

Govorjeni proti pisnemu ali katera leksika je »tipično govorjena«

Darinka Verdonik

Abstract

This paper aims to resolve two key issues concerning the usage of the GOS speech corpus as a data source for lexicographic descriptions: (i) the words which are especially common in the GOS corpus; and (ii) whether it is appropriate to mark such words or some of their usages as »spoken«. Based on the key-word corpus analysis performed, seven different groups of words were defined as being common in the GOS corpus: non-verbal expressions; discourse markers; stylistically marked lexis; deixes; expressions for speech acts that are common in spoken usage; expressions that are »favoured« in spoken usage; and forms that are more favoured in spoken than in written usage. The discussions conclude with the argument that marking these words or some of their usages as »spoken« is too superficial an annotation because the factors which encourage their usage are connected with different levels of discourse, such as formality, spontaneity, common speech acts, etc. We suggest more specific register-stylistic marks, such as »non-formal«, »in conversation«, »expression of attitude«, »regionally specific usages«, etc.

Keywords: speech, speech corpus, spoken lexis

Ključne besede: govor, govorni korpus, govorjena leksika

1 UVOD

V zadnjem desetletju in več lahko sledimo porastu govornih korpusov za številne jezike (gl. Verdonik et al. 2013). Pogosto so bili uporabljeni v različnih diskurzni in pragmatičnih študijah (npr. Aijmer 1996; Martinez 2011; pri nas npr. Verdonik in Kačič 2012), tisti, ki so bili spodbujeni z govornimi tehnologijami, so pomenili ključen dejavnik predvsem v razvoju razpoznavanja tekočega govora (Žgank et al. 2008; Golik et al. 2013), svojo nišo išče t. i. korpusno jezikoslovje govora (Adolphs in Carter 2013), zlasti v angleškem jezikoslovju pa je že več desetletij aktualno tudi vprašanje, kako ravnati z govornimi korpusi kot leksikografskimi viri (Svartvik 1992; Moon 1998; De Cock 2002). Tudi v zadnjem desetletju zasledimo vsaj dve vidnejši, eksplicitni obravnavi tega vprašanja v mednarodnih leksikografskih publikacijah.

Trap-Jensen (2004) na podlagi izkušenj z danskim slovarjem poudarja, da so iz govornega dela korpusa, uporabljenega za izdelavo slovarja, dobili velik del informacij, vključenih v slovar. Med leksiko, ki so jo označili kot rabljeno »zlasti v govorjenem jeziku«, naštevajo medmete in onomatopeje¹ (*mhm, aha, eee, kikiriki*), diskurzne označevalce (*veš, mislim, pol*), pragmatične fraze (*oprostite, ti povem, kaj ne poveš, daj no*), deiktčne zaimke in prislove (*ta, oni*), kletvice (*presneto, hudiča*) ter slengovske in nestandardne izraze (*ali pa kaj takega, daj nehaj*), vključno z regionalizmi in narečno leksiko, ki je razširjena čez lokalne meje svojega izvora. Ob naštetih skupinah od deiktčnih izrazov naprej že avtor sam izrazi pomislek, ali jih je bolje označiti kot »neformalne« ali kot »značilne za govorjeno rabo«, pri regionalizmih pa se sprašuje, ali jih ne bi bilo bolje označiti glede na regionalno razširjenost. Ob naštetem ni zanemarljivo, da navedeni danski slovar temelji na korpusu obsega (samo) 40 mio. besed, od tega 8 mio. v govorjeni rabi in 32 mio. v pisni. Za slednje je korpus takega obsega z današnjega gledišča (prim. Gigafida – 1 milijarda besed) zelo majhen. Trap-Jensen (2004) argumentacija, da brez 315 eksplicitnih navedkov, ki so jih vključili na podlagi govorjenega gradiva v njihovem korpusu, izdelani danski slovar ne bi bil popoln, je zaradi tega lahko nekoliko manj prepričljiva.

Siepmann (2015), nasprotno, nima težav s pomanjkanjem gradiva po današnjih standardih: govorni korpus, ki ga uporablja, CRFC – Referenčni korpus sodobne francoščine (Siepmann in Bürgel 2015), obsega 155 mio. besed v govorjeni rabi, podatke v njem pa primerja z opisi v 10 različnih obstoječih in najširše uporabljenih eno- in dvojezičnih slovarjih francoskega jezika. Ob tem dokazuje, da nobeden od obstoječih francoskih slovarjev za izbrane analizirane

1 Navajamo sorodne slovenske zglede, ki smo jih izbrali na podlagi danskih zgledov v Trap-Jensen (2004). Dobesedno prevajanje danskih zgledov ne bi imelo smiselne učinka, saj so v različnih jezikih za opravljanje podobnih pragmatičnih funkcij rabljeni različni izrazi. Zgledi so navedeni izključno z namenom, da si bralec bolje predstavlja, kaj v danskem slovarju razumejo pod navedenimi poimenovanji.

