

Some steps towards the UNIDIVE BLARK: revising Latvian language resources and tools

Inguna Skadiņa, Baiba Saulīte, Laura Rituma and Lauma Pretkalniņa

How to measure universality and diversity?

Corpora and corpus annotation

- Latvian National Corpora Collection (Multext-EAST)
- Latvian treebank (UD & hybrid)
- FrameNet
- PropBank
- AMR

Lexicons

- tezaurs.lv (TEI)
- WordNet (LMF)
- National Terminology portal (TBX)
- EurotermBank (TBX)

Language processing tools

- LV-Pipe
- UD parsers (BERT based)
- WSD solutions
- NE identification, extraction, linking
- MWE identification