
Ana Zwitter Vitez

Javni diskurz in profil govorcev: oblikoskladenjska in leksikalna analiza korpusa Gos

objavljeno v:

Hotimir Tivadar (ur.): *Slovenski javni govor in jezikovno-kulturna (samo)zavest.*

Obdobja 38. Ljubljana: Znanstvena založba Filozofske fakultete, 2019.

<https://centerslo.si/simpozij-obdobja/zborniki/obdobja-38/>

© Univerza v Ljubljani, Filozofska fakulteta, 2019.

JAVNI DISKURZ IN PROFIL GOVORCEV: OBLIKOSKLADENJSKA IN LEKSIKALNA ANALIZA KORPUSA GOS

Ana Zwitter Vitez

Fakulteta za humanistične študije, Koper in Filozofska fakulteta, Ljubljana
ana.zwitter@guest.arnes.si

DOI:10.4312/Obdobja.38.255-267

Profiliranje govorcev je interdisciplinarno raziskovalno področje, ki omogoča boljše razumevanje dejanskega stanja javnega diskurza v slovenskem jezikovnem prostoru. V raziskavi s korpusnimi metodami analiziramo govor različnih profilov govorcev javnega diskurza korpusa govorjene slovenščine Gos (moški, ženske, starejši, mlajši, višja in nižja izobrazba). Rezultati na ravni frekvenčnih seznamov kažejo, da so razlike med posameznimi profili govorcev minimalne. Analiza ključnih oblikoskladenjskih oznak in tipičnih lem pa pokaže izrazite, pogosto celo stereotipne razlike med posameznimi profili govorcev.

javni diskurz, govorjena slovenščina, profiliranje avtorjev besedil, korpusna analiza

Author profiling is an interdisciplinary research field which enables a better understanding of the actual state of Slovene public discourse. In the present paper corpus methods are used to analyse the linguistic characteristics of various profiles of speakers included in the public discourse of the corpus of spoken Slovene Gos (men, women, seniors, the young, higher and lower education). The results of frequency analysis show that the differences between the individual speaker profiles are minimal. On the other hand, the analysis of key morphosyntactic tags and key lemmas reveals important, even stereotypical differences between the analysed speaker profiles.

spoken discourse, spoken Slovene, author profiling, corpus analysis

1 Uvod

Če javni diskurz predstavlja celoto javno dostopnih jezikovnih intervencij, potem je v središče razprave o pomenu in prihodnosti javnega diskurza v slovenščini smiselno postaviti njegove ustvarjalce oz. nosilce. Govor v medijih ni homogena entiteta – gradijo ga vsakokratne okoliščine sporazumevalnega položaja in različni govorci, ki v njem nastopajo. Če želimo razpravljati o dejanskem stanju javnega diskurza, ga je treba sistematično raziskovati glede na okoliščine tvorjenja in glede na specifične tipičnih profilov udeležencev javnih sporazumevalnih položajev.

Specifične diskurza različnih sociodemografskih profilov govorcev so zanimive tako v okviru razprave o prihodnosti slovenščine kot za forenzične preiskave in raziskave javnega mnenja, v zadnjih letih pa profiliranje govorcev doživlja silovit razmah iz izrazito tržno naravnanih motivov. Na družbenih omrežjih vsak dan nastaja ogromna

količina pisnih besedil, ki jih je mogoče avtomatsko odbirati, zajemati, analizirati in kategorizirati, uporabimo pa jih lahko tudi za potrebe profiliranja uporabnikov. V zameno za uporabo brezplačnih platform za govorno komunikacijo (npr. Skype, WhatsApp, Facebook Messenger; gl. Liu 2008) uporabniki privolijo v snemanje pogovorov, ki so podjetjem v pomoč pri gradnji modelov za avtomatsko analizo govora in profiliranja govorcev. Podjetja imajo tako na voljo podatke, učinkovita orodja za njihovo obdelavo in izpeljano znanje o profilih jezikovnih uporabnikov. Sistematično jezikoslovno preučevanje različnih profilov govorcev bo omogočilo boljše razumevanje dejanskega stanja javnega diskurza v slovenskem jezikovnem prostoru, s čimer bo jezikoslovje pomembno prispevalo k analizi in pojasnjevanju stereotipov, povezanih z jezikom opazovanih sociodemografskih profilov govorcev.

2 Študije javnega govora in profila govorcev

Raziskava se dotika področij analize javnega govora in profiliranja avtorjev besedil. Da stanje v slovenskem javnem diskurzu ni idealno, opozarja Tivadar (2012), ki analizira govor v javnosti pogosto prisotnih vplivnih govorcev in utemeljuje potrebo po izdaji sodobnega pravorečja, oblikovanega na osnovi kodifikacijske tradicije, sodobnih raziskav in opisa govora iz sodobnih javnih besedil (Tivadar 2018). Zaradi izpostavljene potrebe po sistematični preučitvi dejanskega stanja slovenskega javnega govora so nastale številne empirične raziskave. Huber (2017) analizira besednovrstno predvidljivost poudarkov medijskega diskurza, Scuteri (2018) preučuje razvoj televizijskega diskurza in ugotavlja, da so določene ritmično-prozodične značilnosti, tipične za pretekli televizijski informativni govor, prisotne tudi v današnjih televizijskih oddajah. S korpusom govornjene slovenščine Gos (Verdonik, Zwitter Vitez 2011) smo dobili možnost sistematičnega preučevanja spontanega govora. Primerjava različnih tipov diskurza glede na celoto zajetih besedil v korpusu Gos (Verdonik, Kosem 2012) je pokazala nekatere specifične posameznih tipov diskurza na besednovrstni ravni, raziskava Zwitter Vitez (2015) pa je zaznala upadanje deleža standardne izgovarjave od javnega informativnega diskurza proti nejavnemu zasebnemu diskurzu.

