

TONI SUUTARI & ULRIIKKA PUURA  
Kotimaisten kielten keskus

## Kotimaisten kielten keskuksen itämerensuomalaiset aineistot

### 1. Johdanto

Kotimaisten kielten keskuksella eli Kotuksella on mittavat yli sadan vuoden aikana kootut arkistot. Murteita ja nimistöä alettiin kerätä systemaattisesti 1800-luvun loppupuolella, ja murteiden laaja tallentaminen äänitteiksi alkoi avokelanauhatekniikan yleistyttyä 1950-luvun lopussa. Kokoelmista ja niiden kartuttamisesta vastasivat eri säätiöt ja organisaatiot, tunnetuimpana Sanakirjasäätiö, kunnes vuonna 1976 perustettiin Kotimaisten kielten tutkimuskeskus. Arkistot koottiin tuolloin yhteen, samoin aineistoihin liittyvät sanakirja- ja tutkimushankkeet sekä kielenhuoltotyö.

Kotuksen arkistojen laajuus on nykyisin noin 2,2 hyllykilometriä. Suurin osa arkistotiloissa säilytteillä olevista kokoelmista on noin A6-kokoisia käsin tai koneella kirjoitettuja sana- ja nimilippuja. Sähköiset aineistot puolestaan koostuvat yhteensä satojen miljoonien sanojen laajuisista korpuksista, miljoonista kuvamuotoon skannatuista arkistolipuista ja muista dokumenteista sekä noin 17 000 tunnista digitoituja tai suoraan digitaaliseen muotoon kerättyjä av-tallenteita. Kaikkiaan äänitteitä ja videoita on noin 24 000 tuntia.

Kotus soveltaa arkistoihin ja aineistoihin avoimuuden periaatetta. Arkistojen kokoelmat ovat pääosin asiakkaiden vapaasti käytettävissä, ja verkossa aineistoja tarjotaan käyttöön mahdollisimman avoimesti ja maksutta. Toki on myös sellaisia aineistoja, joiden käyttöä joudutaan rajoittamaan henkilösuojan tai tekijänoikeuden perusteella.

## 2. Murteita ja nimistöä

**Suomen murteiden sana-arkistoa** (SMSA) alettiin kartuttaa 1900-luvun alussa *Suomen murteiden sanakirjan* toimitustyötä silmällä pitäen. Keruutyön taustalla on E. N. Setälän sanakirjaohjelma vuodelta 1896. Keruista huolehti alkuun Suomalaisen Kirjallisuuden Seura, vuosina 1916–1924 Suomen kielen sanakirjaosakeyhtiö ja vuosina 1924–1976 Sanakirjasäätiö, jolta kokoelmat siirtyivät Kotukselle.

Murresanastoa ovat tallettaneet koulutettujen kerääjien ohella merkittävässä määrin myös maallikkokirjeenvaihtajat, jotka ovat vastanneet lehdissä, erityisesti vuosina 1927–1990 julkaistussa *Sanastajassa*, esitettyihin kyselyihin. Myös murteenkeruukilpailuja on järjestetty. Suomen murteiden sanakirjan tavoitteena on 1900-luvun alun kielimuodon kuvaus, joten arkiston sana-aineistoa ei enää kartuteta vaan nykymurteita tallennetaan muihin kokoelmiin. Painettuna murre-sanakirjasta on ilmestynyt kahdeksan osaa, ja nykyisin sanakirja on siirtynyt ilmestymään verkossa. Sanakirjassa arkistoon kerätyt tiedot ovat toimitettuina helposti ja tiiviisti käytettävissä osin painettuna, osin osoitteessa <http://kaino.kotus.fi/sms/> – jossa vuonna 2019 on saatavissa sanaväli *a–lysmä*.

