	si je basala pa je *nateknila en del *salote *ko ni bil ravno strebljen ne.
1ned25Vg	001Mz	*e ja tu 'maš ampak tip papirja se pa dol tud' lahk' da ne nastavit tam *ko 'maš tiskanje.
1ned25Vg	003Mz	*pje *ko je *daro!
1ned25Vg	003Mz	a veš tista *ko st- na *maji *akni stoji ... tista *ko je taka bolj ... () mi je Milan enkrat prinesel a veš ...
1ned25Vg	003Mz	tiste *ee nageljčke *ko si mi jih presadila ... najlepši so na Barbarini *akni tja na dvorišče ko sploh ne prisije sonce.
1ned25Vg	003Mz	jaz *s'n pa tam *kuj *ko je b'lo *draft *text ...
1ned25Vg	003Mz	se je zgodilo da je da je printer na printer na ta vel'kem printerju *ko so fakture pisali pa vse ka' je () b'lo ...
1ned25Vg	003Mz	še domo' *s'n si nosil. še *toti papirji *ko so tu *zgira so take vrste.
1ned35Dg	001Mz	*ahm veš kaj če lahk' počakaš tam nasproti železniške postaje tam *ko je oni Rudolf_*Maister.
1ned35Dg	001Mz	al' pa ne ne ne n-! tam na zadnji strani *ahm ... *e tam *ko je Vilharjeva tam *ko je Mobitel?
1ned35Dg	001Mz	*e ja. na tisti strani *ko se gre dol proti *televiziji ne?
1ned35Dg	001Mz	*ko je Kolodvorska pa to.
1ned35Dg	001Mz	tam *ko je en park oziroma *e na robu parka ob cesti je pa kip Rudolfa_*Maistra.
1ned36Vg	002Žz	mhm *ko je dobro.
1ned36Vg	002Žz	tisti pulover bi tud' lahko nosil *ko 'ma *taki puli a veš ovrtnik.
1ned36Vg	002Žz	kđaj drugič ko boš 'mel več *placa noter nosit' boš pa za () *vzemeš *ko so v zamrzovalniku. če ne tak' nimaš *placa *jel'?
1ned57Dg	001Mz	*e *pol pa tam *ko se zavije proti *e podvozu pod železniško.
1ned57Dg	001Mz	ta še ne ampak naslednja. tu *ko se *dreve ven v- ... tale ja.
1ned57Dg	001Mz	ta *ko sem si ga sposodil pr' en'mu ki ni je malo novejši dve leti star pa m- 'ma *UeSBepriključek ...
1ned70Vg	001Mz	ta ženska *ko o 'jem govori je ena pač umetnostna zgodovinarka.
1ned70Vg	058Mz	*ko en ...
1ned70Vg	058Mz	*ko en nogomet se igra praktično v vseh državah po svetu ne?
1ned71Dg	001Mz	mislim pač taki *ko so pri nas kandidirali samo mislim da kje dr'gje b'lo tud' mišljeno ne?

PRILOGA 6: NAJPOGOSTEJŠE N-KRATNICE BESED V KORPUSU, N= [2, 3, 4, 5]

S sivim ozadjem so označene n-kratnice, ki se pojavljajo v prostem govoru, z belo pa navedki iz pesmi, molitev ipd. Za vsak *n* so navedeni vsi primeri najmanjše absolutne pogostosti.

