APPROVED
with the Decree of the Minister of Education and Research
"Approval of the Language Technology Research and Development Program
“Estonian Language Technology 2018-2027”
of the Ministry of Education and Research”

The Language Technology Research and Development Program “Estonian Language Technology 2018-2027” of the Ministry of Education and Research”

Table of Contents

Introduction

Language technology has a key role to play in ensuring the viability and sustainability of a language. Language technology is no longer a narrow field of research for linguists, its impacts are economic, social, and serve the entire society from the general user to the top-notch specialist.

The world is increasingly interacting with machines in natural language, or by speech, and language technology is one of the necessary components in automated IT systems that increasingly need natural language processing.

The position of Estonian language technology in the languages with the same number of speakers is stable. As a result of consistent work, important basic technologies have been developed, and applications that are actually used by the end-user for speech recognition, speech synthesis, and machine translation have been created. Applications are based on extensive language resources and text analysis tools.

Through the program, the state supports a field where it is not always profitable for the private sector to take on the risks associated with the development of technology for a language with a small number of speakers - as a small number of speakers also means a small market.

The activities of the research and development program "Estonian Language Technology 2018-2027" (hereinafter program) supporting the development of language technology will implement the objectives of two sectoral strategies: the research and development and innovation strategy "Knowledge-based Estonia 2014-2020" and "Estonian Language Strategy 2018-2027".

The purpose and indicators of the program

The purpose of the program is to ensure that the basic components of Estonian language technology comply with good international standards and that the Estonian language technology can be used by a broader target group.[1] To achieve this purpose, new language technology applications will be created with the help of the program, the quality of existing applications will be increased and implemented in as many areas as possible, in the private, public and third sectors.

An indicator of the achievement of the program's objectives is the services, products, and applications supported by Action II, into which the language technology component has been integrated. The effectiveness of the program is assessed in particular through the activities of Action II, since the results of Action I find real use only when integrated into systems and

[1]The evaluation framework is based on the language technological support evaluation matrix of the META-NET network (Estonian Language in the Digital Age, 2012, p. 27). For more information on evaluation see the chapter "Program Evaluation."

applications (Action II), and thus the critical part of the program's socioeconomic impact is expressed in the number of users of applications with Estonian language support. The methodology for calculating the corresponding indicators will be developed by the MER by the year 2019 in co-operation with the Program Board. During this process, the initial and target levels of the indicators will be specified.

Explanation of concepts and intellectual property

Language technology is an interdisciplinary field connecting information technology and linguistics, which deals with the development of language software and language resources necessary for computer language processing of the human language. **Language resources** are electronic databases that are used to develop language software: corpora (aggregates of speech signals and texts), electronic dictionaries and databases.

The types of research and development supported under Actions I and II of the Program are defined by the so-called General Block Exemption Regulation:[2]

1) **Basic research** is an experimental or theoretical work mainly done to acquire new knowledge about the fundamentals of phenomena and the facts without the purpose of direct commercial applications.
2) **Applied research** is planned research or critical research to acquire new knowledge and skills that can be used to develop new products, processes, and services or to improve existing products, processes or services significantly. It involves the creation of components for complex systems necessary for applied research (and in particular for generic technology validation) and may include the construction of prototypes in the laboratory or in the environment simulating the existing systems.
3) **Product development** is the acquisition, integration, design, and use of existing scientific, technological, business and other relevant knowledge and skills, for the development of new or improved products, processes, and services. This may include activities aimed at the conceptual definition, designing and documenting new products, processes or services. Product development may include the creation, demonstration, pilot use, testing and validation of new or improved prototypes of products, processes, or services in a real-life working environment where the main objective is to technically improve the non-ready product, process or service.

The results of the projects supported under Actions I and II of the program are **intellectual property**, the protection method of which is chosen on the basis of the objective of maximizing the use of the created property without pursuing profit, whereas the transfer of exclusive proprietary rights is avoided. The use of the created intellectual property for various purposes (public use, research, business applications) is governed by various types of licenses, which

[2]The European Commission Regulation (EU) No 651/2014 on the application of Articles 107 and 108 of the Treaty on the European Union declaring certain categories of aid compatible with the common market.

may be both free and fee-based and also subject to other legal regulations, such as the protection of personal data.