Murrearkiston pääkokoelmassa, jossa kaikki sanaliput ovat aakkosjärjestyksessä, on reilut 8 miljoonaa sanalippua n. 350 000 eri sanasta. Pelkästään murteissa esiintyvän sanaston lisäksi mukana on runsaasti myös murteille ja yleiskielelle yhteistä sanastoa. Arkistossa on myös sanaston erilliskokoelmia ja muita murteisiin liittyviä dokumentteja. Aiemmin murrearkistoon kuului myös noin 1,5 miljoonan sananparren kokoelma, jonka ylioppilasosakunnat ovat keräyttäneet 1930-luvulla. Kotus luovutti sananparsikokoelman vuonna 2016 Kansallisarkistolle, joka on digitoinut kokoelman ja avannut sen verkossa vapaasti käytettäväksi osoitteessa <http://digi.narc.fi/digi/dosearch.ka?sartun=385077.KA>.

Murrearkiston sanalippuja ei ole ainakaan toistaiseksi digitoitu lukuun ottamatta osaa valokuvia ja piirroksia sisältävistä sanalipuista. Sen sijaan kerääjä- ja vastaajakortistot ovat käytettävissä myös sähköisinä. Hiljan on skannattu kuvamuotoon myös laaja synonyymikortisto, jonka avulla voi selvittää, mitä eri ilmauksia murteissa käytetään. Arkistohan on järjestetty nimenomaan hakusanoittain eikä käsitteittäin, eli

synonyymien tuntemus on tietojen hakemisen kannalta olennaista. Synonyymikortistoa voi käyttää vapaasti verkossa osoitteessa <[https://www.kotus.fi/aineistot/sana-aineistot/suomen\\_murteiden\\_sana-arkisto/kortistot\\_ja\\_luettelot](https://www.kotus.fi/aineistot/sana-aineistot/suomen_murteiden_sana-arkisto/kortistot_ja_luettelot)>.

**Nimiarkisto** sai alkunsa vuonna 1915, kun useat tieteelliset seurat perustivat paikannimitoimikunnan. Tuolloin alkoi suomenkielisen paikannimistön systemaattinen keruu. Suomen ruotsinkielistä paikannimistöä oli alettu kerätä jo 1800-luvun puolella, ja asiakirjanimien poiminnan oli aloittanut tieteellisten seurain paikannimikomitea vuonna 1909. Saamenkielisiä nimiä alettiin kerätä 1930-luvulla, ja 1950-luvulla paikannimien keruun rinnalle tulivat myös henkilönimien ja muiden nimien keruu.

Alkuperäisiä tai niihin rinnastettavia, pitäjittäin järjestettyjä paikannimilippuja Suomesta ja lähialueilta on noin 2,7 miljoonaa, ja kokoelmiin liittyviä keruukarttoja on noin 10 000. Lisäksi kopio-lipuista on muodostettu koko Suomen kattava kaikki nimet samassa aakkostuksessa sisältävä yleiskokoelma. Paikannimikokoelmat ovat pääosin kielitieteellisen koulutuksen saaneiden tutkijoiden ja opiskelijoiden kentältä tallentamaa suullista perimätietoa. Nimen lisäksi on siis tallennettu nimiin ja nimettyihin paikkoihin liittyvää oheistietoa. Jonkin verran tietoa on kerätty nimikilpailujenkin avulla.

Nimiarkiston digitalisointia viimeistellään parhaillaan. Kaikki paikannimiliput, keruukartat ja muutkin Nimiarkiston nimikokoelmat on skannattu kuvatiedostoiksi. Paikannimilipuista on siirretty tietokantamuotoon paikannimen hakuasu, paikannusviite, keruupitäjä, kerääjä, keruuvuosi sekä paikan laji. Sijainti on tarkennettu keruukarttojen avulla pistekoordinaattitasolle. Lisäksi niiden nimien osalta, jotka ovat samoja kuin Maanmittauslaitoksen Paikannimirekisterissä, tietokantaan on lisätty Paikannimirekisterin paikannimi-id. Näin syntynyt paikkatietoaineisto julkaistiin avoimena datana käyttöön vuoden 2017 lopussa osoitteessa <[www.nimiarkisto.fi](http://www.nimiarkisto.fi)>. Sähköisen nimiarkiston julkaisu juhlisti osaltaan satavuotiasta Suomea. Digitaalisesta nimiarkistosta muodostuu kansallinen tietovaranto, sillä samaan kokonaisuuteen yhdistetään ajan myötä myös Suomen ruotsin- ja saamenkieliset paikannimikokoelmat.

