

Darja Fišer,* Kristina Pahor de Maiti**

»Prvič, sem političarka in ne politik, drugič pa ...«

Korpusni pristop k raziskovanju parlamentarnega diskurza

IZVLEČEK

V tem učnem gradivu predstavljamo uporabnost jezikovnih korpusov za analizo družbenokulturnih pojavov, ki jih lahko raziskujemo na podlagi jezikovne rabe v specializiranem, političnem diskurzu. Na primeru korpusne analize govorov poslank in poslancev, kjer nas bo zanimala dinamika zastopanosti žensk v parlamentu in značilnosti njihove rabe jezika, bomo pokazali vrednost bogato označenega diahronnega korpusa slovenskih parlamentarnih razprav za digitalno humanistiko in družboslovje.

Ključne besede: parlamentarne razprave, parlamentarni korpusi, jezik in spol, digitalna humanistika

ABSTRACT

„FIRST, I’M A FEMALE POLITICIAN, NOT A MALE ONE, AND SECOND ...”
A CORPUS APPROACH TO PARLIAMENTARY DISCOURSE RESEARCH

This tutorial shows how corpora can be used to investigate language use and communication practices in a specialized socio-cultural context of political discourse in order to explore socio-cultural phenomena. We will demonstrate the potential of a richly annotated

* Izr. prof., Filozofska fakulteta Univerze v Ljubljani, Inštitut za novejšo zgodovino in Institut Jožef Stefan; darja.fiser@ff.uni-lj.si

** Asist., Filozofska fakulteta Univerze v Ljubljani, Aškerčeva 2, SI-1000 Ljubljana; kristina.pahordemaiti@ff.uni-lj.si

diachronic corpus of Slovenian parliamentary debates for investigating the characteristics and dynamics of the representation of women and their language use in the Slovenian Parliament.

Keywords: Parliamentary proceedings, parliamentary corpora, language and gender, digital humanities

Uvod¹

V parlamentu se odvija glavnina političnih razprav, na podlagi katerih se oblikuje zakonodaja, ki nato neposredno vpliva na vsakdanje življenje vseh prebivalcev in prebivalk določene države. V parlamentarnem diskurzu se tudi odsevajo politični programi, ideološka stališča in institucionalne vloge poslank in poslancev, ki zastopajo interese državljanov in državljanek.² Zato je parlamentarni govor vedno aktualen predmet družboslovnih in humanističnih raziskav.

Zapisi parlamentarnih sej so vse pogosteje dostopni v digitalni obliki, kar poenostavlja njihovo združevanje v strukturirane jezikovne vire oziroma korpuse. Ti so pogosto na voljo prek namenskih spletnih orodij, ki jih imenujemo konkordančniki. Raziskovalci jih uporabljajo pri različnih jezikovnih, slogovnih, kulturnih, družbenih in političnih raziskavah.³

Medtem ko so v jezikoslovju metode korpusne analize že dodobra uveljavljene⁴ in se med drugim pogosto uporabljajo tudi za sociolingvistično proučevanje spolno zaznamovanega jezika,⁵ so v drugih humanističnih in družboslovnih vedah, kjer prav tako številne raziskave temeljijo na jezikovnem gradivu, še razmeroma redke. V tem učnem gradivu zato prikazujemo potencial bogato označenih jezikovnih korpusov in konkordančnikov za raziskave družbenokulturnih pojavov, ki se odražajo v rabi jezika. Za oblikovanje natančnih kvantitativnih poizvedb o jezikovnih podatkih uporabimo dragocene metapodatke o poslankah in poslancih in njihovih govorih, rezultate pa nato še nadgradimo z ročno kvalitativno analizo, s katero pridobimo podroben vpogled v kontekst in zagotovimo ustrezno interpretacijo.

1 Avtorica naslovne izjave z redne seje 4. 6. 2003 je Cvetka Zalokar Oražem (LDS). Gl. *Iskanje*, pridobljeno 19. 2. 2021, <http://hdl.handle.net/11346/siparl-NYUL>.

2 Cornelia Ilie, »Parliamentary Debates,« v: Ruth Wodak in Bernhard Forchtner, ur., *The Routledge Handbook of Language and Politics* (London: Routledge, 2017), 309–25.

3 Lucja Biel, Agnieszka Biernacka in Anna Jopek-Bosiacka, »Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations,« v: *Language and Law* (Cham: Springer, 2018), 249–74. Sylvia Jaworska in Kath Ryan, »Gender and the Language of Pain in Chronic and Terminal Illness: A Corpus-based Discourse Analysis of Patients' Narratives,« *Social Science & Medicine* 215 (2018): 107–14.

4 Douglas Biber in Randi Reppen, *The Cambridge Handbook of English Corpus Linguistics* (Cambridge: Cambridge University Press, 2015). Tony McEnery in Andrew Hardie, *Corpus Linguistics: Method, Theory and Practice* (Cambridge: Cambridge University Press, 2011).

5 Paul Baker, *Using Corpora to Analyze Gender* (London: Bloomsbury, 2014).

S tem učnim gradivom želimo prikazati možnosti, ki jih raziskovalkam in raziskovalcem brez programerskih znanj ponujajo obsežne zbirke parlamentarnih razprav in konkordančniki. To učno gradivo je primerno tudi za uporabnice in uporabnike brez izkušenj z uporabo jezikovnih korpusov in orodij za iskanje po njih. Gradivo temelji na korpusu siParl 2.0,⁶ ki vsebuje parlamentarne razprave Državnega zbora Republike Slovenije od leta 1990 do 2018.

Kako uporabljati učno gradivo

V tem gradivu najprej predstavimo osnove korpusnega jezikoslovja in korpusne analize, čemur sledi uvod v značilnosti specializiranih korpusov parlamentarnih razprav in jedrnat pregled raziskav s področja jezika in spola. Drugi del učnega gradiva je namenjen praktičnemu prikazu najbolj razširjenih tehnik korpusne analize, kot so konkordančni nizi, frekvenčni sezname, ključne besede in kolokacije. Predstavljene tehnike uporabimo za raziskovanje zastopanosti žensk v slovenskem parlamentu, njihovih interesnih področij in govora o ženskah, pri čemer v analizo vključimo primerjavo z značilnostmi, ki veljajo za poslance oziroma skozi čas.

Vsi viri in orodja, uporabljeni v tem gradivu, so brezplačno na voljo v spletu. Za iskanje po korpusu uporabljamo NoSketch Engine,⁷ za dodatne ročne analize in predstavitev rezultatov pa enega od urejevalnikov preglednic (npr. Google Spreadsheet⁸ ali MS Excel⁹).

Besedila v sivih okvirčkih vsebujejo razlago uporabljenih funkcij in povezavo do videoposnetka, ki prikazuje, kako relevantno funkcijo poženemo v konkordančniku. Poleg tega so v okvirčkih tudi povezave do rezultatov v konkordančniku, kar bralkam in bralcem omogoča, da ponovijo analizo na svojem računalniku.

Iskanje po korpusu siParl 2.0 je mogoče prek konkordančnikov NoSketch Engine¹⁰ in KonText¹¹ na spletnem mestu CLARIN.SI¹² – slovenski center v okviru Evropske raziskovalne infrastrukture za jezikovne vire in tehnologije CLARIN ERIC.¹³ Prek

6 Slovenian parliamentary corpus (1990-2018) siParl 2.0, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11356/1300>.

7 NoSketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/4exc2b3e>.

8 Google Preglednice – brezplačno ustvarjajte in urejajte preglednice v spletu, pridobljeno 18. 2. 2021, <https://tinyurl.com/8mdwez3m>.

9 Microsoft Excel, Spreadsheet Software, Excel Free Trial, pridobljeno 18. 2. 2021, <https://tinyurl.com/6sdr9vwj>.

10 NoSketch Engine @ CLARIN.SI, pridobljeno 18. 2. 2021, <https://www.clarin.si/noske>.

11 Konkordance – Obrazec za prvo poizvedbo, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-XOZH>.

12 CLARIN.SI; Slovenska raziskovalna infrastruktura za jezikovne vire in tehnologije, pridobljeno 18. 2. 2021, <https://tinyurl.com/wuhvy98b>.

13 CLARIN – European Research Infrastructure for Language Resources and Technology | CLARIN ERIC, pridobljeno 18. 2. 2021, <https://www.clarin.eu>.

repozitorija CLARIN.SI je korpus siParl 2.0 mogoče tudi prenesti¹⁴ v računalnik in ga nato analizirati z drugimi korpusnimi orodji ali orodji za rudarjenje besedil, npr. z orodjem za korpusno analizo Antconc¹⁵ ali z orodji za osnovno obdelavo besedil v jeziku R.¹⁶

To gradivo je posodobljena različica izvirnega učnega gradiva,¹⁷ v katerem je bila uporabljena prva različica korpusa siParl. Korpus siParl 2.0 v primerjavi s korpusom siParl 1.0¹⁸ vsebuje bogatejše metapodatke o govorkah in govoricah ter sejah, kar omogoča razlikovanje med poslankami in poslanci ter drugimi govorkami in govorci v parlamentu. Poleg tega govori vsebujejo oznake parlamentarnih mandatov, kar omogoča lažjo primerjalno analizo elementov med posameznimi parlamentarnimi sklici. Druga različica korpusa vsebuje tudi dodatne ravni jezikoslovnih oznak, kot so skladenjske oznake in imenske entitete, ki pa jih v naši analizi ne bomo uporabljali, zato jih v gradivu ne obravnavamo.

Če želite tehnike korpusne analize takoj preizkusiti na praktičnih primerih, nadaljujte z nalogami v razdelku Korpusna analiza. Naslednji trije razdelki (Korpusi in konkordančniki, Zapisi parlamentarnih sej in Jezik in spol) so namreč namenjeni poljudnemu pregledu teoretičnih osnov in najpomembnejše terminologije. Čeprav nalogam v šestem razdelku lahko sledite brez predhodnega branja omenjenih treh razdelkov, vas spodbujamo, da se z njihovo vsebino na neki točki seznanite, saj predstavljene osnove zagotavljajo celovito razumevanje in samostojno rabo analitičnih postopkov ter ustrezno interpretacijo rezultatov.

Korpusi in konkordančniki

Korpusi

Besedilni korpusi¹⁹ so obsežne zbirke skrbno izbranih strojno berljivih jezikovnih podatkov, ki jih lahko uporabimo za pripravo jezikovnih opisov ali za preverjanje hipotez o jeziku. Vendar pa so korpusi veliko več kot le zbirke besedil v elektronski obliki. Oblikovani so v enem od standardnih formatov, kot je razširljivi označevalni jezik oziroma XML,²⁰ in kodirani v skladu z vnaprej določeno, a prilagodljivo shemo za predstavitev besedil v digitalni obliki, med katerimi je v jezikoslovju in digitalni humanistiki najbolj uveljavljen standard TEI.²¹

Osnovna enota korpusa je pojavnica,²² ki jo dobimo z avtomatskim postopkom tokenizacije, pri katerem je izbrano besedilo razdeljeno na manjše enote, v našem

14 Slovenian parliamentary corpus (1990-2018) siParl 2.0, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11356/1300>.

15 Corpus Analysis with Antconc, pridobljeno 18. 2. 2021, <https://tinyurl.com/3x85dvvs>.

16 Basic Text Processing in R, pridobljeno 18. 2. 2021, <https://tinyurl.com/f3nj4jt8>.

17 Voices of the Parliament, pridobljeno 18. 2. 2021, <https://tinyurl.com/hb68794d>.

18 Slovenian parliamentary corpus (1990-2018) siParl 2.0, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11356/1236>.

19 Besedilni korpus – Wikipedija, prosta enciklopedija, pridobljeno 18. 2. 2021, <https://tinyurl.com/23zu8r2e>.

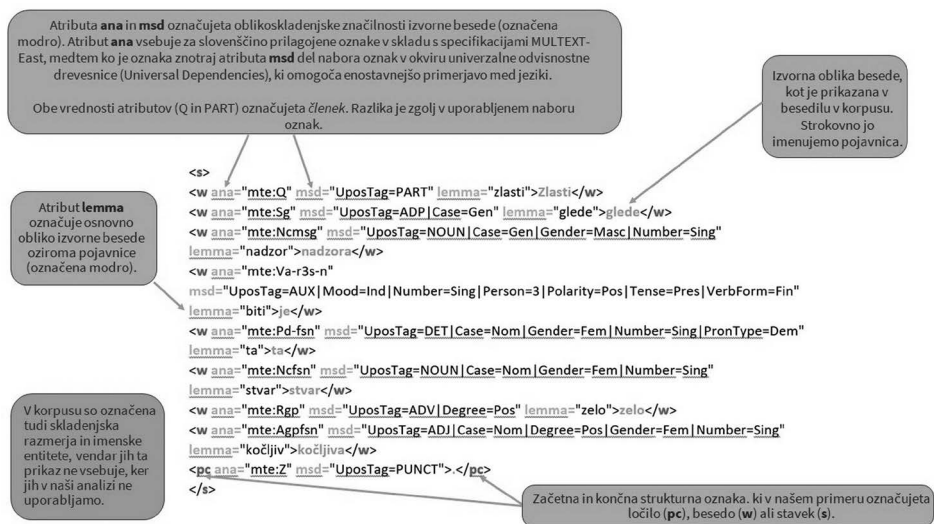
20 XML Tutorial, pridobljeno 18. 2. 2021, <https://www.w3schools.com/xml>.

21 TEI: Text Encoding Initiative, pridobljeno 18. 2. 2021, <https://tei-c.org>.

22 Token | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/48a5j4ru>.

primeru na besede. Običajno se pojavnice določajo glede na presledke med besedami. Za bolj poglobljeno analizo so besedila v korpusu jezikoslovno označena. Najosnovnejši obliki jezikoslovnega označevanja sta oblikoskladenjsko označevanje, s katerim besedam v besedilih pripišemo besedno vrsto (npr. glagol), in lematizacija,²³ pri kateri pojavnice (besednim oblikam v različnih sklonih) pripišemo osnovno slovarsko obliko oz. lemo. Lematizacija je pomembna zlasti za korpus morfološko bogatih jezikov, kot je slovenščina. Poleg jezikoslovnih oznak so v korpus dodani tudi metapodatki o besedilu (npr. datum, vrsta) in govorncu (npr. ime, spol), ki so pomembni za kontekstualizacijo rezultatov, vendar se lahko uporabljajo tudi za izvajanje natančnejših korpusnih poizvedb.

Slika 1. Izvleček iz jezikoslovno označenega korpusa v zapisu TEL.



