**»Prvič, sem političarka in ne politik, drugič pa …«**

**Korpusni pristop k raziskovanju parlamentarnega diskurza**

**Povzetek**

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 diahronega korpusa slovenskih parlamentarnih razprav za digitalno humanistiko in družboslovje.

**Kazalo**

[1 Uvod 2](#_Toc57789620)

[2 Kako uporabljati učno gradivo 2](#_Toc57789621)

[3 Korpusi in konkordančniki 3](#_Toc57789622)

[3.1 Korpusi 3](#_Toc57789623)

[3.2 Konkordančniki 4](#_Toc57789624)

[4 Zapisi parlamentarnih sej 5](#_Toc57789625)

[4.1 Parlamentarni diskurz 5](#_Toc57789626)

[4.2 Zanesljivost sejnih zapisov 5](#_Toc57789627)

[4.3 Poznavanje raziskovalnih podatkov 6](#_Toc57789628)

[5 Jezik in spol 6](#_Toc57789629)

[6 Korpusna analiza 7](#_Toc57789630)

[6.1 Korpus siParl 2.0 7](#_Toc57789631)

[6.2 NALOGA 1: Zastopanost žensk v slovenskem parlamentu 9](#_Toc57789632)

[6.2.1 Izdelava podkorpusov 10](#_Toc57789633)

[6.2.2 Uporaba frekvenčnih seznamov 11](#_Toc57789634)

[6.2.3 Primerjalna analiza 12](#_Toc57789635)

[6.3 NALOGA 2: Ključne teme poslank in poslancev 14](#_Toc57789636)

[6.3.1 Luščenje ključnih besed 15](#_Toc57789637)

[6.3.2 Analiza konkordanc 16](#_Toc57789638)

[6.3.3 Primerjalna analiza 17](#_Toc57789639)

[6.4 NALOGA 3: Obravnava tematik, povezanih z ženskami 21](#_Toc57789640)

[6.4.1 Delo s frekvencami 22](#_Toc57789641)

[6.4.2 Luščenje kolokacij 24](#_Toc57789642)

[6.4.3 Primerjalna analiza 25](#_Toc57789643)

[7 Zaključki 27](#_Toc57789644)

[8 Zahvale 27](#_Toc57789645)

[9 Reference 27](#_Toc57789646)

# Uvod[[1]](#footnote-1)

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 državljank (Ilie, 2017). 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 (Biel idr., 2018; Jaworska in Ryan, 2018).

Medtem ko so v jezikoslovju metode korpusne analize že dodobra uveljavljene (Biber in Reppen, 2015; McEnery in Hardie, 2011) in se med drugim pogosto uporabljajo tudi za sociolingvistično proučevanje spolno zaznamovanega jezika (Baker, 2014), 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.

S tem učnim gradivom želimo prikazati možnosti, ki jih raziskovalcem brez programerskih znanj ponujajo obsežne zbirke parlamentarnih razprav in konkordančniki. To učno gradivo je primerno tudi za uporabnike brez predhodnih izkušenj z uporabo jezikovnih korpusov in orodij za iskanje po njih. Gradivo temelji na [korpusu siParl 2.0](http://hdl.handle.net/11356/1300), 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 seznami, 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 in skozi čas.

Vsi viri in orodja, uporabljeni v tem gradivu, so brezplačno na voljo v spletu. Za iskanje po korpusu uporabljamo [NoSketch Engine](https://nlp.fi.muni.cz/trac/noske), za dodatne ročne analize in predstavitve rezultatov pa enega od urejevalnikov preglednic (npr. [Google Spreadsheet](https://www.google.com/sheets/about/) ali [MS Excel](https://products.office.com/en/excel)).

Besedila v modrih 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 Engine](https://www.clarin.si/noske/) in [KonText](https://www.clarin.si/kontext/first_form?corpname=siparl20) na spletnem mestu [CLARIN.SI](http://www.clarin.si/info/o-projektu/) – slovenski center v okviru Evropske raziskovalne infrastrukture za jezikovne vire in tehnologije [CLARIN ERIC](https://www.clarin.eu/). Prek repozitorija CLARIN.SI je korpus siParl 2.0 mogoče tudi [prenesti](https://www.clarin.si/repository/xmlui/handle/11356/1300?locale-attribute=sl) v računalnik in ga nato analizirati z drugimi korpusnimi orodji ali orodji za rudarjenje besedil, npr. z orodjem [za korpusno analizo Antconc](https://programminghistorian.org/en/lessons/corpus-analysis-with-antconc) ali s postopki za [osnovno obdelavo besedil v jeziku R](https://programminghistorian.org/en/lessons/basic-text-processing-in-r).

To gradivo je posodobljena različica [izvirnega učnega gradiva](https://www.modernlanguagesopen.org/articles/10.3828/mlo.v0i0.295/), v katerem je bila uporabljena prva različica korpusa siParl. Korpus siParl 2.0 v primerjavi s korpusom [siParl 1.0](http://hdl.handle.net/11356/1236) vsebuje bogatejše metapodatke o govorkah in govorcih 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 identitete, ki pa jih v naši analizi ne bomo uporabljali, zato jih v gradivu ne obravnavamo.

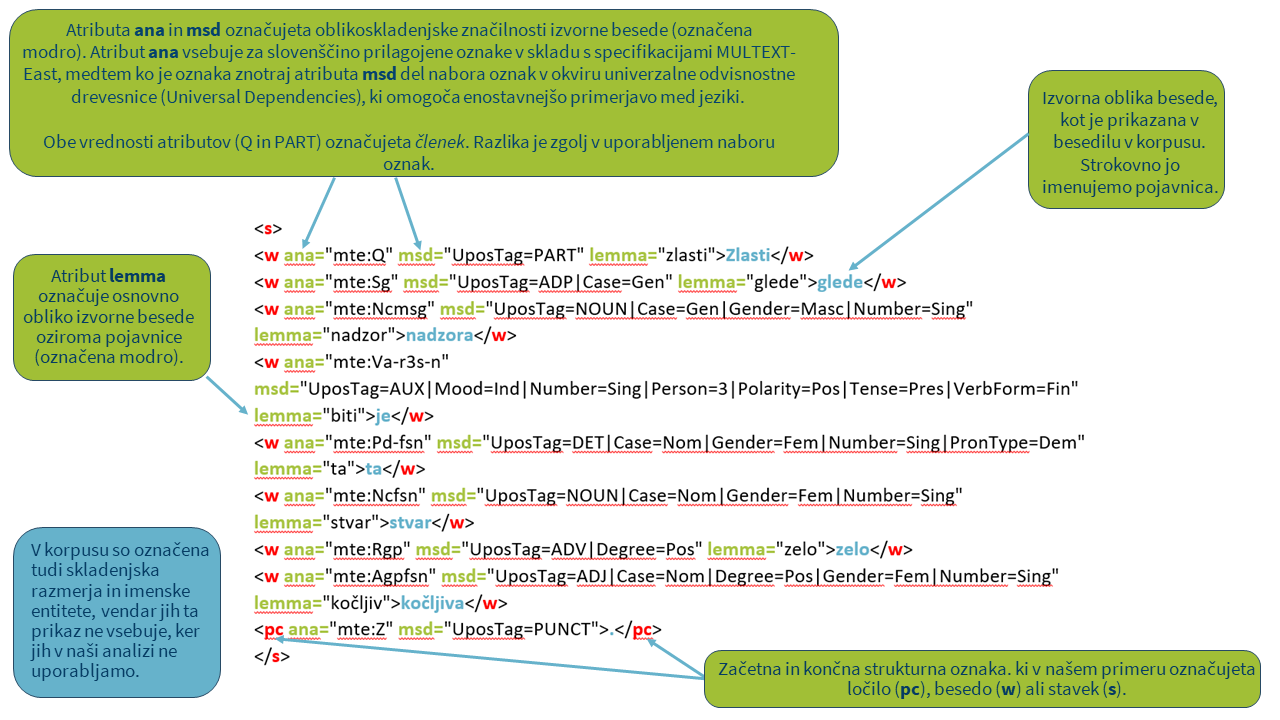
Če želite tehnike korpusne analize takoj preizkusiti na praktičnih primerih, nadaljujte z nalogami v razdelku Korpusna analiza. Naslednji trije razdelki (Korpusi in konkordančniki, Zapisi parlamentarnih sej in Jezik in spol) so namreč namenjeni poljudnemu pregledu teoretičnih osnov in najpomembnejše terminologije. Čeprav nalogam v 6. 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 korpusi](https://sl.wikipedia.org/wiki/Besedilni_korpus) 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](https://www.w3schools.com/xml/), 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](https://tei-c.org/).