<i>n=2</i>	<i>n=3</i>	<i>n=4</i>	<i>n=5</i>
ne vem 173	prosi za nas 50	o Marija o Marija 27	Marija o Marija o Marija 14
ja ja 130	ja ja ja 27	Marija o Marija o 14	o Marija o Marija o 14
je pa 129	Marija o Marija 27	ne ne ne ne 12	Marija kako si lepa ti 11
da bi 125	o Marija o 27	kako si lepa ti 11	Marija o Marija kako si 11
da je 103	aleluja aleluja aleluja 24	Marija kako si lepa 11	o Marija kako si lepa 11
to je 101	ne ne ne 24	Marija o Marija kako 11	o Marija o Marija kako 11
je bil 78	se mi zdi 22	o Marija kako si 11	aleluja *alelu aleluja aleluja aleluja 8
ko je 73	pa ne vem 18	se mi zdi da 10	star song se še vrti 8
tak' da 73	je rekla da 14	aleluja *alelu aleluja aleluja 8	zakaj star song se še 8
pa je 70	mihoho *mihoho 14	alelu aleluja aleluja aleluja 8	Gospod je dal spoznati svoje 6
	*mihoho		
v *redi 67	to je pa 14	song se še vrti 8	Gospod s teboj hočem s 6
da se 66	da bi se 13	star song se še 8	Gospod usmili se Gospod usmili 6
je b'lo 63	hočem s teboj 12	zakaj star song se 8	hočem s teboj Gospod s 6
če bi 62	ne vem če 12	če nikogar ni doma 7	hočem s teboj iti vse 6
al' pa 60	s teboj Gospod 12	mihoho *mihoho *mihoho 7	iti vse dni s teboj 6
	*mihoho		
jaz *s'n 60	kako si lepa 11	dal spoznati svoje rešenje 6	je dal spoznati svoje rešenje 6
se je 59	Marija kako si 11	dni s teboj Gospod 6	ne ne ne ne ne 6
a ne 58	men' se zdi 11	en dva tri štiri 6	s teboj Gospod s teboj 6
aleluja 56	o Marija kako 11	Gospod je dal spoznati 6	s teboj hočem s teboj 6
aleluja			
za nas 56	si lepa ti 11	Gospod s teboj hočem 6	s teboj iti vse dni 6
a veš 55	aleluja *alelu aleluja 10	Gospod usmili se Gospod 6	teboj Gospod s teboj hočem 6
ne ne 52	jaz sem pa 10	hočem s teboj Gospod 6	teboj hočem s teboj iti 6
pa ne 50	mi zdi da 10	hočem s teboj iti 6	teboj iti vse dni s 6
prosi za 50	al' pa kaj 9	iti vse dni s 6	usmili se Gospod usmili se 6
te pa 50	en dva tri 9	ja ja ja ja 6	vse dni s teboj Gospod 6
e ja 48	jaz *s'n pa 9	je dal spoznati svoje 6	eno pesem tiho šepečem hvala 5
mi je 45	ja ne vem 9	je rekla da bi 6	grize če nikogar ni doma 5
pa še 42	je pa res 9	mlada in vroča kri 6	je rdečo rožo v laseh 5
se mi 42	kadar ognji dogorijo 9	se Gospod usmili se 6	ko pa Jurček cicibanček truden 5
v redu 41	ne bi b'lo 9	s teboj Gospod s 6	le eno pesem tiho šepečem 5
je to 40	se mi je 9	s teboj hočem s 6	ne grize če nikogar ni 5
ne bo 40	zakaj star song 9	s teboj iti vse 6	nosila je rdečo rožo v 5
o Marija 40	alelu aleluja aleluja 8	teboj Gospod s teboj 6	pa Jurček cicibanček truden se 5
je b'la 39	da se bo 8	teboj hočem s teboj 6	pesem tiho šepečem hvala ti 5
bi se 37	da se je 8	teboj iti vse dni 6	
ne bi 37	Gospod usmili se 8	usmili se Gospod usmili 6	
bi b'lo 36	je rekel da 8	vse dni s teboj 6	
pa tak' 36	ob pol devetih 8	al' pa kaj takega 5	
se pa 35	rekla da bi 8	eno pesem tiho šepečem 5	
pa se 34	se še vrti 8	grize če nikogar ni 5	
jaz sem 33	song se še 8	je rdečo rožo v 5	
je v 33	star song se 8	Jurček cicibanček truden se 5	
pa tud' 32	če bi se 7	ko pa Jurček cicibanček 5	
da bo 31	če nikogar ni 7	le eno pesem tiho 5	
je že 31	da je bil 7	ne grize če nikogar 5	
a res 30	in vroča kri 7	nosila je rdečo rožo 5	
pa *te 30	ja to je 7	pa Jurček cicibanček truden 5	
pol pa 29	ja v *redi 7	pesem tiho šepečem hvala 5	
tak' ko' 29	je b'lo pa 7	prosimo te usliši nas 5	
pa to 28	ko rano zjutraj 7	rdečo rožo v laseh 5	
s teboj 28	nikogar ni doma 7	te ne moti nič 5	
	tam *ko je 7	tiho šepečem hvala ti 5	

PRILOGA 7: POVEDI V KORPUSU

51 NAJPOGOSTEJŠIH POVEDI

Gre za vse povedi s pogostostjo 7 ali več. 276 pa je v korpusu povedi, ki imajo pogostost

vsaj 2.

ja. 261	aja. 26	zdravo! 19	aja ... 14	*hm. 10	aleluja	*mihoho
mhm. 176	*ahm ... 24	ja? 17	ja ja ... 13	v *redi. 9	aleluja. 7	*mihoho
aha. 57	*e ja. 24	zdravo. 17	*lalalalalalal	*ha? 9	dobro. 7	*mihoho
okej. 49	ne vem. 23	aleluja aleluja	ala. 13	aleluja *alelu	ne ... 7	*mihoho.
prosi za	adijo. 22	aleluja. 1	ja ja ja. 12	aleluja	ne! 7	7
nas! 49	a res? 22	6	v redu. 12	aleluja	prosim. 7	*mmhm. 7
ja ... 38	aha ... 20	kaj? 16	*mhm. 12	aleluja. 8	zakaj star	
hvala. 31	dober	tri štir! 16	*mm. 12	*ahm. 8	song se	
ja ja. 31	dan. 20	aja? 15	adijo! 11	*hehe ... 8	še vrti? 7	
ne. 28	ja! 19	amen. 15	*hehe. 11	aha! 7	*e ... 7	

102 NAJKRAJŠI IN NAJPOGOSTEJŠI POVEDI

Gre za vse povedi z dolžino kvečjemu 4 znakov, označeno pred povedjo, in pogostostjo povedi, označeno za povedjo.

2 a! 1	3 to? 2	4 mhm. 1	4 *e ... 7	4 hm ... 2	4 gre! 1	4 oh ... 1	4 *ah. 1
2 o! 1	3 *m. 2	76	4 aja! 6	4 j- ... 2	4 hja! 1	4 pet. 1	4 *au. 1
2 o. 1	3 ah! 1	4 aha. 57	4 ka? 6	4 n- ... 2	4 jah! 1	4 pol. 1	4 *A ... 1
3 ja. 261	3 ah. 1	4 ja ... 38	4 in ... 4	4 pa ... 2	4 jaz! 1	4 prt! 1	4 *eh. 1
3 ne. 28	3 da! 1	4 aja. 26	4 no ... 4	4 *he. 2	4 joj! 1	4 se ... 1	4 *hm? 1
3 ja! 19	3 mi? 1	4 kaj? 16	4 ah ... 3	4 *ss. 2	4 maj? 1	4 smo? 1	4 *ho! 1
3 ja? 17	3 ni. 1	4 aja? 15	4 kje? 3	4 *uu! 2	4 mhm? 1	4 še ... 1	4 *m ... 1
3 nel 7	3 si? 1	4 *mm. 1	4 nič. 3	4 al'. 1	4 na ... 1	4 tam. 1	4 *ou. 1
3 ne? 5	3 šel 1	2	4 *ha! 3	4 bom! 1	4 nič! 1	4 tam? 1	4 *pš! 1
3 no. 3	3 to! 1	4 *hm. 10	4 *ou! 3	4 bom. 1	4 nič? 1	4 to ... 1	4 *tc. 1
3 a ... 2	3 za? 1	4 *ha? 9	4 dva. 2	4 čao! 1	4 ni ... 1	4 uh ... 1	4 *tš! 1
3 eh! 2	3 *e! 1	4 aha! 7	4 eh ... 2	4 d- ... 1	4 oba? 1	4 ura? 1	4 *uf. 1
3 hm? 2	3 *e. 1	4 ne ... 7	4 evo. 2	4 evo! 1	4 od ... 1	4 vem. 1	4 -la. 1