Konkordančniki

Po korpusih iščemo s specializiranimi orodji, ki se imenujejo konkordančniki.²⁴ Nekatere konkordančnike moramo pred uporabo namestiti na osebni računalnik, drugi pa so dostopni neposredno prek spleta. Uporabljamo jih za iskanje vseh pojavitev pojavnice v korpusu, ki ustrezajo našemu iskalnemu pogoju. Na voljo je veliko konkordančnikov, ki omogočajo podobne funkcije (gl. npr. ta podrobni seznam),²⁵ med najbolj priljubljenimi pa so AntConc²⁶ (brezplačen) in WordSmith Tools²⁷ (potreben je nakup licence), ki ju je mogoče uporabljati brez internetne povezave in

23 Lematizacija – Wikipedija, prosta enciklopedija, pridobljeno 18. 2. 2021, <https://tinyurl.com/u37yyhr>.

24 Corpus Linguistics: Method, theory and practice, pridobljeno 18. 2. 2021, <https://tinyurl.com/27ajekSj>.

25 Concordancers, pridobljeno 18. 2. 2021, <https://tinyurl.com/yn6wcebj>.

26 Laurence Anthony's AntConc, pridobljeno 18. 2. 2021, <https://tinyurl.com/3ntaevtt>.

27 WordSmith Tools home page, pridobljeno 19. 2. 2021, <https://tinyurl.com/4256exts>.

ne vsebujeta predhodno naloženih korpusov, ter spletna konkordančnika BYU²⁸ (za polno delovanje je potrebna brezplačna registracija) in SketchEngine²⁹ (brezplačna akademska licenca), ki vsebujeta številne predhodno naložene korpusne jezike. SketchEngine omogoča tudi ustvarjanje in označevanje lastnih korpusov.

Večina sodobnih konkordančnikov omogoča naslednje funkcije, ki jih bomo uporabili tudi v tem gradivu:

- **Konkordančni niz**³⁰ prikazuje vse zadetke iskane besede ali niza besed v korpusu ter njihovo sobesedilo. Konkordance, posamezne enote konkordančnega niza, so lahko razvrščene naključno ali glede na iskano besedo oziroma niz besed. Razvrstiti jih je mogoče tudi glede na levi ali desni kontekst, kar izpostavi značilne vzorce, v katerih se iskana beseda oziroma niz besed pojavlja.
- **Frekvenčni sezname**³¹ prikazujejo vse zadetke za izbrano korpusno poizvedbo, ki je lahko v obliki besede/besedne zveze ali določenega nabora metapodatkov. Zadetki so lahko razvrščeni po abecedi ali pogostosti, pripisan pa jim je tudi podatek o njihovi frekvenci. V nasprotju s konkordančnim nizom besedni sezname vsebujejo sobesedila zadetkov.
- **Sezname ključnih besed**³² izpostavljajo besede, ki v izbranem korpusu izstopajo v primerjavi z izbranim referenčnim korpusom.
- **Sezname kolokacij**³³ vsebujejo nize besed, ki se v kombinaciji z iskano besedo pojavljajo pogosteje kot ob naključnem pojavljanju.

Na voljo so že številna navodila za uporabo konkordančnikov (npr. SketchEngine Quick Start Guide³⁴) in tečaji korpusnega jezikoslovja (npr. Corpus Linguistics: Method, Analysis, Interpretation³⁵). V tem gradivu se osredotočamo na prikaz uporabe konkordančnika NoSketchEngine in specializiranega korpusa za iskanje odgovorov na konkretna raziskovalna vprašanja na metodološko ustrezen način. Gradivo je namenjeno študentkam in študentom ter raziskovalkam in raziskovalcem s področja (digitalne) humanistike in družboslovja, ki jih zanima raziskovanje sociokulturnih pojavov prek jezikovne rabe.

Zapisi parlamentarnih sej

Parlamentarni korpusi so različnih velikosti in vsebujejo podatke v različnih modalitetah (pisni, govornjeni, gestikulacija), iz različnih časovnih obdobj in v enem ali več jezikih. Čeprav so parlamentarne razprave večinoma dostopne v pisni obliki

28 *English Corpora: most widely used online corpora. Billions of words of data: free online access*, pridobljeno 18. 2. 2021, <https://www.english-corpora.org>.

29 *CQL – Corpus Query Language* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/skdvt638>.

30 *Concordance – most powerful corpus search* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/8zrs9f8n>.

31 *Wordlists – word frequency lists* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/9cn4afkz>.

32 *Keywords and term extraction* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/9sfrjfp3>.

33 *collocation* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/3ket5dps>.

34 *Quick Start Guide - learn to use* | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/3hx8y7px>.

35 *Corpus Linguistics Analysis - Online Course – FutureLearn*, pridobljeno 18. 2. 2021, <https://tinyurl.com/4x9cp5t4>.

(npr. korpus razprav iz danskega parlamenta³⁶), so včasih dostopne tudi kot zvočni posnetki ali videoposnetki skupaj s transkripcijami (npr. korpus razprav iz češkega parlamenta³⁷). Parlamentarni korpusi kot diahroni viri³⁸ nam omogočajo poglobljeno raziskovanje jezikovnih in družbenih sprememb skozi čas. Večina parlamentarnih korpusov vsebuje bogate metapodatke³⁹ o govorih (npr. datum govora, trajanje govora, točka dnevnega reda, v kateri je bil govor podan) ali govorkah in govorcih (npr. spol, starost, izobrazba, politična pripadnost, institucionalna vloga), kar omogoča poglobljen vpogled v kontekst preučevanega pojava.⁴⁰

Parlamentarni diskurz

Parlament je institucija, za katero veljajo specifična pravila in konvencije, kar vpliva tudi na jezikovno rabo, začeni s pravili glede podeljevanja besede, naslavljanja itd. (gl. npr. poslovnik državnega zbora RS⁴¹). Ta pravila in konvencije so se izoblikovale v določenem družbenozgodovinskem okolju, zato niso v vseh parlamentih enake; v grškem parlamentu je na primer skakanje v besedo strogo prepovedano, medtem ko je v britanskem parlamentu prekinjanje govorcev pogosto in običajno ni sankcionirano.⁴² To je seveda treba upoštevati pri oblikovanju raziskovalnih vprašanj, izvedbi raziskave in interpretaciji rezultatov. Zato je pri raziskovanju te vrste diskurza⁴³ še posebej pomembno, da dobro razumemo okoliščine, v katerih je nastal.

Zanesljivost sejnih zapisov

Uradno objavljeni sejni zapisi niso dobesedne transkripcije⁴⁴ parlamentarnih razprav, ampak zapisi v skladu z uveljavljenimi zapisovalnimi praksami institucije. Te prakse se razlikujejo med posameznimi državami in časovnimi obdobji. Med urejanjem zapisov običajno pride do izločanja značilnih prvin govorenega jezika, lahko pa tudi do nekaterih drugih posegov, kot so izločanje očitnih jezikovnih ali vsebinskih napak, narečnih ali pogovornih izrazov ter žaljivega in vulgarnega jezika. Smernice za urejanje sejnih zapisov običajno žal niso javno objavljene, kar lahko bistveno oteži raziskovalno

36 *The Danish Parliament Corpus 2009 - 2017, v1*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/20.500.12115/8>.

37 *Czech Parliament Meetings*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11858/00-097C-0000-0005-CF9C-4>.

38 *Synchrony and diachrony* – *Wikipedia*, pridobljeno 18. 2. 2021, <https://tinyurl.com/s96f8r38>.

39 *Metapodatek* – *Wikipedija, prosta enciklopedija*, pridobljeno 18. 2. 2021, <https://tinyurl.com/yvj88wak>.

40 Pertti Alasutari, Marjaana Rautalin in Jukka Tyrkkö, »The Rise of the Idea of Model in Policymaking: The Case of the British Parliament, 1803-2005,« *European Journal of Sociology/Archives Européennes de Sociologie* 59, št. 3 (2018): 341–63. Jane Demmen, Lesley Jeffries in Brian Walker, »Charting the Semantics of Labour Relations in House of Commons Debates Spanning Two Hundred Years,« *Doing Politics: Discursivity, Performativity and Mediation in Political Discourse* 80 (2018): 81.

41 *Poslovnik državnega zbora*, pridobljeno 18. 2. 2021, <https://tinyurl.com/wmbczcj4>.

42 Cornelia Ilie, »Parliamentary Debates.«

43 *Diskurz* – *Wikipedija, prosta enciklopedija*, pridobljeno 18. 2. 2021, <https://tinyurl.com/tjj76yh>.

44 *Transcription (linguistics)* – *Wikipedia*, pridobljeno 18. 2. 2021, <https://tinyurl.com/e86cheyh>.

delo. Kritiko virov v parlamentarnem kontekstu podrobneje obravnavata Mollin⁴⁵ in Rix.⁴⁶ Drug pomemben vidik pri proučevanju parlamentarnega govora je, da so govori poslank in poslancev pogosto pripravljene vnaprej, in sicer v pisni obliki, kar vpliva na njihove skladenjske in slogovne značilnosti. To sicer ne velja za vsa parlamentarna okolja, niti za vse vrste parlamentarnih nastopov (npr. interpelacije, obrazložitve glasu, poslanska vprašanja). Tudi te posebnosti je poleg družbenozgodovinskega konteksta treba vedno upoštevati pri oblikovanju raziskovalnih vprašanj, izvedbi raziskave in interpretaciji rezultatov.

Slika 2 prikazuje primerjavo uradnega zapisa govora Anje Bah Žibert,⁴⁷ poslanke Državnega zbora RS 2014–2018, s 33. seje Državnega zbora na temo predloga zakona o Prešernovi nagradi (18. točka dnevnega reda⁴⁸) z dobesedno transkripcijo njenega govora, ki smo jo pripravili na podlagi videoposnetka seje (min. 5:00–7:33).⁴⁹ Že na prvi pogled se vidi, da uradni sejni zapisi DZ RS niso najboljši vir za preučevanje značilnosti spontanega govora, kot so napačni začetki, premori, raba mašil ali ponovitev.

Slika 2. Primerjava uradnega sejnega zapisa in dobesedne transkripcije iz Državnega zbora RS.

Uradni sejni zapis

»Če bi zadevo zreducirali, nagrajevanje, v smislu, ki je prvoten, to je se pravi dodeljevanje nagrad za najvišje ali pa največje prispevke v umetnosti. In zdaj bi na drugi strani to zreducirali na spol. Kaj to pomeni? /.../ Jaz mislim, seveda se lepo sliši, le v manjšini so ženske zastopane v družbi na nekih pomembnih funkcijah, ampak, da kdorkoli reče, da je nekdo na tej funkciji zgolj zato, ker je ženskega spola. Jaz mislim, da s tem ne delamo nobene koristi, kvečjemu zbudamo še nek manjvrednosti položaj v smislu „ne gre za sposobnost, gre za to, ker si drugega spola“.«

(Anja Bah Žibert, PS SDS)

Dobesedni prepis

»**Zdej, s tem**, če bi zadevo **zreduciral**, nagrajevanje, v smislu, ki je prvoten, to je, se pravi, dodeljevanje nagrad za najvišje **prispevke al pa največje prispevke** v umetnosti. In **zdej** bi na drugi strani to **rz... am ... zreduciral** na spol. Kaj to pomeni? /.../ Jaz mislim **da, kh...** seveda se lepo sliši, **ko am ... na nek način** le v manjšini so ženske zastopane v družbi na nekih pomembnih funkcijah, ampak, da kdorkoli reče, da je nekdo na tej funkciji zgolj zato, ker je **ne vem** ženskega spola. Jaz mislim, da s tem ne delamo nobene koristi, **kvečjem** zbudamo še nek manjvrednostni položaj v smislu „ne gre za sposobnost, gre za to, ker si drugega spola“.«

(Anja Bah Žibert, PS SDS)

45 Sandra Mollin, »The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts,« *Corpora* 2, št. 2 (2007): 187–210.

46 Kathryn Rix, »'Whatever Passed in Parliament Ought to be Communicated to the Public': Reporting the Proceedings of the Reformed Commons, 1833–50,« *Parliamentary History*, 33, št. 3 (2014): 453–74.

47 Anja Bah Žibert – Wikipedija, prosta enciklopedija, pridobljeno 18. 2. 2021, <https://tinyurl.com/ed7p75r8>.

48 Sejni zapisi Državnega zbora, 33. seja (Ljubljana: Državni zbor, 20218), 128, dostopno na <https://tinyurl.com/2s6r3r54>.

49 33. redna seja Državnega zbora, 4. del, pridobljeno 18. 2. 2021, <https://tinyurl.com/5Stwvm6u>.

Poznavanje raziskovalnih podatkov

Predpogoj za ustrezno oblikovanje raziskovalnih vprašanj, izvedbo raziskave in interpretacijo rezultatov, pa tudi za optimalno izkoriščanje vsega, kar raziskovalni podatki omogočajo, je, da se najprej podrobno seznanimo s tem, kaj izbrani korpus vsebuje, kako je bil zgrajen in označen ter kakšne so njegove omejitve. Ker niso vsi korpusi enako označeni, moramo vedno preveriti podrobnosti o izbranem korpusu. Te informacije so običajno na voljo na domači strani korpusa (npr. korpus historične nizke nemščine⁵⁰), v konkordančniku, prek katerega je korpus na voljo (npr. britanski parlamentarni korpus Hansard v konkordančniku Univerze Brigham Young⁵¹), ali v repozitoriju, kjer je korpus arhiviran (npr. hrvaški parlamentarni korpus ParlaMeter-hr v repozitoriju CLARIN.SI⁵²). Več o kodiranju in označevanju parlamentarnih podatkov lahko preberete v izobraževalnem modulu o zbirkah parlamentarnih podatkov PARTHENOS.⁵³

Jezik in spol

Razlike med jezikom moških in žensk so predmet številnih raziskav na področju sociolingvistike, stilistike, retorike, študij spola, medijskih študij in analize diskurza.⁵⁴ Rezultati kažejo, da so razlike majhne, a sistematične.⁵⁵ Pri raziskavah politične komunikacije pa lahko opazimo, da se tradicionalno posvečajo predvsem diskurzu politikov, medtem ko je diskurz političark šele pred kratkim postal predmet raziskav.⁵⁶

To je pomembno, ker je več avtoric⁵⁷ pokazalo, da se zakonodajalke razlikujejo od svojih sodelavcev pri vprašanih, ki jih naslavljajo, stališčih, ki jih zavzemajo, in pristopih, ki jih uporabljajo pri pripravi zakonodaje. Blaxill in Beelen⁵⁸ sta v analizi zastopnosti žensk v britanskem parlamentu po letu 1945 odkrila, da je v govorih poslank večji poudarek na vprašanih, ki zadevajo ženske (npr. izobraževanje, zdravstvo itd.) in ki se neposredno dotikajo njihove dobrobiti v družbi (npr. varstvo otrok, nasilje v družini,

50 *Corpus of Historical Low German*, pridobljeno 18. 2. 2021, <http://www.chlg.ac.uk/helipad>.

