

# Showcasing the Socio-economic Impact of ERICs

The ParlaMint logo is a light blue circle containing a stylized map of Europe. The map is composed of a grid of dots, with the left side in dark blue and the right side in yellow. The word "ParlaMint" is written in dark blue text above the map.

ParlaMint

Darja Fišer  
Former Vice-executive Director  
CLARIN ERIC  
26 January 2020

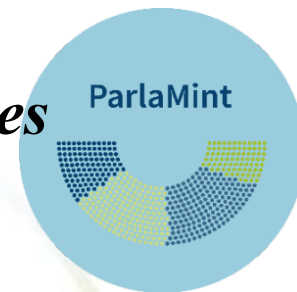
# CLARIN ERIC in 3 bullets



- Infrastructure ... that provides easy and sustainable access for scholars in the **humanities and social sciences** and beyond
- to **digital language data** (in written, spoken, video or multimodal form)
- and **advanced tools** to discover, explore, exploit, annotate, analyse or combine them, wherever they are located

# Some roles of language ...

## ... and the *corresponding audiences*

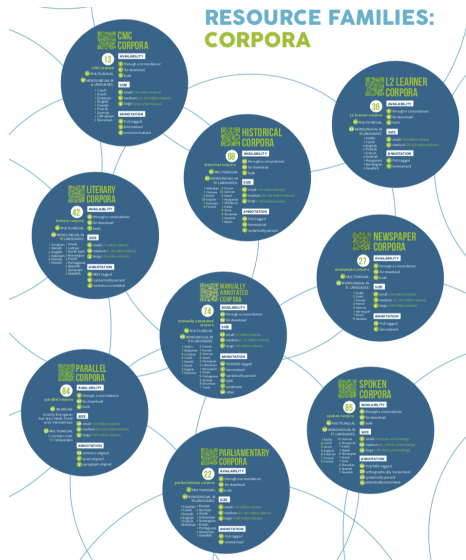


- Carrier of cultural content: *cultural heritage*
- Record of the past: *archaeology, history*
- Main communication instrument within and across societies: *sociology, anthropology*
- Preserving and disseminating our knowledge: *all disciplines*
- Instrument to formulate rules for society: *law, theology*
- Carrier of information: *media studies, journalism*
- Means of human expression: *literary studies, art, psychology*
- Focus of cognitive processes: *brain studies, psychology, neurology*
- Component of national or cultural identity: *political sciences, social sciences*
- Object of study: *linguistics, language learning*
- Object of computer processing: *language and speech technology*

# CLARIN

## RESOURCE FAMILIES

### RESOURCE FAMILIES: CORPORA



### RESOURCE FAMILIES: LEXICAL RESOURCES



- documentation
  - discover
  - reuse
  - replicate
- interoperability
  - combine
  - compare
  - innovate

# Parliamentary records



- great potential for reuse and re-purposing within many fields of study SSH and beyond:
  - is suited for both *close reading* and *distance reading*
  - *Humanities*: history, language change, discourse analysis ...
  - *Social sciences*: social and cultural dynamics, political sciences, economics ...
- considered a rich data type
  - apart from linguistic content, rich in metadata (speaker, party affiliation, age, sex, education, origin, duration of speech)
  - apart from linguistic content, rich in extralinguistic clues (interruptions, voting results)
- but also
  - often presenting itself as messy or noisy data
  - calling for links with data in other modalities than text and speech
  - created under specific circumstances that need to be well understood before strong conclusions can be drawn

# ParlaMint

<https://www.clarin.eu/content/parlamint>



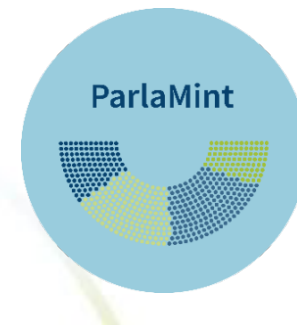
- **Motivation:** Parliamentary data directly reflects the most recent events with a global impact on *human health, social life* and *economics* such as the current **COVID-19 pandemic**.
- **Goal:** Provide policy making resources and tools for focused observations on trends, opinions, decisions on *lockdowns* and *restrictive measures* as well as on *the consequences* with respect to: **health, medical care systems, employment**, etc. during pandemic times.
- **Scope:** Coverage: 16 languages, 20+ million words each



## ParlaMint workflow

1. Getting the parliamentary data and metadata
2. Converting them into the Parla-CLARIN schema
3. Validation (formal and qualitative)
4. Linguistic annotation with syntactic structures of Universal Dependencies and Named Entities
5. Making the corpora available through concordancers and ParlaMeter
6. **On-going:** Adding more parliaments
7. **On-going:** Building use cases in Political Science and Digital Humanities based on the corpus data

