

DEVELOPMENT PLAN OF THE INSTITUTE OF THE ESTONIAN LANGUAGE 2016-2019

1. Mission	2
2. Vision	2
3. Core values	3
4. Strategic documents and cooperation	3
5. Financing	4
6. Emphasis in the near future	6
6.1. Topics in the Centre of Excellence in Estonian Studies.....	7
6.2. Development of speech technology.....	11
6.3. Lexicography and lexicology	13
6.4. Language planning and care	18
6.5. Terminology work	22
6.6. Research of cognate languages and dialects.....	24
6.7. Language technology.....	27
6.8. Language archives of the Institute of the Estonian Language.....	29
7. Strengths, weaknesses, opportunities and threats.....	32
8. Strategic goals, activities and indicators, and new initiatives	35
9. Development of the organisation.....	40
9.1. Organisational structure, culture, and management	40
9.2. Recruitment	41
9.3. Training	42
9.4. Organisation of work.....	42
9.5. Working environment.....	42
9.6. Development of the technical equipment.....	43
10. Intellectual property of the Institute of the Estonian Language	43

The Institute of the Estonian Language is a national research and development institution managed by the Ministry of Education and Research. The Institute was established in 1947 as part of the Academy of Sciences of the Estonian SSR and its name until 1993 was the Institute of Language and Literature.

1. Mission

The Institute of the Estonian Language as a research and development institution with excellent qualifications and extensive experience helps contribute to the survival and development of the Estonian language through time.

The Institute performs a number of public functions: compilation of dictionaries of the Estonian language that are important for the national culture, provision of free public language advice (incl. advice about names and terminology), coordination of nationwide terminology work, services for people with special needs (speech technology, dictionary of sign language), e-services for language learners, collection and development of language archives.

The task of the Institute is to guarantee the norm of the Estonian standard language, which pursuant to law is established by the most recent version of the Dictionary of Standard Estonian (“Rules for the application of the norm of the Estonian standard language”, RT I, 14 June 2011, 3).

The Institute has the status of the Office of Onomastic Expertise (Appointment of the Office of Onomastic Expertise, RTL 2004, 87, 1376; RTL 2005, 28, 397), whose function is to give expert opinions about names and organise targeted research of place names.

2. Vision

The Institute of the Estonian Language is an independent national research and development institution and an important internationally recognised scientific centre that studies, develops, and preserves Estonian and cognate languages.

The Institute is a reliable partner whose research and development activities are known and valued in Estonia and the rest of the world.

3. Core values

Openness. The Institute of the Estonian Language is open to everything that happens in society. The Institute distributes its scientific achievements and considers itself obliged to publish its research outcomes in top scientific publications as well as in popularising formats, both in Estonian and foreign languages.

The Institute defends the freedom of scientific thought, condemns censorship of scientific works and the monopolisation of research trends.

Consistency. The Institute ensures that society values the Estonian language as the main force that holds Estonian society together.

Knowledge base. The Institute proceeds from scientific analysis when planning its activities and bases its cooperation with other stakeholders on this.

Innovation. The Institute values the creation of new knowledge and its implementation in the interests of society.

4. Strategic documents and cooperation

The Institute of the Estonian Language proceeds from the values of the Constitution of the Republic of Estonia (RT 1992, 26, 349).

The Institute designs its strategy on the basis of the following documents:

- “Development Strategy of the Estonian Language 2011–2017”;
- “Language Programme 2016–2019”;
- “Estonian Foreign Language Strategy 2009–2017”;
- The Estonian RD&I strategy “Knowledge-Based Estonia 2014–2020”;
- Development Plan of the Area of Government of the Ministry of Education and Research “Smart and Active People 2015–2018”;
- National Programme “Estonian Language and Cultural Memory II 2014–2018”;
- National Programme “Estonian Language Technology (2011–2017)”;
- National Programme “Estonian Language Terminology 2013–2017”;
- “Kindred People Programme IV 2015–2019”;
- “Estonian Lifelong Learning Strategy 2020”;
- Competitiveness Plan “Estonia 2020”;
- Code of Ethics of Estonian Scientists 2002.

In order to perform its functions, the Institute of the Estonian Language cooperates with the University of Tartu, Tallinn University, the Institute of Cybernetics of the Tallinn University of Technology, the Estonian Literary Museum, Foundation Innove, the Mother Tongue

Society, the Estonian Association of Applied Linguistics, the Fenno-Ugria Institution, and other national organisations.

The Institute of the Estonian Language performs the public functions commissioned to the Institute by different ministries: the Ministry of Education and Research, the Ministry of Defence, the Ministry of Finance, and the Ministry of the Interior.

The Institute is a member of the consortium that forms the Centre of Estonian Language Resources. The Centre of Language Resources in turn belongs to the pan-European cooperation network CLARIN-ERIC.

The Institute is a partner of the Centre of Excellence in Estonian Studies (CEES, 2016–2023). The topics that participate in the Centre of Excellence are Institutional Research Funding IUT35-1 “Speech styles, sentence prosody, phonological variation: description, theory and modelling (2015–2020)” and Personal Research Funding PUT550 “The cultural shift in Estonian in the 17th and 18th century (2014–2017)”.

The most important foreign partners of the Institute of the Estonian Language are the Institute for the Languages of Finland and the Czech company Lexical Computing Ltd. Cooperation is also close with colleagues in Latvia, Russia, Germany, Sweden, England, Slovenia, France, *et al.*

The Institute is a member of several international organisations such as the European Association for Lexicography EURALEX, the European Association for Terminology EAFT, and the plain language movement Clarity.

The Institute participates in two European Cooperation in Science and Technology (COST) projects: IS 1305 “European Network of e-Lexicography” and OC-2016-1-20642 “European Network for Combining Language Learning with Crowdsourcing Techniques”.

The Institute participates in international cooperation in name planning (team of UN place names experts) and terminology planning (NATO Military Committee Terminology Board, MCTB).

The Institute and the Estonian Language Council represent Estonia in the work of the European Federation of National Institutions for Language (EFNIL).

5. Financing

- The activities of the Institute of the Estonian Language are financed on the basis of two models. The activities important for the state and national culture are financed from the funding allocated for the “Development Strategy of the Estonian Language”, the action plan of which is approved with a directive of the Minister of Education and Research. This includes:

- 1) planning, development, and care of the Estonian language;
- 2) giving language advice;
- 3) compilation, editing, and constant updating of dictionaries important for national culture, incl. standard language, explanatory, dialectological and etymological dictionaries;
- 4) the electronic dictionary writing system **EELex** (hereinafter the EELex) and provision of technical support services to lexicographers;
- 5) maintaining and updating databases of the standard Estonian language, Estonian dialects and scientific archives of closely cognate languages, and guaranteeing that they can be used by all local and foreign researchers;
- 6) development of the field of Finno-Ugric languages;
- 7) terminology work and names planning;
- 8) supporting the use of sign language and language use by people with special language needs;
- 9) improving the efficiency of Estonian language instruction in general education schools;
- 10) EU language planning;
- 11) promoting plain language.

According to the regulation of the Government of the Republic, the most recent Dictionary of Standard Estonian compiled in the Institute is the basis of the norm of the standard Estonian language. The Institute has the status of the Office of Onomastic Expertise, whose task is to provide expert opinions in issues concerning names planning.

The remaining activities of the Institute (particularly research) are financed on a project-basis mainly via institutional and personal research funding, national programmes and target funding and orders. The competence of the Institute includes:

- 1) monitoring and research of the structure of contemporary Estonian and its development tendencies;
- 2) monitoring and research of the vocabulary (incl. terminology) of contemporary Estonian and its development tendencies;
- 3) researching the origin and history of the Estonian language;
- 4) researching dialects of the Estonian language and the Finno-Ugric languages;
- 5) promotion of the language planning theory;
- 6) development of the language theory and solution of general linguistic problems;
- 7) development of language technology tools for the Estonian language;
- 8) speech technology applications;
- 9) onomastics;
- 10) participation in terminology projects and terminology consultation;
- 11) expert analyses of language;

- 12) consultations and in-service training in the fields of the main activities of the Institute.

The Institute of the Estonian Language implements the institutional research topic “Speech styles, sentence prosody, phonological variation: description, theory and modelling (2015–2020)” and the personal research topic “The cultural shift in Estonian in the 17th and 18th century (2014–2017)”. The Institute had six doctoral students as of the start of 2016: Tiina Paet (“The shape fixation process and semantics of foreign words in Estonian as a result of language contacts”), Tiina Laansalu (“Place names of Kesk-Harjumaa. Diachronic analysis and transition”), Marit Alas (“The changing of place names in Virumaa from a socio onomastical aspect”), Maria Tuulik (“Polysemy of adjectives in Estonian from a lexicographical and lexicological aspect”), Kairi Tamuri (“The acoustic analysis and modelling of the basic emotions in spoken Estonian”), Kristina Koppel (“Parameters of example sentences for Estonian learners’ dictionaries”).

6. Emphasis in the near future

The Institute of the Estonian Language determines its place on the language landscape and strengthens its positions in the areas of language where it has been the most prominent and where its potential is the biggest. These are:

- 1) topics of the Institute of the Estonian Language in the Centre of Excellence in Estonian Studies (the main areas are the history of the Estonian language and speech research);
- 2) speech technology;
- 3) lexicography and lexicology;
- 4) language planning and care;
- 5) terminology work;
- 6) research of cognate languages and dialects;
- 7) development of language technology, incl. development of dictionary and terminology management software;
- 8) preservation, streamlining, and development of the language archives of the Institute of the Estonian Language, incl. the archives of the Estonian dialects and Finno-Ugric languages (EMSUKA).

The research and development activities of the Institute are described in detail in chapters 6.1 to 6.8.

Abbreviations and concept

Graduate School – the Graduate School of Linguistics, Philosophy and Semiotics 2016–2022 operated by the University of Tartu whose partners are Tallinn University, the Estonian

Literary Museum, and the Institute of the Estonian Language.