The **resources and software** created within the program are **managed, made available free of charge as Open Data and licensed by the Center of Estonian Language Resources (CELR).**

If the project supported by the program results in fully delimited software code or language resources, it will be deposited in the final version (with documentation) at the CELR, and the CELR will arrange for the conclusion of the necessary contracts. Prototypes and resources developed with the support of the program (activities of Action 1) must be in free use without restriction, including their use for commercial purposes. In case of solutions and applications for Action II, fewer open licenses may be allowed in justified cases and in agreement with the Program Board.

Any **publications** issued with the support of the program funds **must be provided with Open Access.** This requirement can be fulfilled both by publishing in an open access journal or by self-archiving (Green Open Access) and making the publication available in the Estonian Research Information System (ETIS).

The program actions

The objectives of the program are supported by three actions: I) the development of basic technologies and resources that have not yet been implemented as part of a system and result in a comprehensive software module or systematized language resource; II) the introduction of language technology involving the application of technology as part of a system or the creation of a stand-alone application, resulting in the operation of language technology as part of a system or the completion of a language-technology focused application; III) support for the Center of Estonian Language Resources (CELR), which results in the functioning of the CELR as an active promoter of the introduction of the Estonian language technology centered infrastructure, the information facilitator and language technology.

Action I contributes to raising the level of language technology solutions and language resources and supports relevant new initiatives. Action II supports the creation of applications and introducing language technologies for as many users as possible, and Action III supports making the results of the program as easily accessible as possible by making the results available as well as introducing the possibilities of language technology to the widest possible audience.

Action I. Development of basic technologies and resources

Estonian language technology consists of basic technologies and resources that can be integrated into different systems and services and that are used by diverse technologies and applications as the underlying material.

Basic technologies include software development, including speech recognition and speech synthesis, technologies of machine translation and textual analysis, including mh (parallel) text corpora, speech corpora, syntax and lexicography resources. For both technologies and resources, their clear usage perspectives or applicability are important, i.e., to which target group, in what way and for what are they useful?

Action I activities are supported through open calls for applications and targeted support. The evaluation of applications is based, in particular, on the quality of project applications, ensuring that the different areas of language technology develop in a comparably uniform manner. In this context, priority is given to the development of the basic technologies of Estonian language technology, namely speech recognition, speech synthesis and machine translation, through Action I.

The aforementioned prioritized technologies are the areas developed in the framework of the existing national language technology programs, which are critical parts of the technical language capability and require special attention in financing. These areas are also part of the META-NET network language technology applications list: speech recognition, speech synthesis, grammatical analysis, semantic analysis, text generation and machine translation.[3]

The support from the Action can be applied by public, private and third sector bodies for basic and applied research.

Action II. Introduction of language technology in solutions and applications

Action II supports the use of the language technologies of the Estonian language in existing and emerging solutions or applications. Activities include, inter alia, the integration of Estonian language support in information systems, the provision for language-sensitive content search, and the creation of innovative and cross-sectoral new applications. The activities of Action II are supported through open calls for applications and targeted support.

The objective of the action is to bring the added value from language technology resources to as many users as possible, so the number of users covered by the support is the main evaluation criterion in planning the activities of the action.

The support from the action can be applied by public, private and third sector bodys for applied research and product development.

[3] The Estonian Language in the Digital Age, 2012, p. 27. http://www.meta-net.eu/whitepapers/e-book/estonian.pdf

Action III. Supporting the Center of Estonian Language Resources (CELR) and international cooperation

Within the framework of the action, support is given to CELR as language technology-oriented infrastructure and information mediator, and to international cooperation in language technology.

Activities supported by the action are:

1) Ensuring the services targeted to the scientific users of CELR infrastructure.
2) Licensing and storing of the language technology program and related consultation, including consultation on the implementation of program results in the CELR.
3) The active demonstration of language technology tools for the public, private and third party sectors. (CELR)
4) Participation in international co-operation, including payment of CLARIN ERIC membership fees.[4]

The task of CELR is to ensure the long-term preservation of the program's results, the publication of all results with the widest possible license and the widest possible user access.

The second program related task of CELR is to promote language technology, to engage in language technology information exchange in Estonia and internationally, and to introduce Estonian language technology opportunities and resources to public, private and third sector institutions.

Estonia is a member of the European Research Infrastructure for Language Resources and Technology (CLARIN) ERIC, an object of the European Strategy Forum for Research Infrastructures (ESFRI). The participation fee for the CLARIN network will be covered by the program's resources.

