Use Case 1
Gender in Parliamentary Discourse

Background

While more women than ever are being elected to parliaments around the world, equality is still a long way off, and current progress is far too slow. Most parliaments are still heavily male-dominated, and some have no women members of parliament at all. Even where women are present in greater numbers, glass ceilings often remain firmly in place. (Source: Women in Parliament by the Inter-Parliamentary Union)

CLARIN Resources and services used

- CLARIN Resource Families – parliamentary corpora
- the CLARIN.SI repository
- the noSketch Engine concordancer

Citizen scientist

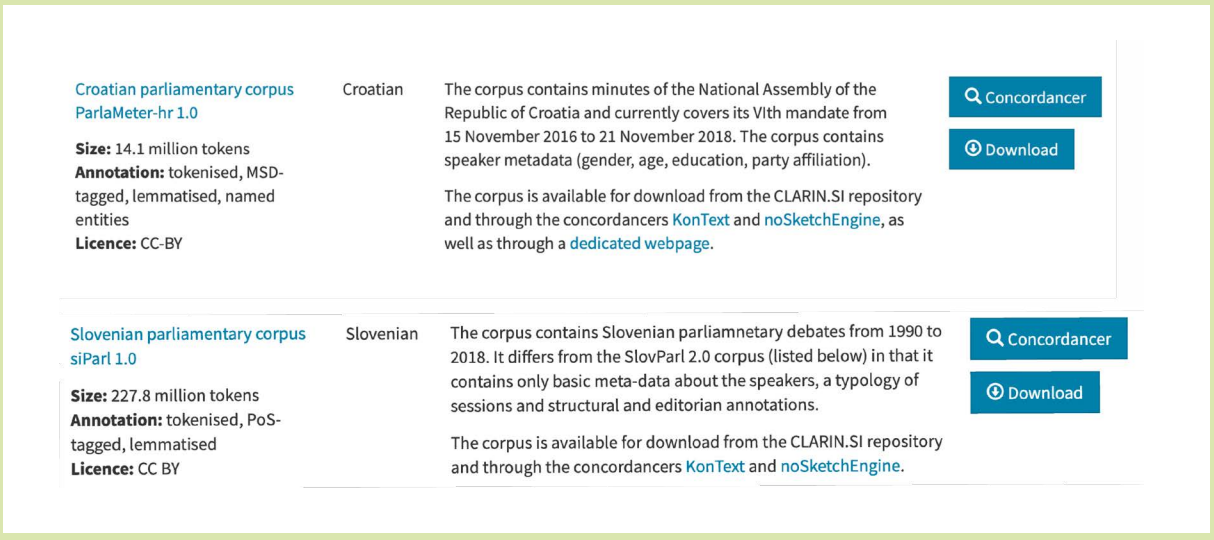
Do female speakers in the Slovenian and Croatian parliaments speak more or less than their male counterparts?

Student

Is the language of female parliamentary speakers similar in Slovenia and Croatia?

Step-by-step guide

1 Search the Parliamentary CLARIN Resource Family for relevant Slovenian and Croatian corpora. In this walkthrough, we'll use the Croatian and Slovenian ParlaMETER corpora, since they are roughly comparable in terms of time span, linguistic annotation and speaker metadata, but you can also use any of the other parliamentary corpora.