ETF – the Estonian Science Foundation grant (until 2015)

IUT – Institutional Research Funding

PUT – Personal Research Funding

SF – Target Funded Research Topic (until 2014)

National programmes:

- EKKM2 – “Estonian Language and Cultural Memory II 2014–2018”
- EKT – “Estonian Language Technology (2011–2017) ”
- HP – “Kindred People Programme IV 2015–2019”
- TERM2 – “Estonian Language Terminology 2013–2017”

6.1. Topics in the Centre of Excellence in Estonian Studies

The Institute of the Estonian Language is a partner of the Centre of Excellence in Estonian Studies (CEES, 2016–2023). The Centre of Excellence is an association of several internationally recognised research teams whose goal is to improve the level and success of scientific research (“Supporting centres of excellence to strengthen international competitiveness and top quality”, RT I, 10 April 2015, 5). The goal of the centre of excellence of humanities is to create the theoretical starting points and scientific bases for solving issues important for the Estonian society. The main areas are: 1) Estonian society and culture; 2) the population and ethnic composition of Estonia, 3) global culture and its local manifestations; 4) modern culture.

The Institute of the Estonian Language participates in activities of the centre of excellence concerning the Estonian language and culture with two research topics: PUT550 “The cultural shift in Estonian in the 17th and 18th century (2014-2017)“ and IUT35-1 “Speech styles, sentence prosody, phonological variation: description, theory and modelling (2015–2020)”. Both research topics stem from the traditional areas of research of the Institute: the first is related to studies of the vocabulary and grammar of historical sublanguages, the second to studies of speech.

6.1.1. Development history of the Estonian language

The goal of this field is to continue the long tradition of studying the development of the vocabulary and grammatical system of the Estonian language and names, enriching it with new theoretical angles and implementing newer corpus-based research methods. The target funded research topic “Word and concept in the development of Estonian vocabulary and

place names”, which was completed from 2010–2014, was related to this research trend. The Personal Research Funding PUT550 “The cultural shift in Estonian in the 17th and 18th century (2014–2017)” is the extended research of the vocabulary and grammar of historical sublanguages. The field of research is closely connected with the research of Estonian dialects and Finno-Ugric languages in the Institute.

1. PUT550 “The cultural shift in Estonian in the 17th and 18th century (2014–2017)”

The project “The cultural shift in Estonian in the 17th and 18th century” is a comparative study of the language used in the runosongs and translations of the hymns of the second half of the 17th and the 18th century. The research project focuses on the vocabulary, morphology, morphosyntactic constructions and phraseology of the two sublanguages. The study is corpus-based and its main goal is to explain the mechanisms of emergence of the modern standard Estonian and outline the relations of “our own” and “foreign” in the developing language. Test corpora of the language of runosongs and translations of hymns will be prepared as a result of the project and the theoretical results will be published in scientific articles. The project includes cooperation in three working groups within the scope of the Centre of Excellence in Estonian Studies with researchers of Baltic literary culture (main partner EKKM2 project EKKM14-336 “History of Baltic literary culture II 2014–2018”), researchers of the history of standard Estonian (main partner PUT475 “Integrated model of the variation of morphosyntax in literary language: pilot study by the example of the Estonian language 2014–2017”) and researchers of runosongs (main partnership programme IUT22-4 “Folklore in the process of cultural communication: ideologies and communities 2014–2019”) as well as with the developers of digital humanities (main partner EKKM2 project EKKM14-327 “Estonian runosong database and academic publications 2014–2018”). An overview of Baltic literary culture will be published in cooperation with partners. International cooperation with Latvian and Finnish researchers is developed in the interests of the project. Analysis of the similar development traits of two literary languages in cooperation with the University in Latvia will continue and the plan is to gather the results in the common publication “Common Steps: from Oral into Written World”. The Institute also cooperates with the Finnish Literature Society within the scope of the project “Letters and Songs: Registers of beliefs and expressions in the Early Modern North”.

2. Extended research of the vocabulary and grammar of historical sublanguages

Extended research is associated with the described PUT, which for a long time has received funding from national programmes, recently from the EKKM2 project EKKM14-316 “Completion of the concordance of Estonian Bible translations 2014–2018”. The vocabulary and grammar of runosongs will be studied in cooperation with the Estonian Literary Museum in addition to the historical sublanguages of the translations of the Bible and hymns. The test corpus to be completed within the scope of the PUT described in the previous point will be

used as the example on the basis of which a broader runosongs corpus for studying the vocabulary and grammar of runosongs will be prepared under the aegis of the Centre of Excellence. Extensive corpora of Bible translations, hymn translations and runosongs will be completed and made accessible to the general public, which in turn will allow for making vocabulary and grammar queries about the three sublanguages. A monograph from the field of language planning thought of the 17th century will be published (2017).

6.1.2. Speech research

The main goal of the field is to use basic speech research (speech styles, sentence prosody, phonological variation) to develop both theoretical and actually functioning models for speech technology applications.

The Institute has been working on speech research for almost fifty years. The present Institutional Research Funding IUT351 “Speech styles, sentence prosody, phonological variation: description, theory and modelling (2015–2020)” is the logical continuation of the target financed research topic “Modelling intermodular phenomena in Estonian (2009–2014)” and ETF research grants “Dynamic aspects of conceptualisation of emotions (2007–2010)” and “Reflection of the rhythm of Estonian speech in grammatical and lexical structures (and *vice versa*) (2009–2012)”.

The four sub-topics of the target financed research topic “Modelling intermodular phenomena in Estonian” were connected with speech research. The research topic proceeded from the hypothesis that an intermodular approach to Estonian would allow to obtain new knowledge about the Estonian language, move forward the general linguistic discussion of language modularity and the connections between modules, develop language technological applications, and solve the present theoretical, descriptive and operational problems. Researching the connections between prosody, language levels and factors outside language focused on the aspects of language and language-related phenomena that are related to prosody: duration, intensity, intonation, rhythm of speech and stress. The research moved from the level of words to the level of sentences and text, covering thereby the connection points of various levels (phonology, vocabulary, morphology, syntax, semantics, pragmatics). The topics that were broadly studied included the time structure of speech and the three-way quantity distinction as guarantors of the rhythm of speech; impact of connections between collocations and duration of speech units on fluctuations in the speech rate; connection between morphology, syntax and word types with prosody. The expression of emotions in the Estonian language was studied in order to recognise the emotionality of speech and written text and the features of text types were analysed.

The object of the research funding IUT35-1 “Speech styles, sentence prosody and phonological variation: description, theory and modelling (2015–2020)” is speech, which is

studied from three aspects: 1) speech styles and their features, 2) prosodic structure and distribution of sentence stresses in an utterance, 3) phonological variation.

Its goal is to fill the gaps in the description of the Estonian language in these areas and broaden the theoretical understanding, particularly of the following issues: the role of expressiveness and (in)formality in speech, factors of phonological variation, prosodic structure and its connections with the distribution of sentence stresses and syntax, and the manifestation of information structure in prosody. The applied goal of the topic is to perform the topical functions of the development of text-speech synthesis, which include modelling speech styles and sentence prosody and solving the problem of phonological variation in the grapheme-phoneme conversion. The research will be carried out with experimental and corpus-based methods using recited and spontaneous speech, recordings of speech styles and material collected experimentally and during field work interviews as the data.

IUT35-1 was successful in its first year (2015). The main activities in the field of speech styles were collection and recording of speech styles and expression tagging. The sources included spontaneous conversations, recited fictional short stories, crime and romantic novels, lectures (from the lecture corpus of the Institute of Cybernetics of the Tallinn University of Technology). Testing principles were developed and an online test environment was created for perception tests. Testing will be used to determine whether and which types of texts can be distinguished by listening. The impact of language and culture on the perception of emotion and the impact of the content of text on the acoustics of speech emotions was ascertained.

In the field of the sentence prosody of the Estonian language, the focus was on the acoustic correlates and distribution of sentence stresses, the expression of information structure and the hypotheses and research questions approaching the features of prosodic phrases. It was found that the emphatic correlates of sentence stress that signal the sentence-final narrow focus are duration and to a lesser extent intensity, whilst the fundamental frequency has a marginal role in comparison with other languages (and contrary to the hypothesis).

A corpus of 52 sentences was prepared for researching phonological variation, which contains 48 words of critical or varying degree of quantity, 50 words beginning with *h* of different frequency of use (*h* in positions that does and does not differentiate meaning) and 48 different (CC) \bar{V} C-type words with *i*-root that end with different consonants. The first recital experiments were carried out. The analysis of the test study based on the data of 11 language managers indicated that *h* at the beginning of the word was pronounced in 95% of the cases during the recital tests, and a correlation between the duration of the *h* and the frequency of the word was found as a result of a detailed analysis of duration: the duration of the *h* in words that used more frequently is shorter.

The basic emotions (joy, sadness, anger) were modelled in the field of speech technology in

parametric speech-text synthesis, relying on earlier results about the acoustic correlates of basic emotions. Test persons were asked to determine emotions in perception tests. Sadness was the most often recognised and joy the least often recognised in synthetic speech.

The special issue “Kõneuurimise suundi. Aspects of Speech Studies“ (ESUKA–JEFUL 2015, no 3) of the “Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics” was prepared on the basis of the outcomes of the first year of the research topic.

The IUT35-1 research team is a partner of the Centre of Excellence in Estonian Studies, and cooperates with the Institutional Research Funding IUT22-4 “Folklore in the process of cultural communication: ideologies and communities 2014–2019” and IUT12-1 “Performative aspects of music 2013-2018”).

6.2. Development of speech technology

The Institute is strong in the development of text-to-speech synthesis and emotion processing.

The following topics have been studied in the field of speech technology:

- interfaces for speech synthesis (2011–2014);
- statistical models of the emotionality of speech and written text (EKT 2011–2014);
- system of spoken subtitles for Estonian television (EKT 2012– 2013);
- audio books generator and Digar audio interface (EKT 2014–2015);
- speech synthesis improvements and extension of applications (2015-2017);
- publication of e-books in multimedia format (2016–2017).

The ongoing IUT35-1 “Speech styles, sentence prosody, phonological variation: description, theory, and modelling (2015–2020)” is also directly connected with the development of speech technology.

6.2.1. Speech synthesis

The Institute has consistently worked on the development of speech synthesis since 1997: we have moved on from diphone synthesis to corpus-based statistical-parametric speech synthesis, rule-based prosody models have been replaced with machine-learning models based on statistical methods, prototypes of speech synthesisers and interfaces for integrating Estonian speech synthesis modules in different voice applications have been developed and various speech synthesis applications (smart phone applications, spoken subtitles, Digar audio interface, audiobook generation and much more) have been created. The theoretical formalisms, speech style definitions, phonological variation research outcomes in grapheme-phoneme conversion and speech prosody test models of the research topic “Speech styles, sentence prosody, phonological variation: description, theory and modelling” are tested and implemented in order to improve the naturalness and credibility of synthesised speech.

