WE COOPERATE WITH...

The National Programme for Estonian Language Technology (NPELT, see more at www.keeletehnoloogia.ee), a government supported research funding initiative aimed at developing computational technologies for the Estonian language. All results of the NPELT projects (language resources and software prototypes) are released in public domain. CELR has an obligation to deposit all such resources and tools for preservation and long-term access.

EKT

EESTI KEELETEHNOLOOGIA
NATIONAL PROGRAMME FOR ESTONIAN
LANGUAGE TECHNOLOGY

CELR is the official consortium representing Estonia in CLARIN ERIC (Common Language Resources and Technology Infrastructure of European Research Infrastructure Consortium - see more at www.clarin.eu) - a large-scale pan-European collaborative effort to create, coordinate and make language resources and technology available and readily useable via an integrated and interoperable research infrastructure.

CLARIN ERIC
Common Language Resources and Technology Infrastructure

CELR also cooperates with META-NET (see more at www.meta-net.eu), a Network of Excellence dedicated to building the technological foundations of a multilingual European information society. A number of our resources are registered and documented in META-SHARE nodes forming network of harvestable repositories all over the Europe. Maintenance of local META-SHARE register and repository system as node (metashare.ut.ee) in bigger network gives us more experience and suitable environment for cataloging our LT resources.

META-NET

CONTACTS

Center of Estonian Language Resources
J.Liivi 2-340, 50409, Tartu, Estonia
Phone: +372 7376 433
E-mail: info@keeleressursid.ee

PEOPLE

Kadri Vider
Executive Manager of CELR
CLARIN ERIC National Coordinator for Estonia
Phone: +372 7376 433
E-mail: kadri.vider@ut.ee

Tiit Roosmaa
Chair of CELR Council
E-mail: tiit.roosmaa@ut.ee

Include Estonian in your research using our resources!

ee.clarin.eu/resources

Haridus- ja Teadusministeerium
Estonian Ministry of Education and Research

TARTU ÜLIKOOL
UNIVERSITAS TARTUENSIS
1632

INSTITUTE OF CYBERNETICS AT TUT
TTÜ KÜBERNEETIKA INSTITUUT

Institute of the Estonian Language

European Union
Regional Development Fund   Investing in your future

Center of Estonian Language Resources

Journalists
Linguists
Historians   Arts and humanities
Terminologists
Translators   Metadata   Phonetics
Data mining
Anthropologists
IT-developers   Mobile applications
Estonian language   Language technology
Lexicographers
Corpus query   Conversation analysis
Media monitoring   Open Data   Teaching
Wordnets
NLP tools   Researchers
Medical science

Cloud of digital language resources in our basket of berries

THE CENTER

The goal of the Center of Estonian Language Resources (CELR) is to create and manage an infrastructure to make the Estonian language digital resources (dictionaries, corpora – both text and speech –, various language databases) and language technology tools (software) available to everyone working with digital language materials. CELR coordinates and organises the documentation and archiving of the resources as well as develops language technology standards and draws up necessary legal contracts and licences for different types of users (public, academic, commercial, etc.). In addition to collecting language resources, a system will be launched for introducing the resources to, informing and educating the potential users.

The main users of CELR are researchers from Estonian R&D institutions and Social Sciences and Humanities researchers all over the world via the CLARIN ERIC network of similar centers in Europe.

CONSORTIUM

CELR is a consortium of 3 institutions:

🌸 The University of Tartu is the leading establishment in the field of corpora, language software and communication modelling. The main processors of the respective resources are the linguists and language technologists working in the computational linguistics study group of the University (see more at www.cl.ut.ee).

🌸 The Institute of Cybernetics at Tallinn University of Technology is the leading establishment in the field of speech technology. The respective resources are processed in their laboratory of phonetics and speech technology (see more at www.phon.ioc.ee).

🌸 The Institute of the Estonian Language is the leading establishment where various lexical resources are processed (see more at www.eki.ee).

RESOURCES

CELR offers more than 50 language resources.

Text corpora

The largest and most representative text corpora for the Estonian language – from historical texts to new media, from different subject fields, with annotation levels up to syntax, emotions or word senses.

e.g. 🌸 Estonian Reference Corpus

🌸 Estonian-English parallel corpus

🌸 Estonian Treebank

Lexical resources

Frequency lists, MWE and place name databases, mono- and multilingual dictionaries.

e.g. 🌸 Estonian-Russian-Estonian dictionaries

🌸 Estonian WordNet

Speech databases

Speech data and corpora for both speech synthesis and recognition.

e.g. 🌸 Estonian Dialect Corpus

🌸 Phonetic Corpus of Estonian Spontaneous Speech

🌸 Estonian Emotional Speech Corpus

Speech processing tools

Downloadable and online tools and mobile applications for speech synthesis and recognition.

e.g. 🌸 Text-to-Speech Synthesis for Estonian

🌸 Real-time speech recognition web service

🌸 Voice recorder and transcriber for Android (Diktofon)

Text processing tools

Downloadable or online tools for annotation, summarizing and edutainment.

e.g. 🌸 Morphological Toolset for Estonian

🌸 EstCG shallow parser for Estonian

SERVICES AND SUPPORT

Archiving and depositing

... texts and corpora, audio and video recordings, lexical data and terminologies, tools for NLP (Natural Language Processing) – every piece of linguistic research data is worth archiving and CELR can do it virtually without limits.

Access to resources

... once the data is collected and archived, users need access to the Estonian language resources. CELR can make them handily available according to the specified terms and conditions for use.

Expertise for users

... access to the data takes you only halfway towards the goal. Whether you get the answers to your questions depends on your knowledge on how to use the huge amount of data in a more efficient way. CELR can share this knowledge in the user guides and FAQ provided on the webpage, but also in workshops and responses to individual queries.

Environment for using and creating tools

...when using the Estonian data, competent users need an environment for searching, analysing and annotating language resources as well as developing new NLP tools – CELR can provide such an environment.

Look for more at
ee.clarin.eu/resources