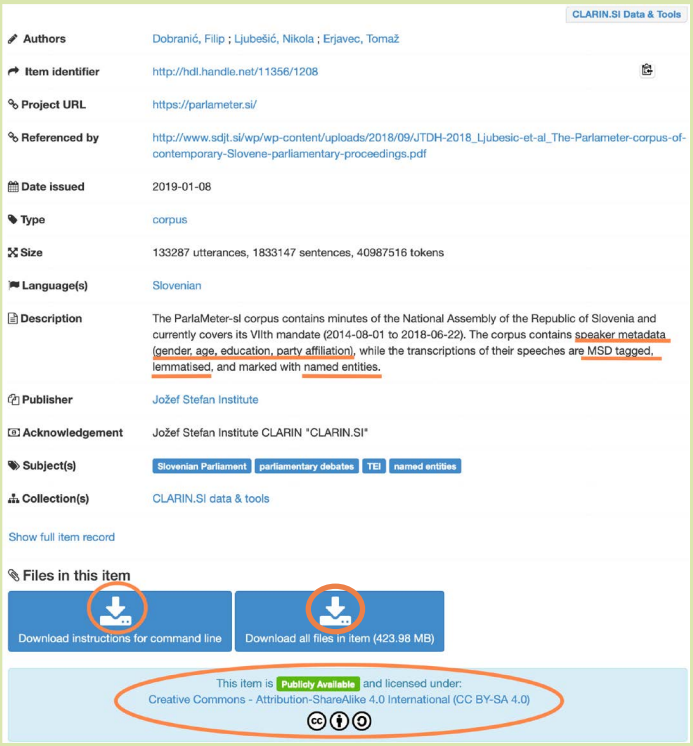
Croatian parliamentary corpus ParlaMeter-hr 1.0 **Size:** 14.1 million tokens **Annotation:** tokenised, MSD-tagged, lemmatised, named entities **Licence:** CC-BY	Croatian	The corpus contains minutes of the National Assembly of the Republic of Croatia and currently covers its VIth mandate from 15 November 2016 to 21 November 2018. The corpus contains speaker metadata (gender, age, education, party affiliation). The corpus is available for download from the CLARIN.SI repository and through the concordancers KonText and noSketchEngine, as well as through a dedicated webpage.	Concordancer Download
Slovenian parliamentary corpus siParl 1.0 **Size:** 227.8 million tokens **Annotation:** tokenised, PoS-tagged, lemmatised **Licence:** CC BY	Slovenian	The corpus contains Slovenian parliamnetary debates from 1990 to 2018. It differs from the SlovParl 2.0 corpus (listed below) in that it contains only basic meta-data about the speakers, a typology of sessions and structural and editorian annotations. The corpus is available for download from the CLARIN.SI repository and through the concordancers KonText and noSketchEngine.	Concordancer Download

CLARIN
Common Language Resources and Technology Infrastructure

SSHOC
social sciences & humanities open cloud

2 For both corpora, check their descriptions to see that:

- In terms of linguistic annotation, both corpora are **annotated for syntactic and morphological features** ("MSD-tagged"), **lemmatized** and **marked for named entities**.
- In terms of extra-linguistic annotation, both corpora are marked for **speaker metadata (gender, age, education, party affiliation).**
- The CC-BY licence shows that the corpus is publicly available, either for **download** or **on-line querying**.



3 Let's start by analysing the Slovenian corpus. First click on Slovenian parliamentary corpus ParlaMeter-sl 1.0 in the CLARIN Resource Families. This takes you to the record for this corpus in the CLARIN.SI repository:

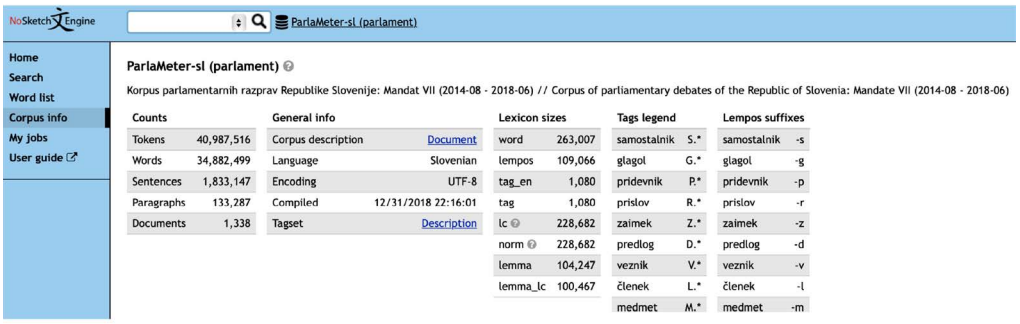
4 The CLARIN.SI repository shows how the corpus has to be cited to ensure proper author-ship attribution, and offers a **persistent identifier** for the resource – http://hdl.handle.net/11356/1208.