Osnovna enota korpusa je [pojavnica](https://www.sketchengine.eu/my_keywords/token/), ki jo dobimo z avtomatskim postopkom tokenizacije, pri katerem je izbrano besedilo razdeljeno na manjše enote, v našem 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](https://en.wikipedia.org/wiki/Lemmatisation), pri katerem pojavnicam (besednim oblikam v različnih sklonih) pripišemo osnovno slovarsko obliko oz. lemo. Lematizacija je pomembna zlasti za korpuse morfološko bogatih jezikov, kot je slovenščina. Poleg jezikoslovnih oznak so v korpus dodani tudi metapodatki o besedilu (npr. datum, vrsta) in govorcu (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 TEI.**

## Konkordančniki

Po korpusih iščemo s specializiranimi orodji, ki se imenujejo [konkordančniki](http://corpora.lancs.ac.uk/clmtp/2-conc.php). Nekatere konkordančnike moramo pred uporabo namestiti na osebni računalnik, drugi pa so dostopni neposredno prek spleta. Uporabljamo jih za iskanje vseh pojavitev pojavnic v korpusu, ki ustrezajo našemu iskalnemu pogoju. Na voljo je veliko konkordančnikov, ki omogočajo podobne funkcije (glej npr. [ta podrobni seznam](http://martinweisser.org/corpora_site/concordancers.html)), med najbolj priljubljenimi pa so [AntConc](http://www.laurenceanthony.net/software/antconc/) (brezplačen) in [WordSmith Tools](http://www.lexically.net/wordsmith/index.html) (potreben je nakup licence), ki ju je mogoče uporabljati brez internetne povezave in ne vsebujeta predhodno naloženih korpusov, ter spletna konkordančnika [BYU](https://www.english-corpora.org/) (za polno delovanje je potrebna brezplačna registracija) in [SketchEngine](https://www.sketchengine.eu/documentation/corpus-querying/) (brezplačna akademska licenca), ki vsebujeta številne predhodno naložene korpuse za različne 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**](https://www.sketchengine.eu/guide/concordance-a-tool-to-search-a-corpus/) 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 seznami**](https://www.sketchengine.eu/user-guide/word-list/)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 seznami ne vsebujejo sobesedila zadetkov.
* [**Seznami ključnih besed**](https://www.sketchengine.eu/guide/keywords-and-term-extraction/) izpostavljajo besede, ki v izbranem korpusu izstopajo v primerjavi z izbranim referenčnim korpusom.
* [**Seznami kolokacij**](https://www.sketchengine.eu/my_keywords/collocation/) vsebujejo nize besed, ki se v kombinaciji z iskano besedo pojavljajo pogosteje od naključnega.

Na voljo so že številna navodila za uporabo konkordančnikov (npr. [SketchEngine Quick Start Guide](https://www.sketchengine.eu/quick-start-guide/)) in tečaji korpusnega jezikoslovja (npr. [Corpus Linguistics: Method, Analysis, Interpretation](https://www.futurelearn.com/courses/corpus-linguistics)). 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, govorjeni, gestikulacija), iz različnih časovnih obdobij in v enem ali več jezikih. Čeprav so parlamentarne razprave večinoma dostopne v besedilni obliki (npr. [korpus razprav iz danskega parlamenta](https://repository.clarin.dk/repository/xmlui/handle/20.500.12115/8)), so včasih dostopne tudi kot zvočni posnetki ali videoposnetki skupaj s transkripcijami (npr. [korpus razprav iz češkega parlamenta](https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0005-CF9C-4)). Parlamentarni korpusi kot [diahroni](https://en.wikipedia.org/wiki/Synchrony_and_diachrony) viri nam omogočajo poglobljeno raziskovanje jezikovnih in družbenih sprememb skozi čas. Večina parlamentarnih korpusov vsebuje bogate [metapodatke](https://sl.wikipedia.org/wiki/Metapodatek) 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 (glej Alasuutari idr., 2018; Demmen idr., 2018).

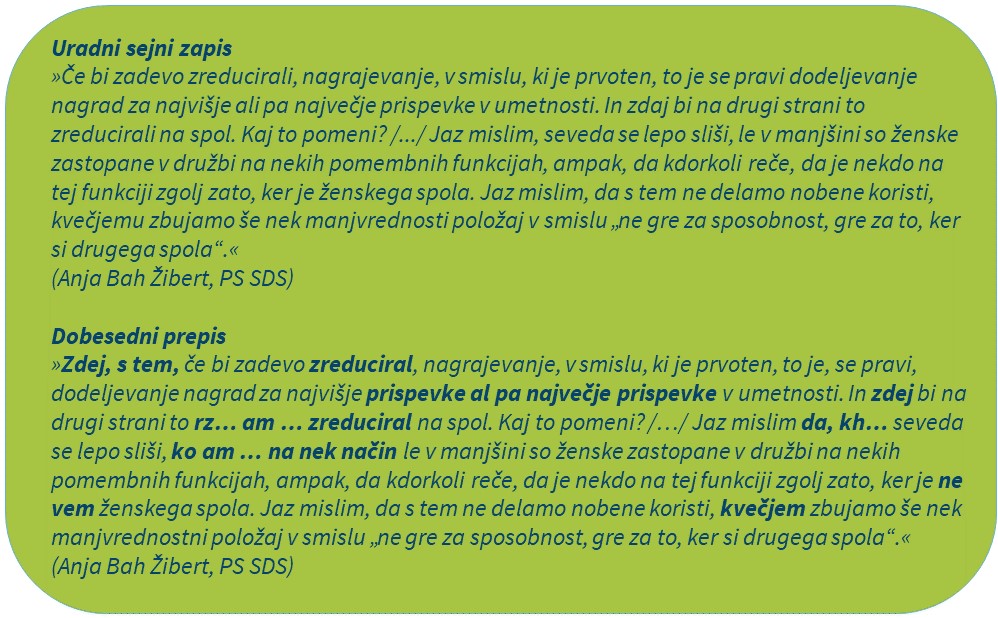
## Parlamentarni diskurz

Parlament je institucija, za katero veljajo specifična pravila in konvencije, kar vpliva tudi na jezikovno rabo, začenši s pravili glede podeljevanja besede, naslavljanja itd. (glej npr. [poslovnik državnega zbora RS](http://www.pisrs.si/Pis.web/pregledPredpisa?id=POSL34)). 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 (Ilie, 2017). To je seveda treba upoštevati pri oblikovanju raziskovalnih vprašanj, izvedbi raziskave in interpretaciji rezultatov. Zato je pri raziskovanju te vrste [diskurza](https://en.wikipedia.org/wiki/Discourse) še posebej pomembno, da dobro razumemo okoliščine, v katerih je nastal.

## Zanesljivost sejnih zapisov

Uradno objavljeni sejni zapisi niso dobesedne [transkripcije](https://en.wikipedia.org/wiki/Transcription_(linguistics)) 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 govorjenega 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 delo. Kritiko virov v parlamentarnem kontekstu podrobneje obravnavata Molli (2007) in Rix (2014). Drug pomemben vidik pri proučevanju parlamentarnega govora je, da so govori poslank in poslancev pogosto pripravljeni 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](https://sl.wikipedia.org/wiki/Anja_Bah_%C5%BDibert), poslanke Državnega zbora RS 2014–2018, [s 33. seje Državnega zbora na temo predloga zakona o Prešernovi nagradi (18. točka dnevnega reda, str. 128)](https://fotogalerija.dz-rs.si/datoteke/Publikacije/Sejni_zapisi_Drzavnega_zbora/2014-2018/2017_09_18_S_33.pdf) z dobesedno transkripcijo njenega govora, ki smo jo pripravili na podlagi [videoposnetka seje (min. 5:00–7:33)](https://4d.rtvslo.si/arhiv/redne-seje-drzavnega-zbora/174491899). 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.**

## 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 nemščine](http://www.chlg.ac.uk/helipad/)), v konkordančniku, prek katerega je korpus na voljo (npr. britanski parlamentarni korpus [Hansard v konkordančniku Univerze Brigham Young](https://www.english-corpora.org/hansard/)), ali v repozitoriju, kjer je korpus arhiviran, (npr. hrvaški parlamentarni korpus [ParlaMeter-hr v repozitoriju CLARIN.SI](https://www.clarin.si/repository/xmlui/handle/11356/1209)). Več o kodiranju in označevanju parlamentarnih podatkov lahko preberete v izobraževalnem modulu o zbirkah parlamentarnih podatkov [Parthenos](https://training.parthenos-project.eu/sample-page/digital-humanities-research-questions-and-methods/researching-parliamentary-records-in-the-digital-humanities/working-with-parliamentary-corpora/).