9 NAJDALJŠIH POVEDI PO ZNAKIH TRANSKRIBCIJE

Gre za vse povedi, daljše od 200 znakov transkripcije. Prosti govor je posivljen.

znakov/besed	poved	govorec	besedilo
241 / 42	blagoslovi Gospod ta prstana ki ju v tvojem imenu blagoslavljam da bosta ta dva ki ju bosta nosila s tem ohranila med seboj popolno zvestobo ostala v tvojem miru in tvoji volji živela vedno v vzajemni ljubezni po Kristusu našem Gospodu amen.	054Mj, nastopajoči	1ned69Vg,
228 / 35	po poroki z Ano se je Pistor vpisal še na *minhensko akademijo v letih tisoč osemsto petindevetdeset do tisoč osemsto osemindvetdeset v razred takrat znamenitega profesorja (*Lihtegerja) ki ga je Pistor že prej močno občudoval.	052Mj, glavni bralec spremne besede	TV-dokumentarec, biografski, delno igran
215 / 35	(v *Minhen) se je posebej želel () postati center () s katere *e da tam doživi čisto nove poglede na umetnost kajti dunajska akademija je veljala za zelo klasično akademijo čeprav je na Dunaju končal tudi specialko.	053Žj, strokovni komentator	
233 / 44	*vun *mar m *spucat zaradi tega ker če bomo ne par dni po uporabljali stranišče v greznico moram *pol drugič spet oni- *oniga dobit da mi bo *vun *spumpal al pa bom moral sam *e *spucat pa če kame je not leti nim- ni dobro ne ...	003Mz, oče, gospodar	6pet77Vr, družinski pogovor
224 / 42	in v enega Gospoda Jezusa Kristusa edinorojenega Sina božjega ki je iz Očeta rojen pred vsemi vek in je Bog od Boga Luč od Luči pravi Bog od pravega Boga rojen neustvarjen enega bistva z Očetom in je po njem vse ustvarjeno.	021Vn, občestvo pri maši	1ned13Vg, obred
221 / 36	podpiraj nas milostno iz nebes o mogočni naš varuh v tem boju zoper moči teme in kakor si nekdan hotel oteti dete Jezusa iz smrtne nevarnosti obrani zdaj sveto cerkev božjo sovražnega zalezovanja in vsakega nasprotovanja.	020Mz, maševalec	nedeljske maše

<i>znakov/besed</i>	<i>poved</i>	<i>govorec</i>	<i>besedilo</i>
216 / 41	vendar zaradi stene med *obenoma prekatoma ki je bila t'ko pod ploščo *teko da je *črez teklo je drekarija stala v tisti cevi ki gre iz hiše tak' visoko da je bil isti nivo kot je višina vmesne stene med prekatoma.	003Mz, oče, gospodar	6pet78Vg, družinski pogovor
208 / 27	za dodatne informacije lahko pokličete (na) telefon štiristo_trideset_(devetindvajset)_devetindvajset ali pa jih obiščete na Štihovi dvajset v Zupančičevi jami vsak delovnik od štirinajstih do dvaindvajsetih.	070Mn, bralec reklame	6pet40Vg, radijska reklama
203 / 48	ja saj to vem *e sta b'la brat pa pa žena letos na m- na medenih tednih na Krfu ne pa sta rekla da sta ste z *e sta se z *e mopedom z vespo okol' voz'la in da sta pač v enem dnevu pol otoka prevoz'la ne.	001Mz, sostanovalec študent	1ned71Dg, pogovor „cimrov“ ob snidenju, pred spanjem

STATISTIKA DOLŽIN POVEDI

Prim. Slika 26, str. 153. Vsaka peta vrstica v spodnji tabeli je posivljena zaradi preglednosti.

<i>vrsta povedi po dolžini v besedah</i>	<i>št. teh povedi</i>	<i>delež med povedmi (8.423 povedi) [%]</i>	<i>vsota besed v povedih te dolžine</i>	<i>delež povedi te dolžine v celotnem korpusu (48.398 besed) [%]</i>
1	1717	20.38	1717	3.55
2	1084	12.87	2168	4.48
3	891	10.58	2673	5.52
4	783	9.30	3132	6.47
5	699	8.30	3495	7.22
6	536	6.36	3216	6.64
7	477	5.66	3339	6.90
8	411	4.88	3288	6.79
9	295	3.50	2655	5.49
10	256	3.04	2560	5.29
11	199	2.36	2189	4.52
12	202	2.40	2424	5.01
13	117	1.39	1521	3.14
14	117	1.39	1638	3.38
15	103	1.22	1545	3.19
16	73	0.87	1168	2.41
17	77	0.91	1309	2.70
18	57	0.68	1026	2.12
19	56	0.66	1064	2.20
20	51	0.61	1020	2.11
21	26	0.31	546	1.13
22	45	0.53	990	2.05
23	23	0.27	529	1.09
24	27	0.32	648	1.34
25	17	0.20	425	0.88
26	12	0.14	312	0.64
27	13	0.15	351	0.73
28	13	0.15	364	0.75
29	8	0.09	232	0.48
30	6	0.07	180	0.37
31	6	0.07	186	0.38
32	5	0.06	160	0.33
33	3	0.04	99	0.20
34	3	0.04	102	0.21
35	4	0.05	140	0.29
36	4	0.05	144	0.30
38	1	0.01	38	0.08
41	2	0.02	82	0.17
42	2	0.02	84	0.17
44	1	0.01	44	0.09
48	1	0.01	48	0.10