51 *Corpus of British Parliament speeches (Hansard)*, pridobljeno 18. 2. 2021, <https://www.english-corpora.org/hansard>.

52 *Croatian parliamentary corpus ParlaMeter-hr 1.0*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11356/1209>.

53 *Working with parliamentary corpora; Parthenos training*, pridobljeno 18. 2. 2021, <https://tinyurl.com/yypffemm>.

54 Penelope Eckert in Sally McConnell-Ginet, *Language and Gender* (Cambridge: Cambridge University Press, 2013). Ruth Wodak, »Pragmatics and Critical Discourse Analysis: A Cross-disciplinary Inquiry«, *Pragmatics & Cognition* 15 (2017): 203–25.

55 Matthew L. Newman, Clara J. Groom, Lori D. Handelman in James W. Pennebaker, »Gender Differences in Language Use: An Analysis of 14,000 Text Samples«, *Discourse Processes* 45, št. 3 (2008): 211–36.

56 Catherine Marshall, »Policy Discourse Analysis: Negotiating Gender Equity«, *Journal of Education Policy* 15, št. 2 (2000): 125–56.

57 Milica Antić Gaber in Gabriella Ilonszki, *Women in Parliamentary Politics: Hungarian and Slovene Cases Comparar* (Ljubljana: Peace Institute, Institute for Contemporary Social and Political Studies, 2003). Monique Leijenaar, *How to Create a Gender Balance in Political Decision-making: A Guide to Implementing Policies for Increasing the Participation of Women in Political Decision-making* (Luxembourg: Office for Official Publications of the European Communities, 1997). Christina Wolbrecht, »Female Legislators and the Women's Rights Agenda: From Feminine Mystique to Feminist Era«, v: Cindy S. Rosenthal, ur., *Women Transforming Congress* (Norman: University of Oklahoma Press, 2002), 170–97.

58 Luke Blaxill in Kaspar Beelen, »A Feminized Language of Democracy? The Representation of Women at Westminster since 1945«, *Twentieth Century British History* 27, št. 3 (2016): 412–49.

enakost itd.).⁵⁹ Tudi Bäck idr.,⁶⁰ Hansen idr.⁶¹ ter Mensah in Wood,⁶² ki so raziskovali korpuse razprav iz švedskega, danskega oz. ganskega parlamenta, so odkrili, da ženske pogosteje kot moški govorijo o temah, ki se tičejo t. i. *mehkih zakonodajnih področij* (npr. makroekonomija, energetika, promet, bančništvo, finance, veselje, znanost in tehnologija ter komunikacija⁶³), medtem ko moški pogosteje razpravljajo o temah s t. i. *trdih zakonodajnih področij* (npr. zdravstvo, delo, zaposlitev in priseljevanje, izobraževanje ter socialno varstvo).

Upoštevati je treba, da na uporabo jezika vplivajo različni dejavniki (npr. družbeni razred, kontekst, starost, hierarhija), pri čemer je spol le eden od mnogih.⁶⁴ Bing in Bergvall⁶⁵ poudarjata, da so podobnosti v uporabi jezika pri različnih spolih običajno spregledane, čeprav so bolj izrazite kot razlike. Tudi Blaxill in Beelen⁶⁶ sta odkrila podobno težnjo v okviru raziskav parlamentarnega diskurza. Če torej spol govorca poznamo vnaprej, kot to velja v našem primeru, je treba to dejstvo nujno upoštevati, da se izognemo napačni interpretaciji rezultatov in pretiranemu poudarjanju razlik.⁶⁷

Korpusna analiza

V tem razdelku bomo na obsežnem korpusu slovenskih parlamentarnih razprav pokazali, kako lahko osnovne tehnike korpusne analize (za več o tem gl. razdelek 3 Korpusi in konkordančniki) uporabimo za iskanje odgovorov na naslednja raziskovalna vprašanja:

- V Nalogi 1 bomo *analizirali zastopnost žensk v slovenskem parlamentu*. Najprej si bomo ogledali, **kako ustvarimo podkorpuse**. Nato se bomo naučili, **kako sestaviti frekvenčne sezname**, ki prikazujejo število govorcev oziroma govork in njihovih govorov v podkorpusah.
- V Nalogi 2 bomo *identificirali tematike, o katerih so skozi čas razpravljale poslanke v primerjavi s poslanci*. Pri tem se bomo naučili, **kako izluščimo ključne besede** iz podkorpusov, in si nato ogledali, **kako prek analize konkordanc** preverimo njihov kontekst in rabo.

59 Gl. Tracey L. Osborn, *How Women Represent Women: Political parties, gender and representation in the state legislatures* (Oxford: Oxford University Press, 2012).

60 Hanna Bäck, Marc Debus in Jochen Müller, »Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag«, *Political Research Quarterly* 67, št. 3 (2014): 504–18.

61 Dorte H. Hansen, Costanza Navarretta in Lene Offersgaard, »A Pilot Gender Study of the Danish Parliament Corpus«, v: Darja Fišer, Maria Eskevich in Franciska de Jong, ur., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (Miyazaki: LREC, 2018).

62 Eric O. Mensah in Sandra F. Wood, »Articulations of Feminine Voices in Ghana's Parliament: A Study of the Hansard from 2010-2011«, *AFRREV LALIGENS: An International Journal of Language, Literature and Gender Studies* 7, št. 2 (2018): 61–77.

63 Lena Wängneru, »Intressen kontra stereotyper. Varför finns det kvinnliga och manliga politikområden i riksdagen?«, *Statsvetenskaplig Tidskrift* 99, št. 2 (1996).

64 Jennifer Coates, *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language* (London: Longman, 1997). Lia Litosseliti, *Gender and Language: Theory and Practice* (London: Hodder Arnold, 2006).

65 Janet M. Bing in Victoria L. Bergvall, »The Question of Questions: Beyond Binary Thinking«, v: Jennifer Coates, ur., *Language and Gender: A Reader. Vol. 1* (Oxford: Blackwell, 1998), 496–510.

66 Blaxill in Beelen, »A Feminized Language of Democracy?«.

67 Angela Goddard in Lindsey M. Patterson, *Language and Gender* (London: Routledge, 2015).

- V Nalogi 3 bomo raziskali, kako so poslanke in poslanci v minulih 25 letih razpravljali o tematikah, povezanih z ženskami. Ogleдали si bomo, **kako uporabljamo podatke o relativni frekvenci** v različno velikih podkorpusih in **kako izluščimo kolokacije** izbranih samostalnikov v podkorpusih.

Korpus siParl 2.0

Korpus siParl 2.0⁶⁸ vsebuje sejne zapise 11. mandata Skupščine socialistične republike Slovenije za obdobje 1990–1992, sejne zapise Državnega zbora Republike Slovenije⁶⁹ za obdobje 1992–2018 (1.–7. mandat), sejne zapise delovnih teles Državnega zbora Republike Slovenije za obdobje 1996–2018 (2.–7. mandat) in sejne zapise Sveta predsednika Državnega zbora za obdobje 1996–2018 (2.–7. mandat).

Korpus vsebuje metapodatke o govorkah in govorcih (npr. ime, spol, vloga, politična stranka), tipologijo parlamentarnih zasedanj (npr. redna seja, izredna seja itd.) ter strukturne in uredniške oznake (npr. zakonodajno obdobje, leto/naslov seje, imenske entitete).⁷⁰ Korpus je tudi oblikoskladenjsko označen in lematiziran. Poleg tega vsebuje nekatere druge jezikoslovne oznake (npr. oznake univerzalne odvisnostne drevesnice (Universal Dependencies) in skladdenjsko razčlenjevanje), ki pa jih v tem gradivu ne uporabljamo. Korpus obsega več kot milijon govovor oziroma 200 milijonov besed, ki jih je prispevalo skoraj 8.500 različnih govork in govorcev (npr. poslanke in poslanci, članice in člani vlade, predstavnice in predstavniki ministrstev, strokovnih organizacij, nevladnih organizacij in interesnih skupin). Splošne informacije o korpusih in o zapisih parlamentarnih sej najdete v razdelku 3 Korpusi in konkordančniki in v razdelku 4 Zapisi parlamentarnih sej.

Te informacije o korpusu so na voljo tudi v konkordančniku na strani CLARIN.SI⁷¹ (gl. Sliki 3a in 3b):⁷²

- Kvantitativne informacije o velikosti korpusa so na voljo v razdelku *Zadetki* (gl. številko 1).
- Osnovne oznake za besedne vrste so navedene v *Legendi oznak* (gl. številko 2), opis celotnega nabora oznak pa je na voljo prek povezave *Tagset/Opis*⁷³ (gl. številko 3).
- Korpus je organiziran prek strukturnih atributov. Atribut *session/seja* zamejuje posamezne seje (gl. številko 4) in vsebuje informacije, kot so datum, vrsta in naslov seje (gl. Sliki 3b, številko 4).

68 *Slovenian Parliamentary Corpus (1990–2018) siParl 2.0*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11356/1300>, gl. Andrej Pančur, Tomaž Erjavec, Mihael Ojsteršek, Mojca Sorn in Neja Blaj Hribar, *Slovenian Parliamentary Corpus (1990–2018) siParl 2.0*. Slovenian language resource repository CLARIN.SI, 2020.

69 *Državni zbor Republike Slovenije – Wikipedija, prosta enciklopedija*, pridobljeno 18. 2. 2021, <https://tinyurl.com/8ke5j554>.

70 *Named entity – Wikipedia*, pridobljeno 18. 2. 2021, <https://tinyurl.com/3jc7ftsd>.

71 *Corpus info: siParl 2.0 (parlament 1990–2018)*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-OLBP>.

72 Slovensko različico konkordančnika izberete v zgornjem desnem kotu s klikom simbola zobnika in izbiro ustreznega jezika s spustnega seznama.

73 Zaradi izida nove različice konkordančnika je ta trenutno zgolj delno preveden v slovenščino.

- Atribut *speech/govor* zamejuje posamezne govore (gl. številko 5) in vsebuje informacije o imenu, spolu, vlogi in strankarski pripadnosti govorce oz. govorca (gl. Sliko 3b, številko 5).

Slika 3a: Osnovne informacije o korpusu siParl 2.0.

The screenshot shows the siParl 2.0 interface with the following sections:

- Home**: siParl 2.0 (parlament 1990-2018)
- Home**: Korpus parlamentarnih razprav Državnega zbora Republike Slovenije 1990-2018 v2.0 // Corpus of parliamentary debates of the National Assembly of the Republic of Slovenia 1990-2018 v2.0
- Zadeki** (1):

Število pojavnic	239,749,733
Besede	202,902,663
Stavki	11,537,491
Odstavki	3,326,865
Dokumenti	11,967
- Splošne informacije** (3):

Corpus description	Document
Jezik	Slovenian
Kodiranje	UTF-8
Zbrano	04/10/2020 13:11:51
Tagset	Opis
- Velikost leksikona**:

word	647,958
lc	563,284
lempos	309,557
lemma	301,124
lemma_lc	290,311
tag_en	1,163
tag	1,163
ud_pos	17
ud_feats	1,905
id	1,222
ud_dep	33
ud_dep_head_lemma	202,058
ud_dep_head_tag_en	1,104
ud_dep_head_tag	1,104
ud_dep_head_ud_pos	18
ud_dep_head_ud_feats	1,508
ud_dep_head_id	1,109
- Legenda oznak** (2):

samostalnik	S.*	samostalnik	-s
glagol	G.*	glagol	-g
pridevnik	P.*	pridevnik	-p
prislov	R.*	prislov	-r
zaimek	Z.*	zaimek	-z
predlog	D.*	predlog	-d
veznik	V.*	veznik	-v
členek	L.*	členek	-l
medmet	M.*	medmet	-m
- Končnice msd-jev**:

samostalnik	-s
glagol	-g
pridevnik	-p
prislov	-r
zaimek	-z
predlog	-d
veznik	-v
členek	-l
medmet	-m
- Strukture in atributi** (4):

name	4,070,050
p	3,326,865
session	11,967
speech	1,134,540

Slika 3b: Razširjene informacije o strukturnih atributih v korpusu siParl 2.0.

Strukture in atributi

name 4,070,050

type 5

p 3,326,865

d URL to digital library 3,326,865

id 3,326,865

session 11,967 (4)

mandate_en Legislative period (in English) 8

mandate Legislative period (in Slovene) 8

id Session ID 11,967

title Session title 11,700

type_en Type of session (in English) 7

type Type of session (in Slovene) 7

organ Working body of the National Assembly 246

date 4,327

month 329

wordcount 9,696

year 29

speech 1,134,540 (5)

party_init Acronym of speaker's party 38

party_en Name of speaker's party (in English) 44

party Name of speaker's party (in Slovene) 44

role_en Role of speaker (in English) 2

role Role of speaker (in Slovene) 2

who Speaker ID 8,414

name Speaker name 8,288

type_en Type of speaker (in English) 2

type Type of speaker (in Slovene) 2

birth 69

gender 3

NALOGA 1: Zastopanost žensk v slovenskem parlamentu

Državni zbor Republike Slovenije ima 90 poslank in poslancev, vključno s predstavnico oz. predstavnikom italijanske in madžarske manjšine, ki so trenutno del devetih političnih strank.⁷⁴ Slovenija je ena od najmlajših držav Evropske unije, ki je v preteklih 30 letih doživela dramatične spremembe na področju enakosti spolov. Leta 1986, ko je bila Slovenija še vedno del Socialistične federativne republike Jugoslavije, so poslanke zasedale skoraj četrtno vseh poslanskih sedežev, leta 1992, ko je bil prvič sklican Državni zbor Republike Slovenija, pa se je ta delež občutno zmanjšal, saj so poslanke zasedale le še vsaki deseti poslanski sedež.⁷⁵ Med postopkom tranzicije, ko so se družbeni, politični, gospodarski in vrednostni sistem bistveno spremenili, so ženske v Sloveniji izgubile več gospodarskih in družbenih prednosti, pridobljenih v času socializma, in so bile skoraj popolnoma izrinjene iz političnih institucij. V sedmem mandatu (1. 8. 2014–21. 6. 2018), ki je zadnji mandat, vključen v korpus siParl 2.0, so poslanke zasedle malo več kot 40 % sedežev, večinoma zaradi zakonodajnih ukrepov, ki predpisujejo spolne kvote.⁷⁶ Glede na EU indeks vrednosti enakosti spolov in politične moči v Evropski uniji za leto 2017⁷⁷ je bila Slovenija uvrščena med prvih 10 držav Evropske unije po deležu poslank.

V Nalogi 1 bomo primerjali govore v korpusu siParl 2.0 s trendi na parlamentarnih volitvah in v slovenski družbi.

