

LIIS ERMUS, MARI-LIIS KALVIK &  
TIINA LAANSALU  
Institute of the Estonian Language

## The Archive of Estonian Dialects and Finno-Ugric Languages at the Institute of the Estonian Language

**Abstract** This report gives an overview of the materials in the Archive of Estonian Dialects and Finno-Ugric Languages (AEDFUL) at the Institute of the Estonian Language (IEL). The AEDFUL holds the world's largest collection of Estonian dialect examples as well as many other materials on other Finno-Ugric languages. Materials in the AEDFUL have been collected by researchers from the IEL and the Mother Tongue Society during the 20<sup>th</sup> century. All the Estonian dialect areas as well as all of the Finnic languages are represented in written and/or recorded form. Especially large amounts of language materials have been collected for Livonian, Ingrian, and Votic. At the beginning of the 21<sup>st</sup> century when active collecting work ended, a new era began focused on digitization and making these materials publicly available. At present, electronic databases and dictionaries are available via the Internet.

## 1. Brief overview

The language collections of the Institute of the Estonian Language (IEL; in Estonian, *Eesti Keele Instituut*, EKI) are located in the Archive of Estonian Dialects and Finno-Ugric Languages (AEDFUL; in Estonian, *Eesti murrete ja soome-ugri keelte arhiiv*, EMSUKA).

In 1947, the Institute of Language and Literature of the Academy of Sciences of the Estonian SSR was established in Tallinn. In the same year, a dialect archive was established based on the collections of the Mother Tongue Society (MTS; in Estonian, *Emakeele Selts*) and the University of Tartu. These materials were collected during 1920–1946 and moved to Tallinn with the establishment of the new archive. In 1947, systematic collection work was begun by the researchers of the Department of Estonian Dialects and the Department of Finno-Ugric Languages of the IEL and the amount of archive materials increased. Later collections from different departments of IEL were incorporated into the Institute's archive: the Archive of Standard Estonian (Est. *Eesti kirjakeele arhiiv*; collected in the Department of Lexicology during 1955–2003), the Archive of Etymological Data (Est. *Etümoloogiline kartoteek*; collecting started in the Department of Finno-Ugric Languages in 1970), the collection of the materials of related languages, etc. During the 1970s and 1980s the students of Tallinn Pedagogical Institute (from 2005 to present Tallinn University) also gathered dialect materials during expeditions (mostly to the islands in western Estonia and to Tartumaa in southern Estonia). Active collection work by the IEL (for the archive) ended in the 1990s and most slip collections were completed in 2002. By the turn of the century, the AEDFUL had become the largest Estonian dialect archive in the world.

The main subcollections of the AEDFUL are the Archive of Estonian Dialects (Est. *Eesti murrete arhiiv*; 5.5 million slips with dialect vocabulary and sentence examples and 2 345 volumes / 153 000 pages, both co-owned by the IEL and the Mother Tongue Society), the Archive of Recordings of Estonian Dialects and Finno-Ugric Languages (Est. *Eesti murrete ja soome-ugri keelte heliarhiiv*; 4 355 hours of original recordings), and the slip collection of dialect vocabulary of Finno-Ugric languages (281 500 slips).

## 2. Composition of the AEDFUL collections

### 2.1. Manuscripts

The manuscript collection consists of scientific overviews, dialect transcriptions, expedition diaries as well as dialect and ethnographical overviews written by researchers of the IEL and members of the MTS. The manuscripts are mostly descriptions or overviews of Estonian dialects or of specific phenomena within dialects; there are also some works on other Finno-Ugric languages (dissertations, scientific overviews, expedition diaries). There are altogether about 153 000 pages of manuscripts.

The main series are:

- Dialect texts in Finno-Ugric phonetic transcription: series ES MT and KKI MT
- Expedition diaries: series ES P and KKI P
- Overviews: series AES, ES, KKI (with subseries), SU, KT

Series marked with ES, AES, and KT contain works of MTS members (also works made under support or supervision of the MTS); series marked with KKI mostly contain works of IEL researchers. There are also subseries under the KKI series containing copies of some dialect-related studies (bachelor's and master's theses) from the University of Tartu Archives of Estonian Dialects and Kindred Languages.