The aforementioned activities will be financed by the program budget at the request of the CELR, and the activity plan will be approved by the Program Board. Support for activities under Action 3 represents a maximum of 15% of the annual budget of the program. The Center of Estonian Language Resources cannot submit applications in Actions I and II.

State aid

If the support for the activities of this program constitutes State aid within the meaning of § 30 (1) of the Competition Act, granting of the support will be based on European Commission Regulation (EU) No 651/2014 on the application of Articles 107 and 108 of the Treaty establishing the European Community declaring certain types of aid compatible with the internal market (OJ L 187 26 June 2014, p. 1-78), as amended by Commission Regulation (EU) 2017/1084 (OJ L 156, 20.06.2017, p. 1-18) (hereinafter referred to as the "General Block

[4]CLARIN ERIC (Common Language Resources and Technology Infrastructure) is an international research infrastructure in the field of language technology (https://www.clarin.eu/). In Estonia, the Center for Estonian Language Resources is the center and contact point for CLARIN.

Exemption Regulation"), from Article 25, and it is governed by provisions of the said Regulation and § 34^2 of the Competition Act.

Where support for the activities of this program is de minimis aid, granting of the support is based on the Commission Regulation (EU) No 1407/2013 on the application of Articles 107 and 108 of the Treaty on the Functioning of the European Union to de minimis aid (OJ L 352, 24.12.2013, p. 1-8) (hereinafter referred to as the de minimis aid Regulation), and it is governed by the provisions of the said Regulation and the provisions of § 33 of the Competition Act.

The support granted to a research and development institution within the framework of an activity does not constitute state aid for research and development if the project of the research and development institution is related to non-economic activities.

If the support is state aid or de minimis aid, the support will not be awarded to partners and final beneficiaries for whom a decision has been taken to recover the aid on the basis of a prior decision of the European Commission in which the aid was found to be unlawful and incompatible with the internal market, whereas the decision has not been executed.

This program does not apply, in addition to that referred to in the preceding paragraph, to other cases provided for in Article 1 (2), (3), (4) and (5) of the General Block Exemption Regulation and Article 1 (1) of the de minimis aid Regulation.

Program funding, management and implementation

Program funding

From 1 January 2018, the activities of the program will be funded by the Ministry of Education and Research (MER) from the state budget.

The budgets for projects to be implemented under the program consist of direct costs and overhead costs. **Direct costs** are personnel costs (salaries plus all national taxes, payments and statutory benefits, student research scholarships) and operating expenses (mission expenses, costs of acquiring basic or minor assets directly related to the execution of the project, outsourced research and development services, costs related to publishing and disseminating the results of a research project and other costs necessary for the implementation of the project, depending on the specific features of the project). **The overhead cost** is the grant holder’s costs related to the management of the project agency and the provision of a high quality environment for the execution of the project. **The overhead cost rate** of projects of Actions I and II is up to 20% of direct costs, except for outsourced research and development services.

The annual budget of the program is divided as follows:

1) Management costs are fixed in the program's annual budget;
2) Action III: up to 15% of the annual budget of the program;

3) The remaining funds fall into the activities of Measure I and II, and their detailed breakdown is confirmed by the Program Board every year.

Management

The implementation of the program is directed by a Board of seven members, whose membership is approved by the Minister of Education and Research for a period of up to five years. The Board consists of representatives of the Ministry of Education and Research, the Ministry of Economic Affairs and Communications and the private sector and experts. The Board may, if necessary, include additional experts in its work.

The Board:

1. approves the annual budget of the program together with the activity plan and program implementation reports;
2. establishes the conditions for applying for project support together with the MER;
3. makes a proposal to the MER for financing projects;
4. advises MER on the implementation, co-ordination, and evaluation of the activities, as well as in the design of language technology policy.

Implementation of the program

The program is implemented by the Ministry of the Education and Research in co-operation with the Program Board.

In addition to the open calls for applications for funding projects and the activity plan based support to CELR**,** MER has the right to initiate open calls for applications or provide targeted support for the development of specific language technology solutions or applications, participation in international co-operation (for example, for cost-sharing in co-operation projects) and international joint initiatives or implementation of other activities.

Evaluation of the program

In the middle (2022-2023) and at the end (2027) of the programming period (2018-2027), the program's objectives will be assessed. In Action I, the evaluation framework is a matrix for assessing language support for the META-NET network (Estonian Language in the Digital Age, 2012, p. 27). For Action II, the number of users covered over the period is assessed. Action III assesses the success of the activities of the Center of Estonian Language Resources.