štiri enobesedne in tri večbesedne leksikalne enote ne opisuje nekaterih najbolj pogostih, vsakdanjih leksikalnih vzorcev in kolokacij, ki jih najdemo le v govornih korpusih, v pisnih, še tako velikih, pa ne. Čeprav Siepmann (2015) meni, da bi slovaropisci te značilnosti govorjene rabe lahko opisali na podlagi lastne jezikovne intuicije, pa je po drugi strani jasno, da je dodajanje opisov pomenov ali zgledov na podlagi lastne jezikovne intuicije v slovaropisju (vedno bolj) nesprejemljivo. Prav zaradi tega pa ima Siepmannov članek še večjo težo, s tem ko dokazuje, da lahko (1) srednje veliki korpusi, kot je CRFC, osvetlijo leksikalne vzorce in kolokacije, o katerih iz še tako velikih pisnih korpusov ne izvemo nič, in da (2) imajo t. i. govorjene (angl. *colloquial*) leksikalne enote svoje lastne pomenske odtenke, frazeologijo in pragmatiko ter nikakor niso slogovno manj vredni nadomestki za bolj formalne enote (povzeto iz Siepmann 2015). Tovrstna leksika in vzorci njene rabe so po mnenju Siepmanna izredno globoko zakoreninjani v rutinah vsakdanjega življenja, hkrati pa v veliki meri še neopisani.

Oba predstavljena članka skušata dokazati, da se v govorjeni rabi z leksikalnega vidika pojavljajo rabe, ki bi sodile v slovarski opis, a tam pogosto manjkajo. Medtem ko je Trap-Jensen (2004) nakazoval, da lahko vrsto leksikalnih enot, od neverbalnih prek deiktčnih do nestandardnih in regionalno značilnih izrazov, označimo kot tipično govorjene same po sebi, se je Siepmann (2015) osredotočal predvsem na nove pomenske in funkcijske odtenke v govorjeni rabi ter tipično govorjene kolokacije za nekaj izbranih, pogosto rabljenih izrazov.

V tem prispevku bomo vprašanje, kaj je tipično za govorjeno rabo, dodatno raziskali. Ob korpusih, ki so na voljo za sodobno slovenščino, nas bo zanimalo, katera leksika se v govorjeni rabi pojavlja bolj pogosto kot v pisni, kaj so razlogi za to in ali jo smemo samo zaradi njene večje frekventnosti v slovarju označiti kot »tipično za govorjeno rabo«. Pri tem se bomo omejili na enobesedne leksikalne enote.

2 JEZIK V GOVORJENI IN PISNI RABI

Raziskave, ki primerjajo jezik v obeh kanalih, govornem in pisnem, rade izpostavljajo razlike. Vse od 20-ih let 20. stoletja so se vrstile analize (pretežno na angleščini), ki so dokazovale kvantitativne razlike v dolžini besed, povedi, besedil, variabilnosti slovarja, kasneje, v drugi polovici 20. stoletja, tudi razlike v težavnosti, abstraktnosti, zgoščenosti in razumljivosti besedil v enem in drugem kanalu ter v distribuciji besed v njih. Brez dvoma prehod iz govornega v pisni kanal in obratno spremljajo tudi spremembe v vzorcih leksikalne izbire, povzame pregled teh raziskav Akinnaso (1982: 103).

Tudi v 80-ih letih in vse do danes se nadaljujejo predvsem kvantitativne primerjave govornih in pisnih besedil. Avtor številnih je Biber (npr. 2009; 2012; et al.), ki s korpusnojeziškoslovnim pristopom dokazuje, da so znatne razlike v leksikalnih vzorcih med neformalnim pogovorom in akademskim/znanstvenim pisanjem. Med drugim opozarja, da splošni opis leksikalno-skladenjskih vzorcev, kot ga najdemo v Longmanovi slovnici govornih in pisnih angleščine (Biber et al. 1999), vključuje nekatere značilnosti, ki jih ne najdemo niti v neformalnem pogovoru na eni strani niti v znanstvenih besedilih na drugi strani, ob tem pa ne vključuje nekaterih pomembnih značilnosti, ki jih najdemo v enem ali drugem tipu diskurza (Biber 2012).

Toda faktorjev, ki vplivajo na razlike v jezikovni rabi v neformalnem pogovoru na eni in znanstvenih besedilih na drugi strani, je veliko, ne samo različen kanal. To bi lahko vrglo nekaj dvoma na zaključke Biberjevih raziskav. Vendar tudi raziskave z manj diametralno nasprotnim gradivom dokazujejo pomembne razlike. Chafe in Danielewich (1987) primerjata jezikovne vzorce v angleščini v različnih žanrih in kanalih komunikacije v akademskem okolju. V raziskavo zajameta vedno istih 20 govorcev, zastopani žanri pa so: predavanja, pogovori v univerzitetnem kampusu, pisma in znanstveni članki. Na ravni leksike ugotovita, da je tako v predavanjih kot v pogovorih (torej obeh žanrih govornih rabe) slovar relativno omejen v primerjavi s pisnima žanroma, uporaba omejevalcev (angl. *hedges*) je večja in referenčna neeksplicitnost pogostejša. Heinrichsen in Allwood (2005) primerjata neformalne pogovore in sociolingvistične intervjuje s časopisnimi besedili v danščini in švedščini ter trdita, da je več podobnosti med govorno rabo v danščini in švedščini (torej dveh sorodnih, a različnih jezikih) kot med pisno in govorno rabo v istem jeziku, danskem ali švedskem. Razlike opazujeta v razpršenosti slovarja, natančnosti izražanja, distribuciji besed in besednih vrst itd.