There are over 53 000 pages of dialect transcriptions in Finno-Ugric phonetic transcription (FUPT), collected between 1916 and 1998. Older transcriptions were done by ear in real time during the consultation, but from 1956 onward most of the transcriptions are based on sound recordings. Recording-based transcriptions are written in full Finno-Ugric phonetic transcription, transcriptions done by ear in real time are usually written in a simplified form.

Expedition diaries (544 volumes with almost 11 000 pages) are from 1921 to 1987 and mostly derive from dialect-collecting expeditions; however, a small amount of this material is from Estonians outside Estonia (e.g., in exile and in colonies) and from Finno-Ugric expeditions. These diaries also offer information about the places visited, the language consultants (sometimes with photos), and an overview of the local environment and circumstances.

A series of overviews (886 volumes, 56 000 pages) are written by researchers of the IEL and members of the MTS, some university dissertations are also included in these series. These works are mostly devoted to different aspects of Estonian dialects – phonetics, morphology, syntax, and lexicon. In many texts written especially by MTS correspondents, aspects of everyday life, local traditions, and ethnographical materials are reflected.

Works on other Finno-Ugric languages include the dissertations of the researchers of the Department of Finno-Ugric Languages of the IEL (for example, Arvo Laanest, Elna Pajusalu (Adler), Paul Kokla), there are also some expedition diaries (e.g., the diaries of Arvo Laanest's Ingrian expeditions), grammatical overviews, manuscripts of dictionaries, etc. Most written materials are about Ingrian and Karelian. Works are mostly in Estonian or Russian.

## 2.2. Slip and card archives

The main slip archives at the AEDFUL are:

- Wiedemann's Archive (in Estonian, *Wiedemanni sõnavarakogu*)
- Archive of Conceptual Lexis (*Mõisteline sõnavarakogu*)
- MTS Correspondents' Archive of Dialect Vocabulary (*Murdekorrespondentide sõnavarakogu*)
- Additional Archive (*Täiendav sõnavarakogu*)
- Aggregated Alphabetical Archive of Dialect Vocabulary (*Murdesõnavara koondkartoteek*)
- Place Names Archive (*Kohanimekartoteek*)
- Archive of Persons' and Animals' Names (*Isiku- ja loomanimedekogu*)
- Archive of Finno-Ugric Languages (*Soome-ugri keelte arhiiv*)
- Etymological Archive (*Etümolooiline kartoteek*)
- Archive of Jargon (*Argosõnavara kogu*)
- Archive of Standard Estonian (*Eesti kirjakeele arhiiv*)
- Archive of Terminology (*Terminoloogiakartoteek*)

A typical slip collection at the AEDFUL consists of a number of paper slips with standard measurements of ca. 10 x 8 cm. Usually, the following types of data are found on the front side of the slips: dialect words in transcription, major morphological forms, and usage example(s) in sentence(s). Information about the collector and the time of collection

can be marked either on the front or back side of the slip. Data on slips are written by hand.

The main dialect archives are usually organized by Estonian historical parishes excluding the Archive of Conceptual Lexis (in Estonian, *Mõisteline sõnavarakogu*) and the Aggregated Alphabetical Archive of Dialect Vocabulary (Est. *Murdesõnavara koondkartoteek*). Most other slip archives are organized by the alphabetical order of the headwords. Dialect archives also include some smaller collections (around a few thousand slips) like those of Jaan Jõgever, the Estonian Literature Society, Elmar Wrager, Herman Sergo, the Archive of Grammatical Content, etc. Dialect slips are mostly written in FUPT, except those of MTS correspondents that are in simplified transcription.

The core of the dialect slip archives is Wiedemann's Archive (Est. *Wiedemanni sõnavarakogu*). The materials in this archive were collected by MTS correspondents on the basis of F. J. Wiedemann's *Ehstnisch-deutsches Wörterbuch* (1869) during the period between 1922 and 1949. The basic archive contains 715 000 slips from 46 parishes from all dialect groups, additionally also material from the Lutsi and Kraasna linguistic enclaves (located in Latvia and Russia, respectively). The number of slips from one parish varies from a few thousand to 60 000. The biggest collections are from Hargla, Karksi, Kihelkonna, Kodavere, Lüganuse, Martna, and Muhu parishes (each of them contains over 20 000 slips). Wiedemann's Archive also contains a sub-archive of additional vocabulary (about 90 000 slips) composed of lesser-known words that are not presented in Wiedemann's dictionary.