Digitaalisessa arkistossa ei ole kyse vain olemassa olevan tiedon käytettävyyden parantamisesta, vaan sähköinen toimintaympäristö

tarjoaa tärkeitä mahdollisuuksia tietojen rikastamiseen ja täydentämiseen. Kansalaiset voivat tuottaa tietoa siitä, ovatko arkistoon tallennetut nimet vielä käytössä, onko nimi korvautunut uudella, onko nimettyä paikkaa enää olemassakaan tai mitä kokonaan uutta nimistöä on syntynyt. Samoin voidaan täydentää paikkoihin ja nimiin liittyvää taustatietoa sekä muuntaa nimiin liittyvää tietoa kuvatiedostoista tietokantamuotoon.

Nimiarkistossa on aineistoja myös Suomen nykyisen ja luovutetun alueen ulkopuolelta. Aunuksen ja Vienan alueelta on pitäjittäiset paikannimikokoelmat sekä erilliskokoelmina suomalais-karjalaiset ja venäläiset kunnan- ja kylännimet sekä tiedot talojen ja asukkaiden lukumääristä vuodelta 1905. Itä-Karjalan yleiskokoelma käsittää Vienan ja Aunuksen nimet. Inkerin alueelta on pitäjittäiset paikannimikokoelmat sekä yleiskokoelma. Arkangelin aluearkistosta kopioituna on käytössä Vienan Karjalan ja Vienanmeren paikannimikokoelma (myös sähköisenä muttei verkossa). Kokoelmiin kuuluvat myös Inkerin Pähkinälinnan läänin ja Karjalan asiakirjanimikokoelmat.