Z razmahom raziskav diskurzivnih označevalcev se pojavijo tudi razprave, ki utemeljujejo, da lahko tovrstne vloge leksikalnih enot najdemo predvsem v govornih rabi in da predstavljajo na leksikalni ravni eno najbolj izrazitih razlik v razmerju do pisne rabe. Adolphs in Carter (2003) tako na primeru angleške besede *like* dokazujeta, da nam uporaba govornega korpusa pomaga odkriti značilnosti govornih rabe, ki pred tem niso bile sistematično identificirane. Podobno dokazuje Siepmann (2015) za sedem izbranih izrazov v francoščini, kot smo že predstavili v uvodu.

Čeprav od Akinnasovega (1982) pregleda raziskav, ki primerjajo govorno in pisno rabo, mineva več kot trideset let, so njegovi zaključki še vedno aktualni. Ozko kvantitativna primerjava leksikalnih (in skladenjskih) enot lahko da popačeno sliko, trdi, in argumentira, da je treba razlike opazovati iz širšega, diskurzivnega stališča, saj lahko nanje vpliva veliko faktorjev, ne samo govornost

oz. pisnost: formalnost proti neformalnosti, tema pogovora, narava komunikacijskih ciljev itd. Govorjena in pisna raba jezika sta v mnogo pogledih komplementarni, zatrjuje Akinnaso: nekatere komunikacijske potrebe večinoma zadovoljujemo skozi govorno rabo, druge prevladujoče skozi pisno (1982). Te funkcijske razlike vplivajo na leksikalne, skladenjske in semantično-pragmatične lastnosti besedil v enem in drugem načinu jezikovne rabe. Ob tem, da predstavljene raziskave brez dvoma dokazujejo opazne razlike med obema načinoma jezikovne rabe, je zato treba temeljito premisliti, kaj so najpomembnejši faktorji, ki v resnici vplivajo na to.

3 GRADIVO IN METODA

Za primerjavo razlik v leksiki med govorno in pisno rabo smo uporabili korpusa Gos (Verdonik et al. 2013) in Gigafida (Logar Berginc et al. 2012). Oba sta dostopna tudi prek orodja NoSketchEngine (Erjavec 2013; Kilgariff et al. 2014), kjer so mogoče tudi nekatere statistične primerjave besedil v njih.

Analiza ključnih besed (Scott 1997) je dobro znana korpusna metoda, po kateri naredimo seznam leksike, ki po pogostosti rabe v analiziranem korpusu najbolj odstopa od rabe v referenčnem korpusu, iz česar sklepamo, da je za analizirani korpus ključna – tipična. Nas zanima, katera leksika je ključna za korpus govorne slovenščine Gos, iz česar bi lahko sklepali, da je tipična za govorno rabo. Gigafida nam pri tem služi kot referenčni korpus. Za vsako različnico izračunamo, kolikokrat se pojavi na 1 milijon pojavnic v analiziranem (Gos) in referenčnem korpusu (Gigafida). Da se izognemo matematičnemu problemu, kadar se katera od različnic v enem od korpusov ne pojavi niti enkrat, prištejemo obema vrednostma parameter n ($n = 100$), nato pa delimo število pojavitev v analiziranem korpusu s številom pojavitev v referenčnem korpusu in rangiramo različnice glede na dobljeni faktor (Lexical Computing Ltd. 2013).

Ker nas zanima leksika z vidika slovarske obravnave, ta pa poteka znotraj slovarskih iztočnic, smo izračune delali na lemah, ne pa na besednih oblikah. Lematizacija obeh korpusov je bila avtomatska, zato smo pri tem zajeli tudi napake, ki nastanejo pri lematizaciji. Naslednji korak po izdelavi prvih seznamov ključnih lem je tako bil popravljanje napačno lematiziranih enot, ki so se pojavljale na prvih seznamih.

Ko smo dobili sezname ključnih lem, na katerih ni bilo več očitno napačno lematiziranih enot, smo opravili ročno analizo rezultatov. Za vsako od prvih 100 ključnih lem smo pregledali konkordance v korpusu Gos, po potrebi primerjali s konkordancami v korpusu Gigafida, iskali morebitne izstopajoče kolokacije,

pregledovali vrste rab in jih primerjali z rabo najbližjih sopomenk. Nazadnje smo skušali med prvimi 100 ključnimi lemami definirati skupine, ki so povezane s skupnimi vzroki za večjo frekventnost v korpusu Gos.

Nadaljevanje seznama ključnih lem od 100 naprej, s poudarkom na prvi polovici seznama, smo samo pregledali ter iskali morebitne dodatne, med prvimi 100 ključnimi lemami neopažene, a vseeno pogosto izstopajoče skupine in vzorce.

4 REZULTATI

Definiranje skupin lem, ki so povezane s skupnim vzrokom za večjo frekventnost v korpusu Gos, ni mogoče brez nekaj posploševanja. Rabe posamezne leme so pogosto različne, enako razlogi zanje. Upoštevali smo predvsem tisto, kar je najbolj izrazito in pogosto. Nekatere leme smo tako uvrstili v več skupin.