The modules of Estonian speech synthesis Mbrola-et, UnitSelection-et, eSpeak-et and HTS-et were created by the end of 2015 and developed further with the support of the “Interfaces for speech synthesis (2011–2014)” and “Perfection of speech synthesis and expansion of uses (2015–2017)” projects of the national programme “Estonian language technology 2011–2017”. The main focus in development work was on synthesis based on the hidden Markov models and the applications created on the basis of this. The Institute plans to test the perfected HTS-synthesis with special modelling of the GlottHMM voice source and to create speech synthesisers on the basis of Estonian dialects.

The everyday users of speech synthesis are visually impaired and dyslexic people. The project “A system of spoken subtitles for Estonian television” was launched in 2012 in cooperation with the Estonian Public Broadcasting and the Estonian Association of the Blind. It is possible to have the synthesiser read out the subtitles on ETV channels since 2013.

(<http://heli.eki.ee/koduleht/index.php/rakendus/>). The voicing interface of the system for presentation of the material of the Estonian National Library (ENA), which makes it possible to listen to magazines and books as audio files in the ENA’s web environment, was prepared in cooperation with the ENA in late 2015.

An audiobook as a public service generator will be completed in 2016, which allows people with special needs (visually impaired and dyslexic people) and others to upload random texts or books to the Institute’s server and download their audio files <http://heliraamat.eki.ee/voxpathuli/>. The Institute plans to develop a system for publication of e-books on the basis of Estonian speech synthesisers from 2016–2017, which will make it possible to add audio and video files to e-books in ePub3 format.

The expression of the basic emotions in parametric speech synthesis has been tested. The plan is to consider various speech styles in reading out different types of texts in speech synthesis and the creation of adaptive HTS-speech models for the expression of basic emotions.

The existing and future software will be distributed on the basis of a licence as freeware via the Centre of Estonian Language Resources.

The ongoing research topic IUT35-1 “Speech styles, sentence prosody, phonological variation: description, theory and modelling (2015–2020)” also contributes directly to the perfection and development of Estonian speech synthesis. The research outcomes of the topic are expected to have been implemented in specific speech synthesis applications by 2019: in a subtitle voicing application, newsreader, web interfaces of dictionaries and an audiobook generator.

6.2.2. Emotion processing

The Institute has been working on speech emotion processing since 2006. The Estonian emotional speech corpus (<http://peeter.eki.ee:5000/>) was created within the scope of the

programme “Language technology support for the Estonian language”. The corpus contains 1234 Estonian sentences. Female voice, 44.1 KHz, 16Bit, Mono. The sentences in the corpus have passed listening and reading tests and on the basis of the outcomes, they were classified as joy, anger, sadness, and neutral; on the valence dimension as negative, positive and neutral, and on the activity dimension as active and passive. The sound and text of the sentences can be downloaded. Three levels have been segmented and tagged: sentences, words, phonemes. The corpus has been documented and is accessible publicly and free of charge. The material of the corpus is used to study the perception of emotions and acoustics, and to train the automatic emotion recogniser.

Emotion recognition from speech and written text was studied from 2011–2014 within the scope of the national language technology programme “Statistical models of the emotionality of speech and written text”.

The prototype of a speaker-specific emotion recogniser, which recognises the positivity-negativity and activity-passivity of speech in real time, was completed (see <https://github.com/EKT1/emotional/>). Since speech emotion processing is becoming of the strengths of the Institute, it is necessary to guarantee that the development continues.

Emotionality in written text is recognised both on the basis of the lexicon and the statistical method.

Recognition accuracy ranges from 60–85% depending on the text (see <http://peeter.eki.ee:5000/valence/> and <https://github.com/EKT1/valence/>). A valence corpus has been created for training statistical models. It contains orthographic paragraphs from newspaper articles whose emotionality (positive, negative, contradictory, and neutral) have been tagged (see <http://peeter.eki.ee:5000/valence/paragraphsquery/>). Different types of text must be added to the valence corpus.

The plan is to make the technological solutions of emotion recognition more convenient for users and to make the applications accessible with cloud service.

The cooperation between the Institute of the Estonian Language, the computational linguists and language technologists of the University of Tartu and the scientists of the Institute of Cybernetics of the Tallinn University of Technology makes it possible to connect the speech synthesis system that connects different speech styles and emotionality of speech and the emotion recognition system created in the Institute with the dialogue systems developed in the University of Tartu and the speech recognition of the Institute of Cybernetics of the Tallinn University of Technology, which creates the precondition for the creation of a fully adaptive human-machine communication systems.

6.3. Lexicography and lexicology

The Institute of the Estonian Language has developed into the only institution in Estonia who

is compiling, collecting and preserving large academic dictionaries, as well as creating and developing the dictionary writing system. All of the work related to dictionaries are done in the Institute's online Dictionary Writing System EELex (see point 6.7), which is developed by lexicographers and language technologists. All of the language resources (dictionaries and databases plus language archives) important for national culture are stored at the Institute.

In 2015, the publishing house Valgus handed the manuscripts of several large (unfinished) bilingual dictionaries to the Institute, which will be included in the dictionary system of the Institute of the Estonian Language as soon as finances become available (planned as one of the activities of the project "Integrated content search of the Centre of Estonian Language Resources").

The dictionaries of the Institute of the Estonian Language are based on lexicological research. They are compiled and edited on the basis of a text corpus (since 2013 also on the corpus collected automatically from the web), which is why they can be considered exhaustive and reliable. The Institute's dictionaries are compiled by experts several of whom have or are studying for a doctoral degree in lexical semantics. The Institute's language ware has been used in research and as an important database for the creation of other language archives.

The language archives of the Institute of the Estonian language contain (as at the start of 2016) approximately 70 dictionaries and databases, approximately 1.5 million words or concept-based terms and approximately 10 million word slips and cards. Users make about 7.8 million queries per year on the Institute's website.

The Institute of the Estonian Language compiles, collects and updates various types of dictionaries:

- dictionaries of contemporary Estonian (e.g. Explanatory Dictionary of the Estonian Language, Dictionary of Standard Estonian, learner's dictionaries for various levels of language proficiency, Dictionary of Foreign Words);
- large bilingual dictionaries (e.g. Russian-Estonian and Estonian-Russian, Finnish-Estonian and Estonian-Finnish dictionaries; the databases taken over from publishing house Valgus in 2015, e.g. Estonian-French, Estonian-Swedish, Estonian-German, will follow);
- dictionaries of Estonian dialects and Finno-Ugric languages (e.g. the large scientific dictionary of Estonian dialects, Estonian Etymological Dictionary, Hiiu dictionary, and Estonian-Udmurt dictionary (published in 2015) and the Eastern dialect dictionary (published in 2016));
- dictionaries of historical sublanguages (e.g. the historical database of Estonian Bible translations).

The compilation and editing of dictionaries and making them accessible online as a public

service belong almost fully among the activities the performance of which is funded by the Ministry of Education and Research. Some dictionary works have received (partial) funding from national programmes: EKKM2 – Estonian Collocations Dictionary; HP – Finno-Ugric dictionaries; TERM2 – terminological dictionaries and databases. The works related to the Dictionary Writing System EELEX have so far mainly been financed via the projects of the EKT programme, but since this can be qualified as dictionary software management, then it has partially been funded by the Ministry of Education and Research as well since 2016.

The lexicographers of the Institute have been using the corpus processing tool Sketch Engine (in cooperation with the Czech company Lexical Computing Ltd.) on the basis of a licence since 2010. The scope of the biggest national corpus EstonianNC is about 560 million text words. The corpus covers the text corpus of the University of Tartu (250 million text words) and the online corpus etTenTen (250 million text words, online texts until 2013). The corpus has been morphologically tagged (ESTMORF, Kaalep 1998). Formalised grammar (Kallas 2013), which is constantly updated, has been created in the Institute for analysing various syntagmatic relationships. Cooperation with the University of Tartu and Lexical Computing Ltd. will continue from 2016–2019 for the expansion of the possibilities of analysing Estonian languages corpora and lexicographic analysis.

The Institute cooperates with the Estonian Language Foundation in the publication of dictionaries.

Work on the following dictionaries will start or continue from 2016–2019 (the planned year of completion is in the brackets).

6.3.1. Estonian language dictionaries

- **Dictionary of the Estonian Language (one volume, 2018/2019)**

The dictionaries follow the type of the main dictionaries of several languages of culture: it contains *ca* 100,000 headwords (printed as one volume) and focuses primarily on the language use in the first decades of the 21st century and the meaning of the words. The meaning that is the most common in modern language is usually given as the first one. Examples of use are given in addition to explanations of meaning (phrases, sample sentences, also compound words – on average, one example per meaning). Information about the origin of the word is provided.

The dictionary is compiled and edited with the corpus processing tool Sketch Engine. Once the dictionary is complete, it will be tied to many other important databases of the Institute via web links, especially the Explanatory Dictionary of Estonian (2009), that describes language use in the 20th century, the Dictionary of Foreign Words

(online since 2015), the Handbook on Standard Estonian and the Dictionary of Standard Estonian (both will be completed in 2018) and The Dictionary of Estonian

Word Families (2012).

- **The Dictionary of Foreign Words**

The database of the Dictionary of Foreign Words (*ca* 40,000 most common foreign words, loanwords, abbreviations, quotes and expressions) belongs to the Institute since 2015.

The online version of the Dictionary of Foreign Words (released in 2015) is based on the eighth edition (2012) of the Dictionary of Foreign Words by Richard Kleis, Johannes Silvet, and Eduard Vääri, which has a long history. The Institute plans constantly update the word-list of the dictionary (in cooperation with the working group of the one-volume explanatory dictionary) and develop the web interface.

- **The Estonian Collocations Dictionary (2018)**

The database of the lexical and grammatical collocations of the Estonian language is being prepared on the basis of Estonian National Corpus. Project started in 2014 and is partly financed from the national programme EKKM2. The dictionary is compiled and edited with the Corpus Query System Sketch Engine. The dictionary will be completed in 2018.

The target group of the dictionary are Estonian language learners (proficiency levels B2 and C1, both as native language or a second language or foreign language). There are *ca* 10,000 headwords, the frequency of the words in corpus and their necessity in everyday life are considered.

- **Database of new words (constant work)**

The database of new words and meanings is connected with all modern language dictionaries created in the Institute via the dictionary writing EELex software, especially with the Dictionary of Standard Estonian, the Dictionary of Foreign Words and the one-volume explanatory dictionary.