# Male vs. Female Key Lemmas



*ParlaMint-SI 1.0 (parliament: COVID) : Male*

lemma	frequency	frequency/mill ?
levičar	<a href="#">364</a>	21.3
navzoč	<a href="#">4,326</a>	252.9
penez	<a href="#">167</a>	9.8
esklopa	<a href="#">212</a>	12.4
Iran	<a href="#">260</a>	15.2
gorica	<a href="#">168</a>	9.8
počenjati	<a href="#">182</a>	10.6
islam	<a href="#">180</a>	10.5
levičarski	<a href="#">163</a>	9.5
Palestina	<a href="#">111</a>	6.5

*ParlaMint-SI 1.0 (parliament: COVID) : Female*

lemma	frequency	frequency/mill ?
Medical	<a href="#">107</a>	16.7
KB	<a href="#">56</a>	8.7
pacientov	<a href="#">210</a>	32.7
enaročanje	<a href="#">75</a>	11.7
ennapotnica	<a href="#">42</a>	6.5
žilen	<a href="#">417</a>	64.9
psihiatričen	<a href="#">69</a>	10.7
roditeljski	<a href="#">80</a>	12.5
kardiolog	<a href="#">80</a>	12.5
zaznamba	<a href="#">100</a>	15.6
rejništvo	<a href="#">206</a>	32.1
opornica	<a href="#">422</a>	65.7
posvojitev	<a href="#">328</a>	51.1



# COVID Period Key Lemmas in ParlaMint-SI 1.0

NoSketch Engine

Q

ParlaMint-SI 1.0 (parliament: COVID) / COVID

Home

Search

Word list

Corpus info

My jobs

User guide

Save

Change options

Word list

Corpus: ParlaMint-SI 1.0 (parliament: COVID)

Subcorpus: COVID

Reference corpus: ParlaMint-SI 1.0 (parliament: COVID)

Reference subcorpus: == the rest of the corpus ==

Switch focus and reference (sub)corpus

Page 1

Go

Next >

ParlaMint-SI 1.0 (parliament: COVID) : COVID			ParlaMint-SI 1.0 (parliament: COVID) : == the rest of the corpus ==		
lemma	frequency	frequency/mill	frequency	frequency/mill	Score
epidemija	1,772	639.0	15	0.7	371.5
covid-19	411	148.2	0	0.0	149.2
koronavirus	348	125.5	0	0.0	126.5
ventilator	408	147.1	4	0.2	124.2
karantena	240	86.5	1	0.0	83.5
korona	195	70.3	0	0.0	71.3
virus	423	152.5	24	1.2	71.2
maska	687	247.7	58	2.8	65.5
pandemija	142	51.2	0	0.0	52.2
protikoronski	128	46.2	0	0.0	47.2
varda	146	52.6	3	0.1	46.9
protikoron	108	38.9	0	0.0	39.9
Gale	103	37.1	4	0.2	32.0
okužba	296	106.7	50	2.4	31.6
respirator	81	29.2	0	0.0	30.2
covid	76	27.4	0	0.0	28.4
Trata	73	26.3	1	0.0	26.1

# Beyond language

<b>TEXT.SUBCORPUS</b> <input type="checkbox"/> COVID <input type="checkbox"/> Reference <input type="button" value="Select All"/>	<b>TEXT.FROM</b> <input type="text"/>	<b>TEXT.TO</b> <input type="text"/>
<b>SPEECH.SPEAKER_TYPE</b> <input type="checkbox"/> MP <input type="checkbox"/> notMP <input type="button" value="Select All"/>	<b>SPEECH.SPEAKER_ROLE</b> <input type="checkbox"/> chair <input type="checkbox"/> regular <input type="button" value="Select All"/>	<b>SPEECH.SPEAKER_PARTY</b> <input type="checkbox"/> ATAKA <input type="checkbox"/> BGUn <input type="checkbox"/> VOLYA <input type="checkbox"/> _ <input type="checkbox"/> ABB <input type="checkbox"/> БСП <input type="checkbox"/> ГЕРБ <input type="checkbox"/> ДПС <input type="checkbox"/> ОП <input type="checkbox"/> ПФ <input type="checkbox"/> РБ <input type="button" value="Select All"/>
<b>SPEECH.SPEAKER_ID</b> <input type="text"/>	<b>SPEECH.SPEAKER_GENDER</b> <input type="checkbox"/> ? <input type="checkbox"/> F <input type="checkbox"/> M <input type="button" value="Select All"/>	<b>SPEECH.SPEAKER_BIRTH</b> <input type="checkbox"/> 1942 <input type="checkbox"/> 1943 <input type="checkbox"/> 1944 <input type="checkbox"/> 1945

## Word list

Corpus: ParlaMint-BG 1.0 (parliament: COVID)

Total number of items: 17






Total frequency: 34,310

<u>note.type</u>	<u>frequency</u>
vocal:interruption	16,257
kinesic:applause	7,642
kinesic:signal	2,225
vocal:noise	1,718
vocal:exclamat	1,604
vocal:laughter	1,295
kinesic:ringing	1,187
vocal:speaking	738
incident:action	647
kinesic:kinesic	364
incident:incident	220
vocal:shouting	140
incident:leaving	114
incident:entering	62
vocal:greeting	34
vocal:question	32
incident:break	31

