

ParlaCLARIN III Workshop on Creating, Enriching and Using Parliamentary Corpora at LREC2022

1st Call for Papers

Date: Monday 20 June

To be held as part of the 13th edition of the Language Resources and Evaluation Conference (LREC), at the Palais du Pharo, Marseille, France (20–25 June 2022).

Location: Marseille, France

WORKSHOP DESCRIPTION

Parliamentary data is an important source of scholarly and socially relevant content, serving as a verified communication channel between the elected political representatives and members of the society. The development of accessible, comprehensive and well-annotated parliamentary corpora is therefore crucial for the information society, as such corpora help scientists and investigative journalists to ascertain the accuracy of socio-politically relevant information, and to inform the citizens about the trends and insights on the basis of such data explorations. Research-wise, parliamentary corpora are a quintessential resource for a number of disciplines in digital humanities and social sciences, such as political science, sociology, history, and (socio)linguistics.

The distinguishing characteristic of parliamentary data is that it is spoken language produced in controlled circumstances. Such data has traditionally been transcribed in a formal way but is now also increasingly released in the original audio and video formats, which encourages resource and software development and provides research opportunities related to structuring, synchronisation, visualisation, querying and analysis of parliamentary corpora. Therefore, a harmonised approach to data curation practises for this type of data can support the advancement of the field significantly. One of the ways in which the research community is supported in this line of work is through the conversion of existing corpora and further development of new cross-national parliamentary corpora into a highly comparable, harmonised set of multilingual resources. These allow researchers to share comparative perspectives and to perform multidisciplinary research on parliamentary data. We envision that the ParlaCLARIN III workshop, as a venue for knowledge and experience exchange on the topic, will contribute to the development and growth of the field of digital parliamentary science.

This third ParlaCLARIN workshop is a continuation of the 2018 and 2020 editions held at the respective LREC conferences. On the one hand, it continues to bring together developers, curators and researchers of regional, national and international parliamentary debates from across diverse disciplines in the Humanities and Social Sciences. On the other hand, we envisage the appearance of new discussion threads, tasks, and challenges that are partially inspired by or related to the new data releases such as ParlaMint and data formats such as Parla-CLARIN.

We invite unpublished original work focusing on (but not exclusive to)

Compilation, annotation, visualisation and utilisation of parliamentary records
Harmonisation of existing multilingual parliamentary resources, containing either synchronic or diachronic data or both
Linking or comparing of parliamentary records with other datasets of political discourse such as party manifestos, political speeches, political campaign debates, and social media posts, and to other sources of structured knowledge, such as formal ontologies and LOD datasets (in particular for the description of speakers, political parties, etc.)
Special themes for this year's workshop are:

Machine translation of parliamentary proceedings and research using machine translated parliamentary data
Semantic tagging of parliamentary proceedings and research using semantically tagged parliamentary data
Compilation, alignment and annotation of multimodal parliamentary resources and research using multimodal parliamentary data
Apart from the dissemination of the results, the workshop also aims to address the identified obstacles, discuss open issues and coordinate future efforts in this increasingly trans-national and cross-disciplinary community.

TOPICS OF INTEREST

Topics include but are not limited to:

- Creation and annotation of parliamentary data in textual and spoken format
- Enrichment of parliamentary data with semantic and named entity tagging
- Querying and visualisation of parliamentary data
- Text mining over parliamentary and other political language data
- Harmonization of multilingual parliamentary resources
- Adoptions or extensions of the Parla-CLARIN and ParlaMint schema to other parliamentary resources
- Comparative studies of parliamentary corpora
- Parliamentary corpora as a source of political language
- Diachronic studies based on parliamentary corpora
- Studies of parliamentary corpora with particular focus on the debates dedicated to global crises, such as COVID pandemic, climate crisis
- The potential of parliamentary resources beyond academia

SUBMISSION & PUBLICATION

We accept submission of long papers (up to 8 pages), short papers (up to 4 pages) and demo papers (up to 4 pages) to be presented as a long or short oral presentation at the workshop. The papers of the workshop will be published in online proceedings.

When submitting a paper from the START page, authors will be asked to provide essential information about resources (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a result of your research. Moreover, ELRA encourages all LREC authors to share the described LR (data, tools, services, etc.) to enable their

reuse and replicability of experiments (including evaluation ones). For contact data, stylesheets, up-to-date details on submission and the workshop itself, please consult the workshop website.

Submission page: will be communicated by beginning January 2022

IMPORTANT DATES

Paper submission deadline: 15 March 2022

Notification of acceptance: 25 April 2022

Camera-ready paper: 20 May 2022

Workshop date: 20 June 2022

ORGANISING COMMITTEE

- Darja Fišer, Institute of Contemporary History and University of Ljubljana, Slovenia
- Franciska de Jong, CLARIN ERIC, The Netherlands
- Maria Eskevich, CLARIN ERIC, The Netherlands
- Jakob Lenardič, University of Ljubljana, Slovenia

The workshop is supported by the CLARIN research infrastructure. To contact the organisers, please mail clarin@clarin.eu (Subject: [ParlaCLARIN@LREC2022]).

PROGRAMME COMMITTEE

(in alphabetical order)

- María Calzada Pérez, Universitat Jaume I, Spain
- Çağrı Çöltekin, University of Tübingen, Germany
- Tomaž Erjavec, Jožef Stefan Institute, Slovenia
- Francesca Frontini, Istituto di Linguistica Computazionale "A. Zampolli", CNR Pisa, Italy
- Maria Gavriilidou, ILSP/Athena RC, Greece
- Barbora Hladká, Charles University, Czechia
- Haidee Kotze, Utrecht University, The Netherlands
- Nikola Ljubešić, Jožef Stefan Institute, Slovenia
- Bente Maegaard, CST, Department of Nordic Languages and Linguistics, University of Copenhagen
- Laura Morales, Sciences Po, France
- Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences, Poland
- Petya Osenova, IICT-BAS and Sofia University "St. Kl. Ohridski", Bulgaria
- Stelios Piperidis, ILSP/Athena RC, Greece
- Paul Rayson, Lancaster University, United Kingdom

IDENTIFY, DESCRIBE AND SHARE YOUR LRS!

When submitting a paper from the START page, authors will be asked to provide essential information about resources (in a broad sense, i.e. also technologies, standards, evaluation kits, etc.) that have been used for the work described in the paper or are a new result of your research. Moreover, ELRA encourages all

CfP-ParlaCLARIN-III.txt

LREC authors to share the described LRs (data, tools, services, etc.) to enable their reuse and replicability of experiments (including evaluation ones).