PRILOGA 8: XML-STRUKTURA TRANSKRIBCIJSKE DATOTEKE, EXMARALDA

Program EXMARALDA Partitur Editor, basic transcription. Pred popravki v programu Elan (torej brez zamika časovnih referenc). Za vizualno upodobitev transkripcije gl. Slika 17, str. 135.

```
<?xml version="1.0" encoding="UTF-8"?>          <!-- (c) http://www.rz.uni-hamburg.de/exmaralda -->
<basic-transcription>
<head>
  <meta-information>
    <project-name></project-name>
    <transcription-name>pet-2MD=1</transcription-name>
    <referenced-file url="file:/D:/StAnId/1Snemanje/6petek/pet-2MD=1.mp3"/>
    <ud-meta-information></ud-meta-information>    <comment></comment>
    <transcription-convention></transcription-convention>
  </meta-information>
  <speakertable>
    <speaker id="SPK0">
      <abbreviation>060Žn</abbreviation>
      <sex value="f"/>
      <languages-used></languages-used>          <1></1>          <2></2>
      <ud-speaker-information></ud-speaker-information> <comment></comment>
    </speaker>
    <speaker id="SPK1">
      <abbreviation>001Mz</abbreviation>
      <sex value="m"/>
      <languages-used></languages-used>          <1></1>          <2></2>
      <ud-speaker-information></ud-speaker-information> <comment></comment>
    </speaker>
  </speakertable>
</head>
<basic-body>
  <common-timeline>
    <tli id="T68" time="304.0"/>
    <tli id="T7" time="309.5002593477592"/>
    <tli id="T8"/>
    <tli id="T9"/>
    <tli id="T10"/>
    <tli id="T11"/>
    <tli id="T12"/>
    <tli id="T13" time="316.0"/>
  </common-timeline>
  <tier id="TIE1" category="Ozadje" type="d" display-name="[Ozadje]">
    <event start="T68" end="T7">rez=6pet09Dm</event>
    <event start="T7" end="T8">navzočnost: Srečanje mimoidočih pešcev.</event>
    <event start="T12" end="T13">6pet09Dm=rez</event>
  </tier>
  <tier id="TIE0" speaker="SPK0" category="v" type="t" display-name="060Žn [v]">
    <event start="T7" end="T8">kol'k' je ura?</event>
    <event start="T9" end="T10">ja.</event>
    <event start="T11" end="T12">aha. hvala.</event>
  </tier>
  <tier id="TIE2" speaker="SPK1" category="v" type="t" display-name="001Mz [v]">
    <event start="T8" end="T9">ura?</event>
    <event start="T10" end="T11">deset do dvanajst.</event>
    <event start="T12" end="T13">prosim.</event>
  </tier>
</basic-body>
</basic-transcription>
```

PRILOGA 9: XML-STRUKTURA TRANSKRIBCIJSKE DATOTEKE, ELAN

PROGRAM ELAN

Gre za partituro besedila 6pet09Dm, ki ima glede na tisto iz prejšnje priloge dodate trakove (angl. tier) z metaoznaki besed: [POJAVNICA], [BV], [LEMA]. Kakšna je vizualna upodobitev te partiture, gl. Slika 8, str. 84. Entitete za posebne znake so v spodnjem besedilu zamenjane, npr. $\langle \' \rangle$ v $\langle ' \rangle$ (opuščaj), prav tako je prostorsko skrčena običajna XML-struktura, da je zapis krajši.

```
<?xml version="1.0" encoding="UTF-8"?>
<ANNOTATION_DOCUMENT xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:noNamespaceSchemaLocation="http://www.mpi.nl/tools/elan/EAFv2.3.xsd"
  DATE="2007-08-18T10:42:21+01:00" AUTHOR="" VERSION="2.3" FORMAT="2.3">
<HEADER MEDIA_FILE="" TIME_UNITS="milliseconds">
  <MEDIA_DESCRIPTOR MEDIA_URL="file:///D:/StAnId/4Besedila/Elan/6pet09Dm.mp3" MIME_TYPE="audio/*"/>
</HEADER>

<TIME_ORDER>
  <TIME_SLOT TIME_SLOT_ID="ts1" TIME_VALUE="0"/>
  <TIME_SLOT TIME_SLOT_ID="ts2" TIME_VALUE="5500"/>
  <TIME_SLOT TIME_SLOT_ID="ts3" TIME_VALUE="6584"/>
  <TIME_SLOT TIME_SLOT_ID="ts4" TIME_VALUE="7667"/>
  <TIME_SLOT TIME_SLOT_ID="ts5" TIME_VALUE="8750"/>
  <TIME_SLOT TIME_SLOT_ID="ts6" TIME_VALUE="9833"/>
  <TIME_SLOT TIME_SLOT_ID="ts7" TIME_VALUE="10917"/>
  <TIME_SLOT TIME_SLOT_ID="ts8" TIME_VALUE="12000"/>
</TIME_ORDER>

<TIER TIER_ID="[Ozadje]" PARTICIPANT="" LINGUISTIC_TYPE_REF="Ozadje" DEFAULT_LOCALE="sl">
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a1" TIME_SLOT_REF1="ts1" TIME_SLOT_REF2="ts2">
    <ANNOTATION_VALUE>rez=6pet09Dm</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a2" TIME_SLOT_REF1="ts2" TIME_SLOT_REF2="ts3">
    <ANNOTATION_VALUE>navzočnost: Srečanje mimoidočih pešcev.</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION>
  </ANNOTATION>
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a3" TIME_SLOT_REF1="ts7" TIME_SLOT_REF2="ts8">
    <ANNOTATION_VALUE>6pet09Dm=rez</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
</TIER>

<TIER TIER_ID="060Žn [v]" PARTICIPANT="060Žn" LINGUISTIC_TYPE_REF="v" DEFAULT_LOCALE="sl">
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a4" TIME_SLOT_REF1="ts2" TIME_SLOT_REF2="ts3">
    <ANNOTATION_VALUE>kol'k' je ura?</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a5" TIME_SLOT_REF1="ts4" TIME_SLOT_REF2="ts5">
```