# 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 (glej Eckert in McConnell-Ginet, 2013; Wodak, 2007). Rezultati kažejo, da so razlike majhne, a sistematične (Newman idr., 2008). 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 (glej Marshall, 2000).

To je pomembno, ker so številni avtorji (Antić Gaber in Ilonszki, 2003; Leijenaar, 1997; Wolbrecht, 2002) pokazali, da se zakonodajalke razlikujejo od svojih sodelavcev pri vprašanjih, ki jih naslavljajo, stališčih, ki jih zavzemajo, in pristopih, ki jih uporabljajo pri pripravi zakonodaje. Blaxill in Beelen (2016) sta v analizi zastopanosti žensk v britanskem parlamentu po letu 1945 odkrila, da je v govorih poslank večji poudarek na vprašanjih, ki zadevajo ženske (npr. izobraževanje, zdravstvo itd.) in ki se neposredno dotikajo njihove dobrobiti v družbi (npr. varstvo otrok, nasilje v družini, enakost itd.) (glej Osborn, 2012). Tudi raziskovalci Bäck idr. (2014), Hansen idr. (2018) ter Mensah in Wood (2018), ki so raziskovali korpuse razprav iz švedskega, danskega oz. ganskega parlamenta, so odkrili, da so ženske pogosteje kot moški govorile o temah, ki se tičejo t. i. *mehkih zakonodajnih področij* (po Wängnerud, 1996, npr. makroekonomija, energetika, promet, bančništvo, finance, vesolje, znanost in tehnologija ter komunikacija), 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 (Coates, 1997; Litosseliti, 2006). Bing in Bergvall (1998) poudarjata, da so podobnosti v uporabi jezika pri različnih spolih običajno spregledane, čeprav so bolj izrazite kot razlike. Tudi Blaxill in Beelen (2016) 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 (Goddard in Patterson, 2015).

# Korpusna analiza

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

* V Nalogi 1 bomo *analizirali zastopanost 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 podkorpusih.
* 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 s pomočjo analize konkordanc** preverimo njihov kontekst in rabo.
* V Nalogi 3 bomo *raziskali, kako so poslanke in poslanci v minulih 25 letih razpravljali o tematikah, povezanih z ženskami.* Ogledali si bomo, **kako uporabljamo podatke o relativni frekvenci** v različno velikih podkorpusih in **kako izluščimo kolokacije** izbranih samostalnikov v podkorpusih.

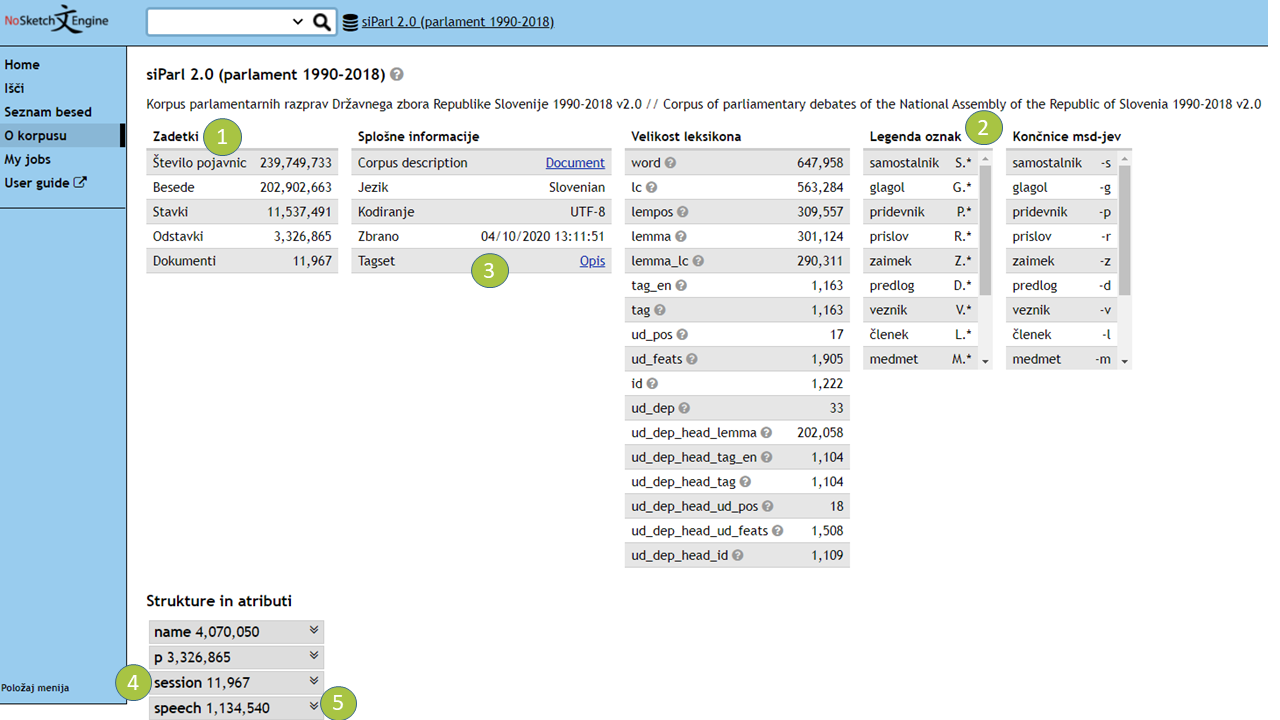
## Korpus siParl 2.0

[Korpus siParl 2.0](http://hdl.handle.net/11356/1300) (Pančur idr., 2020) vsebuje sejne zapise 11. mandata Skupščine socialistične republike Slovenije za obdobje 1990-1992, sejne zapise [Državnega zbora](https://sl.wikipedia.org/wiki/Dr%C5%BEavni_zbor_Republike_Slovenije) Republike Slovenije 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).

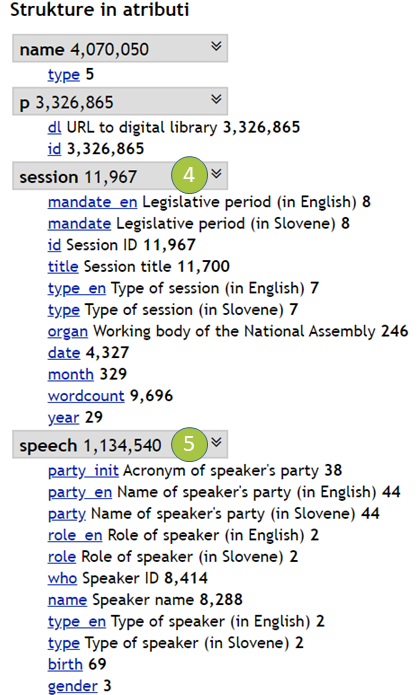
Korpus vsebuje metapodatke o govorkah in govorcih (npr. ime, spol, vloga, politična stranka), tipologijo parlamentarnih zasedanj (npr. redna seja, izredna seja itd.) ter strukturne in uredniške oznake (npr. zakonodajno obdobje, leto/naslov seje, [imenske entitete](https://en.wikipedia.org/wiki/Named_entity)). Korpus je tudi oblikoskladenjsko označen in lematiziran. Poleg tega vsebuje nekatere druge jezikoslovne oznake (npr. oznake univerzalne odvisnostne drevesnice (Universal Dependencies) in skladenjsko razčlenjevanje), ki pa jih v tem gradivu ne uporabljamo. Korpus obsega več kot milijon govorov 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](https://www.clarin.si/noske/run.cgi/corp_info?corpname=siparl20&struct_attr_stats=1) na strani [CLARIN.SI](http://www.clarin.si/info/o-projektu/) (glej Sliki 3a in 3b)[[2]](#footnote-2):

* Kvantitativne informacije o velikosti korpusa so na voljo v razdelku *Zadetki* (glej številko 1).
* Osnovne oznake za besedne vrste so navedene v *Legendi oznak* (glej številko 2), opis celotnega nabora oznak pa je na voljo prek povezave *Tagset/Opis****[[3]](#footnote-3)***(glej številko 3).
* Korpus je organiziran prek strukturnih atributov. Atribut *session/seja* zamejuje posamezne seje(glej številko 4) in vsebuje informacije, kot so datum, vrsta in naslov seje (glej Sliko 3b, številko 4).
* Atribut *speech/govor* zamejuje posamezne govore (glej številko 5) in vsebuje informacije o imenu, spolu, vlogi in strankarski pripadnosti govorke oz. govorca (glej Sliko 3b, številko 5).