The Archive of Conceptual Lexis contains 150 000 slips arranged according to 48 conceptual areas. The idea of collecting material by conceptual areas came from Andrus Saareste. The material was gathered in the 1920s and 1930s and later was also published in Saareste's four-volume *Eesti keele mõisteline sõnaraamat*, Conceptual dictionary of Estonian (Stockholm, 1958–1968).

The Place Names Archive (Est. *Kohanimekartoteek*) contains 650 000 text slips and small maps or charts of Estonian parishes and towns (e.g., Tallinn, Valga) and also a small amount on Livonian (60 slips) and Votic (957 slips) place names. The material is collected by MTS correspondents and IEL researchers. The Place Names Archive also contains the personal collections of Endel Varep and Gustav Vilbaste. Varep's collection (over 111 000 sheets) is about settlement history; it

contains information about manors and villages. Vilbaste's collection contains personal notes and excerpts from publications about certain parishes (Harju-Jaani, Jõelähtme, Kose, Kuusalu) and islands in northern Estonia. The material from the Place Names Archive is the basis for *Eesti kohanimeraamat*, the Dictionary of Estonian place names (KNR).

Other larger dialect slip archives are the MTS Correspondents' Archive of Dialect Vocabulary (Est. *Murdekorrespondentide sõnavarakogu*; 400 000 slips, collected from 1940) and the so-called Additional Archive (Est. *Täiendav sõnavarakogu*; 300 000 slips, collected between 1947 and 2002 by IEL researchers with the purpose of supplementing the existing collections of the MTS).

The Aggregated Alphabetical Archive of Dialect Vocabulary (Est. *Murdesõnavara koondkartoteek*) gathers the copies of all basic parish-based dialect slip archives into an alphabetically arranged archive where material is organized according to headwords. Most of the hand-copy work was done during 1951–1963. This type of slip catalog serves as a user-friendly base for both lexicographical and scientific work. Based on the materials in this archive, *Väike murdesõnastik*, Dialectological dictionary of Estonian (VMS) I–II and *Eesti murrete sõnaraamat*, Dictionary of Estonian dialects (EMS), and also many other subdialect dictionaries have been compiled. The Aggregated Archive contains over 2.7 million slips and it is still actively used.

The Archive of Finno-Ugric Languages (Est. *Soome-ugri keelte arhiiv*) contains over 281 500 slips and consists of archives of Votic, Ingrian, Karelian, and Veps. The largest of these is the Votic collection (228 000 slips).

24 000 slips of Votic vocabulary was inherited from the MTS when the Department of Finno-Ugric Languages was established. In the 1950s, the IEL started its own collection. 204 000 slips were added based on the expeditions of the researchers of the Department of Finno-Ugric Languages. An especially large number of expeditions took place during the years 1959–1965. Particularly important subcollections are the collections gathered by Paul Ariste (44 100 slips) and Dmitri Tsvetkov (18 500 slips of Jõgõperä vocabulary, collected in 1920).

Material from the Votic archive was used to compile the Dictionary of Votic (VKS, 1959–2011), later published as a one-volume dictionary (ed. Grünberg 2013). From 2019, the dictionary is also available at <http://www.eki.ee/dict/vadja/>.

The second largest Finno-Ugric slip archive is the Ingrian archive. It contains over 60 000 slips devoted to Ingrian vocabulary and grammar. Most of the materials were gathered by Arvo Laanest on his expeditions during 1960–1980. The vocabulary is partially divided into subarchives by dialect area. The largest dialect vocabulary collection is from the Hevaha dialect (almost 9 000 slips); collections from Oredež, Soikkola, and Lower Luga contain fewer than 2 000 slips each. The collection of Hevaha dialect material was used to compile *Isuri keele Hevaha murde sõnastik*, Vocabulary of the Ingrian Hevaha dialect (Laanest 1997).

There are also about 10 000 slips of Karelian vocabulary collected by Konstantin Manzhin in the archive. In addition, Aime Kährik has compiled a systematic collection of southern Veps verbs (about 2 500 slips).

Conceptual lexis collections of Finno-Ugric languages include two collections: a collection of Votic and Ingrian plant names, collected by Gustav Vilbaste (over 4 000 slips), and a collection of Votic maritime vocabulary, collected by Tiit-Rein Viitso (over 2 000 slips).