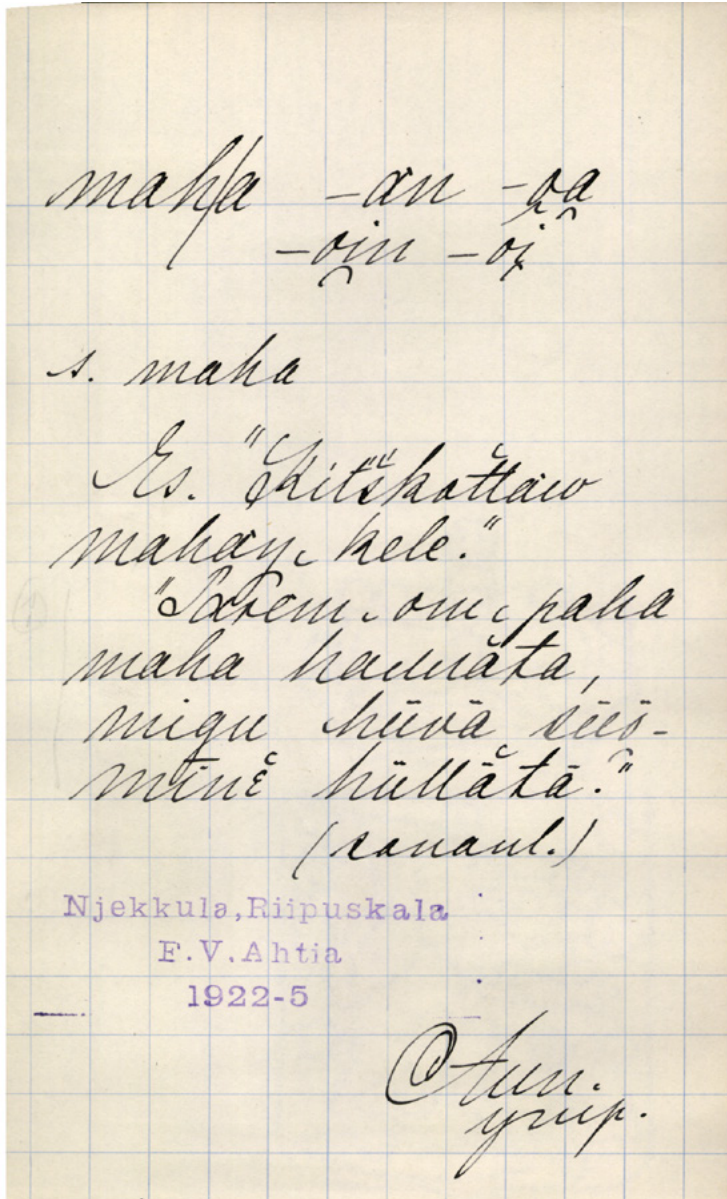
Viron paikannimistöä on Lauri Kettusen laatima 1500-sivuisen luettelo, joka perustuu M. J. Eisenin Tartossa säilytettävään kokoelmaan, ja luettelon pohjalta tehdyt nimiliput. Luettelo on digitoitu mutta se ei ole vielä verkon kautta saatavilla. Kokoelmissa on myös Pohjois-Viron talonnimien luettelo vuodelta 1887. Niin ikään vepsäläisistä paikannimistä on aakkosellinen kokoelma (kerääjinä Viljo Nissilä 1942–1944 ja Jussi Rainio 1967). Nimistötietoja sisältyy myös moniin Kotukseen arkistoituihin henkilöarkistoihin, joista seuraavassa on nostettu esiin keskeisimmät. **Heikki Ojansuun henkilöarkisto** sisältää paikannimiä ja asiakirjapöytäkirjoja Suomesta, Virossa (mm. Fr. Kuhlbarsin Viron paikannimien kokoelma) ja liiviläisalueelta. **Viljo Nissilän henkilöarkistossa** on paikannimiin liittyviä asiakirjapöytäkirjoja, muistiinpanoja ja nimiluetteloita mm. vepsäläisestä asutuksesta. Nissilän aineistoihin kuuluu myös laaja ns. Päämajan paikannimistöimikunnan Itä-Karjalan kokoelma. **O. A. Louhelaisen arkistossa** on suomalaista ja karjalaista sekä Inkerin nimistöä, erityisesti historiallisia tietoja 1500–1800-luvuilta.

### 3. Karjalaa monessa muodossa

*Karjalan kielen sanakirja* ilmestyi kuusiosaisena vuosina 1968–2005 Kotimaisten kielten (tutkimus)keskuksen ja Suomalais-Ugrilaisen Seuran yhteistyönä. Vuonna 2009 sanakirja julkaistiin kokonaisuudessaan verkossa. Sanakirja perustuu **Karjalan kielen sana-arkiston** aineistoon, joka koostuu reilusta puolesta miljoonasta sanalipusta. Pääosa aineistosta on yhtenä aakkosiin järjestettynä kokoelmana. Erikseen ovat Juho Kujolan tverinkarjalan sekä Valdain ja Tihvinän karjalaismurteiden sanaliput sekä Helmi ja Pertti Virtarannan keräämät vienankarjalan ja tverinkarjalan sanaliput. Vanhin aineisto on 1800-luvun lopulta ja perustuu Arvid Genetzin muistiinpanoihin sekä K. F. Karjalaisen keruisiin. E. V. Ahtia on kerännyt arkiston tiedoista lähes puolet vuosina 1898–1936. Sodan jälkeen tietoja keräsivät Raja-Karjalan karjalaisilta ja Suomeen siirtyneiltä Itä-Karjalan pakolaisilta Eino Leskinen, R. E. Nirvi ja Aimo Turunen.