Profiliranje avtorjev besedil je izrazito ciljno naravnano področje, zato se ga različne raziskovalne vede lotevajo z različnimi pristopi. Jezikoslovci preučujejo jezikovne specifične različnih starostnih skupin avtorjev besedil (Kemper, Summer 2001), jezikovne razlike glede na spol (Tannen 1990) in regionalno pripadnost govorcev (Trudgill 2002), kar je pomembno za razumevanje družbene variantnosti jezika. Raziskovalci s področja računalništva se prav tako posvečajo značilnostim različnih profilov piscev in z uporabo metod strojnega učenja dosegajo odlične rezultate pri avtomatski prepoznavi različnih profilov avtorjev besedil (Koppel idr. 2008). V študiji s področja avtomatskega ugotavljanja spola avtorjev besedil v slovenščini Škrjanec idr. (2018) ugotavljajo, da med moškimi avtorji in ženskimi avtoricami tvitov obstajajo razlike na slogovni in leksikalni ravni: moški uporabljajo več kletvic, števnikov in izkazujejo izrazito leksemsko specifično, ženske pa uporabljajo več zaimkov, emotikonov in podvojenih znakov.

Na podlagi predstavljenih raziskav ugotavljamo, da je treba slovenski javni govor sistematično raziskati ne le z vidika posameznih jezikovnih prvin (intonacija, pavze,

retorične prvine), temveč tudi z vidika profila govorcev, ki predstavljajo temeljne oblikovalce javnega diskurza.

3 Metodologija

V raziskavi bomo analizirali govor različnih profilov govorcev radia in televizije, ki so vključeni v korpus govorne slovenščine Gos (Verdonik, Zwitter Vitez 2011). Govorce bomo razvrstili na podkorpuse glede na obstoječe metapodatke o njihovem spolu, starosti in izobrazbi ter poskusili identificirati tipične jezikovne prvine opazovanih profilov govorcev. Vsebine diskurza, zajetega v korpus Gos, ne moremo analizirati, saj ni označena z metapodatki.

Podkorpus	Število pojavnic	Število besed	Delež glede na ostale podkorpuse
Ženske	543.240	505.272	48 %
Moški	594.156	552.629	52 %
Nad 35 let	232.793	216.522	68 %
Pod 35 let	109.523	101.868	32 %
Fakulteta ali več	235.786	219.306	55 %
Srednja šola ali manj	189.374	176.138	45 %

Tabela 1: Analizirani podkorpusi korpusa Gos.

Za pridobivanje podatkov bomo uporabili orodje Sketch Engine¹ (Kilgariff idr. 2004), ki omogoča vpogled v konkordance, gradnjo podkorpusev in njihovo kontrastivno primerjavo. Pri oblikoslovni analizi bomo najprej pregledali zastopanost posameznih besednih vrst v analiziranih podkorpuseh, kar nam bo omogočilo celovit vpogled v rabo besednih vrst analiziranih profilov govorcev. Nato bomo s pomočjo seznamov ključnih oblikoskladenjskih oznak (Scott 1997) opazovali tipične specifične posameznih podkorpusev.

Na leksikalni ravni bomo pregledali frekvenčne seznane lem za vsakega od podkorpusev ter opazovali podobnosti in razlike med izbranimi profili govorcev. Nato bomo opazovali seznane ključnih lem (Scott 1997), ki so tipične za en profil govorcev in izrazito nepogoste za primerjani profil govorcev. Pri kategorizaciji se bomo opirali na metodologijo, ki vključuje dva nabora krovnih pojmov: seznam ved ARRS (*humanistika, družboslovje, naravoslovje, tehnika, medicina*) in tematske zavihke portala MMC RTV Slovenija (*Slovenija, Svet, Šport, Kultura, Življenjski slog, Svet zabave, Gospodarstvo*). Izbrani nabor krovnih pojmov je zasnovan na podlagi kriterijev za zajem besedil javnega diskurza korpusa Gos (Verdonik, Zwitter Vitez 2011), ki pokriva tako medijski kot pedagoški diskurz. K izbranemu naboru kategorij smo dodali še kategorijo *Profil*, ki vključuje tiste leme, za katere na podlagi konteksta ocenimo, da so tipične za določen profil govorca (npr. *frend* za mlajše govorce). V primeru dvoma pri razvrščanju bomo preučili kontekst, ki ga ponujajo konkordance

¹ www.sketchengine.co.uk.