4.1 Neverbalni izrazi

Na vrhu seznama za govorno rabo ključnih lem se pojavljajo nekateri neverbalni elementi, tipični za spontano govorno produkcijo: to so *eee*, *eem*, *mmm*, *aaa*, *nnn*, pogosto označeni kot označevalci vrzeli, lahko pa jim pripišemo tudi pomembne metadiskurzne vloge (npr. pri menjavanju vlog). Ti izrazi se v pisni rabi praviloma ne pojavljajo in zanje niti še ni splošno sprejete standardizacije zapisa – ta se šele poskuša vzpostaviti prek korpusa Gos. Tudi če najdemo iste nize v pisnem korpusu Gigafida, imajo zato praviloma drugačno vlogo in pomen: lahko so napaka v lematizaciji, kratica, v enaki vlogi kot v govorni rabi pa le izjemoma, pri premem govoru.

S slovarskega vidika se pri teh izrazih postavlja vprašanje njihovega mesta v slovarju: kot jezikovni elementi, ki niti niso jasno verbalizirani, so zelo mejne enote in jih do zdaj slovarji niso vključevali v svoje opise.

4.2 Diskurzni označevalci

Naslednja skupina lem, ki izstopajo po pogostosti v govorni rabi, so tiste, ki lahko v posameznih oblikah nastopajo tudi kot t. i. diskurzni označevalci. Na seznamu prvih 100 ključnih lem jih je precej: *mhm*, *aha*, *aja*, *veš/veste* (*vedeti*), *okej*, *v bistvu* (*bistvo*), *glej/te* (*gledati*), *poglejte* (*pogledati*), *zdaj*, *mislim* (*misliti*), *vidiš/te* (*videti*), *v redu* (*red*), *nič*, *dobro*, *jah*.

Za diskurzne označevalce (Verdonik 2007) je značilno, da ne prispevajo k propozicijski vsebini, ampak imajo prevladujoče različne metadiskurzne vloge: vzpostavljajo povezave z vsebino predhodnega ali/in prihodnjega diskurza, pomagajo vzpostavljati in vzdrževati odnos med sogovornikoma/i, izražajo odnos/držo govorca do propozicijske vsebine oz. pomagajo organizirati potek diskurza (menjavanje vlog, menjavanje tem, zaključevanje pogovora) – pri čemer niso specializirani, ampak lahko posamezen diskurzni označevalec opravlja več vlog hkrati, velikokrat pa obstaja tudi močna vez s predstavnim pomenom in enoumno razlikovanje med rabo v diskurzni vlogah in propozicijski vsebini ni mogoče. Diskurzni označevalci so bili najbolj eksplicitno opaženi prav v govornem rabi (Schiffrin 1987), kar pa ne pomeni, da jezikovnih sredstev z (delno) podobnimi funkcijami ni tudi v pisni rabi, kjer so znani predvsem kot metabesedilo (prim. Pisanski Peterlin 2011).

4.3 Slogovno zaznamovana leksika

Kot posebno skupino lem smo označili tiste, ki jih slovensko normativno jezikoslovje označuje kot tako ali drugače jezikovno slogovno zaznamovane – po Toporišču (2000) so označene kot neknjižne zvrsti in interesne govornice. Čeprav bi morda pričakovali, da bodo seznam za govorno rabo ključnih lem prevladujoče sestavljale leme iz te skupine, se v resnici njihovo število giblje nekaj nad eno desetino vseh ključnih lem. Na seznamu prvih 100 ključnih lem je namreč 8 takih, ki so v slovenskih referenčnih slovarjih (pravopisni in SSKJ2) že same po sebi označene kot slogovno zaznamovane: *ful*, *glj*, *fajn*, *evo*, *pizda*, *cajt*, *izgledati*, *okej*. Še dodatnih 6 lem je takih, ki se v govornem korpusu pogosto rabijo v pomenih, ki so označeni kot taki: *pol* v pomenu 'potem', *rabiti* v pomenu 'potrebovati, biti treba'; *znati* v pomenu 'vedeti'; *drugače* v pomenu 'sicer/sicer pa'; *enkrat* v pomenu 'nekoč', *moči* v pomenu 'morati' (tip *še bomo mogli vadet*).

Med navedenimi lemami so tri take, ki so v Gosu rabljene prevladujoče (ne pa izključno!) pri mladih: *ful*, *pizda* in *okej*. Na nadaljnjem seznamu ključnih lem je takih še več, npr. *kul*, *kurec*, *kurba*, *fak*. Pri tem pa je treba upoštevati, da Gos kot referenčni korpus majhnega obsega ni zadosten vir za zanesljivo prepoznavanje prvih govornic mladih. Tako ta kot tudi druge t. i. interesne govornice so sicer pogosto predmet zanimanja, vseeno pa v slovenskem jeziku še nimamo ustreznega obsežnega in reprezentativnega gradiva za te zvrsti, zato je označevanje tovrstnih, tj. slengovskih in žargonskih prvih v slovarju lahko precej nezanesljivo in v veliki meri temelji na intuiciji.