Kaikki karjalan sanalippukokoelmat on skannattu kuvamuotoon, ja aakkosten alkupäästä nimiliput on linkitetty verkkosanakirjaan. Koko skannattu aineisto on verkossa ladattavissa osoitteessa <[https://www.kotus.fi/aineistot/sana-aineistot/karjalan\\_kielen\\_sana-arkisto](https://www.kotus.fi/aineistot/sana-aineistot/karjalan_kielen_sana-arkisto)>. Kuvamuotoisia sanalippuja linkitetään vähitellen sanakirjan sana-artikkeleihin. Alkuperäiset sanaliput on deponoitu vuonna 2016 Itä-Suomen yliopistolle karjalan kielen ja kulttuurin oppiaineen haltuun.

Karjalan kielen sanaston suurkerääjä E. V. Ahtia laati myös kieliopeja. *Karjalan kielioppi* (äänne- ja sanaoppi) julkaistiin vuonna 1938. Lauseoppi ja johto-oppi jäivät aikanaan julkaisematta mutta säilyivät arkistossa. Lauseopin alkuperäinen käsikirjoitus on digitoitu ja käytettävissä Kotuksen verkkosivuilla (Ahtia, ei vuosilukua). Lisäksi verkkosivuilla ovat Karjalan Kielen Seuran tuottama konekirjoitettu versio sekä lauseopista että johto-opista, samoin näköisversio painetusta äänne- ja sanaopista (Ahtia 1938, 2014a, 2014b).



Kuva 1. E. V. Ahtian karjalan kielen sanalippu. Kuva Kotuksen kokoelmista.

#### 4. Muiden pienten itämerensuomalaisten kielten sana-aineistoja

**Vepsän verkkosanasto** osoitteessa <http://kaino.kotus.fi/sanat/vepsa/> sisältää Lauri Kettusen laatiman kenttämuistiinpanoihin perustuvan sanakirjan käsikirjoituksen aineiston. Tiedot on siirretty tietokantaan sellaisenaan, mutta pitäjänlyhenteet on korvattu pitäjien koko nimillä käytön helpottamiseksi. Kettunen teki kaksi matkaa vepsäläiskyliin. Vuosina 1917–1918 hän liikkui yksinomaan etelävepsäläisten alueella, mutta vuonna 1934 hän eteni etelävepsäläisistä kylistä keski-vepsäläisen alueen kautta äänis- eli pohjoisvepsäläisiin kyliin. Ensimmäisen matkansa hän teki yksin, toisella mukana olivat myös Lauri Posti ja Paavo Siro. Sanastoa Kettunen keräsi kummallakin matkalla. Ensimmäisellä matkalla hän myös äänitti vepsää, ja tämäkin aineisto on Kotuksessa arkistoituna ja digitoituna.

Kotuksessa on säilytteillä laaja **Reino Peltolan vepsän sanalippukokoelma**, 58 lippaallista, jossa mukana on myös Setälän, Postin, Kettusen ja Tunkelon keräämää aineistoa. Kokoelmiin kuuluu myös **inkeroismurretta** (lähinnä Soikkolasta ja Laukaansuusta), vähän Gunvor Löppösen ja R. E. Nirvin keräämää **vatjaa** sekä Tarton yliopiston kokoelmasta kopioitua **etelävirolaista aineistoa**. Lisäksi Eesti Keele Instituutin viron murteiden kokoelmista on vanhat mikrofilmikopiot. Sähköisessä muodossa ovat **lyydin, liivin ja vatjan käänteissanastot**, jotka on aikanaan julkaistu painettuina (Elomaa & Laakso toim. 1986; Laakso toim. 1988, 1989). Nämä pyritään saamaan myös verkon kautta vapaasti käytettäviksi.