(npr. lemo *situla* bi lahko uvrstili tako na področje kulture kot humanistike, kontekst konkordanc pa pokaže, da je uporabljena pri pouku zgodovine, kar lemo uvrsti na področje humanistike).

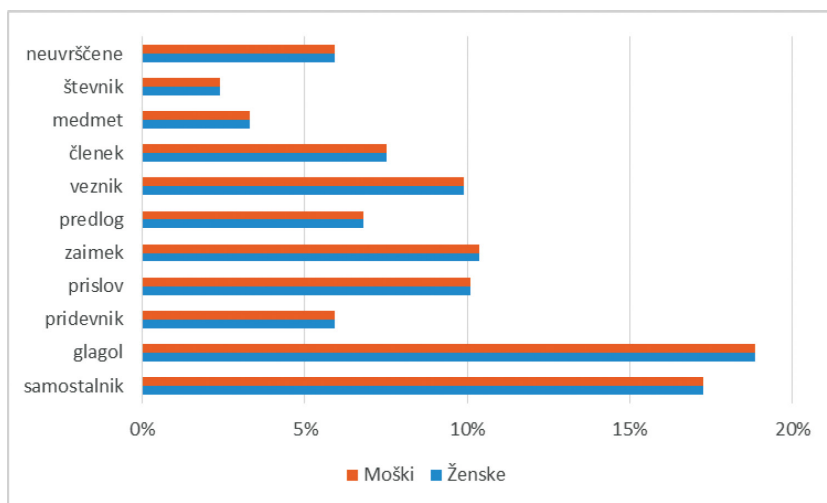
Pred začetkom analize se ustavimo pri omejitvah raziskave. Korpus Gos z milijonom besed ni reprezentativen korpus, zato ugotovitev te raziskave ne moremo posplošiti na celotno populacijo. Prav tako se zavedamo, da na lastnosti slehernega diskurza pomembno vpliva tema pogovora, česar ne moremo sistematično analizirati, saj vsebina pogovorov, zajetih v korpus Gos, ni označena z metapodatki. Na koncu je smiselno upoštevati tudi dejstvo, da na diskurz govorcev nikoli ne vpliva samo en sociodemografski dejavnik. Objektivno primerjavo med profili govorcev bi lahko izvedli zgolj tako, da bi vse kategorije govorcev tvorile iste vrste besedil na iste teme, česar v korpusu Gos zaradi narave gradiva ne moremo najti. Predpostavljamo pa, da na podlagi analiziranega gradiva korpusa Gos lahko govorimo o prevladujočem vplivu preučevanega dejavnika kljub različnim vplivom ostalih dejavnikov.

4 Analiza

Specifike posameznih profilov govorcev javnega diskurza (ženske, moški, starejši od 35 let, mlajši od 35 let, višja izobrazba, nižja izobrazba) analiziramo na oblikoslovni in leksikalni ravni.

4.1 Oblikoslovna raven

Najprej izluščimo frekvenčne sezname za vse besedne vrste po klasifikaciji JOS (Erjavec idr. 2010). Graf 1 predstavlja razporeditev posameznih besednih vrst glede na spol govorca javnega diskurza.



Graf 1: Zastopanost besednih vrst pri moških govornikih in ženskih govorkah javnega diskurza.

Iz Grafa 1 je razvidno, da je zastopanost besednih vrst pri obeh spolih govorcev javnega diskurza popolnoma prekrivna. To pomeni, da razporeditev besednih vrst ni

indikator, prek katerega bi lahko določali specifične posameznega profila govorca.

V nadaljevanju želimo ugotoviti, kakšne so oblikoskladenjske specifične moških in ženskih govorcev. To naredimo s pomočjo seznamov ključnih oblikoskladenjskih oznak, pri katerih opazujemo prve tri oblikoskladenjske oznake, ki imajo frekvenco več kot 100. Prve tri take tipične oblikoskladenjske oznake za ženske govorce predstavljata Tabela 2.

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Zspzeie	<i>moja</i>	237	1,3
Zop-er	<i>mene</i>	296	1,2
Zop-et--k	<i>me</i>	924	1,1

Tabela 2: Tipične oblikoskladenjske oznake ženskih govork javnega diskurza v korpusu Gos.

Iz Tabele 2 razberemo, da so predstavljene tipične oblikoskladenjske oznake ženskih govork zaimki. Gre za svojilni zaimek ženskega spola ednine v imenovalniku ednine (*moja*), prvoosebni osebni zaimek v rodilniku ednine (*mene*) in naslonsko obliko osebnega zaimka v prvi osebi ednine v tožilniku (*me*). Upoštevati je treba tudi dejstvo, da je pri avtomatskem oblikoskladenjskem označevanju lahko pod rodilniško oznako zajeta tudi homonimna tožilniška oblika.

