CLARIN.SI:
The Slovenian node of CLARIN

https://www.clarin.si/

CLARIN.SI
Common Language Resources and Technology Infrastructure, Slovenia

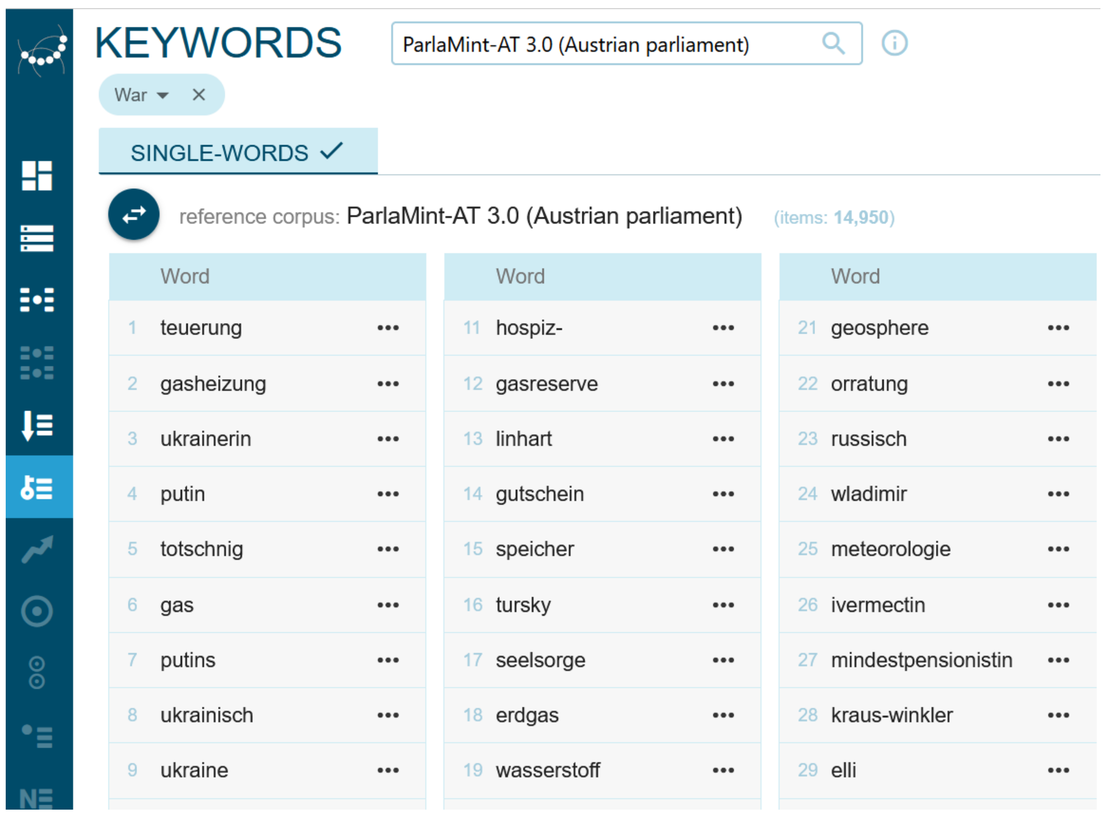
Repository

⇒ 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



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)

Jožef Stefan Institute
University of Ljubljana
ZRC SAZU
inštitut za novejšo zgodovino
ZRS
nuk
University of Maribor
alpineon
University of Nova Gorica
AMEBIS
Univerza na Primorskem
trojína ZAVOD ZA UPORABNO SLOVENISTIKO
SDJT Slovensko društvo za jezikovne tehnologije