Na seznamu ključnih lem od 100 naprej sledi še veliko takih, ki so v dosedanji slovenski slovarpisni praksi označeni za slogovno zaznamovane (npr. *kao*, *probati*, *ajde*, *valjda*, *ziher*, *ratati*, *pasati*, *kul*, *ovi*, *fora*, *zrihtati*, *fejst*, *mej*, *špilati*, *tavžent*,

dec, zastopiti), vendar zaradi nizke pojavnosti o njihovi podrobnejši slogovno-zvrstni opredelitvi večinoma ni mogoče zanesljivo sklepati. Vseeno opozorimo vsaj na to, da se začne pojavljati tudi leksika, ki je v rabi regionalno precej ozko omejena, npr. *gučati, ovi, tipo, hambrt, usikdar, gvišno*. Za manjši delež tovrstne leksike, ki se v Gosu dovolj pogosto pojavlja, bi se dalo zelo na splošno sklepati o njeni regionalni razširjenosti, vendar je velika težava z leksikografskega vidika pogosta napačna lematizacija, ki jo je treba v drugi verziji korpusa Gos kolikor mogoče odpraviti. Za večji del tovrstne leksike (npr. *farba, murka, ahtati, štuk, šimfati, štrihati, zalimati*) pa je Gos premajhen korpus, da bi lahko zanesljivo sklepali o njeni regionalni razširjenosti.

Kot nekoliko posebno, vseeno pa najbližjo temu sklopu označimo za govorno rabo povsem specifično lemo *ne_biti*, ki je pripisana pojavnici, v katerih sta v govornem rabi dve lemi fonetično združeni v eno: *nem, nam, nevš, nev, nam, neb, neo, nem, nav, nevmo, nevte* ... Tukaj izgovorni elementi v korpusni strukturi prehajajo na raven lem; z vidika naše analize bi take strukture bolj ustrezno obravnavali po njihovih sestavnih elementih.

4.4 Deiktiki

Naslednja skupina za govorno rabo ključnih lem, ki smo jo definirali, so izrazi, s katerimi se orientiramo v prostoru in času ter imajo (tudi) deiktično funkcijo. Na seznamu prvih 100 najbolj ključnih lem je takih kar nekaj, pojavljajo se v parih: *poll/potem – zdaj/zdajle, noter/notri – ven, tukaj/tule – tam, gor – dol, sem – (tja – 122. mesto na seznamu ključnih lem), nazaj – (naprej – 111. mesto)*. Da so v govornem rabi bolj pogoste kot v pisni, lahko pojasnimo z večjo vpetostjo govorne rabe v prostor in čas kot komunikacijski okoliščini (kot tudi nasploh z večjo vpetostjo govorne rabe v kontekst).

Tesno povezani z izrazi za orientacijo v prostoru in času so kazalni zaimki. Na seznamu prvih 100 ključnih lem so, poleg že prej naštetih *tukaj, tule* in *tam*, še: *tale, tak, takole, tako, tisti*. Razlaganje njihove uvrstitve med 100 za govorno rabo najbolj ključnih lem je težavnejše, saj so rabe precej raznovrstne. Vseeno lahko ugibamo, da podobno kot za izraze, s katerimi se orientiramo v prostoru in času, na to veliko vpliva njihova deiktična funkcija.

Prav tako sta na seznamu prvih 100 ključnih lem še osebna zaimka *jaz* in *oni*, kjer podobno pogosto izstopa deiktična funkcija. Vsaj pri zaimku *jaz* vpliva na frekventnost rabe tudi odsvetovano eksplicitno izpostavljanje prve osebe v pisni normi (tip *sej jz sn odgovarjal* proti *saj sem odgovarjal*) – s tega vidika je povezan s skupino slogovno zaznamovane leksike.

4.5 Izrazi za govorna dejanja, pogosta v govornjeni rabi

Poleg številnih kazalnih in dveh osebnih zaimkov se na seznamu 100 najbolj ključnih lem za govornjeno rabo znajde še kar nekaj vprašalnih zaimkov: *kaj, koliko, kje, kakšen, kako, kdaj*. Predpostavljamo, da je glavni razlog njihova raba v vprašanjih – to govorno dejanje je v interaktivni govorni komunikaciji pogostejše kot v pisni, ki je prevladujoče neinteraktivna.

Ostale leme, povezane z govornimi dejanji, bolj pogostimi v govornjeni kot pisni rabi, so še: *jutro (dobro jutro – pozdravljanje)*, hvaliti (*hvala – zahvaljevanje*), *gospod (nazivanje)*, *prositi (prosim – vljudnostne konvencije)*.

Izstopa tudi skupina glagolov: to so glagoli premikanja (*iti, priti, hoditi*), rekanja (*reči, praviti*), dajanja (*dati*), delanja (*narediti, delati*), poslušanja (*čuti; slišati* na 135. mestu), gledanja (*videti, gledati, pogledati*), vedenja (*vedeti, znati*) in mišljenja (*misлити*) ter glagol *imeti*. Nekateri od teh (zlasti *vedeti, misлити, gledati, pogledati, videti*) dosegajo visoko frekventnost tudi ali celo zlasti zaradi diskurzno-označevalnih vlog, za ostale pa je verjetno razlog, da so aktivnosti oz. stanja, ki jih ti glagoli opisujejo, v govornjeni rabi pogostejše predmet zanimanja udeležencev komunikacije.