Izdelava podkorpusov

Na podlagi metapodatkov, ki so na voljo v korpusu (gl. Sliko 3b), korpus razdelimo na več delov, t. i. podkorpuse, in sicer glede na naslednja merila:

- **Spol.** Govorke in govorniki so v korpusu označeni z eno od naslednjih kategorij v skladu s podatki iz uradnih sejnih zapisov državnega zbora: moški, ženska ali ni podatka (v primerih, ko so zapisi metapodatkov nepopolni). V tem gradivu uporabljamo le prvi dve kategoriji (moški, ženska), medtem ko tistih, za katere ni podatka o spolu, v analizo ne vključujemo.
- **Vloga.** Govorkam in govornicem je v korpusu pripisana ena od dveh vlog, in sicer član parlamenta oziroma *zunanjí govornik*⁷⁸ (ta vloga označuje povabljenе govorke in govornice, ki niso del parlamenta (npr. članice in člani vlade, predstavnice in

74 Portal DZ - Poslanske skupine, pridobljeno 18. 2. 2021, <https://tinyurl.com/yh6brs32>.

75 Milica Antić Gaber in Irena Selišnik, »The Slovene Version of a ‚Fast Track‘ to Political Equality,« *Teorija in Praksa* 54, št. 2 (2017): 337.

76 Delež temelji na uradnem številu vseh poslank, kot ga navaja Simona K. Lipicer (2017), nato pa je deljen s številom vseh poslank in poslancev (tj. 90). Ta delež je rahlo višji od deleža, navedenega v naši analizi, v kateri smo dosledno uporabljali le podatke iz korpusa. Ker le redko vsi izvoljeni poslanci in poslanke ostanejo na istem položaju celoten mandat, jih nato zamenjajo drugi poslanci oziroma poslanke. To se odraža v korpusu, ki vključuje vse poslanke in poslance, in ne le tistih, ki so bili izvoljeni na volitvah. Glede na podatke iz korpusa je skupno število poslancev in poslank na mandat tako večje od 90.

77 European Institute for Gender Equality Home, pridobljeno 18. 2. 2021, <https://tinyurl.com/84embf66>.

78 Pri oznakah v korpusu je uporabljen generični moški spol.

- predstavniki ministrstev ali nevladnih organizacij). V tem gradivu upoštevamo le govorce in govorko z oznako član parlamenta, ker se želimo osredotočiti na prispevke izvoljenih predstavnic in predstavnikov državljanov Slovenije.
- **Tip.** Poleg vloge je govorkam in govorcem v korpusu pripisan tudi njihov tip, ki ima eno od naslednjih dveh vrednosti: *redni govornik* ali *predsedujoči*. Prva kategorija označuje govorko in govorca v parlamentu, ki so v skladu s poslovnikom izrecno dobili besedo. V naši analizi upoštevamo le to kategorijo. Tiste z oznako *predsedujoči* smo namerno izključili, saj je večina njihovih izjav namreč pogojena s postopkovnimi predpisi in dogovori in torej nanje ne vplivajo strankarska pripadnost, spol ali drugi dejavniki.
 - **Mandat.** Vsak govor v korpusu je kategoriziran glede na parlamentarno obdobje, v katerem je bil izrečen (1.–7. mandat). Ker to presega meje naše raziskave, smo iz analize izključili obdobje odcepitve od Jugoslavije,⁷⁹ ki je v korpusu označeno kot 11. sklic.
 - **Delovni organ.** Korpus siParl 2.0 vsebuje govore s plenarnih zasedanj, z zasedanj delovnih organov državnega zbora (npr. Komisija za peticije, človekove pravice in enake možnosti) in z zasedanj Sveta predsednika Državnega zbora. V tem gradivu analiziramo le govore s plenarnih zasedanj.

Na podlagi teh meril izdelamo 14 podkorpusov, po dva za vsakega od sedmih mandatov, pri čemer eden vsebuje govore poslank, drugi pa govore poslancev s pripisano vlogo *redni govornik*. Poleg teh 14 podkorpusov izdelamo tudi podkorpuse, ki vsebujejo vse govore poslank in poslancev za vsak posamezni mandat in za vseh sedem mandatov skupaj.

Videoposnetek, ki prikazuje izdelavo podkorpusa v konkordančniku NoSketch Engine, je na voljo na <https://sidih.si/20.500.12325/120>.

Ustvarjeni podkorpusi in informacije o njihovi velikosti so na voljo tukaj.⁸⁰

Spodaj je naveden tudi primer zahtevnejšega iskalnega izraza v jeziku CQL, ki ga je treba uporabiti pri izdelavi spodnjega podkorpusa:

Mandat1-Ženske:⁸¹

```
<speech gender="F" & role_en="MP" & type_en="Regular speaker"/>
within <session mandate_en="Term 1 ." *" & organ="Državni zbor Republike
Slovenije"/>
```

⁷⁹ *Socialistična republika Slovenija – Wikipedija, prosta enciklopedija*, pridobljeno 18. 2. 2021, <https://tinyurl.com/k3mafd47>.

⁸⁰ *Subcorpus information*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/sipar1-KGFY>.

⁸¹ Ker je to gradivo na voljo v slovenščini in angleščini, so poimenovanja podkorpusov v konkordančniku dvojezična. Vendar pa v tem gradivu zaradi jasnosti in jedrnatosti uporabljamo zgolj slovenski del imena podkorpusov. Povezava do angleških videov: https://youtube.com/playlist?list=PLhZ_k72iCcm0n0HVgd4O3UcwRIs-W-R0R.

S tem ukazom poiščemo vse izjave (atribut *speech/govor*) govork (atribut *F/Ž*), ki so članice parlamenta (atribut *MP/Član parlamenta*) in redne govork (atribut *Regular speaker/Redni govornik*), in sicer v govorih iz prvega mandata (atribut *Term 1.* /1. Mandat.**), ki so bili izrečeni v Državnem zboru Republike Slovenije (atribut *Državni zbor Republike Slovenije*).

Če želimo ustvariti podkorporus govora poslancev, moramo spremeniti le spol prek atributa *speech gender/spol* (*speech gender="M"*). Za podkorporus, ki vsebuje govore poslank in poslancev iz vseh mandatov, pa moramo najprej vključiti moški in ženski spol prek atributa *speech gender/spol* (*speech gender="F|M"*), nato pa še vse mandate prek atributa *session mandate/mandat seje* (*session mandate_en="Term.*"*).

Uporaba frekvenčnih seznamov

V podkorporusih, ustvarjenih v prejšnjem koraku, bomo analizirali govore poslank in poslancev skozi čas, in sicer na podlagi podatkov s frekvenčnih seznamov,⁸² ki so ena od najosnovnejših korpusnih tehnik. Frekvenčni sezname prikazujejo rezultate od najbolj do najmanj pogostega. To tehniko lahko na primer uporabimo za sestavljanje frekvenčnega seznama vseh besed, ki so bile izrečene v parlamentu oziroma v govorih izbrane poslanke ali izbranega poslanca. V tej nalogi bomo s to tehniko pridobili podatke o številu vseh poslank in poslancev v posameznem mandatu ter številu njihovih govorov, vključno s skupnim številom pojavnic v teh govorih.

Videoposnetek, ki prikazuje izdelavo frekvenčnega seznama v konkordančniku NoSketch Engine, je na voljo na <https://youtu.be/LrYmwK51Fbc>.

Slika 4 prikazuje primer dveh frekvenčnih seznamov.

Kot prikazuje Slika 4, je bilo v prvem mandatu prisotnih 14 poslank, ki so skupno izrekle več kot 3.500 govorov, kar je le malo manj kot milijon pojavnic. Povprečno je to skoraj 260 govorov oziroma 69.000 pojavnic na poslanko. Vendar pa je porazdelitev med poslankami zelo neenakomerna, pri čemer sega od skoraj 700 govorov oziroma več kot 170.000 pojavnic do manj kot 40 govorov oziroma 9.000 pojavnic.

82 Wordlists – word frequency lists | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/v4ta5zbb>.

Slika 4. Frekvenčna seznama za poslanke iz prvega mandata. Na levi⁸³ je prikazano število govorov, ki jih je posamezna poslanka prispevala, na desni⁸⁴ pa je število pojavnic v teh govorih.

Seznam besed		Seznam besed	
Korpus: siParl 2.0 (parlament 1990-2018)		Korpus: siParl 2.0 (parlament 1990-2018)	
Podkorpus: Mandat1-Ženske--Term1-Female		Podkorpus: Mandat1-Ženske--Term1-Female	
Total number of items: 14		Total number of items: 14	
Total frequency: 3,606		Total frequency: 963,984	
<u>speech.name</u>	<u>document frequency</u>	<u>speech.name</u>	<u>frekvenca</u>
Simšič, Danica	680	Karner Lukač, Metka	173,067
Pečan, Breda	547	Simšič, Danica	171,366
Karner Lukač, Metka	460	Pečan, Breda	166,970
Skuk, Nada	283	Skuk, Nada	93,810
Primožič, Jana	273	Kožuh Novak, Mateja	72,451
Potočnik, Viktorija	247	Primožič, Jana	53,652
Kožuh Novak, Mateja	245	Potočnik, Viktorija	52,183
Šturm Kocjan, Jadranka	231	Oman, Irena	50,151
Oman, Irena	226	Šturm Kocjan, Jadranka	47,921
Pozsonec, Maria	163	Pozsonec, Maria	22,754
Dobrajc, Polonca	87	Dobrajc, Polonca	20,323
Čadonič Špelič, Vida	78	Čadonič Špelič, Vida	18,103
Bizilj, Ljerka	47	Bizilj, Ljerka	12,140
Logar, Mihaela	39	Logar, Mihaela	9,093

Po številu govorov je na prvem mestu Danica Simšič, članica Demokratske stranke Slovenije, manjše opozicijske stranke iz prvega mandata. Ta poslanka je prispevala 680 govorov ali skoraj 19 % vseh govorov v celotnem podkorpusu, kar je skoraj 3-kratno povprečje pri ženskah. Podatek o deležu govorov smo pridobili tako, da smo število njenih govorov delili s številom vseh govorov poslank (gl. Slika 4, levo: *Total frequency* in *Simšič, Danica*), podatek o razmerju glede na povprečje pa smo pridobili tako, da smo število poslankinih govorov delili s povprečjem števila vseh govorov, ki ga dobimo tako, da skupno število govorov poslank delimo s številom vseh poslank (gl. Slika 4, levo: *Total frequency* in *Total number of items*).

Po številu pojavnic je poslanka z največjim prispevkom v tem podkorpusu Metka Karner Lukač, članica Slovenske ljudske stranke, najstarejše stranke v Sloveniji. Njeni govori vsebujejo 173.067 pojavnic ali 18 % celotnega podkorpusa, kar je skoraj 3-kratnik povprečja. Poslanka z najnižjim številom govorov in pojavnic je Mihaela Logar, prav tako članica Slovenske ljudske stranke, ki je v svojem 4-letnem mandatu spregovorila le 39-krat, v njenih govorih pa je nekaj več kot 9.000 pojavnic, kar je skoraj 20-krat manj od prvuvrščene govorke. Poleg tega je še zanimivo, da približno polovica vseh prispevkov pripada le trem poslankam, in sicer že omenjenima poslankama ter Bredi Pečan, članici Socialnih demokratov, katere govori so bili v povprečju tudi najdaljši. Ta podatek razberemo iz podatkov o povprečni dolžini govorov za posamezno poslanko, pri čemer povprečno dolžino govora izračunamo tako, da število vseh pojavnic za posamezno poslanko (gl. Slika 4, desno) delimo s številom vseh govorov te iste poslanke (gl. Slika 4, levo).

83 Word list, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-WR9T>.

84 Word list, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-GSW5>.

Primerjalna analiza

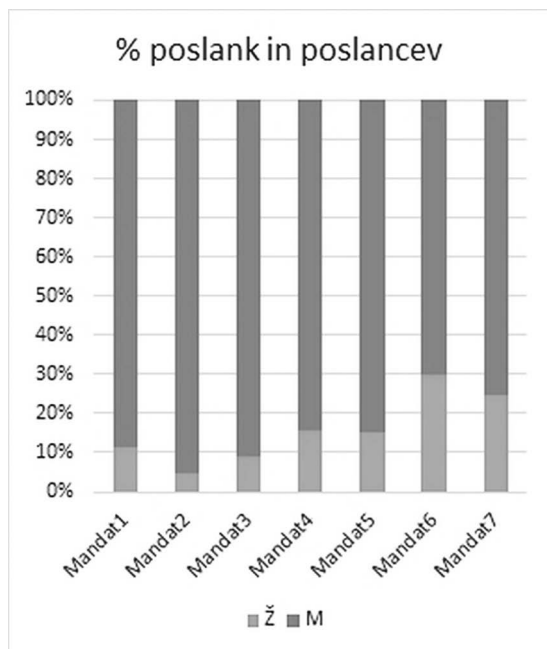
Za potrebe primerjalne analize zastopanosti moških in žensk v slovenskem parlamentu skozi čas zabeležimo število poslancev in poslank ter število pojavnic v vsakem od 14 podkorpusov (gl. razdelek 6.2.1 Izdelava podkorpusov). Ker zbirnih tabel s podatki ne dobimo neposredno v konkordančniku, si pomagamo tako, da podatke ročno vnesemo v preglednico, kot prikazuje Tabela 1. Grafični prikaz rezultatov si lahko ogledate na Sliki 5 in Sliki 6.