<ANNOTATION_VALUE>ja.</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a6" TIME_SLOT_REF1="ts6" TIME_SLOT_REF2="ts7">
 <ANNOTATION_VALUE>aha. hvala.</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
 </TIER>

<TIER TIER_ID="060Žn [v] [POJAVNICA]" PARTICIPANT="060Žn [v]" LINGUISTIC_TYPE_REF="POJAVNICA" DEFAULT_LOCALE="sl"
 PARENT_REF="060Žn [v]">
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a7" ANNOTATION_REF="a4">
 <ANNOTATION_VALUE>kol'k'</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a8" ANNOTATION_REF="a4" PREVIOUS_ANNOTATION="a7">
 <ANNOTATION_VALUE>je</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a9" ANNOTATION_REF="a4" PREVIOUS_ANNOTATION="a8">
 <ANNOTATION_VALUE>ura?</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a10" ANNOTATION_REF="a5">
 <ANNOTATION_VALUE>ja.</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a11" ANNOTATION_REF="a6">
 <ANNOTATION_VALUE>aha.</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a12" ANNOTATION_REF="a6" PREVIOUS_ANNOTATION="a11">
 <ANNOTATION_VALUE>hvala.</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 </TIER>

<TIER TIER_ID="060Žn [v] [BV]" PARTICIPANT="060Žn [v]" LINGUISTIC_TYPE_REF="BV" DEFAULT_LOCALE="sl" PARENT_REF="060Žn [v]
 [POJAVNICA]">
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a13" ANNOTATION_REF="a7">
 <ANNOTATION_VALUE>A</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a14" ANNOTATION_REF="a8">
 <ANNOTATION_VALUE>GP</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a15" ANNOTATION_REF="a9">
 <ANNOTATION_VALUE>S</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a16" ANNOTATION_REF="a10">
 <ANNOTATION_VALUE>A</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a17" ANNOTATION_REF="a11">
 <ANNOTATION_VALUE>M</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a18" ANNOTATION_REF="a12">
 <ANNOTATION_VALUE>M</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
 </TIER>

<TIER TIER_ID="060Žn [v] [LEMA]" PARTICIPANT="060Žn [v]" LINGUISTIC_TYPE_REF="LEMA" DEFAULT_LOCALE="sl" PARENT_REF="060Žn [v]
 [POJAVNICA]">
 <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a19" ANNOTATION_REF="a7">

```

      <ANNOTATION_VALUE>koliko</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
<ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a20" ANNOTATION_REF="a8">
      <ANNOTATION_VALUE>biti</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
<ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a21" ANNOTATION_REF="a9">
      <ANNOTATION_VALUE>ura</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
<ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a22" ANNOTATION_REF="a10">
      <ANNOTATION_VALUE>ja</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
<ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a23" ANNOTATION_REF="a11">
      <ANNOTATION_VALUE>aha</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
<ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a24" ANNOTATION_REF="a12">
      <ANNOTATION_VALUE>hvala</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>

```

</TIER>

<TIER TIER_ID="001Mz [v]" PARTICIPANT="001Mz" LINGUISTIC_TYPE_REF="v" DEFAULT_LOCALE="sl">

```

  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a25" TIME_SLOT_REF1="ts3" TIME_SLOT_REF2="ts4">
    <ANNOTATION_VALUE>ura?</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a26" TIME_SLOT_REF1="ts5" TIME_SLOT_REF2="ts6">
    <ANNOTATION_VALUE>deset do dvanajst.</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <ALIGNABLE_ANNOTATION ANNOTATION_ID="a27" TIME_SLOT_REF1="ts7" TIME_SLOT_REF2="ts8">
    <ANNOTATION_VALUE>prosim.</ANNOTATION_VALUE> </ALIGNABLE_ANNOTATION> </ANNOTATION>

```

</TIER>

<TIER TIER_ID="001Mz [POJAVNICA]" PARTICIPANT="001Mz" LINGUISTIC_TYPE_REF="POJAVNICA" DEFAULT_LOCALE="sl" PARENT_REF="001Mz [v]">

```

  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a28" ANNOTATION_REF="a25">
    <ANNOTATION_VALUE>ura?</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a29" ANNOTATION_REF="a26">
    <ANNOTATION_VALUE>deset</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a30" ANNOTATION_REF="a26" PREVIOUS_ANNOTATION="a29">
    <ANNOTATION_VALUE>do</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a31" ANNOTATION_REF="a26" PREVIOUS_ANNOTATION="a30">
    <ANNOTATION_VALUE>dvanajst.</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>
  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a32" ANNOTATION_REF="a27">
    <ANNOTATION_VALUE>prosim.</ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>

```

</TIER>

<TIER TIER_ID="001Mz [BV]" PARTICIPANT="001Mz" LINGUISTIC_TYPE_REF="BV" DEFAULT_LOCALE="sl" PARENT_REF="001Mz [POJAVNICA]">

```

  <ANNOTATION> <REF_ANNOTATION ANNOTATION_ID="a33" ANNOTATION_REF="a28">
    <ANNOTATION_VALUE></ANNOTATION_VALUE> </REF_ANNOTATION> </ANNOTATION>