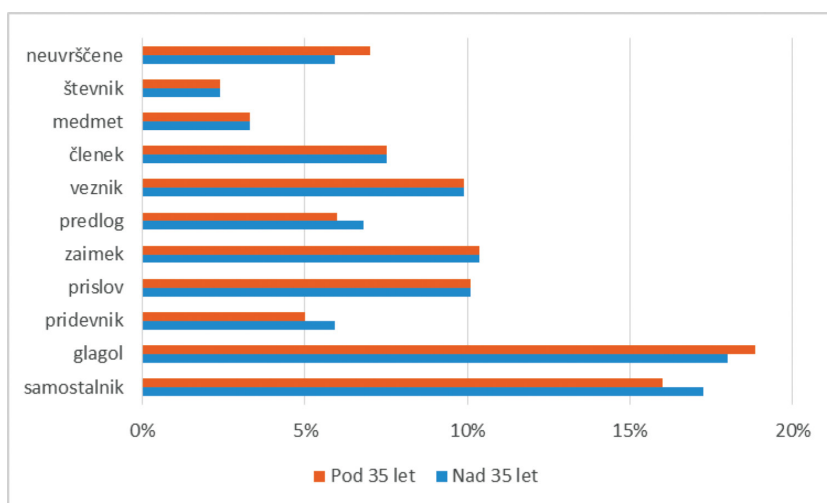
Tabela 3 predstavlja prve tri ključne oblikoskladenjske oznake moških govorcev korpusa Gos.

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Psnzei	<i>Poutiainenova</i>	265	1,6
Slmmt	<i>Palestince</i>	171	1,5
Kbvzei	<i>tretja</i>	449	1,3

Tabela 3: Tipične oblikoskladenjske oznake moških govorcev javnega diskurza v korpusu Gos.

Iz Tabele 3 beremo, da so najbolj tipične oblikoskladenjske oznake moških govorcev javnega diskurza svojilni zaimek ženskega spola v imenovalniku ednine (*Poutiainenova*), lastno ime moškega spola v tožilniku množine (*Palestince*) in vrstilni števniki ženskega spola v imenovalniku ednine (*tretja*).

Frekvenčni seznam besednih vrst pri govorcih nad 35 let in pod 35 let pokaže, da so besedne vrste pri starejših in mlajših govoricah enakovredno razporejene (Graf 2).



Graf 2: Zastopanost besednih vrst pri govornikih javnega diskurza nad 35 in pod 35 let.

Preverimo, kakšne so tipične oblikoskladenjske oznake, ki se pogosto pojavljajo pri govornikih, starejših od 35 let, in so obenem izrazito nepogoste pri govornikih, mlajših od 35 let (Tabela 4).

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Slmmi	<i>Habsburžani</i>	837	2,9
Ggnsdm-n	<i>imate</i>	622	2,3
Kbgzdi	<i>dve</i>	455	1,8

Tabela 4: Ključne oblikoskladenjske oznake pri govornikih javnega diskurza, starejših od 35 let.

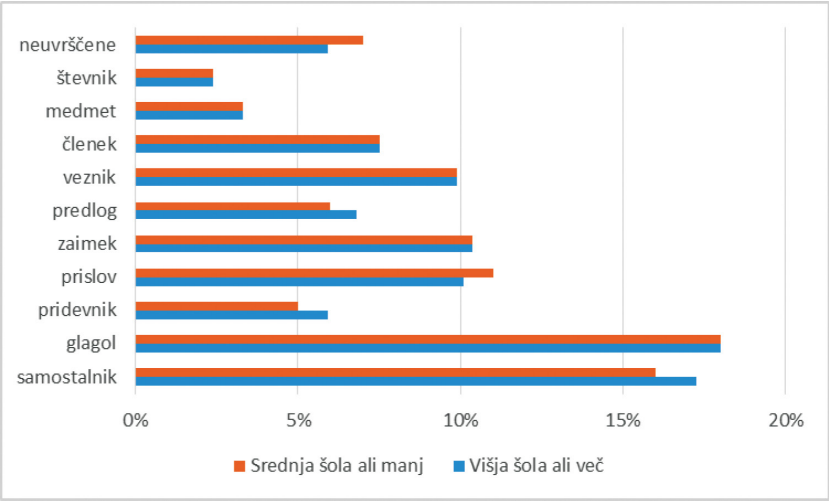
Seznam ključnih oblikoskladenjskih oznak, tipičnih za govorce, starejše od 35 let (Tabela 4), izpostavi oblike lastnega imena moškega spola v imenovalniku množine (*Habsburžani*), nedovršni glagol v sedanjiku in drugi osebi množine (*imate*) in glavni števniki ženskega spola v imenovalniku dvojine (*dve*). Analiza konkordanc pokaže, da so tovrstne oblike zastopane predvsem v pedagoškem diskurzu korpusa Gos.

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Ggnvde	<i>glej</i>	1743,9	2,0
Zop-et--k	<i>me</i>	1150,4	1,9
Gp-spe-n	<i>sem</i>	5578,7	1,9

Tabela 5: Ključne oblikoskladenjske oznake pri govornikih javnega diskurza, mlajših od 35 let.

Analiza ključnih oblikoskladenjskih oznak govorcev, mlajših od 35 let, pokaže, da pri mlajših govornikih izstopajo nedovršni glagol v velelnem naklonu in drugi osebi ednine (*glej*), naslonska oblika osebnega zaimka v prvi osebi ednine v tožilniku (*me*) in pomožni glagol *biti* v sedanjiku prve osebe ednine (*sem*).