Tabela 1. Pregled velikosti podkorpusov siParl 2.0.⁸⁵

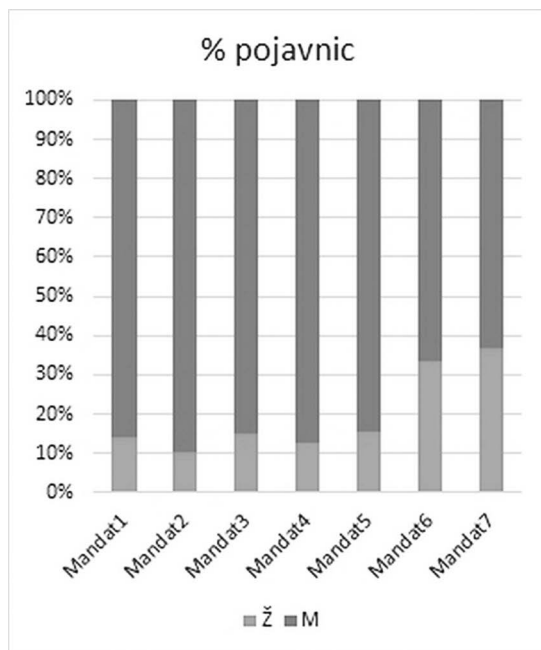
Podkorpus	# poslank in poslancev	% poslank in poslancev	# pojavnic	% pojavnic
Mandat1-Ženske	14	14%	963.984	11%
Mandat1-Moški	87	86%	7.808.562	89%
Mandat1-Skupaj	101	100%	8.772.546	100%
Mandat2-Ženske	11	10%	762.172	5%
Mandat2-Moški	95	90%	15.767.432	95%
Mandat2-Skupaj	106	100%	16.529.604	100%
Mandat3-Ženske	16	15%	1.502.758	9%
Mandat3-Moški	90	85%	15.092.725	91%
Mandat3-Skupaj	106	100%	16.595.483	100%
Mandat4-Ženske	13	13%	3.636.968	16%
Mandat4-Moški	90	87%	19.365.000	84%
Mandat4-Skupaj	103	100%	23.001.968	100%
Mandat5-Ženske	16	16%	3.069.515	15%
Mandat5-Moški	87	84%	16.980.714	85%
Mandat5-Skupaj	103	100%	20.050.229	100%
Mandat6-Ženske	38	33%	4.626.013	30%
Mandat6-Moški	76	67%	10.945.878	70%
Mandat6-Skupaj	114	100%	15.571.891	100%
Mandat7-Ženske	37	37%	6.644.314	25%
Mandat7-Moški	64	63%	20.389.258	75%
Mandat7-Skupaj	101	100%	27.033.572	100%
VsiMandati-Ženske	104	23%	21.205.724	17%
VsiMandati-Moški	356	77%	106.349.569	83%
VsiMandati-Skupaj	460	100%	127.555.293	100%

⁸⁵ Zaradi napake v metapodatkih korpusa, ki so bili pripisani enemu od poslancev (in sicer poslancu z imenom Geza, ki mu je bil napačno samodejno pripisan ženski spol najverjetneje zaradi kombinacije končnice -a, ki je sicer tipična za samostalnike ženskega spola, in redkosti imena), podatki za šesti mandat v času analize ne odražajo uradnega števila poslank, ki je po navedbah Simone K. Lipicer (2017) 39, in ne 38, kot prikazuje korpus.

Slika 5: Delež poslank in poslancev v korpusu siParl 2.0 skozi čas.



Slika 6: Prispevek poslank in poslancev v korpusu siParl 2.0, izmerjen v izrečenih pojavnih skozi čas.



Rezultati kažejo, da delež pojavnic, ki sestavljajo govore poslank, postopoma narašča skupaj z zastopanostjo žensk v parlamentu. Precejšen skok v zastopanosti žensk je mogoče opaziti predvsem v zadnjih dveh mandatih, vključenih v korpus siParl 2.0, kar je najverjetneje posledica sprejema spolnih kvot, v skladu s katerimi mora biti od leta 2011 na volilnih seznamih 35 % žensk.⁸⁶ Pomembna pa je ugotovitev, da kljub temu, da so kvote eden najuspešnejših ukrepov za povečanje števila poslank,⁸⁷ naša korpusna analiza prikazuje velika odstopanja med številom govork in številom pojavnic v njihovih govorih. Rezultati volitev so enostavno dostopni in raziskovalci in analitiki jih pogosto uporabljajo, število in vsebina govorov pa sta veliko težje dostopna, zato ju raziskovalci izven korpusnega jezikoslovja precej manj izčrpno preučujejo, kar predstavlja vrzel, ki jo želimo s to analizo premostiti. Podatki kažejo jasen trend, da poslanke prispevajo manjši delež vsebine, kot bi pričakovali glede na njihov delež v izbranem mandatu. To velja zlasti za drugi mandat, kjer govori poslank vsebujejo dvakrat manj pojavnic glede na njihovo število. Pravzaprav je le v četrtem mandatu prispevek žensk nekoliko večji od njihovega deleža v parlamentu. Zato lahko sklepamo, da več izvoljenih poslank ne pomeni nujno, da bo njihov glas tudi bolj prisoten v Državnem zboru. Vseeno pa se je v preteklih letih z večanjem števila poslank v parlamentu začelo povečevati tudi število njihovih govorov. Prispevek poslank se je s picle desetine vseh govorov v parlamentu v prvih petih mandatih povečal na skoraj tretjino v zadnjih dveh mandatih, ki jih pokriva naš korpus.

NALOGA 2: Ključne teme poslank in poslancev

Raziskave političnega diskurza žensk so pokazale, da ženske običajno razpravljajo o drugačnih temah kot moški. Več o vplivu spola na jezik si lahko preberete v razdelku 5 Jezik in spol.

V Nalogi 2 želimo na podlagi govorov iz korpusa siParl 2.0 primerjati teme, o katerih razpravljajo poslanke in poslanci. Za nekatere parlamentarne korpuse je klasifikacija tem že na voljo,⁸⁸ za korpus siParl 2.0 pa to ne velja. Klasifikacijo tem je mogoče izvesti samodejno, kot v članku *Analysis of policy agendas*,⁸⁹ vendar pa je cilj tega gradiva pokazati učinkovitost raziskav parlamentarnih korpusov zgolj prek uporabe koncordančnikov, torej brez kakršnegakoli znanja programiranja, zaradi česar smo izbrali ročni pristop v dveh korakih. Ročni pristopi so še posebej primerni za zelo specifično razvrščanje tem, kot to velja za našo analizo, za katero še ne obstajajo računalniški modeli ali učne množice za razvoj takih modelov. Pri tem je treba opozoriti, da lahko

86 Antić Gaber in Selišnik, »The Slovene Version of a „Fast Track“«, 337.

87 Ibid.

88 Gl. Federico Nanni, Mahmoud Osman, Yi-Ru Cheng, Simone Paolo Ponzetto in Laura Dietz, »UKParl: A Semantified and Topically Organized Corpus of Political Speeches«, v: Darja Fišer, Maria Eskevich in Franciska de Jong, ur., *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (Miyazaki: LREC 2018).

89 Mladen Karan, Jan Šnajder, Daniela Širinić in Goran Glavaš, »Analysis of Policy Agendas: Lessons Learned from Automatic Topic Classification of Croatian Political Texts«, *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (2016): 12–21.

tako pri samodejnem kot pri ročnem razvrščanju pride do napak, vendar lahko pri ročnem označevanju število napak, ki so posledica nenatančnosti ali pa tudi pristranskosti označevalca oziroma označevalke, zmanjšamo z večjim številom označevalk oz. označevalcev, s čimer pridobimo vpogled v stopnjo ujemanja med njimi, za nadaljnje delo pa lahko upoštevamo temo, ki jo je izbrala večina, oz. vse pripisane teme.

Luščenje ključnih besed

V tej nalogi bomo uporabili uveljavljeno tehniko korpusne analize, ki ji rečemo luščenje ključnih besed.⁹⁰ S to tehniko izbrani korpus primerjamo z referenčnim korpusom in tako pridobimo najbolj značilno besedišče izbranega korpusa, torej (pod)korpusa, ki ga analiziramo. Referenčni korpus pa je običajno velik reprezentativni korpus ustreznega jezika, lahko pa je tudi katerikoli drug (pod)korpus, ki ga želimo uporabiti kot referenčno točko za primerjavo. V tej nalogi bomo primerjali podkorpusa govorov poslank in poslancev za posamezni mandat. Tako bomo pridobili podatke, na podlagi katerih bomo lahko opredelili najbolj izstopajoče teme v slovenskem parlamentu.

Videoposnetek, ki prikazuje izdelavo seznama ključnih besed v konkordančniku NoSketchEngine, je na voljo na <https://youtu.be/DMiZP6f4qI8>.

Ustvarjeni sezname ključnih besed, prikazanih kot leme s poenoteno malo začetnico, so na voljo tukaj:

- Mandat1-Ženske⁹¹
- Mandat1-Moški⁹²
- Mandat7-Ženske⁹³
- Mandat7-Moški⁹⁴

Sezname ključnih besed za vse štiri podkorpuse izvozimo v preglednico in jih ročno označimo (gl. razdelek 6.3.3 Primerjalna analiza).

Sliki 7 in 8 prikazujeta prvih 20 ključnih besed poslank in poslancev v sedmem mandatu.⁹⁵ Opazimo lahko velike razlike. Večina vseh prikazanih ključnih besed poslank je povezana s področjem zdravstvenega varstva, medtem ko ključne besede govorov poslancev pripadajo področjem zunanjih zadev, infrastrukture in pravosodja.

90 *Keyword (linguistics)* – Wikipedia, pridobljeno 18. 2. 2021, <https://tinyurl.com/3eu8u66s>.

91 *Word list*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/sipar1-VBX5>.

92 *Word list*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/sipar1-74HT>.

93 *Word list*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/sipar1-AZM0>.

94 *Word list*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/sipar1-7E3V>.

95 Konkordančnik NoSketchEngine za računanje ključnih besed uporablja formulo Simple Maths.

Slika 7: Prvih 20 ključnih besed poslank v primerjavi s poslanci v prvem mandatu.

Seznam besed

Korpus: siParl 2.0 (parlament 1990-2018)
Podkorpus: Mandat1-Ženske--Term1-Female

Referenčni korpus: siParl 2.0 (parlament 1990-2018)
Referenčni podkorpus: Mandat1-Moški--Term1-Male
Spremeni cilj in referenčni (pod)korpus

Stran [Naslednja >](#)

lemma_lc	siParl 2.0 (parlament 1990-2018) : Mandat1-Ženske--Term1-Female		siParl 2.0 (parlament 1990-2018) : Mandat1-Moški--Term1-Male		Rezultat
	frekvenca	frekvenca/mil @	frekvenca	frekvenca/mil	
porodnišnica	78	80.9	11	1.4	34.0
deti	86	89.2	15	1.9	30.9
zavoljo	35	36.3	4	0.5	24.7
konjerejakovistističen	18	18.7	0	0.0	19.7
umrtjivost	27	28.0	4	0.5	19.2
brest	34	35.3	12	1.5	14.3
porod	18	18.7	4	0.5	13.0
prezgodnji	16	16.6	3	0.4	12.7
učinkovina	12	12.4	1	0.1	11.9
tržič	16	16.6	4	0.5	11.6
porodnica	16	16.6	4	0.5	11.6
učenka	10	10.4	0	0.0	11.4
medicinec	10	10.4	0	0.0	11.4
občanka	17	17.6	5	0.6	11.4
festival	17	17.6	5	0.6	11.4
higienski	11	11.4	1	0.1	11.0
cerknica	30	31.1	16	2.0	10.5
malodane	9	9.3	0	0.0	10.3
davkoplačevalski	12	12.4	3	0.4	9.7
prosenjakovci	14	14.5	3	0.6	9.5

Slika 8: Prvih 20 ključnih besed poslancev v primerjavi s poslankami v prvem mandatu.

Seznam besed

Korpus: siParl 2.0 (parlament 1990-2018)
Podkorpus: Mandat1-Moški--Term1-Male

Referenčni korpus: siParl 2.0 (parlament 1990-2018)
Referenčni podkorpus: Mandat1-Ženske--Term1-Female
Spremeni cilj in referenčni (pod)korpus

Stran [Naslednja >](#)

lemma_lc	siParl 2.0 (parlament 1990-2018) : Mandat1-Moški--Term1-Male		siParl 2.0 (parlament 1990-2018) : Mandat1-Ženske--Term1-Female		Rezultat
	frekvenca	frekvenca/mil @	frekvenca	frekvenca/mil	
rudnik	210	26.9	0	0.0	27.9
proporcionalen	194	24.8	0	0.0	25.8
premog	189	24.2	0	0.0	25.2
kaseta	140	17.9	0	0.0	18.9
tule	137	17.5	0	0.0	18.5
kolesarski	137	17.5	0	0.0	18.5
termoelektrarna	130	16.6	0	0.0	17.6
šoitanj	119	15.2	0	0.0	16.2
dama	116	14.9	0	0.0	15.9
steza	115	14.7	0	0.0	15.7
električen	249	44.7	2	2.1	14.9
gasilski	320	41.0	2	2.1	13.7
navkljub	91	11.7	0	0.0	12.7
vlak	188	24.1	1	1.0	12.3
konceptija	87	11.1	0	0.0	12.1
aluminij	84	10.8	0	0.0	11.8
niz	78	10.0	0	0.0	11.0
zapiranje	163	20.9	1	1.0	10.7
avtocesten	576	73.8	6	6.2	10.3
pledirati	71	9.1	0	0.0	10.1

Analiza konkordanc

Za nadaljnjo analizo izberemo prvih 100 ključnih lem z vsakega od štirih seznamov ključnih besed, pri čemer ne upoštevamo osebnih lastnih imen. Naš cilj je ročno kategorizirati leme, in sicer glede na prevladujočo temo, ki jo prepoznamo pri pregledu konkordanc⁹⁶ za izbrano lemo. Konkordančni niz vsebuje seznam vseh pojavitev iskane pojavnice v kontekstu, kot prikazuje Slika 9:

⁹⁶ Concordance – most powerful corpus search | Sketch Engine, pridobljeno 18. 2. 2021, <https://tinyurl.com/8zrs9f8n>.

- Konkordance je mogoče neposredno prikazati s klikom leme na seznamu ključnih besed. Prikaže se stran, kjer je na vrhu zaslona prikazana izbrana lema z vsemi zadetki (gl. številko 1).
- Besede v rdeči barvi na sredini zaslona (gl. številko 2) so zadetki iskane besede v našem podkorpusu, besedilo v črni barvi (gl. številko 3) pa je kontekst.
- Besedilo v modri barvi na levi (gl. številko 4) so metapodatki. V našem primeru je prikazan podatek o govorniku.
- S klikom zelene konkordance lahko še dodatno razširimo kontekst (gl. številko 5). Na podoben način lahko s klikom govornika pridobimo več metapodatkov.
- Prikaz lahko tudi prilagodimo, in sicer prek *Možnosti prikaza* (gl. številko 6), ter prikažemo več metapodatkov ali širši kontekst.

Slika 9: Razširjeni kontekst za prvi zadetek v konkordančnem nizu⁹⁷ za ključno besedo »proporcionalen«.

The screenshot shows the NoSketch Engine interface. At the top, there is a search bar with the query 'siPart 2.0 (parlament 1990-2018) / Mandat1-Moški--Term1-Male'. Below the search bar, the search results are displayed. The first result is 'Iskalni niz proporcionalen 194 (24.84 na milijon)'. The results are listed in a table with columns for rank, name, and text. The text is color-coded: red for search results, black for context, and blue for metadata. Numbered callouts (1-6) point to various elements: 1 points to the search bar, 2 to the search results, 3 to the context, 4 to the metadata, 5 to the expandable context, and 6 to the filters.

Primerjalna analiza

Parlament je v prvi vrsti zakonodajno telo, v katerem potekajo zakonodajne in proračunske razprave, ki so strukturirane po ministrstvih. Zato nam seznam 14 ministrstev trenutne slovenske vlade⁹⁸ služi kot seznam kategorij za ročno označevanje tem. Kategorije so navedene v Tabeli 2. Lahko bi uporabili tudi številne druge sezname tem, vendar je takšen nabor tem po našem mnenju najbolj naravna izbira v specifičnem okolju parlamentarnega diskurza. Tem 14 vsebinskim kategorijam dodamo še 4 tehnične kategorije za ključne besede, ki jih ni mogoče uvrstiti drugam: *razno* – za ključne besede, uporabljene v razpravah o več različnih temah, *slog* – za očitne pogovorne ali

⁹⁷ *Iskanje*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-VZB9>.

