

The laboratory has been conducting research on natural language processing (<http://nlp.ailab.lv/>) and has been providing access (<http://www.korpuss.lv/>) to different language resources for about 30 years.

The first collection of language resources in the CLARIN-LV repository (<http://repository.clarin.lv/>) is the Language Resources and Tools (LRT (<https://www.clarin.eu/glossary#LRT>)) of the AiLab IMCS UL. The collection demonstrates the variety of Latvian LRTs through the inclusion of:

- Monolingual (e.g., LVK 2018 - the largest balanced corpus of modern Latvian (<https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/3>)) and multilingual (e.g., Lithuanian-Latvian parallel corpus (<https://repository.clarin.lv/repository/xmlui/handle/99999/6>)) corpora;
- Text and audio corpora (e.g., annotated longitudinal corpus of Latvian children's language (<https://repository.clarin.lv/repository/xmlui/handle/99999/7>));
- Lexicons (e.g., tezaurs.lv (<https://repository.clarin.lv/repository/xmlui/handle/99999/9>));
- Browsable and downloadable resources; (<https://repository.clarin.lv/repository/xmlui/handle/99999/6>)
- Tools for language processing (e.g., NLP-pipe (<https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/4>), for details see also article in Tour de Clarin (<https://www.clarin.eu/tags/tour-de-clarin-latvia>)).

Although most LRTs are developed for Latvian, our collection also includes the Latgalian language corpus (<https://repository.clarin.lv/repository/xmlui/handle/99999/8>). (<https://repository.clarin.lv/repository/xmlui/handle/99999/6>)

The aim of CLARIN-LV (<https://www.clarin.lv/en-us/>) for 2020 is to extend the current collection of AiLab IMCS UL language resources (<https://repository.clarin.lv/repository/xmlui/handle/20.500.12574/2>), as well as to add new collections of other language resources developed by different stakeholders in Latvia. The CLARIN-LV language resources and tools can now also be browsed and searched via the Virtual Language Observatory (<https://vlo.clarin.eu/search?fq=dataProvider:CLARIN+Centre+of+Latvian+language+resources+and+tools>).

Speech Synthesis and Recognition Laboratory of UIIP NAS Belarus (Minsk)

By Yuras Hetsevich

The Speech Synthesis and Recognition Laboratory of UIIP NAS Belarus works in the fields of text and speech processing on the basis of human-human, human-machine and machine-machine communications. The Lab has expertise in the building of systems for stationary, mobile and web-based platforms for Belarusian, Russian and English languages.

The special platform has been developed and is being constantly updated further to provide users with a set of 50+ tools (services) for text, voice and other data processing. The developed services are then grouped into thematic domains for more convenient use in specific fields of application.

The approach to the development of each service, is simple, the user can run the service by clicking only on one button, with this action the input test data will be processed and the results will be shown. After this, the user is offered to input his own data and adjust the setting before running the tool. This approach helps students and researchers to get up to speed on NLP and test new hypotheses faster.

The platform provides tools for tokenization, morphological analysis, voiced electronic grammatical dictionary, part-of-speech tagging, frequency counter, spell checking, text-to-speech and many others.

The lab has recently started the process of [CMDI](https://www.clarin.eu/glossary#CMDI) (<https://www.clarin.eu/glossary#CMDI>) metadata creation for all online resources, which means that part of the services are now available via the VLO (<https://vlo.clarin.eu/search?fq=collection:UIIP+NAS+Belarus+Speech+Synthesis+and+Recognition+Laboratory>). All services can also be accessed through the platform directly (<https://corpus.by/?lang=en>). More information is available on the Speech Synthesis and Recognition Laboratory of UIIP NAS Belarus website (<https://ssrlab.by/en/>).

This blog post was written by: Johan Frid (Lund University), Twan Goosen (CLARIN ERIC (<https://www.clarin.eu/glossary#ERIC>)), Yuras Hetsevich (National Academy of Sciences of Belarus), Gerlinde Schneider (University of Graz), Inguna Skadiņa (Institute of Mathematics and Computer Science, University of Latvia)