Frekvenčni seznam besednih vrst pri govornih javnega diskurza z zaključeno višjo oz. visoko šolo, fakulteto ali več v primerjavi z govorniki, ki imajo zaključeno srednjo šolo ali manj, ne pokaže bistvenih razlik (Graf 3). To pomeni, da govorniki korpusa Gos z višjo in nižjo izobrazbo uporabljajo skoraj enaka razmerja med posameznimi besednimi vrstami.



Graf 3: Zastopanost besednih vrst pri govornih javnega diskurza z višjo in nižjo izobrazbo.

Seznam ključnih oblikoskladenjskih oznak pokaže nekaj zanimivih razlik med primerjanima profiloma govorcev. Pri govornih z višjo izobrazbo izstopajo lastno ime moškega spola ednine v mestniku (*Šoštanju*) in v rodilniku (*Chomskega*) ter lastna imena v imenovalniku množine (*Rimljani*). Zaznana uporaba lastnih imen predstavlja povezavo med bolj izobraženimi in starejšimi govorniki (*Habsburžani*).

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Slmem	<i>Šoštanju</i>	894,9	1,8
Slmer	<i>Chomskega</i>	1416,5	1,7
Slmmi	<i>Rimljani</i>	911,8	1,6

Tabela 6: Ključne oblikoskladenjske oznake pri govornih javnega diskurza z višjo izobrazbo.

Pri govornih s končano srednješolsko izobrazbo ali manj izstopajo lastna imena ženskega spola v imenovalniku ednine (*Slovenija*), vrstilni števnik ženskega spola v imenovalniku ednine (*prva*) in druga oseba zaimka v dajalniku ednine (*ti*).

Oznaka	Primer	Frekvenca/milijon	Koeficient ključnosti
Slzei	<i>Slovenija</i>	3205,3	1,9
Kbvzei	<i>prva</i>	871,3	1,7
Zod-ed--k	<i>ti</i>	1510,2	1,4

Tabela 7: Ključne oblikoskladenjske oznake pri govornih javnega diskurza z nižjo izobrazbo.

4. 2 Leksikalna raven

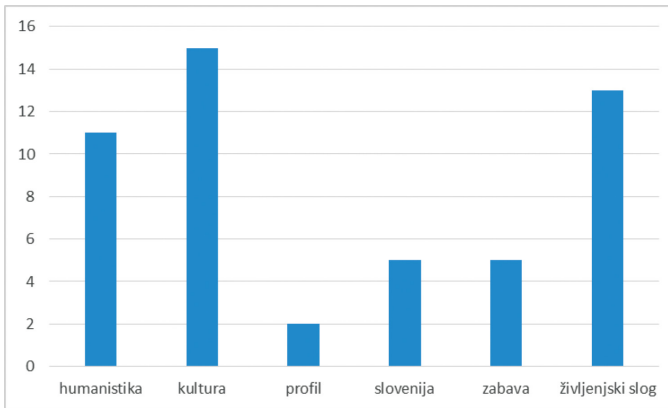
Na leksikalni ravni najprej pregledamo frekvenčne seznime lem za različne profile govorcev javnega diskurza (Tabela 8).

Ženske	Moški	Nad 35 let	Pod 35 let	Višja šola ali več	Srednja šola ali manj
<i>biti</i>	<i>biti</i>	<i>biti</i>	<i>biti</i>	<i>biti</i>	<i>biti</i>
<i>ta</i>	<i>ta</i>	<i>ta</i>	<i>ne</i>	<i>ta</i>	<i>ne</i>
<i>ne</i>	<i>eee</i>	<i>ne</i>	<i>ta</i>	<i>ne</i>	<i>ta</i>
<i>eee</i>	<i>ne</i>	<i>eee</i>	<i>pa</i>	<i>eee</i>	<i>pa</i>
<i>pa</i>	<i>pa</i>	<i>pa</i>	<i>se</i>	<i>pa</i>	<i>ja</i>
<i>se</i>	<i>se</i>	<i>se</i>	<i>in</i>	<i>se</i>	<i>se</i>
<i>da</i>	<i>v</i>	<i>v</i>	<i>ja</i>	<i>v</i>	<i>eee</i>
<i>in</i>	<i>da</i>	<i>da</i>	<i>v</i>	<i>in</i>	<i>v</i>
<i>v</i>	<i>in</i>	<i>in</i>	<i>da</i>	<i>da</i>	<i>in</i>
<i>ja</i>	<i>ja</i>	<i>ja</i>	<i>eee</i>	<i>ja</i>	<i>da</i>

Tabela 8: Najpogostejše leme pri posameznih profilih govorcev javnega diskurza v korpusu Gos.

Kot je razvidno iz Tabele 8, bistvenih razlik med različnimi profili govorcev ni opaziti, kar pomeni, da so najpogostejše leme v korpusu Gos povezane s sistemsko rabo jezika in ne s profilom govorcev. Izpostaviti velja le uporabo diskurznega označevalca *eee*, ki govorcem omogoča sprotno tvorjenje in načrtovanje spontanega diskurza. Uporabljajo ga vsi profili govorcev, najpogosteje se pojavlja pri moških govoricah, najmanj pogosto pa pri govoricah z nižjo izobrazbo.