- **The Dictionary of Standard Estonian (2018)**

See point 6.4.1 (language planning and care).

- **The Scientific Etymological Dictionary (2019/2020)**

See point 6.6.2 (research of cognate languages and dialects).

- **The Dictionary of Estonian Dialects and small dialectological dictionaries**

See points 6.6.3 and 6.6.4 (research of cognate languages and dialects).

6.3.2. Bilingual dictionaries and databases

The Latvian-Estonian, Estonian-Latvian, and Lithuanian-Estonian dictionaries were published in cooperation with the Institute in 2015.

1) The Latvian-Estonian Dictionary. The Estonian-Latvian Dictionary. The dictionaries were compiled within the scope of the Estonian-Latvian programme “Development of Estonian-Latvian and Latvian-Estonian dictionary 2012–2015”. The cooperation partner was the Latvian Language Agency (Latviešu valodas aģentūra). The dictionaries were published in April 2015 and have been accessible online since that time.

The innovative aspect is that this dictionary was compiled in the bilingual database of EELex that is concept-based, i.e. connected with semantic links, which allowed an increase in the uniformity and consistency of the dictionary and the preparation of both language directions at the same time. The Estonian-Latvian and Latvian-Estonian dictionary contain 40,000 headwords for both language directions.

2) The Estonian-Lithuanian Dictionary. The dictionary is based on the database of the Dictionary of Basic Estonian Vocabulary (2014). The dictionary was compiled at the University of Vilnius from 2013–2015 in the Dictionary Writing System EELex and published in 2015.

Work on the following dictionaries will start or continue from 2016–2019 (the planned year of completion is in the brackets).

- **The Large Estonian-Finnish Dictionary VISU 2018**
The database created on the basis of the former Finnish-Estonian Dictionary (2003) from 2008–2011 is updated on the basis of the material of the dictionaries published in the meantime and handed over to the Finnish and Estonian compilers and editors in May 2016. The 100th anniversaries of both the Republic of Finland and the Republic of Estonia will be celebrated with the completion of the dictionary.
- **The Estonian-Mari Dictionary (2017)**
See point 6.6.1. Research of cognate languages.
- **Estonian-X dictionary templates**
Dictionary templates with Estonian source language (lists of headwords with explanations of meanings) for dictionaries of different volume are developed: 5,000 headwords, 15,000 headwords, 40,000 headwords. The dictionary template can be used outside the Institute on the basis of a licence.
All of the compilers and editors work in the Institute’s online Dictionary Writing System EELex.

6.3.3. Lexicological research

The book about emotion words in the Estonian language written by Ene Vainik, Senior Research Fellow at the Institute, will be published in 2016 and it describes the meaning of

such words, their behaviour in text, ties to everyday knowledge, the historical background of the origin and use of the words. The book demonstrates how the Estonian disposition and its linguistic reflections have developed under the influence of various cultural factors, way of life, social organisation and ideologies.

The tradition of researching colour words is long in the Institute. The collection “Color language and color categorization” (Cambridge Scholars Publishing) will be published in 2016. The focus in the coming years will be on the research of the lexical and figurative meanings of the verbs derived from colour words and comparative analysis (of Estonian, Finnish, and Hungarian).

Research of the associative and collocational relationships of the basic vocabulary will continue. The data is collected from empirical field work, experiments and analysis of text corpora. A description of the basic Estonian vocabulary and its structure will be completed as a result of the theoretical and empirical work (association dictionary).

The polysemy of adjectives is studied by Maria Tuulik (PhD student at the University of Tartu since 2013, external student since 2012, supervisors Margit Langemets (Institute of the Estonian Language) and Renate Pajusalu (University of Tartu)).

The parameters of example sentences for Estonian learners’ dictionaries are studied by Kristina Koppel (PhD student at the University of Tartu since 2015, supervisors Jelena Kallas (Institute of the Estonian Language) and Raili Pool (University of Tartu)).

The application for Institutional Research Funding on vocabulary research is planned for 2018–2023. The main theoretical precondition is the interrelation between the mental lexicon and text units. The practical principles and theoretical starting points used in Estonian lexicography will be combined with new research trends that arise from cognitive, corpus and psycholinguistics.

6.4. Language planning and care

The Institute has been a substantive centre of language planning since its establishment: all standard Estonian dictionaries have been prepared and the problems of literary language development and language planning principles have been approached. The importance of the Institute in language planning is emphasised in the Government of the Republic Regulation No 71 “Rules for the application of the norm of the Estonian standard language” of 9 June 2011.

Language care, i.e. the practical measures for improving language use, is one of the most labour-intensive tasks of language planners. The Institute gives free language advice by phone and online (about 6000 phone queries and over 3000 emails per year; over 7600 pieces of

advice on the website), presents opinions about topical language issues (incl. expert opinions on language to courts), prepares practical tools for language users (incl. collections of articles “Keelenõuanne soovitab” (collection 5, 2015), “Õigekirjaspikker” by Maire Raadik (2016), provides linguistic advice and articles on Language Advice website <http://keeleabi.eki.ee/>, and offers training. Language planners explain their positions at public presentations and in publications (incl. the linguistic magazines *Oma Keel* and *Õiguskeel*). In addition to general language, linguistic advice is also given on terminology and names (place names and personal names). The language planners of the Institute constantly participate in the work of several terminology committees. The Institute has the status of the Office of Onomastic Expertise on the basis of the Names Act and the Place Names Act, i.e. gives expert opinions on the request of those who deal with names. The Advisory Dictionary of Document Language (2013) is constantly updated.

6.4.1. Research of standard Estonian, the Dictionary of Standard Estonian (2018)

The Institute continues compiling and updating the dictionaries that are the basis for the norm of the Estonian standard language. The updated version of the Dictionary of Standard Estonian was published in 2013. The next one should be published in 2018. A public list of new words for the Dictionary of Standard Estonian is kept and it is accessible to everyone on the Language Advice website <http://keeleabi.eki.ee/>. Analysing language planning problems and preparing decisions on the basis of them for the Language Board of the Mother Tongue Society will continue.

The compilation of the Handbook on Standard Estonian, which is financed from the national programme “Estonian Language and Cultural Memory II 2014–2018” started in 2014. The Handbook on Standard Estonian will be a supplement to the Dictionary of Standard Estonian, explains its norms and recommendations, and presents the main rules of standard Estonian. The planned volume of the handbook is that of a monograph and it will consist of three main parts: orthographic rules (discussed with the Language Board of the Mother Tongue Society if necessary), systematised language advice (advice about other issues than orthography, e.g. morphology, word formation, syntax, word selection, style, meanings of words, names, formalisation) and source publications of language planning. It should be possible to link the parts of the handbook to the Dictionary of Standard Estonian to form a whole. The outcome of the project will be a combination of the dictionary and the handbook, which will be easier to update in the future and can be restructured if necessary according to new technological possibilities (e.g. to make it more interactive).

The research part requires more development: it is necessary to study the tendencies of standard Estonian and develop language planning principles on the basis of the newest language planning theories. The Institute aims to develop the relevant research trend. The existing cooperation (especially with Finnish and Swedish language planners) is already

strengthened for this purpose.

6.4.2. Development of the language care system

Giving free advice by phone and mail will remain a duty of the Institute and it will also try to expand the options of giving linguistic assistance by implementing other new technical solutions as quickly as possible. The plan is to develop a uniform language advice system in cooperation with terminologists and names researchers, which will allow them to give advice about terminology and names faster than ever before. Considering user-friendliness is important, which is why attempts are made to avoid fragmentation of these services.

High-quality language advice requires the development of the necessary databases.

6.4.3. Names planning and names research

The Institute has the status of the Office of Onomastic Expertise (Appointment of the Office of Onomastic Expertise, RTL 2004, 87, 1376; RTL 2005, 28, 397), as it is a research institution whose main activities include names planning and which employs research staff that specialise in onomastics.

In the last five years names have been approached in the Target financed research Topic “Word and concept in the development of Estonian vocabulary and place names” (SF 2010–2014) and the project “Compilation of the Estonian Dictionary of Place Names” of the national programme “Estonian Language and Cultural Memory” (will be completed in 2016). The compilation of the “Dictionary of Estonian Surnames” (2016–2018) was launched in 2016. It is planned as an overview of the surnames used in Estonia. The volume is 6,000 articles, which consist of the surname, its pronunciation (quantity distinction and palatalisation), recommendation on declension, frequency of the name in the Population Register; historical data about the surname, incl. the earliest registrations and geographical starting point of the surname (if it only emerged at one location) or initial distribution (if the name emerged at several locations); the main hypotheses about the origin of the surname; the data of the estonianisation or modification of the name; if possible, any lore about the name, and references to sources. A general characterisation of surnames and an overview of current research will be given in the introduction. The initial manuscript of the dictionary of surnames will be completed in 2018. The dictionary is prepared in cooperation with the Võro Institute and the Estonian Literary Museum.

The Institute has the status of the Office of Onomastic Expertise gives expert opinions about names. The Institute has applied its names planning competence in the projects of ministries for streamlining names (e.g. the projects of the Ministry of the Interior “Differentiation of harbour names” (2014) and “Differentiation of the names of public transport stops” (2014), project of the Ministry of Finance “Streamlining of the names of springs, single objects and

bridges” (2015).

Two doctoral students work in the area of names research, more specifically onomastics, at the Institute. The place names of Virumaa are studied by Marit Alas, PhD student at Tallinn University since 2012, supervisors Marja Kallasmaa (Institute of the Estonian Language) and Jüri Viikberg (Institute of the Estonian Language, Tallinn University). The place names of Kesk-Harjumaa are studied by Tiina Laansalu, doctoral student at the University of Tartu since 2014, supervisors Marja Kallasmaa (Institute of the Estonian Language) and Karl Pajusalu (University of Tartu).

6.4.4. EU language planning and EU terminology

The EU language planning specialists (senior language planner and senior terminologist) organise annual language seminars for the Estonian translators and terminologists working at the EU institutions in Brussels and Luxembourg, and also seminars for local translation agencies and freelance translators, editors and officials in Estonia. Support materials for translators are compiled and published regularly (both on paper and online), there is a website for EU language planning, linguistic and terminological advice is given by e-mail.

The EU language planning specialists cooperate with the terminologists of the institutions of the European Union, i.e. have worked as national experts in the Department of the Estonian Language of the European Commission.

The EU language planning specialists have supervised three master’s theses, two about terminology and one about plain language and have given lectures about the EU language usage and plain language at Tallinn and Tartu University.