Večja skupina lem je takih, ki jih lahko pogosto povežemo z izražanjem odnosa, mnenja, stališča ali emocij govorca: *ful, fajn, super, res, joj, pizda, lepo, dobro, jah*. Tudi za izražanje tega predpostavljamo, da v govorni interakciji pogostejše najde prostor kot v pisni rabi.

4.6 Izrazi, »priljubljeni« v govornjeni rabi

Naslednja večja skupina lem je večinoma nepolnomenjska leksika, za katero se zdi, da je – morda tudi zaradi nekoliko specifičnih načinov rab in kolociranja – v govornjeni rabi bolj priljubljena in rabljena pogostejše, v dodatnih, novih pomenih in funkcijah. To so: *pač, malo, čisto, itak, nekako, točno, dosti, res, mogoče, zraven, lepo, sigurno, dejansko, stvar*, v določenih rabah tudi zaimek *kakšen* ter nedoločni zaimki *nekako, nek* in *nekje*. Med razlogi za njihovo večjo frekventnost v govornjeni rabi lahko prepoznamo tudi naslednje:

- pokažejo se nekatere dodatne funkcije oz. se rabe širijo na pomene, kjer bi v skladu z jezikovno normo uporabili kakšne druge sopomenske izraze, in tudi v resnici opazimo v pisnem korpusu obratno razmerje v frekventnosti, preračunano na 1 mio. pojavnic (če ilustriramo: v govornem korpusu je *sigurno* rabljeno pogostejše kot *zagotovo*, v pisnem korpusu je

zagotovo rabljeno pogosteje kot *sigurno*), npr. *malo* vs. *nekoliko*, *na kratko* (najprej bi *malo* ponovili drugo svetovno vojno); *čisto* vs. *povsem*, *popolnoma*, *zelo* (jz sem *čis* lepo povedal kako in kaj); *dosti* vs. *precej*, *dokaj* (ena taka *dost* nevsiljiva regulacija); *mogoče* vs. *morda* (v govornem korpusu je frekventnejši *mogoče*, v pisnem *morda*); *zraven* vs. *poleg* (v govornem korpusu je frekventnejši *zraven*, v pisnem *poleg*); *sigurno* vs. *zagotovo*; *dejansko* vs. v *resnici*;

- mnogi imajo ob tem neke vrste funkcijo relativiziranja, za katero se zdi, da je v govorjeni rabi bolj prisotna kot v pisni (gl. npr. Chafe in Danielewicz 1987): *pač*, *malo*, *nekako*, *dosti*, *nek*, *mogoče*, *dejansko*, *nekje* (*včasih je blo mal težko pridet do avtomobila*; *tko da to so nekak dve bistveni spremembi*; *ja to je verjetn dost tak tip ne*; *edin kar nam manjka je v bistvu mogoč mal elegance*; *ne bo nekih večjih problemov*; *date v mikrovalovko za kakšno minuto*; *hitro po prvem marcu takrat nekje drugega marca*);
- nekaterim lahko pripišemo še kakšne druge diskurzne funkcije, npr. poudarjalno, interakcijsko ipd.: *itak*, *točno*, *res*, *lepo*, *sigurno* (*itak je vse predrago*; *res tis k da nam on sporoča to na ta način*; *tako ti dej lepo tele eee špagete kuhat*; *aja točno ja*);
- ponekod zasledimo tudi posamezne za govorno rabo tipične ali specifične kolokacije ali (pragmatične) frazeme: *čakaj malo*, *a res*, *lepo pozdravljeni*, *lepo prosim*, *lepo se imejte*, *lepo telvas prosim*, *ena stvar*, *prva stvar*, *druga stvar*, *kakšna stvar*, *taka stvar*, *pa te stvari ...*

V tej skupini za govorno rabo ključnih lem torej opazimo tudi nekatere vrste rab, ki so za pisno rabo manj značilne ali celo slogovno ali normativno neprimerne, so pa značilne za govorno rabo.

4.7 Oblike, bolj »priljubljene« v govornem rabi

Na seznamu ključnih lem korpusa Gos opazimo, da se večkrat pojavljajo zaimki na *-le*. Med prvimi 100 ključnimi lemmi so to *tule*, *tale*, *zdajle*, *takole*, v nadaljevanju seznama do 300 se zvrstijo še trije: *tukajle*, *tamle*, *takle*, naprej pa še nekateri (*tistile*, *prejle* ...). Zdi se, da je taka oblika zaimkov v govornem rabi bolj »priljubljena« kot v pisni.

Nižje na seznamu ključnih lem, od 100 naprej, se začnejo večkrat pojavljati tudi različne pomanjševalnice, npr. *pesmica*, *slikica*, *kavica*, *mamica*, *majčka*, *flaška*, *dudka*, *mucka*, *zadevica*, *otroček*, *pobič*, *gvantek*, *kuglica*, *dekica*, *šipica*, *mlekec*, *župica*. Podobno kot pri zaimkih na *-le* se vsiljuje hipoteza, da so pomanjševalnice v govornem rabi nekoliko bolj »priljubljene« kot v pisni.