The Etymological Archive (Est. *Etümolooiline kartoteek*) was established in the 1970s and its intention was to gather data from all scientific publications about the etymologies of Estonian words. This archive contains over 50 000 cards with etymological data; it was used to compile *Eesti etümolooiasõnaraamat*, Etymological dictionary of Estonian (Metsmägi et al. 2012) and currently is being used to compile a new, academic etymological dictionary, as well.

The Archive of Jargon (Est. *Argoosõnavara kogu*) contains about 50 000 slips, collected during 1964–2002 mostly by high school and university students. The majority of the collected material represents language use of high school and university students; however, the jargon of soldiers, prisoners, and some common language use from different parts of Estonia is also found in this archive.

The addition of material to the Archive of Standard Estonian (Est. *Eesti kirjakeele arhiiv*) continued from 1955 until 2000. This archive contains approximately 4.3 million slips with vocabulary examples mostly from Estonian literature. It was also the source used for compiling the Explanatory dictionary of Standard Estonian (first edition as EKKS, 1988–2007, second edition as EKSS 2009).

The Archive of Terminology (Est. *Terminoloogiakartoteek*) contains about 576 000 slips. It has been used for compiling various terminology dictionaries.

### 2.3. Sound archive

The sound archive was established in 1956 with the aim of gathering tape recordings from all Estonian dialects and Finno-Ugric languages. Recordings were made during the IEL's expeditions during 1950–1990. A typical dialect recording contains at least 30 minutes of monologue spoken by a previously selected speaker (typically an older person with limited formal schooling whose family has been living exclusively in that particular historical parish). The archive also contains material from the Estonian diaspora as well as contemporary speech. Systematic recording by the IEL ended in the 1990s as well as other systematic collecting of dialect materials; however, sporadic recording continued until the beginning of the 2000s. A selection of examples of the recordings is available for listening on the website of the IEL's sound archive <<http://eki.ee/murded/fonoteek/>>.

Altogether there are about 4 360 hours of recordings. Most of them are on reel-to-reel tapes and about 25% are also on cassettes. The main series are:

- EMH(K): recordings of Estonian dialects (both on reel-to-reel tapes and cassettes)
- SUH: Finno-Ugric languages on reel-to-reel tapes
- SUHK: Finno-Ugric languages on cassettes

The 2 900 hours of recordings of Estonian dialects contain material from all Estonian historical parishes (except Paide parish), including the Leivu and Lutsi linguistic enclaves in Latvia. There are language examples from minorities in Estonia: Russian from Iisaku parish (8 hours) and Swedish from Vormsi, Noarootsi, and Naissaare (55 hours), recorded during 1956–2001. The oldest recordings of Estonian dialects are from 1938 (made by Paul Ariste, originals on vinyl disks); the AEDFUL has restored copies on reel-to-reel tapes (made in 1986). There are on average 27 hours of recorded material from each parish, with the largest amounts from Kuusalu parish (166 h), Vastseliina parish (162 h), and Setu parish (150 h).

Recordings from the Estonian diaspora include the following countries or regions: Russia (Siberia as well as the Far-East and European parts of Russia), Georgia (Abkhazia), USA, Canada, Finland, and Ukraine. In addition there are recordings from descendants of Estonian

Swedes living in Ukraine (in Gammalsvenskby). This subarchive contains a total of 280 hours of recordings: from Siberia 130 h, Georgia (Abkhazia) 48 h, Russian Far-East 35 h, USA 28 h, Canada 20 h, the European part of Russia 15 h, Finland 3 h, Ukraine 1 h, and Gammalsvenskby 13 h. Materials from Russia and Georgia were recorded between 1968 and 1992 and those from the USA and Canada in 1992; Estonians in Finland were recorded in 1965, Ukrainian recordings were made in 2001, and recordings in Gammalsvenskby in 1982. The material recorded in areas within the former Russian Empire represent language use at the end of the 19<sup>th</sup> and the beginning of 20<sup>th</sup> century. In the USA and Canada, the speakers were Second World War refugees and their descendants. There is no separate series for recordings of the Estonian diaspora; these recordings are filed either in the EMH or SUHK series.

