

The ParlaMint Project: Ever-growing Family of Comparable and Interoperable Parliamentary Corpora

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



ParlaMint II releases



Tomaž Erjavec et al. (2023)
Multilingual comparable corpora of parliamentary debates ParlaMint 3.0
<http://hdl.handle.net/11356/1486>



Taja Kuzman et al. (2023)
Linguistically annotated multilingual comparable corpora of parliamentary debates in English ParlaMint-en.ana 3.0
<http://hdl.handle.net/11356/1810>



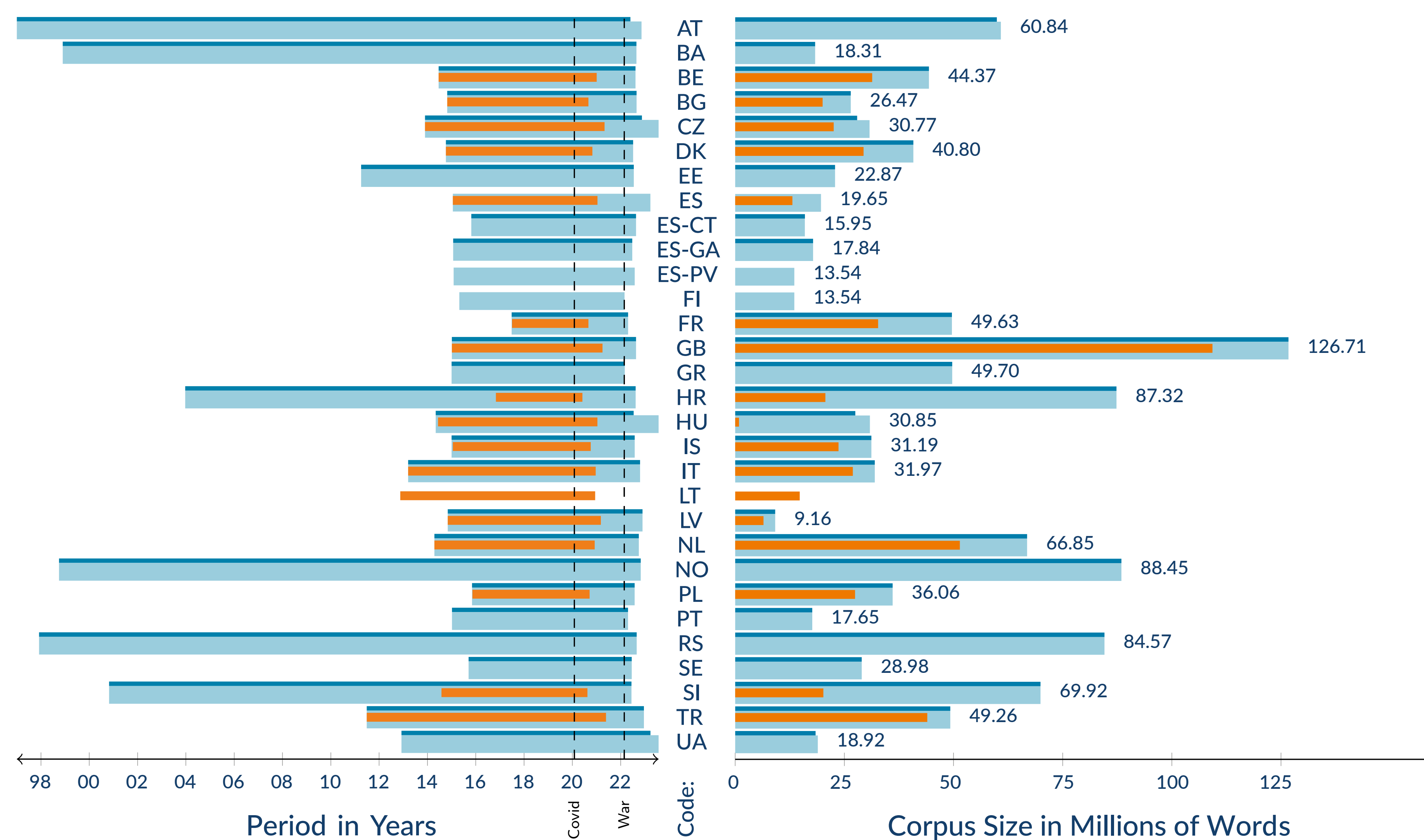
Tomaž Erjavec et al. (2023)
Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 3.0
<http://hdl.handle.net/11356/1488>




Nikola Ljubešić et al. (2022)
ASR training dataset for Croatian ParlaSpeech-HR v1.0
<http://hdl.handle.net/11356/1494>

ParlaMint corpora


- ParlaMint I (v2.1)
- ParlaMint II (v3.0)
- ParlaMint II (v4.0)



New language model and sentiment dataset



XLM-R multilingual model additionally pre-trained on ParlaMint and EuroParl data
Michal Mochtak et al. (2023)
<https://huggingface.co/classla/xlm-r-parla>



First use case of the model – sentiment identification via the ParlaSent dataset
<http://hdl.handle.net/11356/1868>, Michal Mochtak et al. (2023)
– 18,200 sentences in Bosnian, Croatian, Czech, English, Serbian, Slovak, Slovenian
<https://huggingface.co/classla/xlm-r-parlasent>

More to come

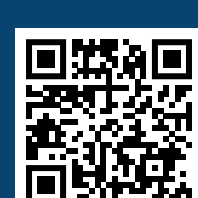
Shared task to be organized within **CLEF 2024** on the topic: **Ideology and Power Identification**

ParlaMint 4.0 release

ParlaMint-en 4.0: also semantically tagged



ParlaMint project is supported by CLARIN ERIC



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

maciej.ogrodniczuk@gmail.com; petya@bultreebank.org; tomaz.erjavec@ijs.si; darja.fiser@inz.si; nikola.ljubesic@ijs.si; ccoltekin@sfs.uni-tuebingen.de; kopp@ufal.mff.cuni.cz; katja.meden@ijs.si; taja.kuzman@ijs.si