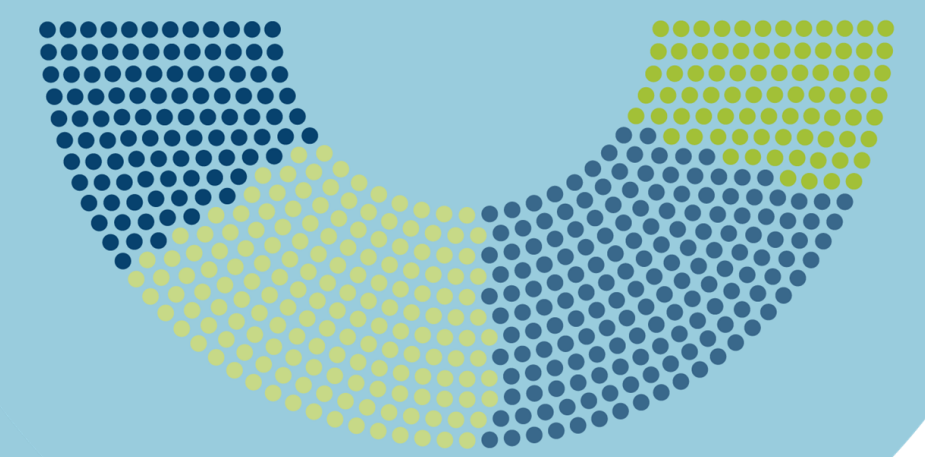


# ParlaMint II: The Show must go on

ParlaMint



Maciej Ogrodniczuk, Petya Osenova, Tomaž Erjavec, Darja Fišer, Nikola Ljubešić, Çağrı Çöltekin, Matyáš Kopp, Katja Meden

## ParlaMint II specifics

- The first CLARIN flagship project
  - Duration: December 2021 – May 2023
- Workplan:
- Upgrading the XML schema and validation
  - Extending existing corpora and enhancing the corpora with additional metadata
  - Adding corpora for new parliaments (together 27 corpora)
  - Improving usability of the corpora

## Partners

Austria, Basque Country, Belgium, Bulgaria, Catalonia, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Greece, Hungary, Iceland, Italy, Latvia, Lithuania, Netherlands, Norway, Poland, Portugal, Romania, Slovenia, Spain, Sweden, Turkey, United Kingdom

## Our tasks

- Schema and Metadata Improvements  
**Leads: Tomaž Erjavec (IJS), Matyáš Kopp (UFAL):** Parla-CLARIN schema to be harmonized according to the ParlaMint one; establish validation procedures; provide elaborate documentation and git management; find common ground in various political systems for adequately adding new metadata.
- Corpus Expansion  
**Lead: Tomaž Erjavec (IJS):** Updating ParlaMint I corpora, adding new ones
- Corpus Enrichment  
**Lead: Nikola Ljubešić (IJS):** Machine translation of corpora into English and semantic tagging with the UCREL Semantic Analysis System (resp. Paul Rayson, UCREL); Alignment with audio signal for selected corpora
- Engagement Activities  
**Leads: Darja Fišer (INZ), Çağrı Çöltekin (TUB):** Tutorial, Showcases, Shared task

## What is already available from ParlaMint I?

- 17 parliamentary corpora deposited in the CLARIN.SI repository and released under CC BY:
    - <http://hdl.handle.net/11356/1432>
    - <http://hdl.handle.net/11356/1431>
  - Mounted on noSketch Engine and KonText concordancers
  - Used in the Helsinki DH Hackathon 2021 and 2022
- + the first freely-available dataset for training ARS for Croatian based on the ParlaMint-HR corpus and video recordings of the Croatian parliament: 1,816 hours (<http://hdl.handle.net/11356/1494>)

## Beyond ParlaMint II

- **More data**
  - from the European Parliament
  - more regional parliaments
  - national parliaments beyond Europe
  - historical data
- **Link with other data sources**
  - e.g. voting results, social media content, newspaper and TV news mentions
- **Further extensions**
  - multimodal corpora
  - gesture annotated corpora
  - monitor corpora



ParlaMint project is supported  
by CLARIN ERIC



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



[maciej.ogrodniczuk@gmail.com](mailto:maciej.ogrodniczuk@gmail.com), [petya@bultreebank.org](mailto:petya@bultreebank.org)