There are about 1 140 hours of recordings from 17 other Finno-Ugric languages: Livonian, Votic, Ingrian, Veps, Finnish, Karelian, Ingrian Finnish, Saami (all Saami languages together), Komi, Erzya, Moksha, Mari, Mansi, Udmurt, Khanty, Selkup, and Nenets. Finnic language recordings make up 74% of all Finno-Ugric recordings. The largest amount of recordings is from Livonian (321 hours), followed by Votic (186 h), Ingrian (123 h), Veps (85 h), Finnish (53 h), Karelian (50 h), and Ingrian Finnish (3 h). The majority of recordings were made during field expeditions in the 1960s, 1970s, and to a lesser extent also in the 1980s. There are also recordings of Votic, Ter Saami, Livonian, and a small amount of Veps and Karelian that was recorded in the sound studio of the IEL. The most productive collectors of material in related languages were Tiit-Rein Viitso and Arvo Laanest. Tiit-Rein Viitso recorded most of the Livonian material; Arvo Laanest was the main collector of Ingrian but has also recorded Finnish and Votic.

During the 1960s and 1970s, most of the original recordings from the EMH, SUH, and SUHK series were also copied onto back-up tapes as well as tapes available for use by those utilizing the archive (about 3 900 hours); back-up tapes and tapes intended for general use are all reel-to-reel tapes.

The sound archive also holds the ES and Ü series, which contain common speech from the second half of the 20<sup>th</sup> century: mostly speeches from language meetings, conferences and seminars, spoken memoirs, etc.

### 3. Digitization and access to digital materials

#### 3.1. Digitizing archive materials

Digitization of the AEDFUL's sound archive subcollections began in 1992 when the first sound recordings were copied onto DAT-cassettes. From 1992 to 1999 about 335 hours of original tapes were copied to DAT format. Since 1999, digitized copies are stored both on the server and on CD/DVD disks. Almost all original tapes are available in digital form as are the back-up tapes of the Finno-Ugric series.

Active digitizing of slip collections began in 2011 when the IEL acquired a fast document scanner. At the beginning of 2019, the Archive of Conceptual Lexis, the Place Names Archive, and Wiedemann's Archive have been digitized. Elmar Wrager's and Herman Sergo's Hiiumaa vocabulary collections are also digitized.

Digitization of manuscripts became large-scale some years later due to the following reasons. First, as sound recordings are less durable, their digitization has been the most urgent. At first, paper materials were scanned only upon request. The second reason was the lack of a proper flat scanner. Acquiring a high-speed document scanner put the focus on digitizing slip collections. From 2010 to 2013 around 1 000 pages of manuscripts were scanned per year. In 2014, the IEL got a new high-speed flat scanner and scanning rates went up to about 5 000 pages per year. From 2016 to 2018 about 20 000 pages were scanned every year. As of October 2019, over 72 000 pages (over 1 800 volumes) have been digitized. All digitized files will be added to the electronic database.

#### 3.2. Electronic databases

Currently, digitized materials of the Archive of Conceptual Lexis and the Place Names Archive are accessible online. The associated web pages are in Estonian. On these web pages, the slips are divided as they are in the original archives.

The Archive of Conceptual Lexis is first divided by concept areas. In one concept area, files are organized according to their location in the catalogs: first by drawers, then by parishes. Files are organized alphabetically or by individual collections. Individual collections

sometimes include additional diary-like materials about the collection work itself or speaker interviews. Files are in groups of 50–200 files that are browsable 10 files at a time; it is also possible to look at the back sides of slips. The website of the Archive of Conceptual Lexis is available at <http://heli.eki.ee/moisteline/>.

The Place Names Archive is divided by parish and then by sub-archive (the guide to the subarchives is given on the website). Within each subarchive, files are arranged alphabetically by headword and are organized into groups of 50–200 files, which are browsable 10 files at a time; it is also possible to look at the back sides of slips. The website of the Place Names Archive has a search engine. Indexing of slips for the search engine database is currently in progress. As of October 2019, approximately 50% of all slips have been indexed: those from Ambla to Pilistvere parishes, and those pertaining to Livonian and Votic place names. The search engine does not cover Endel Varep's personal place names archive. The website of the Place Names Archive is available at <http://heli.eki.ee/kohanimed/>.

Digitization of Wiedemann's Archive was completed in 2016. Digital files will be accessible by the end of 2019 at <http://emsuka.eki.ee>.