⁹⁸ *Ministrstva* | GOV.SI, pridobljeno 18. 2. 2021, <https://tinyurl.com/jbhnhjib>.

žargonske ključne besede, ki so izrazito značilne za posamezne govorce, *ideologija* – za ključne besede, uporabljene za ideološko označevanje (npr. *partijski*, *socialist*), in *interakcija/postopkovnik* – za ključne besede, ki naslavljajo druge poslance ali pa so del postopkovnika. V Tabeli 3 so navedeni ponazoritveni primeri ročno pripisanih tem za prvih 10 ključnih besed poslank in poslancev.

Tabela 2. Kategorije za označevanje ključnih besed. Zadnje štiri kategorije zapisujemo z velikimi tiskanimi črkami, da jih tako ločimo od vsebinskih kategorij.

Teme
Kmetijstvo, gozdarstvo in prehrana
Kultura
Obramba
Gospodarstvo in tehnologija
Izobraževanje, znanost in šport
Okolje in prostor
Finance
Zunanje zadeve
Zdravje
Infrastruktura
Notranje zadeve
Pravosodje
Delo, družina in socialne zadeve
Javna uprava
RAZNO
SLOG
IDEOLOGIJA
INTERAKCIJA/POSTOPKOVNIK

Tabela 3: Ponazoritveni primeri ročno pripisanih oznak glede na temo za prvih 10 ključnih besed poslank in poslancev v prvem mandatu.

Ženske		Moški	
Lema	Tema	Lema	Tema
<i>porodnišnica</i>	Zdravje	<i>rudnik</i>	Infrastruktura
<i>deti</i>	SLOG	<i>proporcionalen</i>	Javna uprava
<i>zavoljo</i>	SLOG	<i>premog</i>	Infrastruktura
<i>konjerejskovisturističen</i>	Gospodarstvo [...]	<i>kaseta</i>	Obramba
<i>umrljivost</i>	Zdravje	<i>tule</i>	SLOG
<i>brest</i>	Gospodarstvo [...]	<i>kolesarski</i>	Infrastruktura
<i>porod</i>	Zdravje	<i>termoelektrarna</i>	Infrastruktura
<i>prezgodnji</i>	Zdravje	<i>šoštanj</i>	Infrastruktura
<i>učinkovina</i>	Zdravje	<i>dama</i>	Interakcija [...]
<i>tržič</i>	Javna uprava	<i>steza</i>	Infrastruktura

Tabeli 4 in 5 vsebujeta povzetek rezultatov ročnega označevanja prvih 100 ključnih lem poslank in poslancev v prvem in sedmem mandatu. Rezultati kažejo, da je nabor tem skozi čas in med spoloma primerljiv. Kljub podobnemu številu opredeljenih vsebinskih tem se poslanci in poslanke pri najbolj izpostavljenih tematikah zelo razlikujejo.

Tabela 4: Teme prvih 100 ključnih besed poslank in poslancev v prvem mandatu.

Pogostnost tem v podkorpusu Mandat1-Ženske	Frek.	Pogostnost tem v podkorpusu Mandat1-Moški	Frek.
Zdravje	33	Infrastruktura	27
Delo, družina in socialne zadeve	21	SLOG	14
Javna uprava	11	Javna uprava	14
Kultura	8	Gospodarstvo in tehnologija	8
Gospodarstvo in tehnologija	6	RAZNO	8
Kmetijstvo, gozdarstvo in prehrana	4	Finance	6
Izobraževanje, znanost in šport	4	Zunanje zadeve	4
RAZNO	4	Kmetijstvo, gozdarstvo in prehrana	3
Finance	3	Okolje in prostor	3
SLOG	3	IDEOLOGIJA	3
Obramba	1	Zdravje	2
Infrastruktura	1	Pravosodje	2
Pravosodje	1	Delo, družina in socialne zadeve	2
		Kultura	1
		Obramba	1
		INTERAKCIJA/POSTOPKOVNIK	1
		Notranje zadeve	1

Tabela 5: Teme prvih 100 ključnih besed poslank in poslancev v sedmem mandatu.

Pogostnost tem v podkorpusu Mandat7-Ženske	Frek.	Pogostnost tem v podkorpusu Mandat7-Moški	Frek.
Zdravje	40	SLOG	42
Delo, družina in socialne zadeve	18	RAZNO	10
Kultura	11	Zunanje zadeve	9
Izobraževanje, znanost in šport	7	Obramba	5
Finance	4	IDEOLOGIJA	5
Pravosodje	4	Gospodarstvo in tehnologija	4
RAZNO	4	Okolje in prostor	4
Kmetijstvo, gozdarstvo in prehrana	3	Finance	4
Javna uprava	3	Infrastruktura	4
SLOG	2	Notranje zadeve	3
Gospodarstvo in tehnologija	1	Pravosodje	3
Zunanje zadeve	1	Javna uprava	3
Infrastruktura	1	Izobraževanje, znanost in šport	2
INTERAKCIJA/POSTOPKOVNIK	1	Kmetijstvo, gozdarstvo in prehrana	1
		INTERAKCIJA/POSTOPKOVNIK	1

V prvem mandatu večina (54 %) vseh analiziranih ključnih besed poslank spada v dve tematski kategoriji: *Zdravje in Delo, družina in socialne zadeve*, pri poslancih pa sta najpogostejši temi s podobnim deležem (55 %) *Infrastruktura in Javna uprava*. V sedmem mandatu sta najpogostejši temi poslank enaki kot v prvem mandatu, vendar predstavljata večji delež ključnih besed (58 %), pri poslancih pa je najpogostejša tema postala *Zunanje zadeve*. Razdelitve tem med spoloma ni mogoče v celoti razložiti s korpusno analizo, vendar pa dobljeni rezultati, ki pri poslankah kažejo izrazito zanimanje za zdravstvene in socialnovarstvene teme, pri poslancih pa premik z infrastrukturnih tematik na zunanjepolitične, zelo nazorno odražajo stanje v družbi. V času neodvisnosti je morala novonastala država oblikovati javno upravo in izgraditi infrastrukturo, za kar so bile v prvem mandatu potrebne intenzivne razprave. Zadnji mandat pa zaznamujejo živahna mednarodna trgovina in večje mednarodne varnostne grožnje, kar vpliva na zakonodajne in proračunske odločitve. V sedmem mandatu lahko še vedno zasledimo poudarjeno zanimanje za socialnovarstvene tematike, poleg tega pa tudi intenzivnejše razprave s področja zdravstva, kar je v veliki meri posledica močnega pritiska na proračun zaradi hude gospodarske krize v tistem obdobju, ki je negativno vplivala tudi na že propadajoči sistem javnega zdravstva.

Opazimo lahko tudi, da v prvem mandatu poslanci in poslanke namenjajo skoraj enako pozornost (glede na delež analiziranih ključnih besed) večjemu številu tem kot v sedmem mandatu (pet skupnih tem z enako pomembnostjo v prvem mandatu oz. tri skupne teme v sedmem mandatu). V prvem mandatu tako poslanci le rahlo pogosteje kot poslanke razpravljajo o *Javni upravi* (M: 14 %, Ž: 11 %), *Gospodarstvu in tehnologiji* (M: 8 %, Ž: 6 %) ter *Pravosodju* (M: 2 %, Ž: 1%), poslanke pa nekaj pogosteje o *Kmetijstvu, gozdarstvu in prehrani* (M: 3 %, Ž: 4 %), medtem ko vsi razpravljajo enako pogosto o *Obrambi* (1 %). V sedmem mandatu so glede na delež analiziranih ključnih besed poslankam in poslancem enako pomembne zgolj tri teme, in sicer *Finance* (4 %), *Pravosodje* (3 %) in *Javna uprava* (3 %). Za poglobitev rezultatov bi bilo zanimivo raziskati, na kakšne načine poslanke in poslanci pristopajo k tem skupnim temam in o njih razpravljajo.

Med temami, ki se pojavijo zgolj v podkorpusu govorov poslank, najdemo *Izobraževanje, znanost in šport* v prvem mandatu ter *Zdravje, Delo, družino in socialne zadeve* ter *Kulturo* v sedmem mandatu. Teme, ki se pojavijo zgolj pri poslancih, pa so *Okolje in prostor* ter *Notranje zadeve* v obeh parlamentarnih mandatih, *Zunanje zadeve* (v prvem mandatu) in *Obramba* (v sedmem mandatu). Presenetljivo je, da se najpogostejše tri teme poslank sploh ne pojavijo na seznamu tem poslancev, medtem ko se najpogostejše teme poslancev (*Infrastruktura, Javna uprava, Zunanje zadeve*) pojavijo na seznamu tem poslank, čeprav je njihov delež majhen. Poleg tega je zanimivo, da tema *Gospodarstvo in tehnologija* pri poslankah izgubi na pomembnosti med prvim in sedmim mandatom, medtem ko to ne velja za poslance.

Rezultati poleg različnih področij zanimanja kažejo tudi na slogovne razlike v razpravah poslank in poslancev. V obeh mandatih so poslanci uporabili precej več slogovno zaznamovanih besed (tj. očitnih pogovornih ali žargonskih besed, ki so izrazito

značilne za posamezne govorce) kot poslanke. Uporaba takšnih besed pri poslancih se je v sedmem mandatu v primerjavi s prvim mandatom potrojila, kar po eni strani kaže na večjo sproščenost poslancev na splošno, po drugi pa označuje spremembo v kulturi razpravljanja, ki je v zadnjih dveh desetletjih postala živahnejša in bolj neformalna. Opazimo lahko tudi, da so ideološke besede, ki običajno delujejo razdvajajoče, značilne le za poslance. To je v skladu z literaturo, kjer avtorji ugotavljajo, da so moške sporazumevalne strategije bolj agresivne in tekmovalne, medtem ko je ženski slog sporazumevanja bolj sodelovalen.⁹⁹ Vendar naj opozorimo, da bi bilo treba te zaključke preveriti s podrobnejšo analizo, saj nekatere sodobne raziskave kažejo, da na sporazumevanje vplivajo številni dejavniki in je zato razlikovanje med moškim in ženskim načinom sporazumevanja pogosto preveč simplistično.¹⁰⁰

Naša analiza kaže tudi na opazne razlike v vlogah in zanimanjih poslank in poslancev, kar je v skladu s prejšnjimi raziskavami, iz katerih je razvidno, da se ženske v primerjavi z moškimi bolj osredotočajo na t. i. mehka zakonodajna področja. Dihrone primerjave razkrivajo, da temi *Zdravje* in *Socialne zadeve* ostajata med prednostnimi zanimanji poslank, osredotočenost poslancev pa se je premaknila s tem *Infrastruktura* in *Javna uprava na Zunanje zadeve*. V prvem mandatu je bila najpogostejša skupna tema *Javna uprava*, v zadnjem mandatu pa so poslanke in poslanci najpogosteje razpravljali o temi *Finance*. To pa ne pomeni, da poslanke zdaj več razpravljajo o t. i. trdih zakonodajnih področjih. Pravzaprav je ravno nasprotno, saj analiza kaže, da so se poslanke v sedmem mandatu celo bolj osredotočale na t. i. mehka zakonodajna področja kot v prvem mandatu.

NALOGA 3: Obravnava tematik, povezanih z ženskami

Ta naloga se naslanja na sorodne raziskave,¹⁰¹ v katerih so raziskovalci ugotavljali, kako pogosto so bile v parlamentarni zgodovini naslovljena vprašanja enakosti spolov (npr. pravice žensk, diskriminacija itd.) ter kdo in na kakšen način je takšna vprašanja odpiral. Zanimivo je, da je vpliv spola očitno opazen tudi v državah z visokim deležem žensk v parlamentu. Ena takšnih je Švedska, za katero so Bäck idr.¹⁰² ugotovili, da poslanke redkeje razpravljajo o vprašanjih s t. i. trdih zakonodajnih področij. Antić Gaber in Ilonszki¹⁰³ pravita, da družba običajno pričakuje, da bodo poslanke dejavne na drugih področjih politike kot moški. Pri tem gre, kot to ugotavljajo politologi,¹⁰⁴ za področja politike, ki so še posebej povezana z ženskami zaradi njihove zgodovinske vloge v družbi ali neposrednega vpliva na življenje žensk.

99 Jennifer Coates, *Women, Men and Language*.

100 Cornelia Ilie, »Gendering Confrontational Rhetoric: Discursive Disorder in the British and Swedish Parliaments,« *Democratization* 20, št. 3 (2013): 501–21.

101 Blaxill in Beelen, »A Feminized Language of Democracy«.

102 Bäck, Debus in Müller, »Who Takes the Parliamentary Floor?«

103 Antić Gaber in Ilonszki, *Women in Parliamentary Politics*.

104 Gl. Tracey L. Osborn, *How Women Represent Women: Political Parties, Gender and Representation in the State Legislatures* (Oxford: Oxford University Press, 2012).

Ker se pri vprašanih enakosti spolov še vedno največkrat govori o neenaki vlogi žensk v družbi in njihovem posebnem položaju, iz katerega izhajajo tako določene pravice in teme, ki se tičejo njihovega življenja, kot tudi diskriminatorne prakse, bomo v Nalogi 3 raziskali, katere tematike, povezane z ženskami, prevladujejo v parlamentarnih razpravah in kako pogosto se o njih razpravlja, pri čemer se bomo osredotočili na uporabo samostalnika »ženska« kot eksplicitnega pokazatelja razprav o teh tematikah.

Delo s frekvencami

V tem delu nas najprej zanima, kako pogosto poslanci in poslanke uporabljajo lemo samostalnika »ženska« v različnih časovnih obdobjih (1992–2018). Izvedemo poizvedbe v vseh podkorporusih in v preglednico zabeležimo podatke o frekvenci, kot prikazuje Tabela 6. Ker želimo primerjati podkorporusa različnih velikosti, moramo namesto absolutne frekvence uporabiti normalizirano frekvenco,¹⁰⁵ pri kateri je uporabljeno enako skupno število vseh besed v (pod)korporusu. Absolutna frekvenca je v primeru primerjave korpusev različnih velikosti namreč lahko zavajajoča. Če na primer pogledamo podatka o absolutni frekvenci izbrane besede za drugi mandat (Ž: 190; M: 794), bi lahko sklepali, da poslanci uporabijo samostalnik »ženska« štirikrat pogosteje kot poslanke. Ker pa je podkorporus govorov poslank v tem parlamentarnem mandatu veliko manjši od podkorporusa govorov poslancev, nam šele normalizirana frekvenca (v našem primeru na milijon besed), pove, da so ta samostalnik poslanke izrekle pravzaprav petkrat pogosteje (249,29) kot poslanci (50,36).