****

**Slika 3a: Osnovne informacije o korpusu siParl 2.0.**

****

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

## 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](https://www.dz-rs.si/wps/portal/Home/ODrzavnemZboru/KdoJeKdo/PoslanskeSkupine). 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 četrtino 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 (Selišnik in Antić Gaber, 2017). 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.[[4]](#footnote-4) Glede na [EU indeks vrednosti enakosti spolov in politične moči v Evropski uniji za leto 2017](https://eige.europa.eu/gender-equality-index/2017/compare-countries/index/bar) 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 (glej Sliko 3b), korpus razdelimo na več delov, t. i. podkorpuse, in sicer glede na naslednja merila:

* **Spol.** Govorke in govorci 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 govorcem je v korpusu pripisana ena od dveh vlog, in sicer *član parlamenta* oziroma *zunanji govornik*[[5]](#footnote-5)(ta vloga označuje povabljene govorke in govorce, ki niso del parlamenta (npr. članice in člani vlade, predstavnice in predstavniki ministrstev ali nevladnih organizacij). V tem gradivu upoštevamo le govorce in govorke 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 govorke in govorce v parlamentu, ki so v skladu s poslovnikom izrecno dobili besedo. V naši analizi upoštevamo le te. 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](https://en.wikipedia.org/wiki/Socialist_Republic_of_Slovenia), 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 tej povezavi.

Ustvarjeni podkorpusi in informacije o njihovi velikosti so na voljo [tukaj](https://www.clarin.si/noske/run.cgi/subcorp?corpname=siparl20).

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

**Mandat1-Ženske[[6]](#footnote-6):**

<speech gender="F" & role\_en="MP" & type\_en="Regular speaker"/> within <session mandate\_en="Term 1 .\*" & organ="Državni zbor Republike Slovenije"/>

S tem ukazom poiščemo vse izjave (atribut *speech*/*govor*)) govork (atribut *F*/Ž), ki so članice parlamenta (atribut *MP*/Član parlamenta) in redne govorke *(*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*).

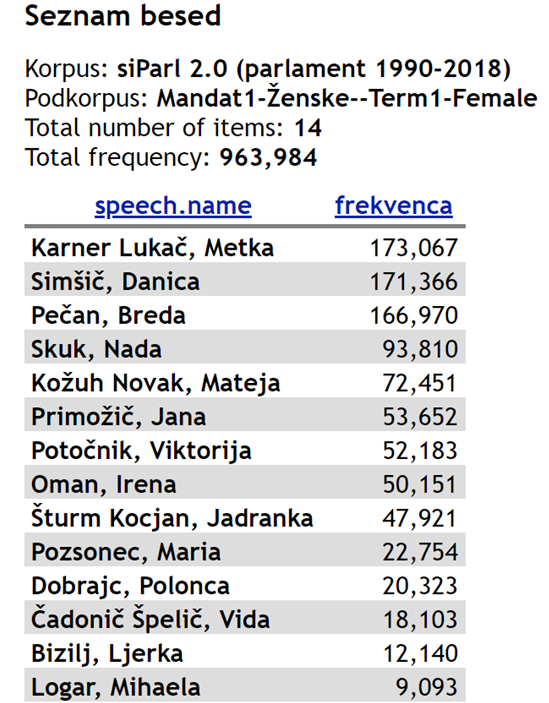
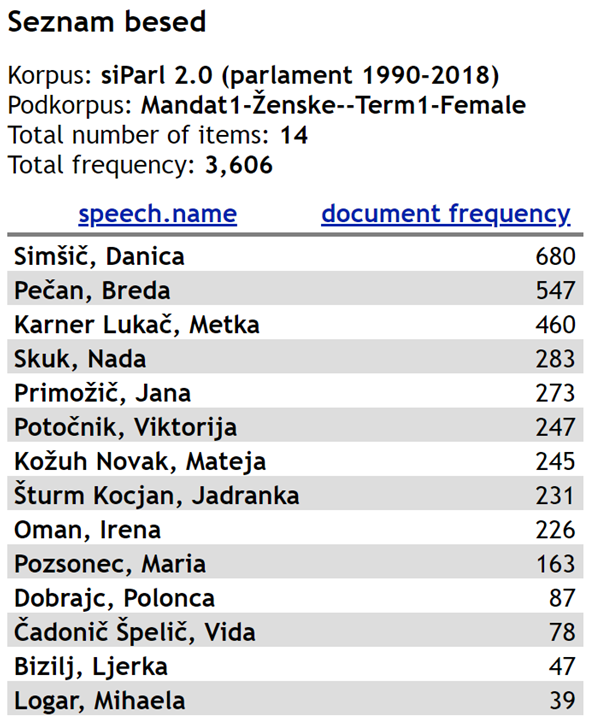
Če želimo ustvariti podkorpus govorov poslancev, moramo spremeniti le spol prek atributa *speech gender/spol* (*speech gender="M"*). Za podkorpus, ki vsebuje govore poslank in poslancev iz vseh mandatov, pa moramo najprej vključiti moški in ženski spol prek atributa *speech gender/spol* (*speech gender="F|M"*),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

### Uporaba frekvenčnih seznamov

V podkorpusih, ustvarjenih v prejšnjem koraku, bomo analizirali govore poslank in poslancev skozi čas, in sicer na podlagi podatkov s [frekvenčnih seznamov](https://www.sketchengine.eu/guide/wordlist-frequency-lists/#toggle-id-2-closed), ki so ena od najosnovnejših korpusnih tehnik. Frekvenčni seznami 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 tej povezavi.

Slika 4 prikazuje primer dveh frekvenčnih seznamov.



**Slika 4. Frekvenčna seznama za poslanke iz prvega mandata.** [**Na levi**](https://www.clarin.si/noske/run.cgi/wordlist?corpname=siparl20&attrs=word&ctxattrs=word&structs=session%2Cspeech%2Cgap%2Cg&refs=%3Dspeech.name&pagesize=50&wlmaxitems=100&wlsort=f&subcnorm=freq&corpname=siparl20&reload=&usesubcorp=Mandat1-%C5%BDenske--Term1-Female&wlattr=speech.name&usengrams=0&ngrams_n=2&ngrams_max_n=2&nest_ngrams=0&wlpat=&wlminfreq=1&wlmaxfreq=0&wlfile=&wlblacklist=&include_nonwords=1&wlnums=docf&wltype=simple) **je prikazano število govorov, ki jih je posamezna poslanka prispevala,** [**na desni**](https://www.clarin.si/noske/run.cgi/wordlist?corpname=siparl20&attrs=word&ctxattrs=word&structs=session%2Cspeech%2Cgap%2Cg&refs=%3Dspeech.name&pagesize=50&wlmaxitems=100&wlsort=f&subcnorm=freq&corpname=siparl20&reload=&usesubcorp=Mandat1-%C5%BDenske--Term1-Female&wlattr=speech.name&usengrams=0&ngrams_n=2&ngrams_max_n=2&nest_ngrams=0&wlpat=&wlminfreq=1&wlmaxfreq=0&wlfile=&wlblacklist=&include_nonwords=1&wlnums=frq&wltype=simple) **pa je število pojavnic v teh govorih.**

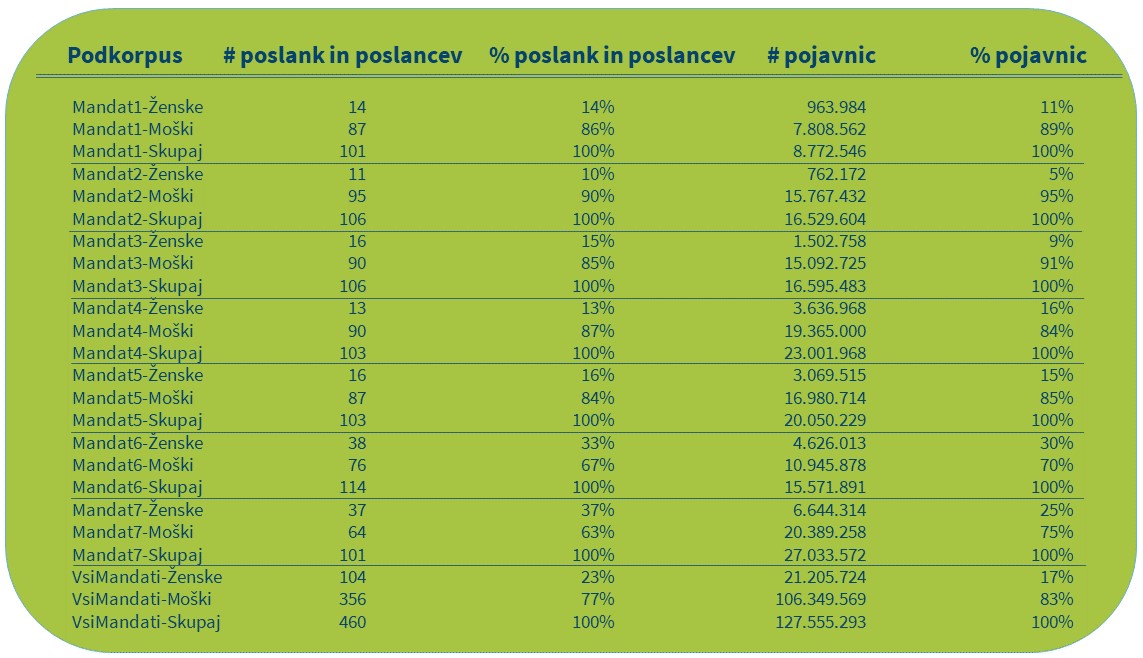
Kot prikazuje Slika 4, je bilo v prvem mandatu prisotnih 14 poslank, ki so skupno izrekle več kot 3.500 govorov, kar je le malo manj kot milijon pojavnic. Povprečno je to skoraj 260 govorov oziroma 69.000 pojavnic na poslanko. Vendar pa je porazdelitev med poslankami zelo neenakomerna, pri čemer sega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