6.4.5. Promoting plain language

The EU language planning specialists have been promoting the principles of plain language in Estonia since 2010, incl. articles, lectures, radio and TV appearances, the website selgesonum.ee and Facebook.

The competition for the Clear Message Prize was launched in 2014, where entries are submitted by private institutions and state agencies every year.

The Institute has participated in the international project IC Clear (2011–2014), whose goal was to create the curriculum for an adviser of clear communication, and it also participates in follow-up projects.

The plain language study module was prepared in 2016 and it will be offered as a seminar and various forms of cooperation to private institutions and state agencies. The EU language planning specialists have applied for and received money for plain language activities from the programmes of the European Commission; however, permanent funding from the state would be necessary for continuing and expanding the activity.

6.4.6. Training

As a part of language care, the Institute of the Estonian Language organises free information days for teachers of general education schools, incl. annual conferences for Estonian language teachers in Tallinn and Tartu, where important topics (orthography, morphological formation, phrasing, style, vocabulary) are discussed.

The Institute also organises free training for interested people, giving them the necessary general knowledge of good language use (orthography, syntax, phrasing, choice of words, etc.) or preparing special courses that proceed from the needs of the institutions that order the training. Training has been more constantly organised since 2004 and there are four training events on average every year. Topical courses (orthography and phrasing) have been organised since 2007.

6.4.7. Cooperation with other institutions

The Language Board of the Mother Tongue Society (which decides on basic language planning issues) and the Centre of Language Care in Tartu (which organises training for state and local government officials and prepares information materials) also deal with the field of language planning and care in Estonia. The representatives of the Institute of the Estonian Language are members of the Language Board; Maire Raadik has been the head of the Board since 2016; the Institute of the Estonian Language is one of the founders of the Centre for Language Care <http://keelehooldekeskus.eki.ee/> and coordinates its work. The Institute also cooperates with universities and the Language Inspectorate (inspects compliance with the Language Act, incl. adherence with the norms of standard Estonian in official language use).

Language planning is one of the central and constant tasks of the Institute, which is why it is mainly financed by the Ministry of Education and Research. The funding from the Ministry of Education and Research is used for the following activities (since 2008): language advice; language care in Tartu; EU language planning; dictionaries of standard Estonian.

6.5. Terminology work

Terminology work is applied work in its nature and some of it is carried out on a project basis. The Institute has been working on extensive follow-up projects in recent years, which cover military, security, and defence policy terminology, education terminology and aviation terminology.

Updating the terminology database Esterm has so far been done by individual terminological entries or concept groups, or by streamlining terminology entries by subjects fields in cooperation with experts and subject fields specialists. Three terminologists have been

consistently streamlining and updating Esterm since 2015.

The terminologists, who are mainly involved in long-term projects, advise new terminology committees and working groups, help them in creating the structure of a terminology database and organising the work of the terminology committee, and attend the meetings of the committees as advisers. The terminologists have also participated in training the working groups funded via the national programme “Estonian Language Terminology 2013-2017” and also organised training for other people interested in terminology work.

The terminologists have explained the principles and methods of terminology work at the appearances and in the publications aimed at the general public or narrower circles of interested persons.

The terminologists have participated in cooperation projects for updating the streamlining EU’s inter-institutional terminology database IATE, e.g. the terminologists have worked by the European Commission as experts. Terminology work is highly international, e.g. the Institute belongs to the European Association for Terminology (EAFT) and a terminologist has participated in the work of the NATO Military Committee Terminology Board (MCTB) for several years. Thus, the terminologists have gained international experience and knowledge about the development trends of theoretical thought.

The working groups or committees established by several agencies, organisations or research institutions in Estonia, which usually only include a language specialist in their work, deal with terminology. The know-how about the structure of a terminology database and its use, the organisation of terminology work, compilation of terminology entries and research, which proceeds from the concept-based approach, is mainly used by the terminology projects coordinated by the Terminology Department.

Terminologists work hand in hand with language planners and both of them participate in the work of many terminology committees.

Terminology work has received funding from the following projects in the last five years:

- project for military, security, and defence policy terminology (Ministry of Defence, since 2006);
- project for aviation terminology (Ministry of Economic Affairs and Communications, since 2008);
- project for education terminology (funded by the Ministry of Education and Research since 2008);
- project for streamlining and updating of the terminology database Esterm (funded by the Ministry of Education and Research since 2008);
- project for integration terminology (Ministry of Culture, in 2013).

The Terminology Department has cooperated with the following committees and institutions in the last five years:

- the working group of archaeological terminology of the Institute of History and Archaeology of the University of Tartu (2014–2016);
- geological terminology committee of the Tallinn University of Technology (2010–2012) and since 2016 of Geological Society of Estonia;
- working group of the Estonian Society of Music Therapy (2015);
- the project for the e-health terminology database of the Estonian eHealth Foundation (2013);
- the terminology project of the Chair of Activity Therapy of the Tallinn Health Care College (2013);
- the expert group for streamlining the vocabulary of behavioural economics (2013);
- working group of medical physics terminology (2014);
- Accounting Terminology Committee of the Ministry of Finance (2015);
- terminology project of the Estonian Disc Golf Association (2015);
- working group of deep neural networks (2015);
- training project of the Language Centre of the Estonian University of Life Sciences (2015);
- terminology project of the Estonian Union of Organisations Involving Assistance and Therapy Animals (2015-2016).

6.6. Research of cognate languages and dialects

The dialects of the Estonian language have been researched at the Institute since its establishment and Finno-Ugric languages have been researched since 1957. Decades of research have laid the foundation for the publication of works important for the Estonian language and making the material of disappearing languages via dictionaries and grammar books in order to better understand the development of the Estonian language and its position in the Baltic language environment.

The research of cognate languages and dialects has received funding via the following topics and projects in the last five years:

- “Word and concept in the development of Estonian vocabulary and place names” (SF 2010–2014);
- compilation of the Dictionary of Estonian Dialects; preparation of the scientific etymological dictionary; preparation of small regional dialectological dictionaries (funded by the Ministry of Education and Research);
- research of Finno-Ugric languages and publication of the Estonian-Finnish Dictionary (funded by the Ministry of Education and Research);
- development of the field of Finno-Ugric languages (funded by the Ministry of Education and Research);
- compilation of the Estonian-Mari Dictionary (funded by the HP);

- compilation of the grammar of Estonian dialects (EKKM2).

6.6.1. Research of cognate languages

The Institute holds a special position in the research of cognate languages, because Finno-Ugric languages are no longer taught or researched at Tallinn University. The research of cognate languages in Tartu focuses on phonetics and to a lesser extent on phonology.

The Institute's goal is to develop a centre for the vocabulary of Finno-Ugric languages whose tasks will include supporting the compilation of bilingual dictionaries of cognate languages, whilst one of the languages must be Estonian.

- The **Estonian-Udmurt Dictionary (2015)** was published. It is only accessible electronically at <http://portaal.eki.ee/dict/eud/>. The dictionary contains approximately 10,000 headwords. The dictionary is based on the Estonian-X dictionary template developed in the Institute of the Estonian Language, which also covers the whole basic vocabulary of Estonian, i.e. 5,000 of the most frequently used words (based on the Dictionary of Basic Estonian Vocabulary completed in 2014). The Explanatory Dictionary of the Estonian Language is the main source of examples. The compilation of the dictionary was supported by the “Kindred People Programme IV”, the Institute of the Estonian Language and the University of Tartu.
- The **Estonian-Mari Dictionary** (expected publication in 2017) is being compiled. The dictionary contains approximately 7,000 headwords (incl. basic vocabulary) with examples of use. The Mari language is one of the largest Finno-Ugric languages in Russia. The goal is to compile a dictionary of basic vocabulary at the contemporary scientific level, which will be freely accessible to everyone on the internet.
- The **electronic database of the Votic language** is being created. The one-volume Dictionary of the Votic Languages, which is a corrected edition of the previously published fascicules I–VII, was published in 2013. The online dictionary of the Votic language, which is based on the one-volume Dictionary of the Votic Language published in 2013, is being compiled.
- The new **Scientific Etymological Dictionary**, which has been under compilation since 2012, should be published in 2019 (see 6.6.2).
- The vocabulary of Finno-Ugric origin in the Estonian language is also studied in addition to the compilation of the etymological dictionary.
- An international seminar or conference about a narrower field of Finno-Ugric linguistics is organised every November or December.
- If possible, the Institute's representatives participate in international projects (e.g. the atlas of European languages “Atlas Linguarum Europae” or the ALE).
- The **electronic database of the Livonian language** has been compiled. Editing the

material of the German-Livonian Dictionary compiled by Sjögren and Wiedemann and entering it in the database was completed in 2015.

The digitisation of the sound archive of Finno-Ugric languages has been completed and displaying the material in an online environment has started (see point 6.8.1).

6.6.2. The Scientific Etymological Dictionary

The Etymological Dictionary as a dictionary that reflects that the origin of a language's vocabulary is an essential part of the lexicographic repository of every language of culture. The Etymological Dictionary of the Estonian Language published in 2012 is a kind of work that has never been published in Estonian before. It reflects the origin of the root words in the Estonian language and gives all the counterparts in cognate languages and the sources of loans in the case of loanwords. The dictionary differs from earlier etymological dictionary of the Estonian language compiled in German by Julius Mägiste, which was not completed, in terms of its compactness and ease of use. The foreign words borrowed into the Estonian language in the 20th century, the origin of which is given in the Dictionary of Foreign Words (2012), are not included in the dictionary.

The Etymological Dictionary was added to the EELEX environment in 2010. After the publication of the one-volume Etymological Dictionary of the Estonian Language in 2012, the work continued with the compilation of a bigger academic edition.

Foreign words and dialect words have been added to the word-list of the **Scientific Etymological Dictionary**. Headwords are approached in greater detail in the dictionary by adding dialectical data, derivations, counterparts in foreign languages and bibliographic references to the origin.

6.6.3. Research of dialects and compilation of the dialectological dictionary

The Institute has researched dialects since its establishment. The research of dialects today is associated with the compilation and editing of the Dictionary of Estonian Dialects. This dictionary is an academic work that covers the entire Estonian dialectic vocabulary. The compilation of the dictionary started in 1964 and the first fascicule was published in 1994. The compiled parts have by now been digitised and the dictionary has been compiled in the Institute's Dictionary Writing System EELEX since 2012 (see 6.7). Words from *lõpetis* to *mütsakas* can be used in the online dictionary as at 1 January 2016. The earlier fascicules (4–26) can be downloaded as PDF files (<http://www.eki.ee/dict/ems/ems.html>), work on converting them to the Dictionary Writing System EELEX continues. The material of fascicules 1–3 has been manually input in the dictionary software due to the absence of the relevant files and the plan is to edit the material and make it usable in the online dictionary with the material of the other earlier fascicules.