```

</TIER>


```

<TIER TIER_ID="001Mz [LEMA]" PARTICIPANT="001Mz" LINGUISTIC_TYPE_REF="LEMA" DEFAULT_LOCALE="sl" PARENT_REF="001Mz
[POJAVNICA]">
  <ANNOTATION>      <REF_ANNOTATION ANNOTATION_ID="a34" ANNOTATION_REF="a28">
    <ANNOTATION_VALUE>ura</ANNOTATION_VALUE>      </REF_ANNOTATION>      </ANNOTATION>
  <ANNOTATION>      <REF_ANNOTATION ANNOTATION_ID="a35" ANNOTATION_REF="a29">
    <ANNOTATION_VALUE>deset</ANNOTATION_VALUE>    </REF_ANNOTATION>    </ANNOTATION>
  <ANNOTATION>      <REF_ANNOTATION ANNOTATION_ID="a36" ANNOTATION_REF="a30">
    <ANNOTATION_VALUE>do</ANNOTATION_VALUE>      </REF_ANNOTATION>    </ANNOTATION>
  <ANNOTATION>      <REF_ANNOTATION ANNOTATION_ID="a37" ANNOTATION_REF="a31">
    <ANNOTATION_VALUE>dvanajst</ANNOTATION_VALUE> </REF_ANNOTATION>    </ANNOTATION>
  <ANNOTATION>      <REF_ANNOTATION ANNOTATION_ID="a38" ANNOTATION_REF="a32">
    <ANNOTATION_VALUE>prosim</ANNOTATION_VALUE>  </REF_ANNOTATION>    </ANNOTATION>
</TIER>

<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="Ozadje" TIME_ALIGNABLE="true" GRAPHIC_REFERENCES="false"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="v" TIME_ALIGNABLE="true" GRAPHIC_REFERENCES="false"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="POJAVNICA" TIME_ALIGNABLE="false" GRAPHIC_REFERENCES="false"
  CONSTRAINTS="Symbolic_Subdivision"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="BV" TIME_ALIGNABLE="false" GRAPHIC_REFERENCES="false" CONSTRAINTS="Symbolic_Association"
  CONTROLLED_VOCABULARY_REF="Besedne vrste StAnId_poskus"/>
<LINGUISTIC_TYPE LINGUISTIC_TYPE_ID="LEMA" TIME_ALIGNABLE="false" GRAPHIC_REFERENCES="false" CONSTRAINTS="Symbolic_Association"/>

<LOCALE LANGUAGE_CODE="sl" COUNTRY_CODE="SI"/>

<CONSTRAINT STEREOTYPE="Time_Subdivision" DESCRIPTION="Time subdivision of parent annotation&apos;s time interval, no time gaps allowed within this interval"/>
<CONSTRAINT STEREOTYPE="Symbolic_Subdivision" DESCRIPTION="Symbolic subdivision of a parent annotation. Annotations referring to the same parent are ordered"/>
<CONSTRAINT STEREOTYPE="Symbolic_Association" DESCRIPTION="1-1 association with a parent annotation"/>
<CONSTRAINT STEREOTYPE="Included_In" DESCRIPTION="Time alignable annotations within the parent annotation&apos;s time interval, gaps are allowed"/>

<CONTROLLED_VOCABULARY CV_ID="Besedne vrste StAnId4" DESCRIPTION="">
  <CV_ENTRY DESCRIPTION="medmet">M</CV_ENTRY>
  <CV_ENTRY DESCRIPTION="glagol">G</CV_ENTRY>
  <CV_ENTRY DESCRIPTION="glagol pomožni biti">GP</CV_ENTRY>
  <CV_ENTRY DESCRIPTION="samostalnik">S</CV_ENTRY>
  <CV_ENTRY DESCRIPTION="prislov">A</CV_ENTRY>
  <CV_ENTRY DESCRIPTION="števnik">Š </CV_ENTRY>
  <CV_ENTRY DESCRIPTION="predlog">E</CV_ENTRY>
</CONTROLLED_VOCABULARY>
</ANNOTATION_DOCUMENT>

```

PRILOGA 10: PRIMER PARTITURE IZ KORPUSA – BESEDILO 1NED19Vg

Partitura transkripcijske datoteke 1ned18Vg.xml je v programu EXMARALDA Partitur Editor izvoževna v dokument formata RTF – ta je prikazan spodaj. Vzporedni posnetek: 1ned19Vg.mp3.

	0 [0.]	1 [2.]	2 [4.9]	3 [7.6]
[Ozadje]	<i>rez=1ned19Vg prostor: Zaklepanje mežnarije, dalje hoja proti domu.</i>			
008Žz [v]	ne kar boš šla peš! *ehehe.		pa šprica neka' dol z megle!	
[aud]	1ned19Vg-posnetek-1			
	4 [9.3]	5 [11.0]	6 [12.8]	7 [22.]
008Žz [v]	z megle prši z megle!		() *vjutro original ko' da bi en *mladi *kuj zahrzal	
[aud]	1ned19Vg-posnetek-2			
	8 [27.2]		9 [30.6]	
001Mz [v]				kak' *te lahk'
008Žz [v]	veš kak' ... ampak ji *pal ni ratalo več samo tak' neka' se je *špilala.			
008Žz [pv]	<i>nvg: smeh</i>			
[aud]				
	10 [31.5]	11 [32.5]	12 [33.5]	13 [34.]
001Mz [v]	žvižga?	tisto je bolj tak': *eijuvhohoa!		
001Mz [pv]	<i>nvg: oponašanje konjskega hrzanja</i>			
008Žz [v]	ži- zažvižgala je tak'!	ja samo veš		
008Žz [pv]				
[aud]				
	..			
008Žz [v]	kak' je b'lo čist' or'iginalno tak' al' tak' samo pač *ko je malo višje. al'. veš kak'			
[aud]				
	14 [40.2]	15 [42.]	16 [43.9]	
008Žz [v]	*frajersko!		seveda *s'n čula *s'n pa *te tam zdrav'n	
008Žz [pv]	<i>nvg: smeh</i>			
019Žz [v]	pa ti si slišala al' ka'?			
[aud]				
	17 [46.5]	18 [48.2]		
008Žz [v]	tebe stala v dnevni sobi.		ja ti si b'la v kuh'nji jaz pa *s'n b'la ene meter *bek	
028Žz [v]	*hehe v dnevni sobi () b'la v kuh'n'i ...			
[aud]				
	19 [50.0]	20 [54.0]	21 [58.]	22 [59.4]
[Ozadje]	<i>signal: Oglašanje ptiča.</i>			
008Žz [v]	od tebe v dnevni sobi ka' z'aj?		kaj je to za *ni *tič?	a je *toto sraka? 'lej
028Žz [v]				
[aud]	1ned19Vg-posnetek-3			
	..			
	24 [71.2]			25 [71.8]
001Mz [v]				26 [72.5]
008Žz [v]				27 [73.5]
[aud]	jako v'lko je tam v oni- onem ta drugem *orehi je! 'lej!			
	28 [74.4]	29 [75.4]	30 [76.4]	31 [77.3]
001Mz [v]	pisana ...		*letan'ce 'ma m- modro-bele.	*mhmhm *hnmhmhm.
008Žz [v]	je al'?		več ka' *s'n vid'la	
[aud]	1ned19Vg-posnetek-4			
	35 [89.9]	36 [91.5]	37 [93.1]	38 [94.2]
001Mz [v]	a res?			
008Žz [v]	zadnjič po ne vem *keko letih. zelen'ga det'la.		a veš kak' redki so že? tak' je	
[aud]				