4.8 Drugo

Na seznamu prvih 100 ključnih lem se pojavi tudi nekaj veznikov: *ampak, samo, ker, saj, torej*. Izražajo večinoma protivnost, vzročnost, posledičnost in sklepalnost. O razlogih za njihovo pogostejšo rabo v govornjeni rabi težko sklepamo, potrebna bi bila celostnejša primerjalna analiza rabe veznikov v govornjeni in pisni rabi.

Na seznamu prvih 100 ključnih sta tudi števnika *trideset* in *dvajset* ter v govornem korpusu zelo pogosto z njima povezan pridevnik *cel* (tip *ena cela sedem*). V nadaljevanju seznama se takoj po prvih 100 ključnih lemah pojavlja še več števnikov (*petdeset, sto, osemdeset, devetdeset* ...). Razlog je najverjetneje v tem, da so v govornem korpusu vsi števniko izpisani in nikoli zapisani s cifro.

5 DISKUSIJA

Potem ko smo poiskali osrednje in pogoste vzroke, zakaj so določene leme v govornjeni rabi frekventnejše kot v pisni, nas zanima, katere od teh lahko dejansko povežemo s samim govornim kanalom ter posledično leksikalne enote in/ali njihove posamezne rabe v slovarju označimo kot tipične za govornjeno rabo. V pregledu literature v razdelku 2 smo povzeli zaključek Akinnasa (1982), da nekatere komunikacijske potrebe večinoma zadovoljujemo skozi govornjeno rabo, druge prevladujoče skozi pisno rabo, in da te funkcijske razlike vplivajo na leksikalne, skladijske in semantično-pragmatske lastnosti besedil v enem in drugem načinu jezikovne rabe. Ali torej lahko ločujemo med tem, kaj je tipično za govornjeno rabo zaradi lastnosti samega govornega kanala, in tem, kaj je na seznamu ključnih lem zaradi funkcijskih razlik med obema kanaloma? Odgovor na to vprašanje smo iskali tako, da smo skušali definirati lastnosti govornega kanala in jih ločiti od diskurzivnih funkcij, pogosto povezanih z govornim kanalom.

Verjetno najbolj prepoznavna fizična lastnost govornega kanala je izgovarjanje. Z njim so na leksikalni ravni povezana popravljanja, neverbalno odzivanje (*eee, mhm* ...) ter zapolnjevanje vrzeli, ki lahko vključuje tudi povsem leksikalizirane elemente (*v bistvu, da tako rečem*) in ni vedno nedvoumno prepoznavno.

S fizičnimi lastnostmi govornega kanala tesno povezana lastnost govornjene rabe je prevladujoča spontanost tvorjenja. Takoj ko nekaj izgovorimo, je to hkrati že preneseno naslovniku: vmes ni veliko časa za razmišljanje in popravljanje. Toda hkrati vemo, da se v javnem diskurzu besedilo, ki bo govornjeno, pogosto

pripravi vnaprej (zapiše), ali pa se vsaj izdelajo vsebinske oporne točke oz. lahko govornik raba poteka ob hkratni pisni ali multimediji, npr. predstavitve ob prosojnicah. Spontanost zato ni več tako vseprisotna lastnost govornika kot izgovarjanje.

Govorni kanal zahteva hkratno prisotnost tvorca in naslovnika, s tem pa omogoča sproten odziv naslovnika in neposredno interakcijo med njim in tvorcem. Lahko bi torej rekli, da je lastnost govornega kanala tudi interaktivnost. Toda v diskurzu prek javnih medijev to, razen izjemoma, ne velja, prav tako ne v drugih oblikah javnega govornega diskurza, kot so npr. gledališče, javne prireditve ipd. Vedno ko imamo na eni strani zelo množičnega naslovnika, je interaktivnost govornika precej ali celo popolnoma okrnjena.

Prav tako bi lahko trdili, da je govornik bolj vpeta v prostor in čas in nasploh v kontekst ter da je tudi to lastnost, ki izhaja iz samega govornega kanala. Toda tudi pisna jezikovna raba ima svoje prostore in čase, v katere je vpeta, le da je ta lastnost nekoliko manj izrazita in eksplicitno prisotna.

Na skrajnem drugem koncu je vprašanje neformalnosti govornika. Trdili bi lahko, da izgovarjanje in poslušanje zahtevata bližino tvorca in naslovnika v prostor-času, zaradi česar govornik raba praviloma poteka med manj udeleženci, torej je posledično pogostejše intimnejša in zato manj formalna. Toda mnogo situacij govornika v javnem diskurzu je prevladujoče formalnih, zaradi družbenih razmerij pa so take tudi nekatere nejavne nezasebne situacije, npr. razgovori za službo, konzultacije pri profesorju, posredovanje informacij, tudi delovni sestanki itd.

Po tem razmisleku lahko sklenemo, da ne moremo enostavno ločevati lastnosti govornika od diskurzivnih funkcij, ki jih skozi to rabo bolj pogosto ali celo praviloma uresničujemo.