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 (glej 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 (glej 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 prvouvrščene 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 (glej Slika 4, desno) delimo s številom vseh govorov te iste poslanke (glej Slika 4, levo).

### 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 (glej 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.[[7]](#footnote-7)**

**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 pojavnicah 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 35 % kandidatov na volitvah žensk (Selišnik in Antić Gaber, 2017). Pomembna pa je ugotovitev, da kljub temu, da so kvote eden najuspešnejših ukrepov za povečanje števila poslank (Selišnik in Antić Gaber, 2017), 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 pičle 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 (glej Nanni idr. 2018), za korpus siParl 2.0 pa to ne velja. Klasifikacijo tem je mogoče izvesti samodejno, kot v Karan idr. (2016), vendar pa je cilj tega gradiva pokazati učinkovitost raziskav parlamentarnih korpusov zgolj prek uporabe konkordančnikov, 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 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](https://en.wikipedia.org/wiki/Keyword_(linguistics)). 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 kateri koli 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.

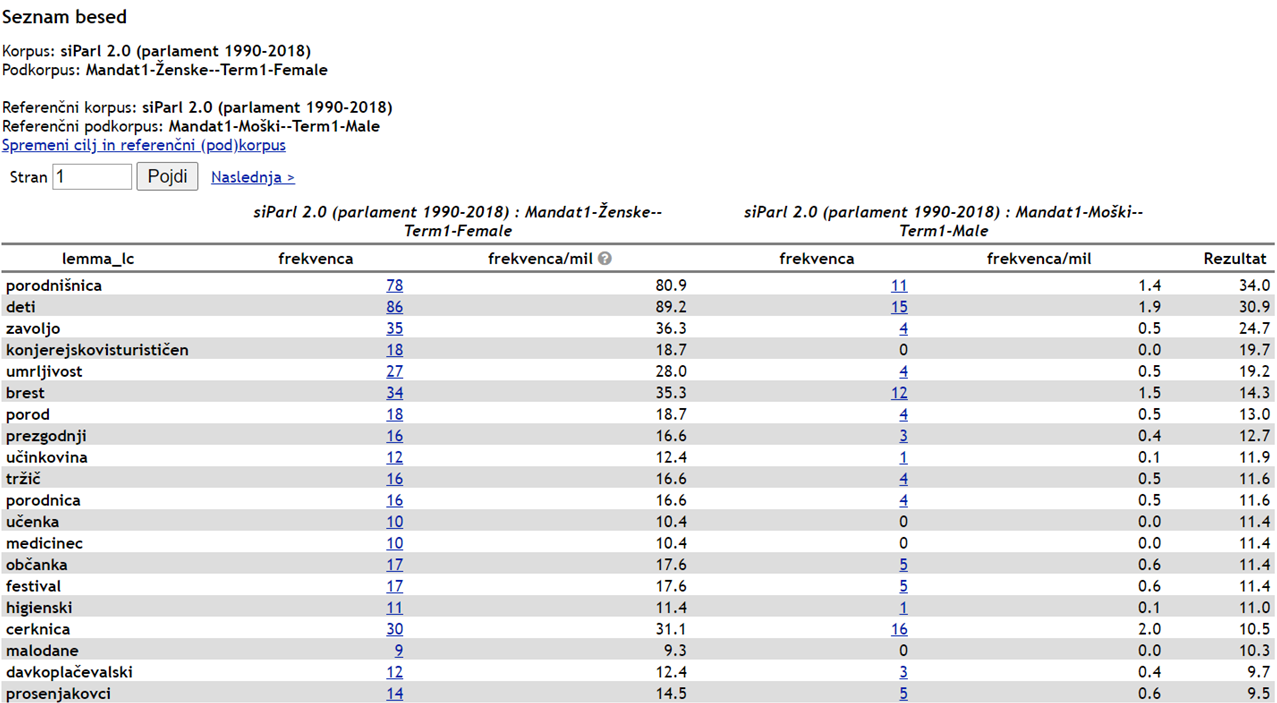
Videoposnetek, ki prikazuje izdelavo seznama ključnih besed v konkordančniku NoSketchEngine, je na voljo na tej povezavi.

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

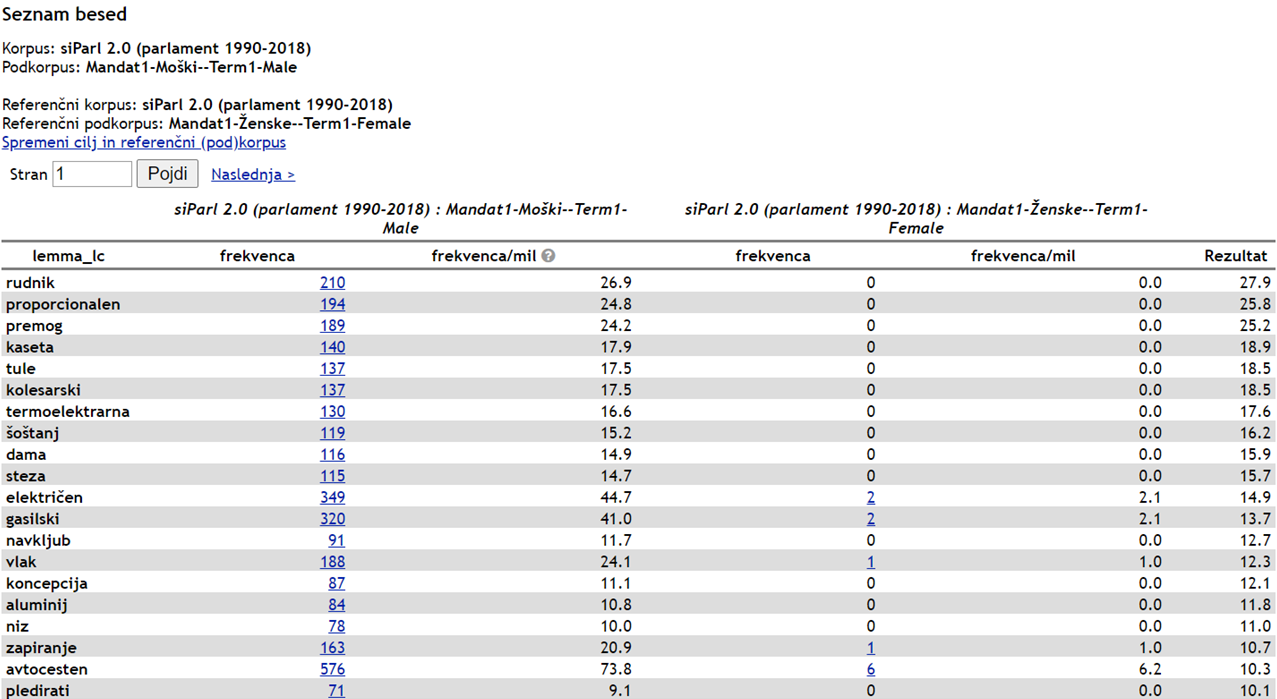
* [Mandat1-Ženske](https://www.clarin.si/noske/run.cgi/wordlist?corpname=siparl20&attrs=word&ctxattrs=word&structs=session%2Cspeech%2Cgap%2Cg&refs=%3Dspeech.name&pagesize=50&wlmaxitems=100&wlsort=f&subcnorm=freq&corpname=siparl20&reload=&usesubcorp=Mandat1-%C5%BDenske--Term1-Female&wlattr=lemma_lc&usengrams=0&ngrams_n=2&ngrams_max_n=2&nest_ngrams=0&wlpat=&wlminfreq=5&wlmaxfreq=0&wlfile=&wlblacklist=&wlnums=frq&wltype=keywords&ref_corpname=siparl20&ref_usesubcorp=Mandat1-Mo%C5%A1ki--Term1-Male&simple_n=1)
* [Mandat1-Moški](https://www.clarin.si/noske/run.cgi/wordlist?;corpname=siparl20;usesubcorp=Mandat1-Mo%C5%A1ki--Term1-Male;wlattr=lemma_lc;wlminfreq=5;wlmaxfreq=0;wlpat=.%2A;wlmaxitems=100;wlsort=f;ref_corpname=siparl20;ref_usesubcorp=Mandat1-%C5%BDenske--Term1-Female;wlcache=;simple_n=1.0;wltype=keywords;wlnums=frq;include_nonwords=0;blcache=;wlpage=1;usengrams=0;ngrams_n=2;ngrams_max_n=2;nest_ngrams=0;complement_subc=0)
* [Mandat7-Ženske](https://www.clarin.si/noske/run.cgi/wordlist?corpname=siparl20&attrs=word&ctxattrs=word&structs=session%2Cspeech%2Cgap%2Cg&refs=%3Dspeech.name&pagesize=50&wlmaxitems=100&wlsort=f&subcnorm=freq&corpname=siparl20&reload=&usesubcorp=Mandat7-%C5%BDenske--Term7-Female&wlattr=lemma_lc&usengrams=0&ngrams_n=2&ngrams_max_n=2&nest_ngrams=0&wlpat=&wlminfreq=5&wlmaxfreq=0&wlfile=&wlblacklist=&wlnums=frq&wltype=keywords&ref_corpname=siparl20&ref_usesubcorp=Mandat7-Mo%C5%A1ki--Term7-Male&simple_n=1)
* [Mandat7-Moški](https://www.clarin.si/noske/run.cgi/wordlist?;corpname=siparl20;usesubcorp=Mandat7-Mo%C5%A1ki--Term7-Male;wlattr=lemma_lc;wlminfreq=5;wlmaxfreq=0;wlpat=.%2A;wlmaxitems=100;wlsort=f;ref_corpname=siparl20;ref_usesubcorp=Mandat7-%C5%BDenske--Term7-Female;wlcache=;simple_n=1.0;wltype=keywords;wlnums=frq;include_nonwords=0;blcache=;wlpage=1;usengrams=0;ngrams_n=2;ngrams_max_n=2;nest_ngrams=0;complement_subc=0)