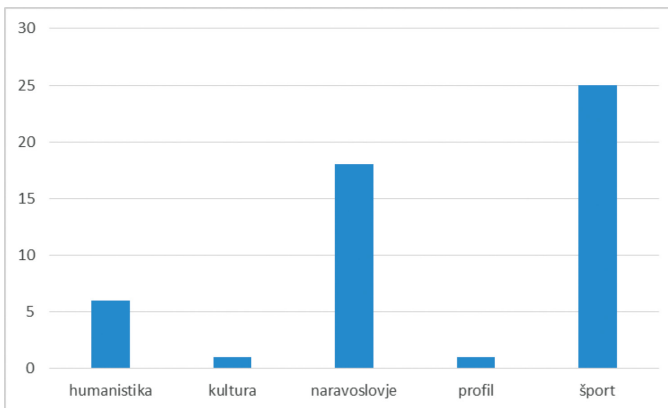
V naslednjem koraku izvedemo leksikalno analizo posameznih profilov govorcev javnega diskurza na podlagi seznamov ključnih lem, ki smo jih kategorizirali glede na področja, predstavljena v poglavju Metodologija. Graf 4 kaže kategorizacijo tistih lem, ki so izrazito pogoste za ženske govorce javnega diskurza in izrazito nepogoste za moške govorce javnega diskurza.



Graf 4: Kategorizacija ključnih lem ženskih govork javnega diskurza korpusa Gos.

Iz Grafa 4 je razvidno, da ključne leme ženskih govork spadajo na področja humanistike (*knežji, situla*), kulture (*poezija, razstava*) in življenjskega sloga (*cimra, požirek*). Pri teh rezultatih velja spomniti na pedagoški diskurz korpusa Gos, ki močno vpliva na teme pogovorov.

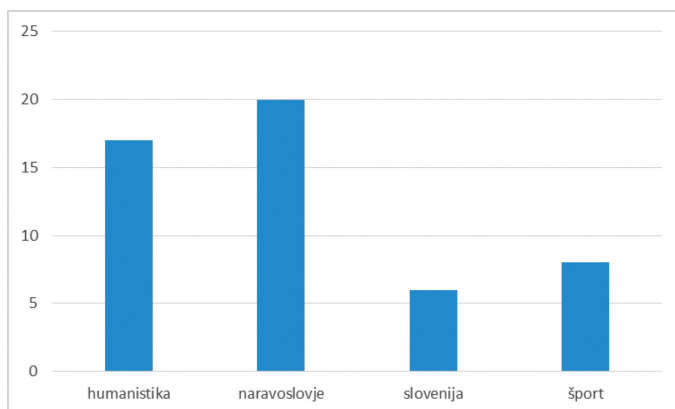
Graf 5 predstavlja kategorizacijo ključnih lem moških govorcev.



Graf 5: Kategorizacija ključnih lem moških govorcev javnega diskurza korpusa Gos.

Kategorizacija ključnih lem moških govorcev javnega diskurza pokaže, da tipične leme moških govorcev spadajo na področji naravoslovja (*mikroval, koordinata*) in športa (*stotinka, zavoj*), kar verjetno govori o povezavi med temo pogovora in spolom voditelja ali vabljenega govorca.

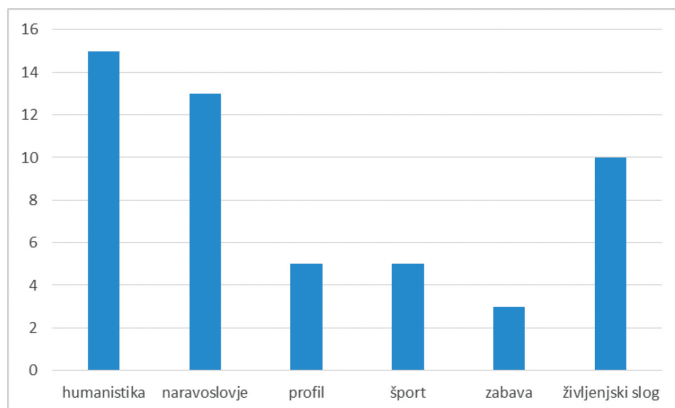
V naslednjem koraku kategoriziramo ključne leme govorcev javnega diskurza, starejših od 35 let (Graf 6).



Graf 6: Kategorizacija ključnih lem govorcev javnega diskurza, starejših od 35 let.

Iz kategorizacije je razvidno, da so tipične leme govorcev javnega diskurza, starejših od 35 let, povezane s področji naravoslovja (*raztopina, zrcaliti*) in humanistike (*povedek, taborišče*). Ponovno je treba upoštevati dejstvo, da gre predvsem za pedagoški diskurz korpusa Gos.

