CLARIN Café

Join Our Parliamentary-flavoured Coffee: ParlaMint

Warsaw/Sofia/Ljubljana, 2020-11-03
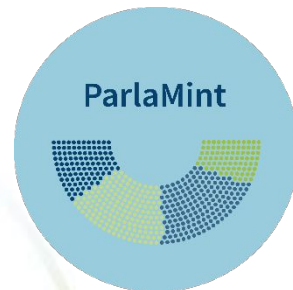
ParlaMint in a Nutshell:

New Comparable Parliamentary Corpora in the Pandemic Era

Maciej Ogrodniczuk and Petya Osenova

What is ParlaMint?

ParlaMint

MiniProject supported by CLARIN-ERIC

Budget: 98,000 €
Duration: Jul 1 2020 – May 30 2021

Motivation: Parliamentary data directly corresponds to the most recent events with global impact on *human health, social life* and *economics* such as the current COVID-19 pandemic.

Goal: Provide 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.

How is ParlaMint implemented?

Phase 1 (*July 2020 – Sep 2020*):

Corpus of *Bulgarian, Croatian, Polish* and *Slovene* parliamentary speeches was created with two parts marked:

- COVID-19 subcorpus (Nov 2019 – 2020)
- reference subcorpus (2015 – Nov 2019).

Phase 2 (*Dec 2020 – May 2020*):

Parliaments for 6 more countries will be created according to the methodology established in Phase 1.
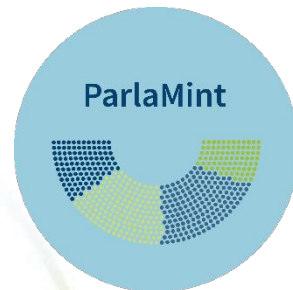(Stay for the last part for more information how to participate)

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. Next: Adding more parliaments
7. Next: Building use cases in Political Sciences and Digital Humanities based on the corpus data.

Meet the participants :)

ParlaMint

- Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences (also: Michał Rudolf)

- Petya Osenova, Institute for Information and Communication Technologies, Bulgarian Academy of Sciences (also: Kiril Simov)

- Tomaž Erjavec, Jožef Stefan Institute, Ljubljana (also: Darja Fišer and Kristina Pahor de Maiti)

- Andrej Pančur, Institute of Contemporary History, Ljubljana

- Nikola Ljubešić, Jožef Stefan Institute, Ljubljana

- Filip Muki Dobranić, Parlameter.org (also: Tomaž Kunst)

- Ruben Ros, Luxembourg University

What is available: first phase results

ParlaMint
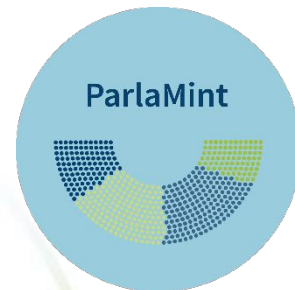
- Parla-CLARIN schema
  (https://github.com/clarin-eric/parla-clarin)

- Corpora to download from CLARIN.SI repository
  (http://hdl.handle.net/11356/1345)

- Corpora in NoSketch Engine
  (https://www.clarin.si/noske/parlamint.cgi)

- Corpora in Kontext
  (https://www.clarin.si/kontext/)

- *Ongoing Now:* Corpora in Parlameter
  (https://parlameter.org/)

- Know-how on dealing with parliamentary data!

Join Us as a Contributor and/or as a User:

Possibilities, Potential and Perspectives

Maciej Ogrodniczuk and Petya Osenova

If you want to join us as a contributor

- We invite proposals to add parliamentary corpora for additional countries to the ParlaMint collection

- Call schedule:

  26 Oct 2020:   Call issued
  16 Nov 2020:   Submission deadline
  20 Nov 2020:   Results announced
  1 Dec 2020:   Projects start
  30 Mar 2021:   Delivery of the corpora

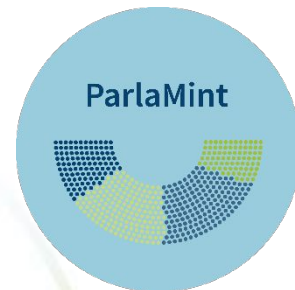**ParlaMint**

If you want to join us as a contributor

We expect:

- Obtaining data for the COVID-19 and reference parts of the corpus
- Conversion of the data into the ParlaMint format
- Linguistic processing of the corpus with Universal Dependencies, preferably including a suitable NER module
- Producing documentation on the provided corpus

We provide:

- Dedicated guidelines on how to prepare the data
- The already existing corpora as models and the ParlaMint Schema:
- Upload to the concordancers
- Upload to Parlameter

More details

Funding: 5,000 Euro per project

Duration: Dec 1, 2020 – Mar 31, 2021 (4 months)

Assessment criteria:

- Qualifications of the team involved
- Status of the available and/or accessible parliamentary corpora
- Diversity of parliaments and languages
- Potential with respect to the project goals

If you want to join as a User

ParlaMint

Three showcases are envisaged:

1. Parlameter showcase: graphical exploration of the corpus
   suited for investigations by political scientists, journalists
   and active citizens

2. Extending the CLARIN tutorial "Voices of the Parliament":
   how corpora can be used to investigate language use and communication
   practices in a specialised socio-cultural context of political discourse.

3. A DH-related showcase prepared by our digital historian.

But you do not have to wait for all this to come.

Start exploring the data **NOW** through the concordancers!!! :)