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

Sliki 7 in 8 prikazujeta prvih 20 ključnih besed poslank in poslancev v sedmem mandatu.[[8]](#footnote-8) 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.



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



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

### 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 [konkordanc](https://www.sketchengine.eu/guide/concordance-a-tool-to-search-a-corpus/) za izbrano lemo. Konkordančni niz vsebuje seznam vseh pojavitev iskane pojavnice v kontekstu, kot prikazuje Slika 9:

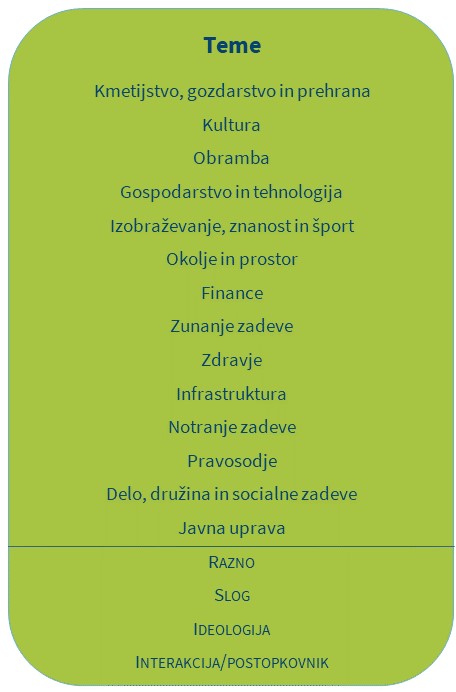
* Konkordance je mogoče neposredno prikazati s klikom leme na seznamu ključnih besed. Prikaže se stran, kjer je na vrhu zaslona prikazana izbrana lema z vsemi zadetki (glej številko 1).
* Besede v rdeči barvi na sredini zaslona (glej številko 2) so zadetki iskane besede v našem podkorpusu, besedilo v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

****

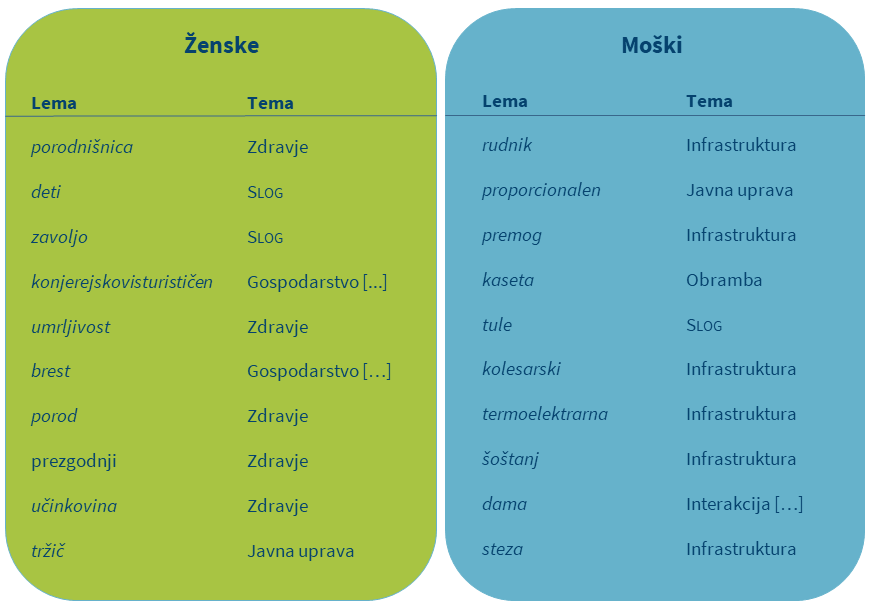
**Slika 9: Razširjeni kontekst za prvi zadetek v** [**konkordančnem nizu**](https://www.clarin.si/noske/run.cgi/view?corpname=siparl20;usesubcorp=Mandat1-Mo%C5%A1ki--Term1-Male;q=q%5Blemma_lc%3D%3D%22proporcionalen%22%5D;complement_subc=0) **za ključno besedo »proporcionalen«.**

### Primerjalna analiza

Parlament je v prvi vrsti zakonodajno telo, v katerem potekajo zakonodajne in proračunske razprave, ki so strukturirane po ministrstvih. Zato nam bo seznam [14 ministrstev trenutne slovenske vlade](https://www.gov.si/drzavni-organi/ministrstva/) služil 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 bomo dodali 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 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.

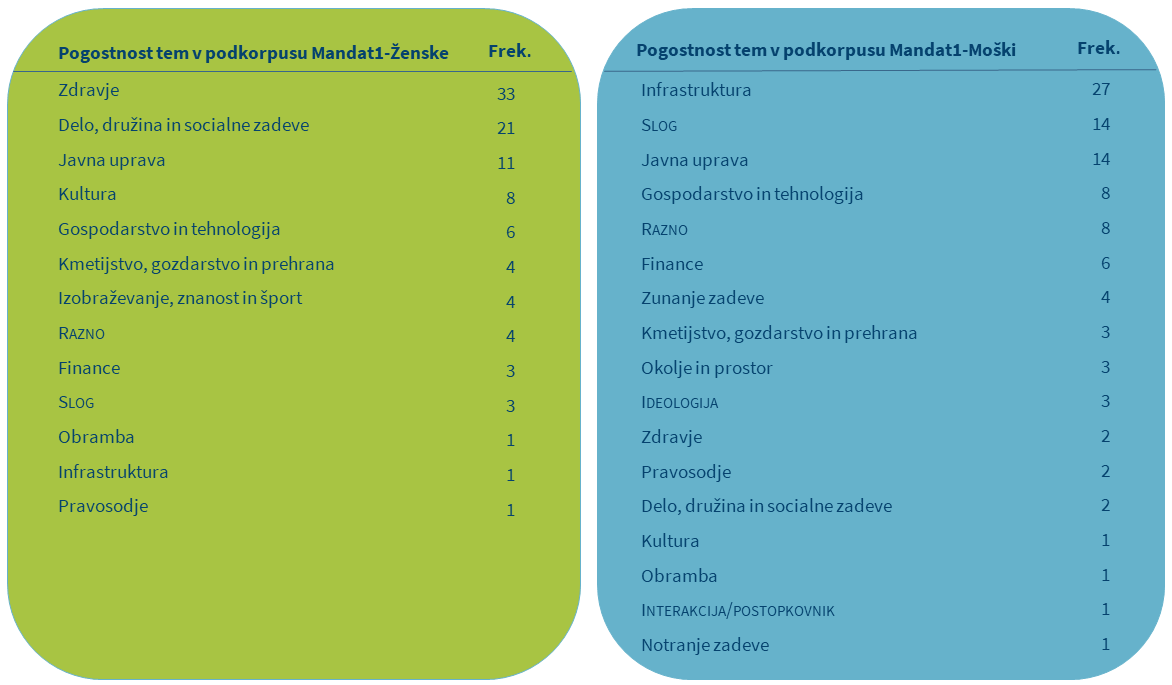
****

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

****

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

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.**



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

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 (5 skupnih tem z enako pomembnostjo v prvem mandatu oz. 3 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 delež teh tem 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 (Coates, 1997). 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 (prim. Ilie, 2013).