Since 2015, data from the AEDFUL have been accessible at <http://emsuka.eki.ee>. The addition of further data and digital files is in progress. Since the beginning of 2017, the metadata for the manuscript series and sound series have been inserted into the system and are available for searching. Manuscript files are in PDF or PDF/A format, sound files are in MP3 (for unregistered users) and WAV (for registered users) format. By the beginning of June 2019, about 4 400 sound objects and over 1 800 PDF objects had been added to the system and are available along with their metadata.

### 3.3. Electronic dictionaries

One of the first uses of the AEDFUL's materials has been to compile and edit dialect dictionaries. Most of these have been compiled and edited using the IEL's electronic environment EELex. The IEL's policy is to make dictionaries available online. By June 2019, nine of such electronic dictionaries have been made accessible.

*Väike murdesõnastik*, Dialectological dictionary of Estonian (VMS), was published in two volumes 1982–1989. It contains 73 397 dialect words, which are or will be presented in the EMS (see below). It contains headwords, their geographical distribution, and basic meanings for each entry. The website <http://www.eki.ee/dict/vms/> contains an application that shows the geographical distribution of any given word on a map of Estonia.

*Eesti murrete sõnaraamat*, Dictionary of Estonian dialects (EMS), is the largest and most complete dictionary of Estonian dialects. It has been published as a series of separate subvolumes (each about 200 pages) since 1994; in November 2019, the 30<sup>th</sup> subvolume was published. By the end of 2019, the electronic dictionary <http://www.eki.ee/dict/ems/> contains material from subvolumes 1–30.

*Vanapärase Võru murde sõnaraamat*, Dictionary of the old Võru dialect (Käsi 2011), presents the old dialect vocabulary from the parishes of Rõuge, Vastseliina, and Setu. The dictionary is available as a PDF file at [http://www.eki.ee/dict/Voru\\_murde\\_sonaraamat.pdf](http://www.eki.ee/dict/Voru_murde_sonaraamat.pdf).

*Mulgi sõnastik*, Vocabulary of the Mulgi dialect (Laande & Todesk 2013), gathers dialect vocabulary from all historical parishes of Mulgimaa: Tarvastu, Paistu, Halliste, Karksi, and Helme. It is available at <http://www.eki.ee/dict/mulgi/>.

*Hiiu sõnaraamat*, Dictionary of the Hiiumaa dialect (Kokla 2015), gathers dialect vocabulary from all historical parishes of Hiiumaa: Emmaste, Käina, Reigi, and Pühalepa. It is available at <http://www.eki.ee/dict/hiiu/>.

*Idamurde sõnastik*, Vocabulary of the Eastern dialect (Pall 2016), is the new edition of Valdek Pall's *Idamurde sõnastik*, Vocabulary of the Eastern dialect from 1994 (Pall 1994). It gathers dialect vocabulary from the historical parishes where the Eastern dialect was spoken: Iisaku (southern part), Torma, Kodavere (northern part), Maarja-Magdaleena, Palamuse, Äksi, and Laiuse. It is available at <http://www.eki.ee/dict/ida/>.

*Seto sõnastik*, Vocabulary of the Setu dialect (Käsi 2016) is based on the Dictionary of the old Võru dialect (Käsi 2011). It gathers vocabulary from Setu parish, mostly from the eastern part of the parish. It is available at <http://www.eki.ee/dict/setosonastik/>.

*Kihnu sõnaraamat*, Dictionary of the Kihnu dialect (Leas et al. 2016), was published in the end of 2016. It presents the vocabulary of the small Kihnu island and is available at <http://www.eki.ee/dict/kihnu/>.

*Kuusalu rannakeele sõnaraamat*, Dictionary of the Coast dialect of Kuusalu (Norvik & Kendra 2019), presents dialect vocabulary from parts of the historical parish of Kuusalu: the coastal area from Kaberneeme to Salmistu and the peninsulas of Juminda and Pärисpea. The dictionary is available at <http://www.eki.ee/dict/kuusalu/>.

*Eesti etimoloogiasõnaraamat*, Etymological dictionary of Estonian (Metsmägi et al. 2012), is based on the word stems presented in *Eesti õigekeelsussõnaraamat ÕS 2006*, The Dictionary of Standard Estonian. This dictionary is based on the Etymological Archive of the IEL. The dictionary compares Estonian words and their meanings with their Finnic and Finno-Ugric cognates. The Etymological dictionary of Estonian is available at <http://www.eki.ee/dict/ety/>.

