

# CLARIN.SI: The Slovenian node of CLARIN

<https://www.clarin.si/>

**CLARIN.SI**  
Common Language Resources and  
Technology Infrastructure, Slovenia



## Repository

Ukrainian web corpus MaCoCu-uk 1.0

Please use the following text to cite this item or export to a predefined format: BIBTEX CMDI

Bañón, Marta; et al., 2023, *Ukrainian web corpus MaCoCu-uk 1.0*, Slovenian language resource repository CLARIN.SI, ISSN 2820-4042, <http://hdl.handle.net/11356/1838>.

Share: Facebook Twitter

CLARIN.SI Data & Tools

Authors: Bañón, Marta; et al. show everyone

Item identifier: <http://hdl.handle.net/11356/1838>

Project URL: <https://macocu.eu/>

Referenced by: <https://hdl.handle.net/11370/685514a8-947e-44f9-83cf-90356c5f1684>

Date issued: 2023-05-24

Type: corpus, text

Size: 21471613 texts, 6181945683 words

- ⇒ CTS and CLARIN certified repository of language data and software
- ⇒ 562 entries (3.4T): Slovenian (278), Croatian (74), Serbian (66), Bulgarian (27), Macedonian (17), Montenegrin (9)
- ⇒ Large annotated corpora, training corpora, lexical resources, language models, ELEXIS catalogue of dictionaries

## Concordancers

KEYWORDS ParlaMint-AT 3.0 (Austrian parliament)

War

SINGLE-WORDS ✓

reference corpus: ParlaMint-AT 3.0 (Austrian parliament) (items: 14,950)

Word	Word	Word
1 teuerung	11 hospiz-	21 geosphere
2 gasheizung	12 gasreserve	22 orratung
3 ukrainerin	13 linhart	23 russisch
4 putin	14 gutschein	24 wladimir
5 totschnig	15 speicher	25 meteorologie
6 gas	16 tursky	26 ivermectin
7 putins	17 seelsorge	27 mindestpensionistin
8 ukrainisch	18 erdgas	28 kraus-winkler
9 ukraine	19 wasserstoff	29 elli

### On-line corpus analysis:

- ⇒ 150 corpora, 35 languages, 28 billion words
- ⇒ Annotations: lemmas, PoS, NER, syntactic dependencies
- ⇒ Several concordancers: KonText, noSketch Engine with(out) log-in
- ⇒ Openly accessible (API!)
- ⇒ Other CLARIN.SI tools: GitLab, WebAnno

## Engagement

### Community:

- CLARIN.SI consortium: 13 main Slovenian language-related organisations
- Co-organiser of biennial LT & DH conferences, support for other events
- Talks, tutorials, LR development support

### Projects:

- Annual CLARIN.SI supported projects: 24 completed (2018–)
- Implementation of European Cohesion: 500k EUR for equipment (2019–2021)
- RDA Node Slovenia (2020–2021)



University of Ljubljana