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. Diahrone 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 (prim. Blaxill in Beelen, 2016), 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. (2014) ugotovili, da poslanke redkeje razpravljajo o vprašanjih s t. i. trdih zakonodajnih področij. Antić Gaber in Ilonszki (2003) pravita, da družba običajno pričakuje, da bodo poslanke dejavne na drugih področjih politike kot moški. Pri tem gre, kot to ugotavljajo politologi (glej Osborn, 2012), za področja politike, ki so še posebej povezana z ženskami zaradi njihove zgodovinske vloge v družbi ali neposrednega vpliva na življenje žensk.

Ker se pri vprašanjih 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.

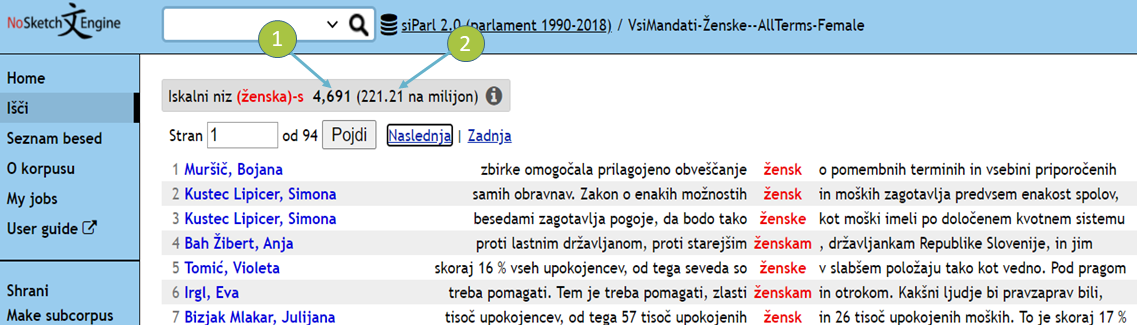
### Delo s frekvencami

Najprej nas zanima, kako pogosto poslanci in poslanke uporabljajo lemo samostalnika »ženska« v različnih časovnih obdobjih (1992–2018). Izvedemo poizvedbe v vseh podkorpusih in v preglednico zabeležimo podatke o frekvenci, kot prikazuje Tabela 6. Ker želimo primerjati podkorpusa različnih velikosti, moramo namesto absolutne frekvence uporabiti [normalizirano frekvenco](http://www.thegrammarlab.com/?p=160), pri kateri je uporabljeno enako skupno število vseh besed v (pod)korpusu. Absolutna frekvenca je v primeru primerjave korpusov 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 podkorpus govorov poslank v tem parlamentarnem mandatu veliko manjši od podkorpusa 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).

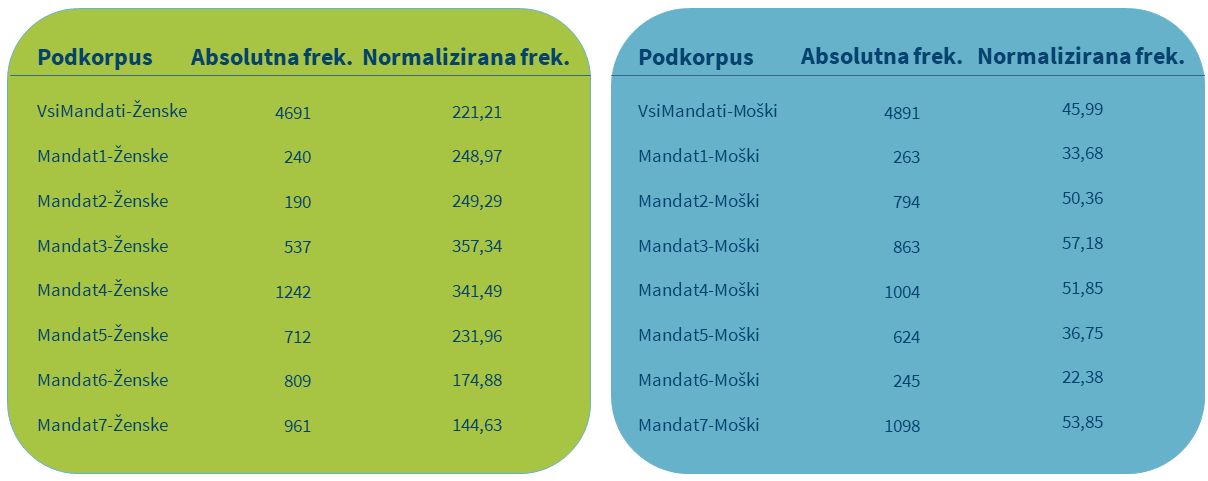
Prikaz poizvedbe v konkordančniku NoSketchEngine za lemo samostalnika »ženska« si lahko ogledate v videoposnetku o luščenju kolokacij (min. 00:10–1:13).

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

* [VsiMandati-Ženske](https://www.clarin.si/noske/run.cgi/first?corpname=siparl20&reload=&iquery=&queryselector=lemmarow&lemma=%C5%BEenska&lpos=-s&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=VsiMandati-%C5%BDenske--AllTerms-Female&fsca_session.organ=&fsca_session.id=&fsca_session.title=&fsca_session.month=&fsca_session.date=&fsca_speech.who=&fsca_speech.name=)



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



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

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 četrtim 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 (glej 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 10: 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](https://en.wikipedia.org/wiki/Collocation). S to tehniko, ki temelji na statističnih testih, pridobimo seznam besed, ki se ob jedrni besedi pojavljajo pogosteje, kot bi bilo to naključno verjetno. 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 tej povezavi.

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.

Za določanje trdnosti kolokacij uporabimo statistično mero [logDice](https://www.sketchengine.eu/my_keywords/logdice/). 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-Ženske](https://www.clarin.si/noske/run.cgi/first?corpname=siparl20&reload=&iquery=&queryselector=lemmarow&lemma=%C5%BEenska&lpos=-s&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=VsiMandati-%C5%BDenske--AllTerms-Female&fsca_session.organ=&fsca_session.id=&fsca_session.title=&fsca_session.month=&fsca_session.date=&fsca_speech.who=&fsca_speech.name=)
* [VsiMandati-Moški](https://www.clarin.si/noske/run.cgi/first?corpname=siparl20&reload=&iquery=&queryselector=lemmarow&lemma=%C5%BEenska&lpos=-s&phrase=&word=&wpos=&char=&cql=&default_attr=word&fc_lemword_window_type=both&fc_lemword_wsize=5&fc_lemword=&fc_lemword_type=all&fc_pos_window_type=both&fc_pos_wsize=5&fc_pos_type=all&usesubcorp=VsiMandati-Mo%C5%A1ki--AllTerms-Male&fsca_session.organ=&fsca_session.id=&fsca_session.title=&fsca_session.month=&fsca_session.date=&fsca_speech.who=&fsca_speech.name=)