*Eesti kohanimeraamat*, the Dictionary of Estonian place names (EKNR), is the first compendious work on Estonian place names. There are over 6 200 place name articles in the dictionary, containing etymological, geographical, and historical information. The dictionary is available at <https://www.eki.ee/dict/knr/>.

### 3.4. Corpus of Estonian dialects

The Corpus of Estonian dialects is compiled in collaboration between the IEL and the Institute of Estonian and General Linguistics at the University of Tartu. It was established in 1998 at the University of Tartu. The corpus contains phonetically transcribed texts (both in FUPT and simplified transcription) as well as sound files from all Estonian dialect areas. The corpus also contains morphologically and syntactically tagged texts. In addition to Estonian dialects, there is also material from the Leivu and Lutsi dialect areas in Latvia as well as materials in Votic and Livonian. The website of the Corpus of Estonian dialects is available at <http://www.murre.ut.ee/murdekorpus/>. It is possible to use a search engine to search the contents of the website <http://www.murre.ut.ee/mkweb/>. For more information about the corpus, see Lindström et al. (2019) in this volume.

## References

- EKKS = *Eesti kirjakeele seletussõnaraamat* [Explanatory dictionary of Standard Estonian] 1–7. 1988–2007. Rudolf Karelson, Valve Kullus (Põlma), Erich Raiet, Mai Tiits, Tiia Valdre & Leidi Veskis (eds). Tallinn: Eesti Keele Instituut.
- EKNR = *Eesti kohanimeraamat* [Dictionary of Estonian place names]. 2016. Marja Kallasmaa & Peeter Päll (eds). Tallinn: Eesti Keele Instituut, Võru Instituut, Eesti Kirjandusmuuseum & Eesti Keele Sihtasutus. Available at: <<https://www.eki.ee/dict/knr/>>
- EKSS = *Eesti keele seletav sõnaraamat* [Explanatory dictionary of Estonian] 1–6. 2009. Margit Langemets, Mai Tiits, Tiia Valdre, Leidi Veskis, Ülle Viks & Piret Voll (eds). Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/ekss/>>
- EMS = *Eesti murrete sõnaraamat* [Dictionary of Estonian dialects] 1–30 (*a-osatama*). 1994–2019. Anu Haak, Evi Juhkam, Mari-Liis Kalvik, Mari Kendla, Tiina Laansalu, Varje Lonn, Helmi Neetar, Ellen Niit, Piret Norvik, Vilja Oja, Valdek Pall, Eevi Ross, Aldi Sepp, Mari-Epp Tirkkonen & Jüri Viikberg (eds). Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/ems/>>
- Grünberg, Silja (ed.) 2013: *Vadja keele sõnaraamat* [Dictionary of Votic]. 2., täiendatud ja parandatud trükk. Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/vadja/>>
- Käsi, Inge 2011: *Vanapärase Võru murde sõnaraamat* [Dictionary of the old Võru dialect]. Helmi Neetar (ed.). Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <[http://www.eki.ee/dict/Voru\\_murde\\_sonaraamat.pdf](http://www.eki.ee/dict/Voru_murde_sonaraamat.pdf)>
- 2016: *Seto sõnastik* [Vocabulary of the Seto dialect]. Maeve Leivo, Nele Reimann & Karl Pajusalu (eds). Tallinn: Eesti Keele Instituut. Available at: <<http://www.eki.ee/dict/setosonastik/>>
- Kokla, Paul 2015: *Hiiu sõnaraamat* [Dictionary of the Hiiumaa dialect]. Meeli Sedrik (ed.). Kärdla: Eesti Keele Instituut, MTÜ Hiiu Öko & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/hiiu/>>
- Laande, Alli & Triin Todesk 2013: *Mulgi sõnastik* [Vocabulary of the Mulgi dialect]. Karl Pajusalu & Urmas Sutrop (eds). Abja-Paluoja–Tallinn–Tartu: Mulgi Kultuuri Instituut, Eesti Keele Instituut, Tartu Ülikooli eesti ja üldkeeleteaduse instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/mulgi/>>
- Laanest, Arvo 1997: *Isuri keele Hevaha murde sõnastik* [Vocabulary of the Ingrian Hevaha dialect]. Tallinn: Eesti Keele Instituut.