		40 [98.7]	41 [101.6]
008Žz [v] [aud]	bil lep! tam *spod za ... čakaj kje smo *te b'li? tu spoda' nekje ... pr' *Garneži men'		
		42 [106.1]	43 [108.8] 44 [114.6]
008Žz [v] [aud]	se zdi smo *zun sedeli al' kje ne vem ... pa je ... le- veš kak' lep je bil! se mi		1ned19Vg
		45 [117.1]	
008Žz [v] 008Žz [pv] [aud]	zdi da prav zimski mraz ej! tak' *s'n se *gir oblekla tak' da mi je jakna premala izg: smeh -posnetek-5		
		46 [122.5]	47 [123.7] 48 [128.9] 49 [130.7]
001Mz [v] 008Žz [v] 008Žz [pv] [aud]	zdej sta pa tiho z'aj sta pa tiho. *gratala! tak' 'mam *spod *gir oble- ... 'lej *joji! ka' pa *te je vidva? :izg 1ned19Vg-posnetek-6		
		50 [132.6] 51 [133.1] 52 [135.]	53 [139.6] 54 [140.3] 55 [141.0]
008Žz [v] 019Žz [v] 028Žz [v] [aud]	ja! *toti mladi rjavi so mu *mogli neka' ... *naret' enkrat. ja. ja. *naret' *ala? ker je		
		56 [141.7] 57 [142.9]	
008Žz [v] 028Žz [v] [aud]	veš kak' se je žulil. pa vedno se je požulil ko smo šli mimo je pr'letel pa se bil cel' *cajt cel' *cajt je bil *ful () ...		
		58 [144.4]	59 [146.7] 60 [149.0]
001Mz [v] 008Žz [v] 028Žz [v] [aud]	ja. veš tisto prijazno ... pa *kuj *naenkret je pa z'aj je pa *taki *hudaben		
		61 [151.6]	62 [153.] 63 [154.5] 64 [155.8]
001Mz [v] 008Žz [v] [aud]	() ko je Monika umrla al' ka'? *gratal. tak' da mu je *mogel eden neka' *naret' veš ko *te je () ... ne ne ne ne		
		65 [156.8]	66 [157.2] 67 [158.1]
001Mz [v] 008Žz [v] [aud]	že dolgo prej? ne *toto je že la- ... že prej je začel. *kuj veš tisto tak' da je po mojem kaki *toti ...		
		68 [163.1]	69 [165.8] 70 [166.1]
001Mz [v] 008Žz [v] [aud]	ja. bom čist' z'aj rekla *froc' *ko so malo bolj žleht ne ... vprašanje ka' mu je		
			71 [170.2] 72 [174.]
[Ozadje] 008Žz [v] 019Žz [v] [aud]	naredil ne. al' mu je vrgel kaj al' ga je mogoče ka' *te veš ... (=028Žz_ime) a 1ned19Vg-posnetek-7		
		73 [175.2]	74 [177.6] 75 [179.2]
[Ozadje] 001Mz [v] 008Žz [v] 019Žz [v] 028Žz [v] [aud]	ja. to je pa sigurno bil ... 'lej dež je začel *jit! si že ti čula da bi () 'mela hlače za *ednajst *jurjov? *ha? () za		
		76 [179.7]	77 [181.8] 78 [184.1] 79 [185.0] 80 [185.4]
008Žz [v]	jaz *s'n mislila da me je *špuknilo. kak' je onemu malemu ime? Tiger. laja!		