CLARIN
Common Language Resources and Technology Infrastructure

SSHOC
social sciences & humanities open cloud

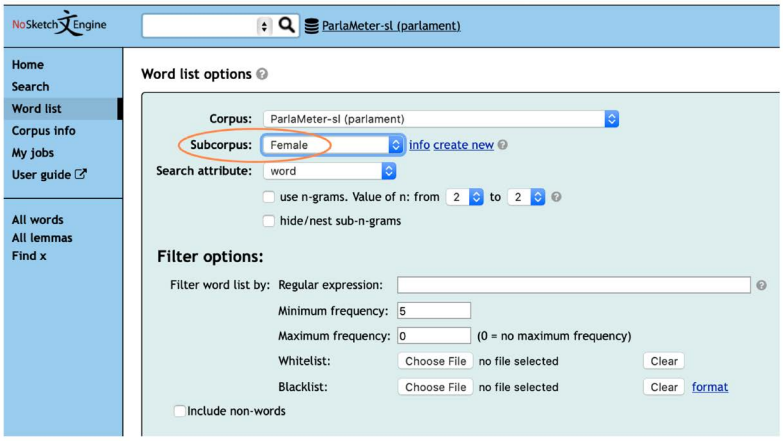
5 The corpus can be queried via two concordancers – **KonText** and **noSketch Engine**. Both offer very versatile search environments in which complex queries can be narrowed down on the basis of the **speaker metadata (age, party affiliation, etc).**

6 Let's query the corpus by using the **noSketch Engine**. In the repository, click on the downward arrow next to "noSketch" and then select search.



7 Let's recall our task: we're interested how parliamentary speakers are represented in the corpus in terms of gender. In other words, how many words of the total 34,882,499 are spoken by female parliament speakers and how many by male speakers?

8 We can figure this out by creating a **word list** and narrowing it down to the "Female" subcorpus.



9 After clicking on **Make word list**, we get the result for female speakers. We repeat the procedure for the "Male" subcorpus and see that the male speakers say 2.5 times more words than their female counterparts.

Repeat the procedure for the Croatian ParlaMeter corpus.

- What is the gender division in terms of words between male and female speakers in this corpus?
- Is the difference greater or smaller than that in the Slovenian corpus?

Additional Task

We can also construct word lists for individual word classes, such as nouns, verbs, adjectives, etc.

Which are the most frequent nouns used by the speakers in the Slovenian corpus?

a. Under search attribute, change from "word" to "tag_en". This specifies that you're searching for parts of speech rather than individual words.

b. In the filter option "Regular expression", write N.*. This specifies that you're searching for all nouns.

c. Under output options, select "lemma" under "Change output attributes". This ensures that all inflectional variants are all accounted for under a single base form of the word.

d. Click here to see the result for such a query.

Repeat the procedure for the Croatian ParlaMeter corpus. Are the results similar to the Slovenian corpus?

Research bite

In the Slovenian ParlaMeter corpus, the most frequent topics among the female speakers are ***health*** and ***labour, family*** and ***social affairs,*** which are followed by ***public administration and education, science and sport.***
Most of the 100 top-ranking keywords uttered by female speakers, on the other hand, could not be classified into a single topic because they were used either to achieve a ***stylistic effect***, were general words that were used in *multiple topics*, such as descriptive adjectives or legal terms, or ***ideological expressions,*** all of which indicate a more discursive, debating ***style*** of the male speakers, but could also stem from the fact that the leading roles in that term were predominantly held by male members of parliament (Source: Parlameter – a Corpus of Contemporary Slovene Parliamentary Proceedings by Darja Fišer, Nikola Ljubešić and Tomaž Erjavec).