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

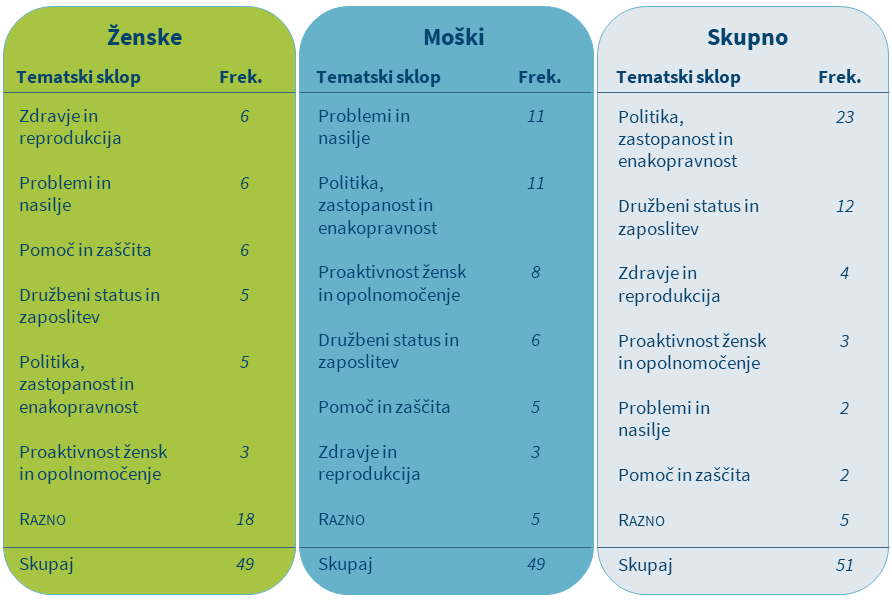
### Primerjalna analiza

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[[9]](#footnote-9). 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.**

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.**

Opazimo lahko tudi, da so v podkorpusu govorov poslank primerljivo zastopani vsi tematski sklopi razen enega, skupaj pa pokrivajo skoraj 60 % kolokacij. Tako je edina izjema tematski sklop *Proaktivnost žensk in opolnomočenje*. Kolokacije, ki spadajo v ta tematski sklop, so približno dvakrat redkejše kot kolokacije iz katerega koli 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. »omogočati«). 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 (2003), 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 prioritetami, 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 (glej 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 (Coates, 1997), več intenzifikatorjev in (predvsem ocenjevalnih) pridevnikov (Poynton, 1989) ter več osebnih zaimkov kot moški, ki pri sporazumevanju uporabljajo več števnikov, členov in predlogov (Newman idr., 2008).

# 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 bolje 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) [korpusih v drugih jezikih](https://www.clarin.eu/resource-families). 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. Cornelii Ilie, Tjaši Cankar, Mladenu Zobcu in Rubenu Rosu za strokovno recenzijo in testiranje učnega gradiva ter vse njihove dragocene komentarje.

# Reference

Alasuutari, P., Rautalin, M., in Tyrkkö, J. (2018). 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*(3), 341–363.

Antić Gaber, M., in Ilonszki, G. (2003). *Women in parliamentary politics: Hungarian and Slovene cases comparur.* Peace Institute, Institute for Contemporary Social and Political Studies.

Antić Gaber, M. in Selišnik, I. (2017). *The Slovene version of a “fast track” to political equality*. Teorija in Praksa, 54(2), 337.

Bäck, Hanna., Debus, Marc., in Müller, Jochen. (2014). *Who Takes the Parliamentary Floor? The Role of Gender in Speech-making in the Swedish Riksdag.*

Baker, P. (2014). *Using corpora to analyze gender*. London: Bloomsbury.

Biber, D., in Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press.

Biel, Ł., Biernacka, A., in Jopek-Bosiacka, A. (2018). Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations. V: *Language and Law* (str. 249–274). Cham: Springer.

Bing, J. M., in Bergvall, V. L. (1998). The question of questions: Beyond binary thinking. In J. Coates (ur.), *Language and Gender: A Reader* (Vol. 1, str. 496–510). Oxford: Blackwell.

Blaxill, L., in Beelen, K. (2016). *Women in Parliament Since 1945: Have they changed the debate?* History in Policy.

Coates, Jennifer. (1997). *Women, men and language: A sociolinguistic account of gender differences in language*. London: Longman.

Demmen, J., Jeffries, L., in Walker, B. (2018). 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*, 81.

Eckert, P., in McConnell-Ginet, S. (2013). *Language and gender*. Cambridge: Cambridge University Press.

Goddard, A., in Patterson, L. M. (2015). *Language and gender*. London: Routledge.

Hansen, D. H., Navarretta, C., in Offersgaard, L. (2018). A Pilot Gender Study of the Danish Parliament Corpus. *Proceedings of the ParlaClarin Workshop at the Eleventh International Conference on Language Resources and Evaluation: (LREC 2018)*.

Ilie, C. (2013). Gendering confrontational rhetoric: Discursive disorder in the British and Swedish parliaments. Democratization, 20(3), 501–521.

Ilie, C. (2017). Parliamentary debates. In R. Wodak & B. Forchtner (Eds.), The Routledge handbook of language and politics (pp. 309–325). Routledge.

Jaworska, S., in Ryan, K. (2018). Gender and the language of pain in chronic and terminal illness: A corpus-based discourse analysis of patients’ narratives. *Social Science & Medicine*, *215*, 107–114.

Karan, M., Šnajder, J., Sirinic, D., in Glavaš, G. (2016). 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*, 12–21.

Kustec, L. S. (ur.) (2017). *Poslanke Državnega zbora: Uradno in osebno od prvega do sedmega mandata (1992-2017)*. Ljubljana: Državni zbor.

Leijenaar, M. (1997). *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*. Office for Official Publications of the European Communities.

Litosseliti, L. (2006). *Gender and Language: Theory and Practice*. Hodder Arnold.

Marshall, C. (2000). Policy discourse analysis: Negotiating gender equity. *Journal of Education Policy*, *15*(2), 125–156.

McEnery, T., in Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

Mensah, E. O., in Wood, S. F. (2018). 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*(2), 61–77.

Mollin, S. (2007). The Hansard hazard: Gauging the accuracy of British parliamentary transcripts. *Corpora*, *2*(2), 187–210.

Nanni, F., Osman, M., Cheng, Y.-R., Ponzetto, S. P., in Dietz, L. (2018). UKParl: A Semantified and Topically Organized Corpus of Political Speeches. V: D. Fišer in F. de Jong (ur.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)* (str. 29–32). European Language Resources Association (ELRA).

Newman, M. L., Groom, C. J., Handelman, L. D., in Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236.

Osborn, T. L. (2012). *How women represent women: Political parties, gender and representation in the state legislatures*. Oxford: Oxford University Press.

Pančur, A., Erjavec, T., Ojsteršek, M., Šorn, M., in Blaj Hribar, N. (2020). *Slovenian parliamentary corpus (1990-2018) siParl 2.0*. Slovenian language resource repository CLARIN.SI.

Poynton, C. (1989). *Language and gender: Making the difference*. Oxford University Press.

Rix, K. (2014). ‘Whatever Passed in Parliament Ought to be Communicated to the Public’: Reporting the Proceedings of the Reformed C ommons, 1833–50. *Parliamentary History*, *33*(3), 453–474.

Wängnerud, L. (1996). Intressen kontra stereotyper. Varför finns det kvinnliga och manliga politikområden i riksdagen? *Statsvetenskaplig Tidskrift*, *99*(2).

Wodak, R. (2007). Pragmatics and Critical Discourse Analysis: A cross-disciplinary inquiry. *Pragmatics & Cognition*, *15*, 203–225.

Wolbrecht, C. (2002). *Female legislators and the women’s rights agenda: From feminine mystique to feminist era*. 170–197.

1. Avtorica [naslovne izjave](https://www.clarin.si/noske/run.cgi/first?iquery=Prvi%C4%8D+%2C%09sem+politi%C4%8Darka%09in+ne+politik&corpname=siparl20&corpus-search-form=true) z redne seje 4. 6. 2003 je Cvetka Zalokar Oražem (LDS). [↑](#footnote-ref-1)
2. Slovensko različico konkordančnika izberete v zgornjem desnem kotu s klikom simbola zobnika in izbiro ustreznega jezika s spustnega seznama. [↑](#footnote-ref-2)
3. Zaradi izida nove različice konkordančnika je ta trenutno zgolj delno preveden v slovenščino. [↑](#footnote-ref-3)
4. Delež temelji na uradnem številu vseh poslank, kot ga navaja Kustec (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. [↑](#footnote-ref-4)
5. Pri oznakah v korpusu je uporabljen generični moški spol. [↑](#footnote-ref-5)
6. 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. [↑](#footnote-ref-6)
7. 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 Kustec (2017) 39, in ne 38, kot prikazuje korpus. [↑](#footnote-ref-7)
8. Konkordančnik NoSketchEngine za računanje ključnih besed uporablja formulo [Simple Maths](https://www.sketchengine.eu/documentation/simple-maths/). [↑](#footnote-ref-8)
9. Kategorija skupno, kot nakazuje njeno ime, združuje kolokacijske kandidate, ki se pojavijo tako na seznamu za podkorpus govorov poslank kot za podkorpus govorov poslancev. [↑](#footnote-ref-9)