The dictionary of dialects is financed within the scope of the support of the Ministry of Education and Research, because it is time consuming and is extremely important for Estonian culture, the financing of which on the basis of competition would not be appropriate.

6.6.4. Compilation of small regional dialectological dictionaries

Dictionaries of dialects with small numbers of people are compiled in cooperation with other institutions (the Estonian Language Foundation, the University of Tartu, regional institutions or institutes, such as the Institute of Kihnu Culture, the Institute of Mulgi Culture, the Seto Institute, etc.).

The following has been published:

- 1) The Dictionary of Old Võru Dialect (2011);
- 2) The Mulgi Dictionary (2013);
- 3) The Hiiu Dictionary (2015);
- 4) The Eastern Dialect Dictionary (2016);

Seto, Kihnu, Mulgi and Saaremaa dictionaries are being compiled.

6.6.5. Compilation of the grammar of Estonian dialects

The grammar of Estonian dialects is being compiled. The goal is to create a comprehensive description of the phonetics and grammatical structure of the Northern and Southern Estonian dialects. The grammatical overviews of individual dialects produced so far (e.g. dialects of the north-eastern coast, central dialect, Võru dialect, Karksi dialect, plus forewords to volumes of texts in dialects) are taken into account. The monograph will provide a general picture that can be used as the basis for comparative analyses of the common part and differences of Estonian dialects and the dialectic content of the grammar of standard Estonian. The dialectic background that can be seen in the public use of the Estonian language will become clearer and it can also be used in making language planning decisions.

6.7. Language technology

The Institute of the Estonian Language develops dictionary and terminology management software on scientific basis and offers them for public use free of charge. The compilation of new and existing dictionaries and terminology databases on scientific basis is supported. Language resources are standardised and concentrated in databases, and public solutions for their public use are created.

The Institute is one of the three partners in the consortium of the Centre of Estonian Language Resources, which makes the digital resources and technologies (language software and language databases: dictionaries, text and speech corpora) accessible to scientists and other people interested in the Estonian language. The Centre of Estonian Language Resources is

one of the sites of the Estonian research infrastructure and its activities are covered by the project “Federated Content Search for Centre of Estonian Language Resources” (2016–2020) financed from the measure “Increasing the international competitiveness of Estonian R&D and participation in pan-European research initiatives” under the activity “Supporting research infrastructure of national importance on the basis of a roadmap”, project EKT69 “Centre of Estonian Language Resources” (2015–2017) of the national programme “Estonian language technology 2011–2017” and the institutional research support IUT20-65 “Centre of Estonian Language Resources” (2016–2020).

The work trend of electronic lexicography emerged at the Institute 40 years ago when inputting dictionaries in the computer, their use in linguistic research and publication with computers was initiated. The development of the online lexicographer’s working environment EELEX started in 2005. It consists of two components: the software or the Dictionary Writing System and lexical language resources or the dictionary databases.

The software and databases have been structured according to international standards, meaning that it is possible to use them outside the Institute and integrate the language resources and software of other language technology projects. Our main advantages over the other similar dictionary software created elsewhere is the existence of the applications required for the Estonian language, flexibility in respect of different types of dictionaries, a large selection of editing tools and usability as freeware.

The development of the Dictionary Writing System EELEX will continue (within the scope of permanent funding), the terminology management software Termeki (<http://term.eki.ee/>) and the Language Hotline e-keelenõu have been created. The new modularised dictionary and terminology management software EELEX2 will be developed if possible, which will allow to offer the persons and databases engaged in Estonian dictionary and terminology creation a single and modern service, and reduce the development and administrative burden. The Institute plans to finance the development of EELEX2 from the institutional development programmes of research and development institutions and institutions of higher education (ASTRA; project application submitted in 2015).

The lexicographer’s working environment can be used for the compilation, editing and layout of dictionaries, and supports group work. The EELEX has developed in close cooperation with the Institute’s lexicographers and is used as an internal and public tool. The majority of the dictionaries previously compiled at the Institute have been converted to XML and transferred to the EELEX environment. Approximately 60 dictionaries and databases have been or are being prepared in the management software as of 2015. Online versions are created of the dictionaries compiled in EELEX: the Dictionary of Standard Estonian (ÕS 2013), the Explanatory Dictionary of the Estonian Language (2009), the Dictionary of Foreign Words (2015), the Dictionary of Basic Estonian Vocabulary (2014) and the Estonian Etymological

Dictionary (2012) are freely in use.

The online versions usually become accessible a year after the publication of the paper dictionary.

The existing dictionaries are combined into a single knowledge base by linking the lemmas of different dictionaries at the level of meaning and morphology within the scope of project EKT61 “Tools of lexical resources” of the national programme “Estonian language technology 2011–2017”. The principles that make it possible to integrate data in the aforementioned knowledge base are implemented upon the compilation of new dictionaries and databases.

The rule-based morphology software of the Institute of the Estonian Language, which enables semi-automatic generation of morphological information about lemmas (main forms, inflection type, word class), has been integrated with EELex;

- the editor of the dictionary considers the full paradigm of the lemmas (all possible inflected forms);
- show all of inflected forms of words in the online version of the dictionary.

The Institute plans to continue developing the prototypes of morphology interfaces to make them reliable and easy to use.

The research outcomes of the targeted topic “Modelling intermodular phenomena in Estonian” will be used upon describing the grammatical and semantic qualities of words in order to improve the substantive quality of dictionaries.

The development and modernisation of the online versions of the dictionaries compiled in the Institute of the Estonian Language and with the Institute’s help will continue. Special attention will be given to the development of the Institute’s website.

The following projects have been financed in the field of electronic lexicography in the last five years:

- “Modification of the Lexicographer’s Workbench” (EKT 2011–2012);
- “Management and development of the electronic dictionary writing system” (funded by the Ministry of Education and Research);
- “Modelling intermodular phenomena in Estonian” (SF 2009–2014);
- development of the terminology management software Termeki within the scope of the project “Centre of Estonian Language Resources” (2012–2015);
- “Language Hotline e-keelenõu” (2013);
- “Tools of lexical resources” (2014-2017).

6.8. Language archives of the Institute of the Estonian Language

The Institute has rich language archives that support research.

I. Paper archives

Archive of Estonian Dialects	5,580,100 slips
Archive of Standard Estonian	4,300,000 slips
Archive of Finno-Ugric Languages	281,500 slips
Archive of Terminology	550,000 slips
Archive of Etymological Data	50,000 cards
Total: Slip and card archives	10,761,600 slips
Manuscript collections	153,000 pages

II. Sound recordings

Recordings of Estonian dialects	2,900 hours
Recordings of cognate languages	1,140 hours
Recordings of Standard Estonian	55 hours
Recordings from the Estonian diaspora	280 hours
Total	4,375 hours

III. Electronic databases (mostly publicly available at portaal.eki.ee)

Database of the Dialectological Dictionary of Estonian	73,400 headwords
Text corpus	10,400,000 tokens
Database of New Words	10,000 entries
Language Advice Database, incl. advice by e-mail	154,000 14,000 entries
Place Names Database	141,000 44,000 entries
Personal Names Database	138,000 surnames 52,000 first names
Database of the Archive of Etymological Data	50,000 entries
Estonian Emotional Speech Corpus	1,700 sentences
Concordance of Estonian Bible Translations	87,310 verses
Database of the Place Names Archive	600,000 slips
Database of the Archive of Conceptual Lexis	165,000 slips

6.8.1. Archive of the Estonian Dialects and Finno-Ugric Languages

The greatest value of the Archive of the Estonian Dialects and Finno-Ugric Languages at the global level are the slip and card archives, manuscript collections and sound recordings of Estonian dialects and Finno-Ugric languages. Their value-in-use and cognitive motivation

will only increase over time.

- The language material collected mostly from 1922–1993 represents the old Estonian dialects of the later 19th and early 20th century, which no longer exists in this form and cannot be retroactively collected. The same applies to the language material of Finno-Ugric languages (Votic, Ingrian, Ingrian Finnish, *et al.*) collected in the 20th century.
- The Archive of Place Names (over 0.6 million slips) is the base material for modern names planning and cartography as well as onomastics.
- The archive of the vocabulary of Estonian dialects and the archive copied and systematised on the basis of dialectological dictionaries (2.8 million slips) are the base material for the dictionary to be compiled. As the archive contains considerably more material than the Dictionary of Estonian Dialects, then the value-in-use of the archive will not decrease after the completion of the dialectological dictionary.
- The sound recordings of Estonian dialects and Finno-Ugric languages collected over half a century are examples of past language use, which can no longer be heard today.
- Language material must be recorded using modern technical equipment in order to preserve information and keep it usable.

The situation of the EMSUKA as a whole has improved. Storage conditions (storage rooms that meet requirements) as well as the digitisation and storage equipment (efficient scanners, a photocopying system and a large server have been acquired) have improved with the support of the projects “Construction of stage II of the research archive of the Institute of the Estonian Language” (TAP 41–1) and “Improvement and development of the conditions for scientific use of the research archive and collections and archived materials of the Institute of the Estonian Language” (TAP 54–1). The online database <http://emsuka.eki.ee/> that connects all the archives of the EMSUKA to the search system has been completed.

The meta data of manuscript collections and sound recordings series have been entered in the online database and the addition to digital files has started. The addition of the meta data and files of the slip and card archives has started. The digitisation of slip and card archives and manuscript collections and the addition of files in the database and the server will continue.

The material from the EMSUKA sound archive and manuscript collections have also been added to the dialect corpus of the University of Tartu. The material of the EMSUKA has been used by the Estonian National Museum and smaller museums in their expositions. Dialectological dictionaries and chrestomathies have been published in cooperation with other research institutions and village societies.

6.8.2. Archive of Standard Estonian

Updating the Archive of Standard Estonian started in 2000 (4.3 million slips) and traditional language archives are represented by the “Explanatory Dictionary of the Estonian Language”

(2009). The examples of its use can be regarded as a kind of text corpus. Updating the combined index of terminology ended in 2003 (when 550,000 slips had been added). The activity was continued electronically.

In principle, a slip index cannot become redundant until the material it contains had been made electronic in full.