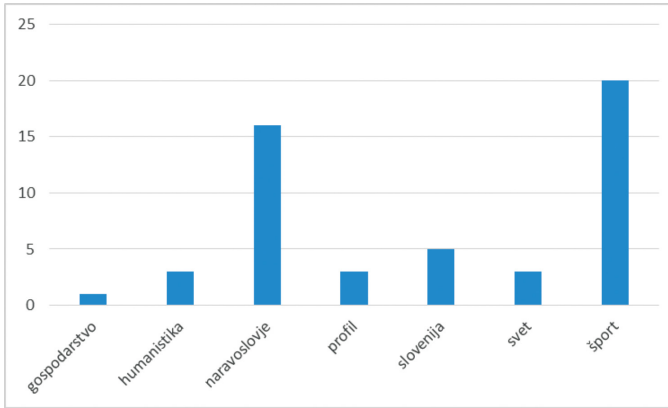
V Grafu 7 predstavljamo kategorizacijo ključnih lem govorcev javnega diskurza, mlajših od 35 let.



Graf 7: Kategorizacija ključnih lem govorcev javnega diskurza, mlajših od 35 let.

Tudi pri mlajših govorcih lahko opazimo prevladovanje humanistike (*Pegam, boršt*) in naravoslovja (*premer, stožec*), kar je posledica velikega deleža pedagoškega diskurza, vključenega v korpus Gos. Vendar opazimo tudi precej prisotno kategorijo življenjskega sloga (*kombi, televizor, golf*) in nekaj lem, ki smo jih pripisali specifikam profila mlajših govorcev (*car, friend, horror*).

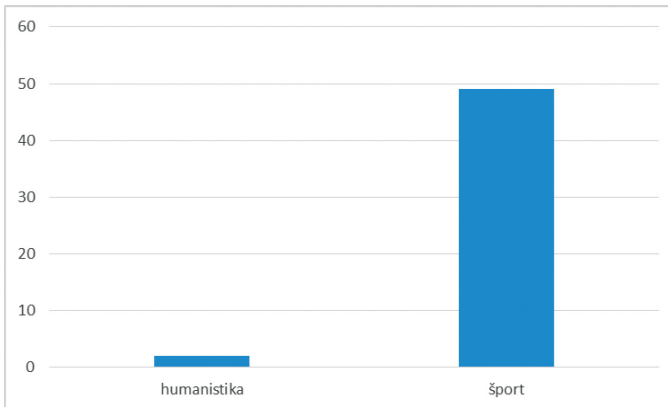
Graf 8 predstavlja kategorizacijo ključnih lem govorcev s končano višješolsko izobrazbo.



Graf 8: Kategorizacija ključnih lem govorcev javnega diskurza z višjo izobrazbo.

Pri govorcih z višjo izobrazbo prevladujejo tipične leme s področja naravoslovja (*modeliranje, rotacijski*) in športa (*polčas, Poljak*). Leme, ki niso odvisne od področja in smo jih pripisali specifikam profila govorcev z višjo izobrazbo, so *ključen, protislovje* in *bržkone*.

V Grafu 9 predstavljamo kategorizacijo ključnih lem govorcev javnega diskurza z zaključeno srednjo šolo ali manj.



Graf 9: Kategorizacija ključnih lem govorcev javnega diskurza z nižjo izobrazbo.

Iz Grafa 9 lahko razberemo izrazito nerazčlenjeno kategorizacijo ključnih lem pri govorcih z nižjo izobrazbo, ki pokriva zgolj področji športa (*ambiciozen, Mazejeva*) in humanistike (*prijateljstvo*), kar je verjetno povezano z vključitvijo šolarjev v pedagoški diskurz.

5 Zaključek

V raziskavi smo predstavili specifike različnih profilov govorcev, ki ustvarjajo javni diskurz korpusa Gos. Rezultati analize kažejo, da frekvenčni sezname ne izpostavijo razlik med govorcami, saj so besedne vrste in najpogostejše leme pri različnih profilih

javnih govorcev skoraj identično razporejene. Sezname ključnih oblikoskladenjskih oznak in ključnih lem pa pokažejo bistvene specifične, ki mestoma utrjujejo stereotipe o nekaterih profilih govorcev. Ženske v analiziranem javnem diskurzu tipično uporabljajo osebne in svojilne zaimke v prvi osebi ednine (*mene, moja*), moški vrstilne števnike (*tretja*) in lastna imena (*Poutiainen*), starejši govorci glagole v drugi osebi množine (*imate*) in lastna imena (*Habsburžani*), mlajši govorci glagole in zaimke v prvi osebi ednine (*me, sem*), govorci z višjo izobrazbo lastna imena v mestniku (*Šoštanju*), govorci z nižjo izobrazbo vrstilne števnike (*prva*) in zaimke v drugi osebi ednine (*ti*).

Na leksikalni ravni sezname ključnih lem posameznih profilov govorcev kažejo naslednje specifične: pri ženskah prevladujejo leme s področja kulture (*poezija*) in življenjskega sloga (*cimra*), pri moških leme s področja športa (*stotinka*) in naravoslovja (*mikroval*), za starejše in mlajše govorce so značilne leme s področij humanistike (*povedek*) in naravoslovja (*raztopina*), pri mlajših pa opazimo nekaj lem, tipičnih za njihov profil (*horror, car*). Bolj izobraženi tipično uporabljajo leme s področja naravoslovja (*rotacijski*) in športa (*polčas*) ter nekatere specifične leme (*ključen, protislovje* in *bržkone*), pri manj izobraženih pa izrazito izstopajo leme s področja športa (*tekma*).