019Žz [v]						Tiger.
028Žz [v]	enajst *jurjov.					
[aud]		1ned19Vg-posnetek-8				
	81 [186.3]	82 [188.1]	83 [189.9]	84 [192.0]	85 [192.9]	
001Mz [v]						*keriga?
008Žz [v]						a si videl onega *pesa? Tiger mu je ime!
008Žz [pv]	nvg: smeh					
019Žz [v]						*a *a Tiger ne zna lajat'!
028Žz [v]	pa meni se ni ().					
[aud]						
	86 [193.7]	87 [196.0]	88 [197.7]			
[Ozadje]	šum: Vožnja avtomobila mimo.					
001Mz [v]						ne ne vem.
008Žz [v]	Brezočnikovega.					Brezočniki () Rolando ne 'majo nemškega ovčarja pa
[aud]						
	..	89 [199.4]	90 [204.8]	91 [208.]		
001Mz [v]						*ovi *ko so se *toti teden z bagrom *kapoli
008Žz [v]	en'ga psa *ko mu je ime Tiger. in *ovi Tiger pr'leti pa je taki mali ...					
008Žz [pv]						nvg: smeh
[aud]						
	..	92 [208.8]	93 [210.5]	94 [211.1]	95 [211.9]	96 [212.7]
001Mz [v]	*e v petek al' ka' je že?	ja.				*ehe ja ...
001Mz [pv]						nvg: smeh
008Žz [v]						*ehe ... taki mal' ...
008Žz [pv]						nvg: smeh
019Žz [v]						pa *toti ni več *taki lepi!
028Žz [v]		ja.				
[aud]			1ned19Vg-posnetek-9			
	98 [214.3]		99 [216.9]	100 [219.6]	101 [220.1]	
001Mz [v]						
008Žz [v]	ja mali je bil *ful *luškani. ko je bil še dojenček še ...					
019Žz [v]	() ni več taki lep. ja mali je bil samo z'aj ...					(=028Žz_ime)! a veš
028Žz [v]						ja.
[aud]						
	..	103 [220.8]	104 [223.3]			
008Žz [v]	tisti je z'aj lajal. si videl kako laja'je!					*bevkaepepev!
008Žz [pv]						nvg: oponašanje lajanja - bevskanja
019Žz [v]	da je () ...					mama a veš zrastel sploh ni ...
[aud]						
	105 [225.9]	106 [227.4]	107 [229.]			
008Žz [v]						ja spremenil se je!
008Žz [pv]						
019Žz [v]	zrastel sploh ni ampak z'aj zgle'a zato *teko ... ne! zgle'a *teko večki zaradi tega					
[aud]						
	..	108 [233.2]	109 [234.2]	110 [235.7]		
019Žz [v]	ker se mu je dlaka baj' mu je za pet centimetrov z'aj zrastla.					je rekla Mariša
028Žz [v]	ja pa ne pa dr'gačen je čisto!					glava je čist' dr'
[aud]						
	..	111 [238.0]	112 [240.4]			
[Ozadje]						<i>razno: Dež pada vedno</i>
008Žz [v]						pa tak' je bil ko'
019Žz [v]	da so ga kopa- ... ko so ga kopali je baj' tak' kot podgana zgle'al!					
028Žz [v]	gačna ...					
[aud]			1ned19Vg-posnetek-10			
	..	113 [244.]	114 [245.]			

[Ozadje]	<i>bolj gosto.</i>		<i>razno: (=001Mz) odpre dežnik.</i>
008Žz [v]	smo mi našega *mačeka skopali ma ...		ja se pa *te Uha tud'
028Žz [v]		pa mama dež gre!	
[aud]			
..		115 [248.9] 116 [249.5]	
[Ozadje]			
008Žz [v]	samo Uha je bil še debeli *te.	pa vseeno ker kratka dlaka pa ko si ga zmočil	
019Žz [v]		() jaz ga ne bom 'mela (). jaz *s'n z'aj ().	
[aud]			
..		117 [253.4]	118 [254.9]
008Žz [v]	pa je bil tak' ko' podgana. o hvala (=001Mz_ime).	daj z'aj da ne *boma preveč	
019Žz [v]			
[aud]			
..		119 [257.8] 120 [261.6]	121 [262.9] 122 [266.7]
008Žz [v]	žagala *ehehe.	vredi ga je vžgalo no.	z'aj pa *gir *k'ne zna odpret'
[aud]		<u>1ned19Vg-posnetek-11</u>	
..		123 [269.5]	124 [271.4] 125 [273.0] 126 [274.5]
001Mz [v]		a joj joj!	*kuj ja *kuj *ofnaj jo *kuj
001Mz [pv]		<i>nvg: smeh</i>	
008Žz [v]	marele!	*o *hehehaha!	hvala
008Žz [pv]		<i>nvg: smeh</i>	
019Žz [v]	*eh saj bomo *skoro doma.		
[aud]			
..		127 [275.8]	128 [277.2]
001Mz [v]	*ofnaj jo jaz *s'n jo tudi. *fajn bod'te dober tek *Romija pozdrav'te!		
008Žz [v]	(=001Mz_ime) adijo!		ja () hvala!
[aud]			
..		129 [278.7]	
[Ozadje]	<i>1ned19Vg=rez</i>		
[aud]			

PRILOGA 11: IZJAVA O AVTORSTVU IN PODATKI AVTORJA

Spodaj podpisani Rok Bečan izjavljam,
da je napisana diploma moje samostojno avtorsko delo.
Vsi viri in literatura in vsa programska oprema,
ki sem jo pri delu uporabljal, so pregledno označeni in navedeni.

V Lovrencu na Pohorju, 26. 9. 2007

Rok Bečan, r. 1981
Ob Radoljni 58
2344 Lovrenc na Pohorju, Slovenija
rok.becan@skavt.net
<http://www.soncek.net/rok/>

PRILOGA 12: ZGOŠČENKA: PRIMERI POSNETKOV IDR.

Pregled vsebine zgoščenke je podan v poglavju III.13.3, Tabela 38, str. 146. Tukaj navajam le razloge za primere besedil iz korpusa, ki so na zgoščenki v mapi **\izbraniposnetki** predstavljeni s transkripcijskimi datotekami in in posnetki. Za podrobnejše podatke o posameznem besedilu gl. Priloga 1.

<i>besedilo</i>	<i>razlog za izbor – njegova zanimivost</i>
1ned04Vg	primer pogovora govorcev, ki v korpusu spregovorijo največ besed in kot družina na domu predstavljajo najtipičnejše sporazumevalne okoliščine v korpusu
1ned17Vg	primer večgovora (7 govorcev) in besedilo z najpogostejšim govorom v korpusu (3,73 besede/s)
1ned19Vg	besedilo, katerega partitura je predstavljena v Prilogi 10
1ned57Dg	primer dialoga med vožnjo (z avtom), hkrati pa edino besedilo v korpusu, v kateri se sogovorec nosilca mikrofona zaveda snemanja oz. razkrije snemanje proti koncu pogovora
6pet42Dg	primer dialoga po mobitelu