The Institute continues with the preservation and streamlining of archives. The main activities are:

- digitisation of slip indexes and manuscript collections;
- re-recording sound recordings to the server;
- development of electronic databases. The Institute has three databases: EMSUKA global database (<http://emsuka.eki.ee>), database of the conceptual archive (<http://heli.eki.ee/moisteline/>), database of the place names archive (<http://heli.eki.ee/kohanimed/>).

The following activities are planned in terms of archives (until 2019):

- to continue digitalising and streamlining the existing material and adding it to the database, and headword labelling the material added to the database of slip and card archives;
- to continue developing and perfecting the web interface of the EMSUKA global database(<http://emsuka.eki.ee>);
- to continue headword labelling the place names archive (<http://heli.eki.ee/kohanimed/>);
- to start headword labelling the slips in the conceptual archive <http://heli.eki.ee/moisteline/>;
- to start headword labelling the Wiedemann archive in EMSUKA to make it usable via the general database;
- to create the English page of the database with the relevant meta data (the activity is necessary for promoting the international use of the archived material);
- to train additional staff in order to speed up the work related to archives;
- to increase cooperation with other institutions (public broadcasting, museums, universities, local governments, etc.).

7. Strengths, weaknesses, opportunities and threats

A SWOT analysis was prepared in early 2016 to assess the status of the Institute, which brought out the strengths and weaknesses of the organisation and the opportunities and threats arising from the external environment. The main achievements of the Institute in research and development and in the area of developing the organisation are given at the end of the chapter.

Strengths of the organisation

- The Institute of the Estonian Language is a centre of research and development of the Estonian language with long traditions. The Institute will celebrate its 70th year of operations in 2017. Operating as an independent institution for a long time has made the Institute of the Estonian Language known nationwide as the centre of the Estonian language;
- The Institute of the Estonian Language performs the public functions commissioned from the Institute by different ministries: the Ministry of Education and Research, the Ministry of Defence, the Ministry of Finance, and the Ministry of the Interior;
- the Institute of the Estonian Language has a professional and diverse staff that has both development and research competency. The Institute has 63 employees (as of 1 April 2016), 16 of whom have doctoral degrees. Three people have defended their doctoral degrees in the last two years. The Institute has six doctoral students as of 1 April 2016;
- The Institute of the Estonian Language studies speech and develops speech technology, and **plays a central role in the development of speech synthesis in Estonia**. Analysis and recognition of emotional speech have also emerged in the last decade. The infrastructure required for speech processing exists;
- The Institute of the Estonian Language belongs to the consortium of the Centre of Estonian Language Resources, which makes the digital resources and technologies accessible. The main resources of the Institute are (online) dictionaries, terminology databases, the software for their compilation and management, voice processing software and speech corpora (e.g. speech synthesis software, the Estonian Emotional Speech Corpus);
- The specialists working at the Institute of the Estonian Language can enrich the study programmes of universities with special courses related to the unique development activities of the Institute (e.g. special courses in lexicography, onomastics, language planning, EU language, plain language).
- The Institute of the Estonian Language organises systematic in-service training for adults, incl. training in the areas of language care, names planning, and terminology;
- supervising doctoral and master's degree students at universities;
- several institutions that deal with the Estonian language and culture have their offices in the building of the Institute of the Estonian Language on Roosikrantsi street: the Estonian Language Foundation, the Mother Tongue Society, Estonian Association of Applied Linguistics, the Estonian Publishers Association, the magazines *Keel ja Kirjandus*, *Linguistica Uralica* and *Täheke*, the Under and Tuglas Literature Centre of the Estonian Academy of Sciences;
- improvement of the reputation and awareness of the Institute: spontaneous awareness

of the Institute has increased from 18% (2010) to 33% (2015) and aided awareness from 62% (2010) to 75% (2015). 15% of the respondents had come in contact with the Institute and its services (6% in 2010). The Institute is considered necessary (81%), professional (67%), and academic (58%), opinions of the institution have improved considerably in comparison with 2010;

- the Institute's Language Hotline e-keelenõu and free language advice won the main award at the Estonian Language Deed competition in 2015;
- the Institute is strong in international cooperation and the organisation of international conferences.

Weaknesses of the organisation

- The uncertainty of project-based financing. Applying for projects every year is time-consuming and the results have recently been unpredictable: even excellent applications for financing are rejected, which makes budgeting difficult. Many specialists, including people with doctoral degrees, have had to leave the Institute because financing has not been received for some projects;
- there is no next generation of research staff and top specialists in some areas of research;
- low mobility of researchers;
- the durability of the outcomes of completed IT projects (maintenance and development work) is not guaranteed due to the lack of funds;
- the durability of the project-based outcomes of language technology (maintenance and development work) is not guaranteed due to the lack of funds;
- the limited funds mean that the Institute cannot hire young specialists (masters, doctors), even though there are people among them who would be very interested in working at the Institute;
- there are no funds for the establishment of a modern online environment that would allow us to provide public services better.

Opportunities for the organisation

- The initiatives of new development activities, e.g. the development of the Estonian centre of applied linguistics at the Institute of the Estonian Language; one of its main activities would be the development of Estonian language proficiency tests (in cooperation with the Ministry of Education and Research and Foundation Innove);
- opportunities for smart specialisation: contacts with entrepreneurs have been established;
- strengthening the Institute's image building and improving its status;
- organisation of more training events and conferences to guarantee an increase in its

own revenue.

Threats of the organisation

- Dependence on possible administrative reforms. The impact of commercial considerations on the management of research becomes too strong;
- the small share of the institutional base funding threatens the sustainability of research;
- the size of the budget will be determined by April or May in the current year;
- excessive centralisation of human resources work.

8. Strategic goals, activities and indicators, and new initiatives

The Institute has set itself three main goals as a result of its mission.

Goal 1. Research at the Institute of the Estonian Language is excellent and sustainable

Research

- Strengthening international competitiveness and quality via participation in the activities of the Centre of Excellence in Estonian Studies (2016–2023);
- implementation of the institutional research funding IUT35-1 “Speech styles, sentence prosody, phonological variation: description, theory, and modelling” (2015–2020);
- implementation of the EKT84 project “Perfection of speech synthesis and expansion of uses (2015–2017)”;
- connecting language technology development closely with the research outcomes of IUT35-1 (2016–2017);
- implementation of the personal research topic PUT550 “The cultural shift in Estonian in the 17th and 18th century” (2014–2017);
- implementation of the EKT61 project “Tools of lexical resources” (2014–2017);
- development of research infrastructure in CLARIN;
- development of the standards of e-lexicology in European cooperation networks (within the scope of COST);
- improving the efficiency of post-graduate studies and participation in the work of the graduate school.

Indicators

- Articles about the historical development mechanisms of the Estonian language have been published;
- the monograph about the language planning thought of the 17th century has been

- published (2017);
- the project of Concordance of Estonian Bible Translations has been completed (2018);
- the historical database of Lutheran hymn translations has been created (2018);
- the principles for compilation of the Estonian runosongs dictionary have been developed (2018);
- articles about the different aspects of modern Estonian have been published;
- the collection “Color language and color categorization” (Cambridge Scholars Publishing) has been published (2016);
- the monograph “Estonian emotions. Word Portraits” has been published (2016);
- the special issue “Kõneuurimise suundi. Aspects of Speech Studies” of the magazine “Eesti ja soome-ugri keeleteaduse ajakiri. Journal of Estonian and Finno-Ugric Linguistics” has been published (2016);
- articles about speech research have been published;
- the satellite conference Aspects of Speech Research has been organised within the scope of the conference of applied linguistics (2017);
- the collection of articles about speech styles, sentence prosody, and phonological variations has been published (2019);
- the word assessment tool of EELex has been created, the headwords of different dictionaries have been linked at the level of meaning and morphology (2017);
- some doctoral students have defended their theses (2018);
- the annual seminars of the graduate school have been organised (at least one per year).

Goal 2. Development at the Institute of the Estonian Language is excellent and sustainable Development

- Participation in the consortium of the Centre of Estonian Language Resources and its work;
- Implementation of the EKKM2 projects: “The Handbook on Standard Estonian (2014–2018)”, “The Estonian Collocations Dictionary (2014–2018)”, “The Grammar of Estonian Dialects (2014–2017)”, “Completion of the Concordance of Estonian Bible Translations (2014–2017);
- working on language advice and care;
- working on names planning;
- organisation of terminology work;
- organisation of EU language care;
- promoting plain language;
- compilation and editing of dictionaries of Estonian and other languages;
- compilation and editing of dictionaries and databases of Finno-Ugric languages and Estonian dialects;

- compilation of the grammar of Estonian dialects;
- compilation and development of language information (e.g. morphology) databases;
- development of electronic dictionary writing system;
- development of the Institute's web and online dictionaries, incl. the Language Hotline e-keelenõu;
- development of speech technology and creation of speech technology applications;
- creation and development of language and speech corpora;
- development and streamlining of the archives and collections of the Institute of the Estonian Language;
- streamlining, updating, and development of the Esterm terminology base;
- advising new terminology projects and commissions about the organisation of terminology work and application of its principles;
- training people involved and interested in terminology work;
- project-based work with the terminology of three fields – military, security and defence policy, education, and aviation – will continue;
- making the website that showcases terminology work accessible;
- organisation of communication of information and in-service training.