- Leas, Reene, Reti Könninge, Silvi Murulauk & Ellen Niit 2016: *Kihnu sõnaraamat* [Dictionary of the Kihnu dialect]. Karl Pajusalu & Jüri Viikberg (eds). Tallinn: Eesti Keele Instituut, Kihnu Kultuuri Instituut, Tartu Ülikooli eesti ja üldkeeleteaduse instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/kihnu>>
- Lindström, Liina, Pärtel Lippus & Tuuli Tuisk 2019: The online database of the University of Tartu Archives of Estonian Dialects and Kindred Languages and the Corpus of Estonian Dialects. – Sofia Björklöf & Santra Jantunen (eds), *Multilingual Finnic. Language contact and change*. Uralica Helsingiensia 14. Helsinki: Finno-Ugrian Society. 327–350. Available at: <<https://doi.org/10.33341/uh.85040>>
- Metsmägi, Iris, Meeli Sedrik & Sven-Erik Soosaar 2012: *Eesti etümo- loogiasõnaraamat* [Etymological dictionary of Estonian]. Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/ety/>>
- Norvik, Piret & Heli Kendra 2019: *Kuusalu rannakeele sõnaraamat* [Dictionary of Kuusalu coastal language]. Vilja Oja (ed.). Tallinn: Eesti Keele Instituut, Juminda Poolsaare Selts & Emakeele Sihtasutus. Available at: <<http://www.eki.ee/dict/kuusalu/>>
- ÕS 2006 = *Eesti õigekeelsussõnaraamat* ÕS 2006 [Dictionary of Standard Estonian]. Tiiu Ereht, Tiina Leemets, Sirje Mäearu & Maire Raadik (eds). Tallinn: Eesti Keele Sihtasutus.
- Pall, Valdek 1994: *Idamurde sõnastik* [Vocabulary of the Eastern dialect]. Tallinn: Eesti Keele Instituut.
- 2016: *Idamurde sõnastik* [Vocabulary of the Eastern dialect]. 2., täiendatud ja parandatud trükk. Ellen Niit (ed.). Tallinn: Eesti Keele Instituut & Eesti Keele Sihtasutus. Available at: <<http://www.eki.ee/dict/ida/>>
- Saareste, Andrus 1958–1968: *Eesti keele mõisteline sõnaraamat* [Conceptual dictionary of Estonian] 1–4. Stockholm: Vaba Eesti.
- VKS = *Vadja keele sõnaraamat* [Dictionary of Votic] 1–7. 1990–2011. Elna Adler, Merle Leppik & Silja Grünberg (eds). Tallinn: Eesti Keele Instituut.
- VMS = Pall, Valdek (ed.) 1982, 1989. *Väike murdesõnastik* [Dialectological dictionary of Estonian] I–II. Tallinn: Valgus. Available at: <<http://www.eki.ee/dict/vms>>
- Wiedemann, Ferdinand Johann 1869: *Ehstnisch-deutsches Wörterbuch*. St.-Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.

## Eesti murrete ja soome-ugri keelte arhiiv Eesti Keele Instituudis

*Liis Ermus, Mari-Liis Kalvik & Tiina Laansalu*

Eesti murrete ja soome-ugri keelte arhiiv (EMSUKA) koondab endas Eesti Keele Instituudi (varasema Keele ja Kirjanduse Instituudi) uurijate ja Emakeele Seltsi liikmete poolt kogutud keelematerjale nii eesti murrete kui ka soome-ugri keelte kohta. EMSUKA sisaldab käsikirju, sedelkartoteeke ja helisalvestisi. EMSUKA murdearhiiv, mis on maailma suurim eesti murdekeelt käsitlev kogu, on olnud paljude murdesõnastike ja teaduslike murdeülevaadete aluseks. Ka soome-ugri keelte materjalide põhjal on koostatud nii sõnastikke kui ka teaduslikke ülevaateid.

Arhiivis on tallel ka etnograafilisi kirjeldusi ning fotosid. Heliarhiivis leidub salvestisi kogu Eesti murdealalt ja samuti pea kõigist teistest soome-ugri keeltest, sealhulgas kõigist läänemeresoome keeltest ning ka Eestist välja rännanute keelest.

Suur osa EMSUKA materjale on digiteeritud ning praegu on töös nende veebi kaudu kättesaadavaks tegemine, et kõik soovijad neid vabalt interneti teel kasutada võiksid. Samuti saab Eesti Keele Instituudi veebilehe kaudu kasutada järjest suuremat hulka EMSUKA materjalidel põhinevaid murdesõnastikke.