V okviru predstavljene analize so nas zanimale predvsem oblikoskladenjske in leksikalne specifične posameznih profilov govorcev, nismo pa iskali odklona od govornega standarda, ki ga definira pravorečje *Slovenskega pravopisa* (2001) ali *Slovenska zborna izreka* (Šeruga Prek, Antončič 2003). Stopnjo odmika od izgovornega standarda bi bilo v prihodnosti smiselno opazovati na podlagi seznamov ključnih besednih oblik za posamezne profile govorcev, ki bi jih lahko kategorizirali glede na to, ali so izgovorjene z upoštevanjem pravorečne norme ali ne. Druga možnost bi bila, da bi korpus Gos poskušali označiti z metodologijo označevanja stopnje standardnosti korpusa Janes (Fišer idr. 2018). Rezultati predstavljene analize bi skupaj s predlaganimi raziskavami odmika od govornega standarda vzpostavili celovito poznavanje oblikoskladenjskih, leksikalnih in izgovornih specifik različnih profilov govorcev, ki soustvarjajo javni diskurz.

Literatura

- ERJAVEC, Tomaž, KREK, Simon, ARHAR, Špela, FIŠER, Darja, LEDINEK, Nina, SAKSIDA, Amanda, SIVEC, Breda, TREBAR, Blaž, 2010: *Priporočila za oblikoslovno označevanje*. <http://nl.ijs.si/jos/msd/html-sl/msd.index.msds.html> (dostop 1. 2. 2019)
- FIŠER, Darja, LJUBEŠIČ, Nikola, ERJAVEC, Tomaž, 2018: The Janes project: language resources and tools for Slovene user generated content. *Language Resources & Evaluation*. 1–24.
- HUBER, Damjan, 2017: Raziskovalna izhodišča besedilnofonetične analize slovenskega medijskega govora (na primeru besednovrstne predvidljivosti poudarkov). *Slavistična revija* LXV/2. 281–299.
- KEMPER, Susan, SUMMER, Aaron, 2001: The structure of verbal abilities in young and older adults. *Psychology and Aging* XVI/2. 312–322.
- KILGARRIFF, Adam, RYCHLY, Pavel, SMRŽ, Pavel, TUGWELL, David, 2004: The sketch engine. *Information Technology*. 105–116.

- KOPPEL, Moshe, SCHLER, Jonathan, ARGMON, Shlomo, 2008: Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology* LX/1. 9–26.
- LIU, Edward, 2008: Advertising using extracted context sensitive information and data of interest from voice/audio transmissions and recordings. *Google patent US20080109222A1*. <https://patents.google.com/patent/US20080109222A1/en> (dostop 1. 5. 2019)
- SCOTT, Mike, 1997: PC Analysis of Key Words – and Key Key Words. *System* XXV/1. 1–13.
- SCUTERI, Lucia Gaja, 2018: Diahrona makroprozodična raziskava slovenskega govora v televizijskih informativnih oddajah. *Slavistična revija* LXVI/2. 143–157.
- Slovenski pravopis*, 2001. Ljubljana: ZRC-SAZU.
- ŠERUGA PREK, Cvetka, ANTONČIČ, Emica, 2003: *Slovenska zborna izreka. Priročnik z vajami za javne govorce*. Maribor: Aristej.
- ŠKRJANEC, Iza, POLLAK, Senja, VERHOEVEN, Ben, 2018: Gender Profiling for Slovene Twitter communication: the Influence of Gender Marking, Content and Style. *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*. 119–125.
- TANNEN, Deborah, 1990: *You Just Don't Understand: Women and Men in Conversation*. New York: Harper Collins.
- TIVADAR, Hotimir, 2012: Nove usmeritve pri raziskavah govora s pogledom v preteklost. *Slavistična revija* LX/4. 587–601.
- TIVADAR, Hotimir, 2018: Slovensko pravorečje kot samostojna kodifikacijska knjiga? *Slavia Centralis* XI/2. 158–171.
- TRUDGILL, Peter, 2002: *Sociolinguistic Variation and Change*. Baltimore: Georgetown University Press.
- VERDONIK, Darinka, KOSEM, Iztok 2012: Key word analysis of discourses in Slovene speech: differences and similarities. *Linguistica* LII/1. 309–321.
- VERDONIK, Darinka, ZWITTER VITEZ, Ana, 2011: *Slovenski govorni korpus Gos*. Ljubljana: Zavod za uporabno slovenistiko Trojina.
- ZWITTER VITEZ, Ana, 2016: Specifike govorjene slovenščine glede na formalnost sporazumevalnega položaja. Erika Kržišnik, Miran Hladnik (ur.): *Toporišičeva Obdobja. Obdobja 35*. Ljubljana: Znanstvena založba Filozofske fakultete. 351–359.