Indicators

- The new Dictionary and Terminology Management Software EELex2 has been developed within the scope of the project of the institutional development programme of research and development institutions and institutions of higher education (ASTRA) (2020);
- the Dictionary of Standard Estonian has been published (2018);
- the Handbook on Standard Estonian has been published (2018);
- the Estonian Dictionary of Place Names has been published (2016);
- the main part of the Estonian Dictionary of Surnames has been compiled (2018);
- responses to at least 7,500 language advice requests per year are guaranteed;
- the word-list of the Dictionary of Foreign Words has been updated;
- the word-list of the Advisory Dictionary of Document Language has been updated;
- the (one-volume) Dictionary of the Estonian Language has been compiled and edited (will be completed in 2018, expected publication in 2019);
- the Estonian Collocations Dictionary has been completed (2018);
- the Great Estonian-Finnish Dictionary has been published (2017/2018);
- the Dictionary of Hobby Education has been published (2016);
- the Concordance of Estonian Bible Translations has been compiled and updated (2018);
- the Grammar of Estonian Dialects has been compiled (2017);

- the publication of at least one fascicule of the Dictionary of Estonian Dialects per year is guaranteed, the 29th fascicule has been published (2018);
- the academic Dictionary of Estonian Dialects is accessible online (fascicules 1–29 as at 2018);
- the main part of the Scientific Etymological Dictionary has been compiled (initial manuscript completed in 2019);
- small dialectological dictionaries have been published;
- small dictionaries of Finno-Ugric languages have been published;
- databases of Finno-Ugric languages have been compiled and made publicly usable;
- the web interface of the global database Archive of the Estonian Dialects and Finno-Ugric Languages (EMSUKA) has been developed and perfected;
- the place names index has been digitised and equipped with headwords to the extent of 50% (2016);
- the Wiedemann slip and card archive has been digitised and uploaded on the web (2016);
- the annual Finno-Ugric seminar or conference has been organised;
- two information conferences for teachers of the Estonian language have been held every year;
- two training events for European Union officials and translators have been held every year;
- a seminar for Estonian freelance translators and officials has been held;
- EU language collections or other supporting material for translators have been published;
- the clear message competition has been organised (once a year);
- language care activities have been coordinated with the Centre of Language Care in Tartu and the contact network of editors has been constantly updated;
- text-speech synthesis on the basis of the hidden Markov model has been perfected (2017);
- the algorithms adapted to the Estonian language for the generation of synthetic voices have been developed on the basis of corpora (2017);
- the speech synthesiser of Võru dialect for male and female voices has been created (2017);
- the models of speech styles, prosodic prominence, and phonological variability have been tested and applied in text-speech synthesis (2019);
- the selection of emotions in the corpus of emotions has been expanded, the environment of emotion perception tests has been updated (2016);
- the initial versions of the identifier of the emotionality of speech have been created (2016);
- the lexicon-based identifier of the emotionality of written text and machine-learning

versions have been created (2016).

Goal 3. Popularisation of the research and development of the Institute of the Estonian Language

- Organisation of open house days at the Institute;
- preparation of the audiovisual material that showcases the activity of the Institute;
- cooperation agreement with Public Limited Company Postimees Grupp. The agreement covers articles about language in the weekly *Postimees*, in-service training for the staff of *Postimees*, publishing, creation of a text corpus with a scientific purpose and language technology solutions;
- cooperation projects with Estonian Public Broadcasting (TV clips etc.);
- working on the development of value, incl. increasing awareness (spontaneous awareness to 40% and aided awareness to 80%).

New initiatives

- Semi-automatic generation of the databases that describe the paradigmatic relationships of words (synonyms, antonyms, etc.);
- compilation of a new type of dictionary of the Estonian language (association dictionary, *ca* 800 headwords);
- voicing of the unadapted foreign words and expressions in the Dictionary of Foreign Words;
- compilation of the school version of the Dictionary of Foreign Words;
- compilation of a Learner's Dictionary of the Estonian Language for the C1 language proficiency level;
- creation of a database of bilingual dictionaries and development of the respective environment;
- compilation of the great Estonian-Russian and Russian-Estonian dictionary;
- compilation of small (*ca* 5,000 headwords) dictionaries of Finno-Ugric languages (in cooperation with the Kindred People's Programme);
- compilation of the Estonian runic folk song dictionary (in cooperation with the Estonian Literary Museum);
- development of the Estonian language proficiency examinations (in cooperation with Foundation Innove);
- development of terminology advice;
- development of plain language activities: organisation of the clear message award competition every year; further development of the plain language study module; organisation of plain language seminars for officials;
- creation of a modern language learning environment based on the resources of the

Institute of the Estonian Language;

- development of the new Dictionary and Terminology Management Software EELEX2;
- provision of the service of dictionary standardisation;
- perfecting the modules of the dictionary and terminology systems EELEX and EELEX2 (incl. the morphology module);
- development of the database and environment of the small regional dialectological dictionaries;
- creation of voice synthesisers of Estonian dialects;
- compilation of dictionaries of social sublanguages (e.g. dictionaries of slang);
- adaptation of the Good Dictionary Example (GDEX) application to the Estonian language in order to find good examples of use from the text corpus for different types of dictionaries and the creation of automatically generated lexicographic applications (in cooperation with the developers of Sketch Engine);
- development of the software for identification of the emotionality of speech and the likeability of voices;
- joining the European Charter for Researchers.

9. Development of the organisation

Employees represent the values of the Institute of the Estonian Language at work and outside work. The goal of the Institute's management is to guarantee the satisfaction and high motivation of employees.

The employees of the Institute of the Estonian Language are highly qualified research specialists who guarantee the professional level of the Institute with their research and development work. The employees of the support units of the Institute are experienced specialists of their fields who make sure that the Institute functions without problems.

9.1. Organisational structure, culture, and management

An organisation that supports employees is developed in order to implement the strategy and perform the established tasks. The compliance of the organisational structure with the substantive changes in research and development has constantly been monitored at the Institute. The structure has been updated several times to ensure that the organisation functions without problems.

The updates of the organisation have been based on the flat structure (as few management levels as possible) to ensure better movement of information, flexibility, and low administration costs.

The goal of the structure and the corresponding control system at the Institute is:

- coordinating the cooperation of employees to achieve better results;
- motivating employees to be economical and work efficiently;
- guaranteeing an adequate and flexible reaction to updates and the society's needs for employees;
- guaranteeing the movement of information;
- creating opportunities for managers to coordinate activities in such a manner that all the skills and possibilities of the organisations are applied to the full.

The Institute has decided to join the European Charter for Researchers. Employees have been interviewed and an analysis of the organisation was prepared for this purpose. The action plan for the development of the organisation that proceeds from the principles of the European Charter for Researchers will be prepared as the next stage.

Organisational culture has an important role in the functioning of an organisation. The Institute of the Estonian Language has become an innovative and open organisation over the years. The employee satisfaction survey carried out in 2011 gave valuable information in this process. The Institute plans to repeat the survey in 2016.

Perfecting the motivation system will also continue. Performance appraisals are good indicators upon the assessment of the working environment and employee satisfaction, and of help in the improvement of the organisation of work and management.

Attention is still given to improving the movement of information. Perfecting the intranet and website is a priority. Making the website of the Institute easier to use is the next step. Introducing the Institute in ordinary and social media is also important.

Cooperation between the employees of different departments is considered important at the Institute. The open house day organised in 2015 to introduce the activities of the Institute is a good example of inclusion and cooperation and the Institute plans to make it a tradition.

The Institute of the Estonian Language joined the Diversity Charter coordinated by the Estonian Human Rights Centre on 9 December 2014. The Diversity Charter is a voluntary charter and the company or organisation that joins it confirms that they value the principles of diversity in the case of their employees, partners and clients, and follows the principles of equal treatment. The organisations that have joined the charter form a community where they can share experiences and promote the topic of equal treatment in their organisations as well as society.

9.2. Recruitment

The Institute has hired several young and promising employees in recent years (the average age of employees is 46.3 years) and its IT team has grown. The turnover of staff at the Institute has always been very low, but several capable employees were forced to leave in

2015 when applications for financing projects were rejected.

9.3. Training

The cooperation agreement with the Institute of Estonian and General Linguistics of the University of Tartu was renewed in 2015. The Institute participates as a partner in the project of the Graduate School of Linguistics, Philosophy and Semiotics of the University of Tartu, whose goal is to increase the number of doctoral theses defended in this field and improve the quality of work and supervision. Participation in the graduate school has been very useful for the young researchers of the Institute.

The performance level of middle managers has improved. Being aware of new management trends is necessary in the conditions of a changing environment. It is therefore important to consistently plan and organise management and teamwork training for managers. In order to encourage young people to choose a career in science, it is necessary to perfect the motivation package that supports such careers and to support the present research staff in self-improvement (participation in conferences, the work of specialist committees, various cooperation projects) and to encourage research staff and specialists to participate in academic work and public debate.

Training some employees to become project managers capable of applying for funding for projects and managing the specialist projects required for the work of the Institute is essential. Improving the qualification of the employees who provide support services is also extremely important. Internal promotion and empowerment of employees has made the organisation democratic and open. Offering specific professional training to employees is a problem, as there is a shortage of trainers in Estonia.

9.4. Organisation of work

In line with the new development trends in the functioning of organisations, the Institute is also moving towards a more flexible organisation of work in order to use its time, human resources, and money more economically. Flexible working time (incl. part-time work) helps to increase the satisfaction and motivation of employees, improve the balance between working and one's private life, as well as to take into account the characteristics of employees, and to support creativity and the generation of new ideas.

9.5. Working environment

The working environment at the Institute has changed considerably in recent years. Most of the work rooms have been renovated according to modern requirements and employees have high-quality and contemporary equipment at their disposal. The Institute has three well-equipped halls that can be used for conferences and seminars, and the Institute's library has

also been fully renovated and technologically equipped.

An electronic entrance has been built, which allows employees to enter the work rooms at the times suitable for them and allows the Institute to save money on guards.

The building of the Institute was renovated in 2012 and 2013: the roof was changed, the building was insulated, a new heating and ventilation system and boiler plant were installed, and the fire safety systems of the building were modernised. The kitchen and the hallway of Section B were renovated and a new study hall was set up for meetings and events. Two modern research archives were also built on the basement floor of the building. Material depositories and technical equipment was added to the research archives in 2014. Disabled access to the building of the Institute and a new lift were built in 2013.

The old electricity system of the building will be replaced in the coming years when the necessary funds are found (an ASTRA project application was already submitted for this purpose in 2015). Replacing some of the windows and inner doors, and renovating work rooms is also planned.

As the room layout in the building does not favour social communication between employees, a cosy recreation room was set up on the second floor of the building in 2015 and it is actively used.

9.6. Development of the technical equipment

The Institute is technically very well equipped. The lack of equipment or its low quality will not be an obstacle to the completion of any projects. The Institute has kept and will keep the computers of its employees as well as the general or project servers, the main and specific software required for work up to date and guarantee a fast internet connection and WiFi. The Institute has contributed to the implementation of software in Estonian and promoted its use on a larger scale. It plans to continue with the modernisation of office software (Microsoft Office 2016 and Adobe Acrobat Pro DC) when money is obtained.

10. Intellectual property of the Institute of the Estonian Language

The proprietary copyright to the works created by the Institute upon the performance of its functions (incl. dictionaries and their electronic versions, computer software, collections of information: language archives, slip and card archives, sound recordings, databases, etc.) belongs to the Institute unless otherwise agreed.

Other parties may use the intellectual property of the Institute on the basis of a written agreement entered into with the Institute.