Prikaz poizvedbe v konkordančniku NoSketchEngine za lemo samostalnika »ženska« si lahko ogledate v videoposnetku o luščenju kolokacij (min. 00:10–1:13). <https://youtu.be/ENXP8TrH42U>.

Konkordančni niz za lemo samostalnika »ženska« je na voljo tukaj: VsiMandati-Ženske¹⁰⁶

Slika 10: Začetek konkordančnega niza za lemo samostalnika »ženska« z absolutno frekvenco (številka 1) in normalizirano frekvenco (številka 2).

1 2

Iskalni niz (ženska)-s 4,691 (221.21 na milijon)

Stran 1 od 94 Pojdi Nazlednja Zadnja

1	Muršič, Bojana	zbirke omogočala prilagojeno obveščanje	žensk	o pomembnih terminih in vsebini priporočenih
2	Kustec Lipicer, Simona	sami obravnava. Zakon o enakih možnostih	ženske	in moških zagotavlja predvsem enakost spolov,
3	Kustec Lipicer, Simona	besedami zagotavlja pogoje, da bodo tako	ženske	kot moški imeli po določenem kvotnem sistemu
4	Bah Žibert, Anja	proti lastnim državljanom, proti starejšim	ženskam	, državljankam Republike Slovenije, in jim
5	Tomič, Violeta	skoraj 16 % vseh upokojeencev, od tega seveda so	ženske	v slabšem položaju tako kot vedno. Pod pragom
6	Irgl, Eva	treba pomagati. Tem je treba pomagati, zlasti	ženskam	in otrokom. Kakšni ljudje bi pravzaprav bili,
7	Bizjak Mlakar, Julijana	tisoč upokojeencev, od tega 57 tisoč upokojenih	žensk	in 26 tisoč upokojenih moških. To je skoraj 17 %

105 Normalizing word counts – *The Grammar lab*, pridobljeno 18. 2. 2021, <https://tinyurl.com/pkwaaf34>.

106 *Iskanje*, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-YMFX>.

Tabela 6. Podatki o absolutni in normalizirani frekvenci za lemo samostalnika »ženska« v korpusu siParl 2.0.

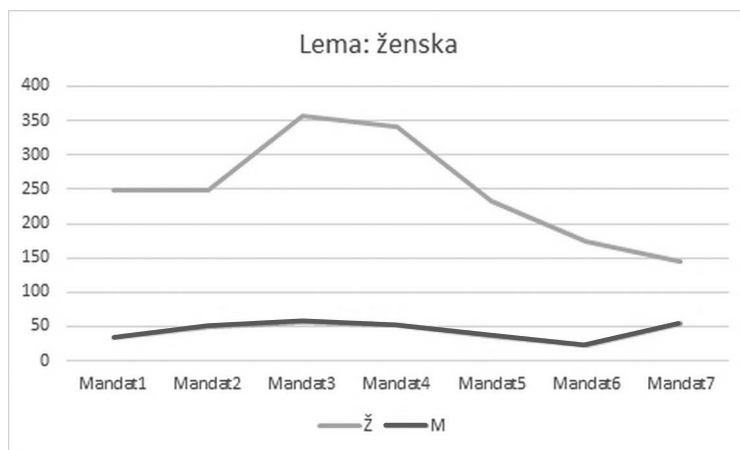
Podkorpus	Absolutna frek.	Normalizirana frek.	Podkorpus	Absolutna frek.	Normalizirana frek.
VsiMandati-Ženske	4691	221,21	VsiMandati-Moški	4891	45,99
Mandat1-Ženske	240	248,97	Mandat1-Moški	263	33,68
Mandat2-Ženske	190	249,29	Mandat2-Moški	794	50,36
Mandat3-Ženske	537	357,34	Mandat3-Moški	863	57,18
Mandat4-Ženske	1242	341,49	Mandat4-Moški	1004	51,85
Mandat5-Ženske	712	231,96	Mandat5-Moški	624	36,75
Mandat6-Ženske	809	174,88	Mandat6-Moški	245	22,38
Mandat7-Ženske	961	144,63	Mandat7-Moški	1098	53,85

Na splošno je normalizirana frekvenca iskane besede za celotno časovno obdobje, ki ga zajema korpus siParl 2.0, skoraj petkrat višja v podkorpusu govorov poslank kot v podkorpusu govorov poslancev (221,21 proti 45,99). Kot je razvidno iz Slike 10, poslanke v svojih govorih pogosteje omenjajo ženske kot poslanci. Čeprav je v obdobju med drugim in četrtem mandatom (1996–2008) mogoče opaziti intenzivnejšo obravnavo teh področij pri poslankah, je v zadnjih mandatih njihovo zanimanje za ta področja izrazito usahnilo.

Najbolj presenetljive rezultate pri poslankah smo odkrili v tretjem in petem mandatu. V tretjem mandatu je normalizirana frekvenca iskane besede dosegla rekordno visoko stopnjo, kar je lahko posledica manjšega števila poslank v prejšnjih dveh parlamentarnih mandatih in neenakopravnega položaja žensk v družbi nasploh. Ker pa je enak trend mogoče opaziti tudi v podkorpusu govorov poslancev, se zdi, da so k temu prispevali tudi drugi razlogi. Za boljši vpogled v ta pojav smo izvedli kvalitativno analizo, pri kateri smo pregledali 50 konkordanc petih najdejavnejših poslancev. Iz konkordanc je razvidno, da se je v tistem času veliko razpravljalo o zakonodaji prav v zvezi z vprašanji enakosti spolov (npr. zakonodaja o enakih možnostih in spolnih kvotah). Skoraj desetletje pozneje, v petem mandatu, se je število omemb iskane besede v podkorpusu govorov poslank nenadoma zmanjšalo in padlo celo pod vrednost frekvence pred letom 2000, kljub precej večjemu številu poslank v parlamentu. Peti mandat je potekal v obdobju velike svetovne gospodarske krize, ki je Slovenijo močno prizadela in je bila verjetno osrednji predmet parlamentarnih razprav, vendar bi bilo to treba potrditi z nadaljnjo raziskavo in kontekstualizacijo ob uporabi kvalitativnih metod, kot je analiza konkordanc.

Zanimivo je tudi, da se v četrtem mandatu kljub drugemu najnižjemu številu poslank samostalnik »ženska« še vedno zelo pogosto pojavlja v primerjavi s šestim in sedmim mandatom, kjer je bilo število poslank precej večje, frekvenca iskane besede pa skoraj dvakrat nižja. Kot smo opazili že pri splošnem doprinosu poslank k razpravam (gl. razdelek 6.2.3 Primerjalna analiza), kjer smo odkrili veliko razhajanje med številom poslank in številom besed, ki so jih izrekle, ti rezultati znova kažejo, da zgolj večje število poslank ne zagotavlja intenzivnejše razprave o tematikah, ki zadevajo ženske.

Slika 11: Normalizirana frekvenca za lemo samostalnika »ženska« v govorih poslank in poslancev v korpusu siParl 2.0.



Luščenje kolokacij

Predstavili bomo še eno priljubljeno tehniko korpusne analize, in sicer luščenje kolokacij.¹⁰⁷ S to tehniko, ki temelji na statističnih testih, pridobimo seznam besed, ki se ob jedrni besedi pojavljajo pogosteje, kot bi se zgolj po naključju. Kolokacije se najpogosteje uporabljajo v leksikografiji in na sorodnih področjih uporabnega jezikoslovja, mi pa jih bomo uporabili kot sredstvo za raziskovanje konceptov ali tem, o katerih se razpravlja v parlamentu.

Da bi lahko ugotovili, o katerih tematikah, ki zadevajo ženske, se v parlamentu največ razpravlja, bomo analizirali kolokacije samostalnika »ženska« v dveh podkorpusih, ki vsebujeta govore poslank oziroma poslancev iz vseh sedmih mandatov, torej podkorpusa *VsiMandati-Ženske* in *VsiMandati-Moški*.

Videoposnetek, ki prikazuje postopek luščenja kolokacij v konkordančniku NoSketchEngine, je na voljo na <https://youtu.be/ENXP8TrH42U>.

Izluščimo kolokacije v razponu ene besede levo in ene besede desno od jedrne besede (tj. leme samostalnika »ženska«), pri čemer se mora kolokacijski kandidat (tj. beseda, ki se tipično pojavlja ob jedrni besedi) v korpusu pojaviti vsaj petkrat, skupaj z jedrno besedo v določenem razponu pa vsaj trikrat. Razpon in najnižjo frekvenco je mogoče nastaviti ročno in sta odvisna od jedrne besede, velikosti korpusa ter cilja naše raziskave. Ker želimo analizo v tem gradivu omejiti na ustaljene besedne zveze, uporabljamo ozek razpon in stroga merila glede frekvence.

¹⁰⁷ Kolokacija – Wikipedija, prosta enciklopedija, pridobljeno 18. 2. 2021, <https://tinyurl.com/5b87rmft>.

Za določanje trdnosti kolokacij uporabimo statistično mero logDice.¹⁰⁸ Čeprav konkordančnik NoSketch Engine ponuja tudi druge mere, po katerih se izračuna trdnost kolokacij, kot sta Mutual Information ali T-score, izberemo mero logDice, ker nanjo ne vpliva velikost korpusa in so zato rezultati uporabni za primerjalno analizo tudi v primeru uporabe podkorpusov različnih velikosti, kot to velja za našo analizo.

Seznama kolokacij za lemo samostalnika »ženska« sta na voljo tukaj:

- VsiMandati-Ženske¹⁰⁹
- VsiMandati-Moški¹¹⁰

Oba seznama kolokacij uvozimo v preglednico in ju ročno analiziramo (gl. razdelek 6.4.3 Primerjalna analiza).

Primerjalna analiza

Z vsakega seznama vzamemo prvih 100 kolokacijskih kandidatov (tj. besed, ki se tipično pojavljajo ob jedrni besedi in torej skupaj z njo tvorijo kolokacijo) in jih ročno razdelimo v tri kategorije: ženske, moški in skupno.¹¹¹ Nato vsako besedo razvrstimo v enega od sedmih tematskih sklopov, ki smo jih oblikovali na podlagi predhodnega pregleda konkordanc, kot prikazuje Tabela 7. Kolokacijskim kandidatom, ki so slovnične besede (na primer predlogi) in niso uporabljeni v konkordancah na eno prevladujočo temo, pripišemo oznako *Razno*.

Tabela 7: Primeri sklopov kolokacij.

Primeri kolokacij			
Tematski sklop	Ženske	Moški	Skupno
Politika, zastopanost in enakopravnost	glas večina /	voliti procent obraz	izvoljen delež pravica
Družbeni status in zaposlitev	upokojen samostojnost	poročen /	brezposeln doba
Zdravje in reprodukcija	zdrav rojevati	/ oploditev	moten neploden
Problemi in nasilje	ostajati posiljen	diskriminirati pobiti	diskriminacija /
Proaktivnost žensk in opolnomočenje	potencial	protest	uspešen
Pomoč in zaščita	omogočati	korist	vključevanje

108 logDice | Sketch Engine, pridobljeno, 18. 2. 2021, <https://tinyurl.com/3dv63nhm>.

109 Collocations, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-QJKC>.

110 Collocations, pridobljeno 18. 2. 2021, <http://hdl.handle.net/11346/siparl-Q5MN>.

111 Kategorija skupno, kot nakazuje njeno ime, združuje kolokacijske kandidate, ki se pojavijo tako na seznamu za podkorpus govorov poslank kot za podkorpus govorov poslancev.

Rezultati, predstavljeni v Tabeli 8, kažejo, da se nekaj več kot polovica kolokacij (51 %) pojavlja tako pri poslankah kot pri poslancih, kar kaže na skupne točke v razumevanju položaja žensk v sodobni družbi. Velika večina skupnih kolokacij (skoraj 70 %) spada v prva dva tematska sklopa: *Politika, zastopanost in enakopravnost* in *Družbeni status in zaposlitev*, in se nanaša na koncepte, povezane z zastopanostjo žensk (npr. »participacija«), njihovim družbenim položajem (npr. »samski«) in enakostjo (npr. »emancipacija«).

Tabela 8: Frekvenca tematskih sklopov za prvih 100 kolokacijskih kandidatov v govorih poslank in poslancev v vseh sedmih mandatih.

Ženske		Moški		Skupno	
Tematski sklop	Frek.	Tematski sklop	Frek.	Tematski sklop	Frek.
Zdravje in reprodukcija	6	Problemi in nasilje	11	Politika, zastopanost in enakopravnost	23
Problemi in nasilje	6	Politika, zastopanost in enakopravnost	11	Družbeni status in zaposlitev	12
Pomoč in zaščita	6	Proaktivnost žensk in opolnomočenje	8	Zdravje in reprodukcija	4
Družbeni status in zaposlitev	5	Družbeni status in zaposlitev	6	Proaktivnost žensk in opolnomočenje	3
Politika, zastopanost in enakopravnost	5	Pomoč in zaščita	5	Problemi in nasilje	2
Proaktivnost žensk in opolnomočenje	3	Zdravje in reprodukcija	3	Pomoč in zaščita	2
RAZNO	18	RAZNO	5	RAZNO	5
Skupaj	49	Skupaj	49	Skupaj	51

Opazimo lahko tudi, da so v podkorpusu govorov poslank primerljivo zastopani vsi tematski sklopi razen enega, skupaj pa pokrivajo skoraj 60 % kolokacij. Edina izjema je torej tematski sklop *Proaktivnost žensk in opolnomočenje*. Kolokacije, ki spadajo v ta tematski sklop, so približno dvakrat redkejše kot kolokacije iz kateregakoli drugega sklopa. V podkorpusu govorov poslancev večina kolokacij (61 %) spada v tri tematske sklope, in sicer *Problemi in nasilje*, *Politika, zastopanost in enakopravnost* in *Proaktivnost žensk in opolnomočenje*. Poleg tega podroben pregled oznak razkriva, da se poslanci osredotočajo predvsem na družbeni status žensk (npr. »poročen«) in njihove težave (npr. »zatiranje«) ter na njihovo proaktivnost in opolnomočenje (npr. »sposoben«), medtem ko težavam, povezanim z zaposlitvijo in zdravjem žensk, ne posvečajo posebne pozornosti. Kolokacije iz podkorpusa govorov poslank se v prvi vrsti nanašajo na področje zdravstva (npr. »zdrav«) in ukrepe, vezane na pomoč in zaščito (npr. »omogočiti«). Tem sledijo kolokacije, ki se nanašajo na družbeni status žensk (npr. »izobrazba«) in nasilje (npr. »posiljen«), kar je spet v nasprotju s podkorpusom govorov poslancev, kjer je večji poudarek na težavah, ki ne vključujejo nasilja (npr. »poniževanje«). Podobno kot v podkorpusu govorov poslancev pa tudi v podkorpusu govorov poslank med prvimi 100 kolokacijskimi kandidati ni besed, povezanih z zaposlitvijo.

