

[Click here to be redirected to the virtual room of the Project Expo.](#)

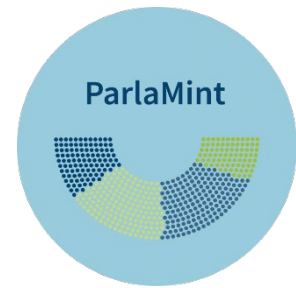
Project abbreviation: ParlaMINT

Project name: ParlaMint: Towards Comparable Parliamentary Corpora

Project coordinator: Maciej Ogrodniczuk (PL-PAS), Petya Osenova (BG-BAS)

Project consortium:

- Main partners: PS-PAS, BG-BAS, SI-IJS, SI-INZ, CZ-UFAL, DE-TUB, UK-UCREL
- Contributing partners: National CLARIN experts: Austria, Basque, Catalonia, Estonia, Finland, Greece, Norway, Portugal, Romania, Sweden, Serbia, Belgium, Croatia, Denmark, France, Hungary, Iceland, Italy, Latvia, Lithuania, Netherlands, Spain, Turkey



Funding: CLARIN ERIC & local in-kind contributions

- Phase 1: 135k EUR
- Phase 2: 163k EUR

Project duration:

- Phase 1: 1 July 2020 - 30 May 2021
- Phase 2: 1 Nov 2021 - 30 Apr 2023

Main key words: parliamentary corpora

Background of the research topic: National parliamentary data is a verified communication channel between the elected political representatives and society members in any democracy. It needs to be made accessible and comprehensive - especially in times of a global crisis. With the recent advances of artificial intelligence, analytics over unstructured parliamentary data for many languages is rapidly becoming a prerequisite for reliable and trustworthy approaches in checking the veracity of information in contemporary society.

One of the most important characteristics of new parliamentary data is its direct correspondence to the most recent events, including the ones with global impact on human health, social life and economics such as the current COVID-19 pandemic. By comparing the data synchronically and diachronically within a cross-lingual context, scientific and civil communities will be able to track pan-European discussion and can be quickly updated on any emerging topic.

Goal of the project: The mission of the ParlaMint project is to turn existing contemporary multilingual and diverse cross-national parliamentary data into resources that are comparable, interpretable and highly communicative with respect to society (NGOs, citizens, researchers, etc.).

The project will provide data 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. in times of emergencies. For the ParlaMint project the emergency case is the COVID-19 pandemic. However, the methodology will be scalable to other events, such as economic crises, etc. Thus, the main aims are:

- to compile a collection of parliamentary datasets (corpora) in a number of languages and in a harmonized format, covering both the current data and older, reference data,
- to process the corpora linguistically,
- to index the data with popular concordancers so that interested parties can search and extract the relevant comparable information,

[Click here to be redirected to the virtual room of the Project Expo.](#)

- to show through appropriate use cases that the CLARIN resources and technology serve societal needs.

For accomplishing this mission a respective-based standard schema has been developed and rich metadata has been added, such as members-of-parliament bios, parties' bios, ruling vs. opposition information and relations among all of them.

Thus, observations over democratic processes are approached through parliaments as digital bodies viewed through the following related strategies:

- speaker and parties statistics (for instance, who spoke more and on which topic; who changed their mind on a certain topic; which party defends/opposes what proposals, etc.)
- topic modeling (which topics are most popular at what time; how topic change and interrelate, etc.)
- time and context-bound social tendencies (tendencies in policy making over time).

Project abstract: Strategy and Data availability: The project will establish a strategy for handling parliamentary data and processing in times of any emergency (COVID-19 is just a showcase). Thus, different reference corpora could be produced with parliamentary records from previous times with global crisis states, e.g. the great economic recession, periods of floods in Europe, the Ebola outbreak etc.

Standard development: The Parla-CLARIN encoding standard will be further developed to cover more detailed and specific metadata across languages and parliaments. The corpora will serve as a baseline for further updates. Such uniform updates across the corpora would strongly support various methods of comparative research.

From showcasing to real applications: The availability of comparable multilingual parliamentary data (also made visible through concordancers and Parlameter) will boost research in the areas of digital humanities, linguistics, politology, sociology, psychology as well as in all the related branches of sciences.

Publications:

| |
|--|
| Corpora: Erjavec T. et al. (2021). <i>Multilingual comparable corpora of parliamentary debates ParlaMint 2.1</i> . Slovenian language resource repository CLARIN.SI. |
| Corpora: Erjavec T. et al. (2021). <i>Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 2.1</i> . Slovenian language resource repository CLARIN.SI . |
| Thesis: Pieters M. (2021). <i>A comparative analysis on the ParlaMint corpus</i> . MSc thesis. |
| Article: Fitsilis, F., Mikros, G. (2021). <i>Development and Validation of a Corpus of Written Parliamentary Questions in the Hellenic Parliament</i> . <i>Journal of Open Humanities Data</i> , 7:18. |
| Article: Erjavec T. et al. (2021). <i>ParlaMint: Comparable Corpora of European Parliamentary Data</i> . <i>Proceedings of CLARIN Annual Conference 2021</i> . |
| Article: Erjavec T. et al. (202x). <i>The ParlaMint corpora of parliamentary proceedings</i> . <i>Language Resources and Evaluation</i> . (submitted) |