Navedene lastnosti govornega kanala skušamo v Tabeli 1 povezati s skupinami izrazov, kot smo jih opisali v razdelku 4. V stolpcu »skupina« navajamo skupine izrazov, definirane v razdelku 4, v stolpcu »funkcije v govorniku rabi« povzemamo najznačilnejše funkcije, ki jih ti izrazi opravljajo v govorniku rabi, v stolpcu »lastnosti govornika rabe« pa skušamo navesti tiste značilnosti govornika rabe, ki spodbujajo njihovo večjo frekventnost, pri čemer na prvem mestu navedemo najvplivnejšo lastnost.

Tabela 1: Tipična leksika govorne rabe, njene funkcije in z njimi povezane lastnosti govornega kanala.

Skupina	Funkcije v govornjeni rabi	Lastnosti govorne rabe
Neverbalni izrazi	neverbalno izražanje	izgovarjanje spontanost interakcija
Diskurzni označevalci	različne metadiskurzne vloge	interaktivnost spontanost
Slogovno zaznamovana leksika	nestandardni jezik	neformalnost
Deiktiki	deiktična funkcija	vpetost v kontekst
Izrazi za govorna dejanja, pogosta v govornjeni rabi:		
• vprašalni zaimki	spraševanje	interaktivnost
• glagoli	diskurzni označevalci izražanje predmetnosti	interaktivnost vpetost v kontekst
• izrazi emocij, drže ...	izražanje emocij, drže ...	neformalnost
Izrazi, »priljubljeni« v govornjeni rabi	novi pomeni in funkcije	neformalnost spontanost interaktivnost
Oblike, bolj »priljubljene« v govornjeni rabi	?	?

Z najbolj prepoznavno fizično lastnostjo govornega kanala, izgovarjanjem, so povezani samo neverbalni izrazi. Diskurzne označevalce in izraze za izražanje vprašanj smo povezali predvsem z interaktivnostjo, ki pa ni značilna za čisto vse oblike govorne rabe in tudi ni ekskluzivna za govorno rabo. Slogovno zaznamovano leksiko, izraze emocij oz. drže ter izraze, »priljubljene« v govornjeni rabi, smo povezali predvsem z neformalnostjo. Tudi tukaj ne gre za lastnost, ki bi bila značilna za čisto vse oblike jezikovne rabe in ki bi bila lastna izključno govornjeni rabi. Deiktike in glagole za izražanje predmetnosti smo povezali z večjo vpetostjo govorne rabe v kontekst. Za oblike, »priljubljene« v govornjeni rabi, nismo nobeni od zgoraj definiranih lastnosti govornega kanala pripisali, da bi posebej spodbujala njihovo večjo frekventnost.

Na podlagi tega razmisleka in poskusne shematizacije v Tabeli 1 je naš sklep, da ne moremo enostavno ločevati, katere ključne leme ali njihove posamezne rabe naj v slovarskem opisu označimo kot tipične za govorno rabo in katerih ne. Naš zaključek je, da je označevanje govornosti proti pisnosti preveč splošen kriterij, da bi bil primeren kot slovarski kvalifikator, razlog pa je ta, da sta tako pisna kot govorna

raba zelo raznovrstni. Predlagamo, da se ustrežnejše slogovno-zvrstne oznake leksikalnih enot in rab iščejo na bolj specifičnih ravneh, na primer: 'neformalno', 'v pogovoru', 'regionalno omejeno', 'za izražanje emocij/drže' in podobno.

6 ZAKLJUČEK

V prispevku smo razpravljali o vprašanih, katera leksika je tipična za govorno rabo in ali lahko tem leksemom ali posameznim načinom njihove rabe pripišemo kvalifikatorsko oznako »v govoru«.

Po pregledu tujih raziskav o razlikah med pisno in govorno rabo smo zavzeli stališče, da so na ravni leksike med obema kanaloma razlike že predhodno dovolj obsežno dokazane, da lahko brez zadržka predpostavimo, da govorni korpus nosi dodatne podatke, zlasti na ravni leksikalnih vzorcev in kolokacij, ki jih iz še tako velikih pisnih korpusov ne moremo pridobiti.

Nato smo predstavili, katere skupine leksemov so v korpusu Gos tipično pogostejše rabljene kot v referenčnem korpusu Gigafida, v diskusiji pa je bila v središču našega zanimanja dilema, ali lahko tem rabam pripišemo kvalifikatorsko oznako govornosti. Dilema izhaja iz dejstva, da je govorna raba pogosto povezana z drugimi funkcijami in situacijami kot pisna ter da razlike med eno in drugo rabo niso nujno pogojene z govornostjo proti pisnosti, ampak z drugimi diskurzivnimi faktorji. Ko smo skušali definirati lastnosti govorne rabe, ki izhajajo iz samega govornega kanala, le-teh nismo mogli jasno ločevati od višjih diskurzivnih lastnosti, npr. neformalnosti, interaktivnosti itd., predvsem pa razen lastnosti izgovaranja nobene druge nismo mogli pripisati celotnemu spektru različnih tipov govorne rabe niti jih nismo mogli prepoznati kot imanentnih izključno govorni rabi. Zato smo zaključili, da govornost proti pisnosti ni ustrezna kategorija za označevanje leksikalnih enot, ker je presplošna, in predlagali iskanje ustreznih oznak na drugih, ožje opredeljenih ravneh, kot so 'neformalno', 'v pogovoru', 'regionalno omejeno', 'za izražanje emocij/drže' in podobno.