Ti rezultati kažejo, da poslanke poskušajo obravnavati številna vprašanja, s katerimi se ženske srečujejo na različnih področjih življenja, in sicer od zdravstvenih vprašanj do nasilja nad ženskami in težav pri političnem delovanju. Poslanci pa se bolj osredotočajo na neenakopravnost žensk v sodobni družbi in ženske vidijo kot dejavne udeleženke v procesu sprememb. Ti rezultati potrjujejo tudi ugotovitve avtoric Antić Gaber in Ilonszki,¹¹² ki sta opazovali dejavnost žensk v slovenskem parlamentu v krajšem obdobju (1996–2004) in z raznolikimi nekorpusnimi metodami analize pokazali, da med poslankami in poslanci obstaja jasna razlika med zakonodajnimi prioriteta, pri čemer sta opredelili podobne teme kot naša analiza.

Naša zadnja ugotovitev zadeva veliko število slovničnih besed v podkorpusu govorov poslank, ki so skoraj štirikrat pogostejše kot v podkorpusu govorov poslancev. Slovnične besede kot kolokacijski kandidati v podkorpusu govorov poslank večinoma vključujejo prislove, predloge, veznike, zaimke in števnike, v podkorpusu govorov poslancev pa prevladujejo le števniki. Podobno kot smo ugotovili pri analizi ključnih besed (gl. razdelek 6.3.3 Primerjalna analiza), to znova kaže na drugačen slog razpravljanja pri poslankah. Ta ugotovitev je skladna s sorodnimi raziskavami, v katerih raziskovalci poročajo, da ženske običajno uporabljajo več jezikovnih sredstev, kot so diskurzni omejevalci, vljudnostne oblike in vprašalni pristavki,¹¹³ več intenzifikatorjev in (predvsem ocenjevalnih) pridevnikov¹¹⁴ ter več osebnih zaimkov kot moški, ki pri sporazumevanju uporabljajo več števnikov, členov in predlogov.¹¹⁵

Zaključki

Cilj učnega gradiva je bil prikazati uporabnost jezikovnih korpusov za analizo družbenokulturnih pojavov, ki jih lahko raziskujemo na podlagi jezikovne rabe v specializiranem diskurzu. Pokazali smo, kako korpusnoanalitične tehnike omogočajo ugotovitve, ki presegajo intuicijo raziskovalke oz. raziskovalca in tako ponujajo večjo preglednost, objektivnost, zanesljivost in ponovljivost raziskav, kar postajajo vse pomembnejša načela tudi na področju humanistike in družboslovja.

Glavni doprinos tega učnega gradiva lahko razdelimo na tri dele. Najprej smo pokazali, kako pomembno je razumeti vsebino in strukturo podatkov, ki jih uporabljamo v raziskavi, da lahko čim bolj izkoristimo njihov potencial. Nato smo predstavili, kako lahko nabor standardnih metod korpusne analize uporabimo za veliko več kot le kvantifikacijo in preproste poizvedbe v korpusu. Namesto tega smo rezultate, pridobljene s konkordančnikom, sistematično uporabili kot odskočno desko za podrobno ročno kvalitativno analizo, ki rezultate previdno umešča v ustrezni družbenojezikovni kontekst. Čeprav smo preučevali razprave iz slovenskega parlamenta, so predstavljeni analitični postopki splošni in jih je mogoče uporabiti tudi na (parlamentarnih)

112 Antić Gaber in Ilonszki, *Women in Parliamentary Politics*.

113 Coates, *Women, Men and Language*.

114 Cate Poynton, *Language and Gender: Making the Difference* (Oxford: Oxford University Press, 1989).

115 Newman, Groom, Handelman in Pennebaker, »Gender Differences in Language Use«.

korpusih v drugih jezikih.¹¹⁶ Nenazadnje smo analizo umestili tudi v resnično raziskovalno okolje, s čimer smo skušali prikazati uporabnost splošnih metod korpusne analize za obravnavo številnih raziskovalnih vprašanj na področju humanistike in družboslovja.

Zahvale

Izdelavo tega učnega gradiva so podprli CLARIN.SI, slovenski del Evropske infrastrukture za jezikovne vire in tehnologije CLARIN, raziskovalni program ARRS P6-0215 (Slovenski jezik – osnovne, kontrastivne in uporabne študije) in projekt H2020 SSHOC – Social Sciences and Humanities Open Cloud (GA 823782). Posebna zahvala gre tudi red. prof. dr. Milici Antić Gaber, doc. dr. Ireni Selišnik, prof. dr. Corneliu Ilie, Tjaši Cankar, Mladenu Zobcu in Rubenu Rosu za strokovno recenzijo in testiranje učnega gradiva ter vse njihove dragocene komentarje.

Viri in literatura

- Alasuutari, Pertti, Marjaana Rautalin in Jukka Tyrkkö. »The Rise of the Idea of Model in Policymaking: The Case of the British Parliament, 1803–2005.« *European Journal of Sociology/Archives Européennes de Sociologie* 59, št. 3 (2018): 341–63. <https://doi.org/10.1017/S0003975618000164>.
- Antić Gaber, Milica in Gabriella Ilonszki. *Women in Parliamentary Politics: Hungarian and Slovene Cases Comparur*. Ljubljana: Peace Institute, Institute for Contemporary Social and Political Studies, 2003.
- Antić Gaber, Milica in Irena Selišnik. »The Slovene Version of a ‚Fast Track‘ to Political Equality.« *Teorija in Praksa* 54, št. 2 (2017): 337.
- Bäck, Hanna, Marc Debus in Jochen Müller. »Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag.« *Political Research Quarterly* 67, št. 3, (2014): 504–18. <https://doi.org/10.1177/1065912914525861>.
- Baker, Paul. *Using Corpora to Analyze Gender*. London: Bloomsbury, 2014. <https://doi.org/10.1177/0075424215578282>.
- Biber, Douglas in Randi Reppen. *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 2015. <https://doi.org/10.1017/CBO9781139764377>.
- Biel, Łucja, Agnieszka Biernacka in Anna Jopek-Bosiacka. »Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations.« V: *Language and Law*, 249–74. Cham: Springer, 2018. https://doi.org/10.1007/978-3-319-90905-9_14.
- Bing, Janet M. in Victoria L. Bergvall. »The Question of Questions: Beyond Binary Thinking.« V: Jennifer Coates (ur.), *Language and Gender: A Reader, Vol. 1*, 496–510. Oxford: Blackwell, 1998.
- Blaxill, Luke in Kaspar Beelen. »A Feminized Language of Democracy? The Representation of Women at Westminster since 1945.« *Twentieth Century British History* 27, št. 3 (2016): 412–49. <https://doi.org/10.1093/tcbh/hww028>.
- Coates, Jennifer. *Women, Men and Language: A Sociolinguistic Account of Gender Differences in Language*. London: Longman, 1997.
- Demmen, Jane, Lesley Jeffries in Brian Walker. »Charting the Semantics of Labour Relations in House of Commons Debates Spanning Two Hundred Years.« *Doing Politics: Discursivity, Performativity and Mediation in Political Discourse* 80 (2018): 81. <https://doi.org/10.1075/dapsac.80.04dem>.

¹¹⁶ Resource families | CLARIN ERIC, pridobljeno 18. 2. 2021, <https://www.clarin.eu/resource-families>.

- Eckert, Penelope in Sally McConnell-Ginet. *Language and Gender*. Cambridge: Cambridge University Press, 2013. <https://doi.org/10.1017/CBO9781139245883>.
- Goddard, Angela in Lindsey M. Patterson. *Language and Gender*. London: Routledge, 2015.
- Hansen, Dorte H., Costanza Navarretta in Lene Offersgaard. »A Pilot Gender Study of the Danish Parliament Corpus.« V: Darja Fišer, Maria Eskevich in Franciska de Jong (ur.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki: LREC 2018.
- Ilie, Cornelia. »Gendering Confrontational Rhetoric: Discursive Disorder in the British and Swedish Parliaments.« *Democratization* 20, št. 3 (2013): 501–21. <https://doi.org/10.1080/13510347.2013.786547>.
- Ilie, Cornelia. »Parliamentary Debates. V: Ruth Wodak in Bernhard Forchtner (ur.), *The Routledge Handbook of Language and Politics*, 309–25. London: Routledge, 2017. <https://doi.org/10.4324/9781315183718>.
- Jaworska, Sylvia in Kath Ryan. »Gender and the Language of Pain in Chronic and Terminal Illness: A Corpus-based Discourse Analysis of Patients' Narratives.« *Social Science & Medicine* 215 (2018): 107–14. <https://doi.org/10.1016/j.socscimed.2018.09.002>.
- Karan, Mladen, Jan Šnajder, Daniela Širinić in Goran Glavaš. »Analysis of Policy Agendas: Lessons Learned from Automatic Topic Classification of Croatian Political Texts.« *Proceedings of the 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities* (2016): 12–21. <https://doi.org/10.18653/v1/W16-2102>.
- Kustec Lipicer, Simona, ur. *Poslanke Državnega zbora: Uradno in osebno od prvega do sedmega mandata (1992-2017)*. Ljubljana: Državni zbor, 2017.
- Leijenaar, Monique. *How to Create a Gender Balance in Political Decision-making: A Guide to Implementing Policies for Increasing the Participation of Women in Political Decision-making*. Luxembourg: Office for Official Publications of the European Communities, 1997.
- Litosseliti, Lia. *Gender and Language: Theory and Practice*. London: Hodder Arnold, 2006. https://doi.org/10.1111/j.1467-9841.2008.00371_10.x.
- Marshall, Catherine. »Policy Discourse Analysis: Negotiating Gender Equity.« *Journal of Education Policy* 15, št. 2 (2000): 125–56. <https://doi.org/10.1080/026809300285863>.
- McEnery, Tony in Andrew Hardie. *Corpus linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2011. <https://doi.org/10.1017/CBO9780511981395>.
- Mensah, Eric O. in Sandra F. Wood. »Articulations of Feminine Voices in Ghana's Parliament: A Study of the Hansard from 2010-2011.« *AFRREV LALIGENS: An International Journal of Language, Literature and Gender Studies* 7, št. 2 (2018): 61–77. <https://doi.org/10.4314/laligens.v7i2.6>.
- Mollin, Sandra. »The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts.« *Corpora* 2, št. 2 (2007): 187–210. <https://doi.org/10.3366/cor.2007.2.2.187>.
- Nanni, Federico, Mahmoud Osman, Yi-Ru Cheng, Simone Paolo Ponzetto in Laura Dietz. »UKParl: A Semantified and Topically Organized Corpus of Political Speeches.« V: Darja Fišer, Maria Eskevich in Franciska de Jong (ur.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*. Miyazaki: LREC 2018.
- Newman, Matthew L., Clara J. Groom, Lori D. Handelman in James W. Pennebaker. »Gender Differences in Language Use: An Analysis of 14,000 Text Samples.« *Discourse Processes* 45, št. 3 (2008), 211–36. <https://doi.org/10.1080/01638530802073712>.
- Osborn, Tracey L. *How Women Represent Women: Political Parties, Gender and Representation in the State Legislatures*. Oxford: Oxford University Press, 2012. <https://doi.org/10.1093/acprof:oso/9780199845347.001.0001>.
- Pančur, Andrej, Tomaž Erjavec, Mihael Ojsteršek, Mojca Šorn in Neja Blaj Hribar. *Slovenian Parliamentary Corpus (1990-2018) siParl 2.0*. Slovenian language resource repository CLARIN.SI, 20202020, <http://hdl.handle.net/11356/1300>.
- Poynton, Cate. *Language and Gender: Making the difference*. Oxford: Oxford University Press, 1989.
- Rix, Kathryn. »'Whatever Passed in Parliament Ought to be Communicated to the Public': Reporting the Proceedings of the Reformed Commons, 1833–50.« *Parliamentary History* 33, št. 3 (2014): 453–74. <https://doi.org/10.1111/1750-0206.12106>.

- Wängnerud, Lena. »Intressen kontra stereotyper. Varför finns det kvinnliga och manliga politikområden i riksdagen?.« *Statsvetenskaplig Tidskrift* 99, št. 2(1996).
- Wodak, Ruth. »Pragmatics and Critical Discourse Analysis: A Cross-disciplinary Inquiry.« *Pragmatics & Cognition* 15 (2007): 203–25. <https://doi.org/10.1075/pc.15.1.13wod>.
- Wolbrecht, Christina. »Female Legislators and the Women's Rights Agenda: From Feminine Mystique to Feminist Era.« V: Cindy S. Rosenthal (ur.), *Women transforming congress*, 170–97. Norman: University of Oklahoma Press, 2002.

Darja Fišer, Kristina Pahor de Maiti

**“FIRST, I’M A FEMALE POLITICIAN, NOT A MALE ONE,
AND SECOND ...” A CORPUS APPROACH
TO PARLIAMENTARY DISCOURSE RESEARCH**

SUMMARY

Parliaments represent the main fora for political debate that shapes legislation which directly impacts our everyday lives. Parliamentary discourse is motivated by a wide range of communicative roles and reveals political, ideological and institutional patterns, which is why it has always been of interest to scholars from a range of disciplines in the humanities and social science. Parliamentary proceedings are being increasingly made available in a digitized form, and turned into structured linguistic resources called corpora for many languages. Researchers use them to perform diverse linguistic, stylistic, cultural, societal and political studies.

This tutorial shows the potential of richly annotated parliamentary corpora for research of the socio-cultural context and changes over time that are reflected through language use. The tutorial encourages students and scholars from digital humanities and social sciences who are interested in the study of socio-cultural phenomena through language, to engage with user-friendly digital tools for the analysis of large text collections. The tutorial is designed in a way that takes full advantage of both linguistic annotations and the available speaker and text metadata to formulate powerful quantitative queries that are then further extended with manual qualitative analysis in order to ensure adequate framing and interpretation of the results.

The tutorial demonstrates the potential of the siParl 2.0 corpus which contains parliamentary debates of the National Assembly of the Republic of Slovenia from 1990 to 2018 via concordancers without the need for programming skills. No prior experience in using language corpora and corpus querying tools is required in order to follow this tutorial.

This tutorial starts with a brief introduction to corpora and corpus analysis, followed by an introduction of the characteristics of specialized corpora of parliamentary debates and an overview of research into language and gender. The second part of the

tutorial is hands-on and demonstrates the potential of some of the best-known corpus analysis techniques, such as concordances, frequency lists, keywords and collocations, to explore the topics female MPs debate in in the Slovenian Parliament over time and to compare and contrast their language use with that of their male counterparts.