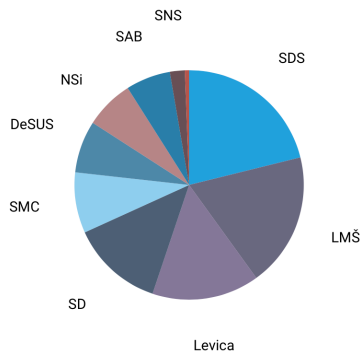
# Beyond research

<https://parlamente.si/>

## Največkrat so pojem uporabili

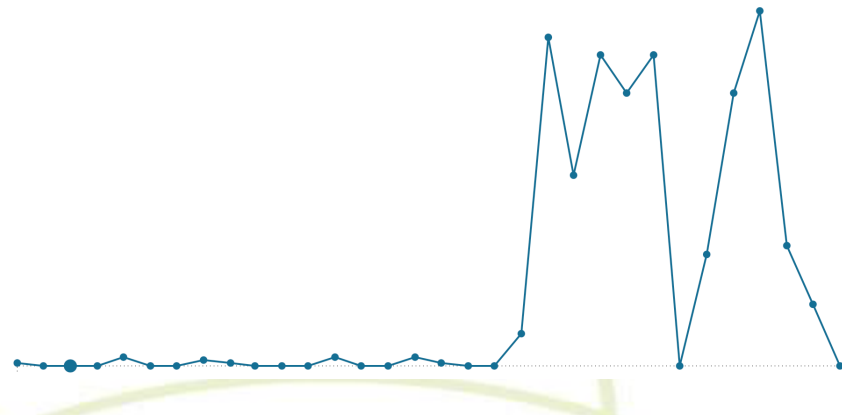
	Robert Polnar DeSUS	35
	Bojana Muršič SD	27
	Andrej Šircelj SDS	27
	Nataša Sukič Levica	25
	Jožef Horvat NSi	24

## Raba po poslanskih skupinah



CEL MANDAT

ZADNJE LETO



## Nastopi, v katerih je bil iskalni niz izrečen

	Rudi Medved LMŠ	30. 12. 2020	...vsem nam na oči v vseh teh protikoronskih zakonih in tudi ta PKP-7 ni nikakršna izjema. Zdaj, samo primer, prisilno upokojevanje je vgrajeno vanj, ne, in kaj bo to pomagalo v boji proti <b>virusu</b> , no? Kaj bo to pomagalo v boji proti <b>virusu</b>
	Jerca Korčec LMŠ	29. 12. 2020	...samo formo in »obličnost«, kako te zakone vi pripravljate. Danes smo slišali s strani kolegov iz SDS, da je korona <b>virus</b> problem svetovnih razsežnosti, s čimer se absolutno vsi strinjamo, ampak neumni ukrepi, pa neumnost te Vlade, to je pa...
	Boštjan Korajžija Levica	29. 12. 2020	...podobno delo, sicer ne na taki ravni kot je recimo medicinske sestre oziroma zdravniki, ampak po izpostavljenosti <b>virusu</b> pa mislim, da zelo podobno. In mislim, da lahko tudi gospa Magajne potrdi in tudi pove kaj na to temo. Zdaj, tudi že...
	Bojana Muršič SD	29. 12. 2020	...mislim pri prvem amandmaju k 1. členu. Razpravljala bom malce širše. Začela bom pa tako, drugi val korona <b>virusa</b> je tej Vladi ušel izpod nadzora. O temu sploh ne rabimo niti ne več govoriti. Okužb je bilo kljub temu, oziroma jih še...
	Robert Polnar DeSUS	29. 12. 2020	...njihovih bolečin, tegob in strahov, soočanje z umrlimi v katerih se prepoznava lastna minljivost. Nikakršen <b>virus</b> pa ne more odstraniti človeške potrebe, da v naravnih in družbenih dejanjih vidi nekakšen smisel. Za vsemi dvomi in negotovostmi...

## Glasovanja, ki vsebujejo iskalni niz



Predlog sklepa (razprava o odgovoru ministra za zdravje Tomaža Gantarja na poslansko vprašanje Marka Bandellija v zvezi s cepivom ...

35

ZA

48

PROTI

0

VZDRŽAN

7

NI

The ParlaMint logo is a light blue circle containing a stylized 'M' shape. The 'M' is composed of a grid of dots, with the left side in dark blue and the right side in yellow. The word 'ParlaMint' is written in a dark blue, sans-serif font across the middle of the circle.

**ParlaMint**

# Thank you!

Questions & comments:

<https://www.clarin.eu>

[darja.fiser@ff.uni-lj.si](mailto:darja.fiser@ff.uni-lj.si)