Syksyllä 2016 Kotuksessa alettiin toimittaa hieman suurempien itämerensuomalaisten nykykielten kaksikielistä sanakirjaa: *Viro-suomi-sanakirja* ilmestyi Suomen ja Viron 100-vuotisitsenäisyyspäivien kunniaksi tammikuussa 2019. Sanakirjaprojekti on virolaisen Eesti Keele Instituutin aloittama ja organisoima yhteistyöhanke. Sanakirja on vapaasti käytettävissä Eesti Keele Instituutin verkkosivuilla. Vuonna 2003 ilmestynyt *Suomi-viro-suursanakirja* puolestaan saatiin verkkoon vapaasti käytettäväksi vuonna 2017.

## 5. Kielikartastot

**Itämerensuomalainen kielikartasto** (IMSK) eli *Atlas Linguarum Fennicarum* (ALFE) on syntynyt suomalais-virolais-karjalaisena yhteistyöhankkeena ja ilmestynyt painettuna kolmessa osassa vuosina 2004, 2007 ja 2010. Toimittamisesta ovat vastanneet Kotimaisten kielten (tutkimus)keskus, Eesti Keele Instituut ja Venäjän Tiedekatemian Karjalan tiedekeskuksen Kielen, kirjallisuuden ja historian instituutti. Tietokantamuotoiseksi aineistoksi kartasto on muokattu Kotimaisten kielten keskuksessa Avoin data ja paikkatieto -projekteissa vuosina 2013 ja 2014. Aineisto on avoimena datana käytettävissä CSC:n ylläpitämässä AVAA-palvelussa osoitteessa <http://avaa.tdata.fi/web/avaa/-/kodus-kielikartastot>. Tarjolla on käyttöliittymä, mutta lisäksi aineiston voi ladata csv- ja shape-muodoissa tai hyödyntää kyselyrajapintapalvelua.

Itämerensuomalainen kielikartasto keskittyy pääasiassa sanasto-ilmioihin, ja sen tavoitteena on kuvata itämerensuomalaisten kielten keskinäisiä suhteita sekä kontakteja naapurikieliin. Kielenilmiöiden levikit kuvastavat itämerensuomalaisten kielten jakautumista nykyiseksi tytärikieliksi. Sanasto liittyy kiinteästi myös kielenulkoisiin kulttuuri-ilmioihin, joten kartasto palvelee myös kulttuurimaantiedettä.

Kartoissa kuvattavat tiedot on kerätty 317 kysymystä sisältävän kyselysarjan perusteella. Aihepiirit kattavat painetun kartaston johdannon sanoin ”koko alkukantaisen inhimillisen elämän alan”. Kysymyksistä valtaosa on onomasiologiaa, eli ne kartoittavat tietyn käsitteen nimityksiä, kuten ilmasuuntien tai viikonpäivien nimityksiä. Osin kartoitetaan alkuperältään saman sanan alueittain vaihtelevia muotoja ja merkityksiä. Kartasto sisältää monen tyyppisiä karttoja, ja kaikkiin karttoihin liittyy yksityiskohtainen nelikielinen kommenttiosuus, joka selittää kartoissa kuvattavia ilmiöitä ja niiden variaatiota. (Verkko-sovellus sisältää vain karttojen sisältämän tiedon, mutta sovelluksen karttatasolta on ladattavissa painetun kartaston 2. ja 3. osasta pdf:nä kartta ja siihen liittyvä seliteosa. 1. osan pdf:t on tarkoitus liittää mukaan tulevaisuudessa.)

Usein kartoituksen tulos on jaettu useaksi kartaksi. Esimerkiksi MAHLA-formatiivi kuvataan kolmessa kartassa, joista kaksi ensimmäistä (117.1 ja 117.2) esittelee eri äänneasut (*mahla* ym.) ja kolmas



(117.3) äänneasujen eri merkitysryhmät; 'lehtipuiden rungossa virtaavan nesteen' lisäksi MAHLA voi tarkoittaa itämerensuomalaisissa kielissä 'nilaa', 'verta' tai 'räkää'.

Kartaston kartat jakautuvat seuraaviin päätyyppeihin:

- 1) Onomasiologiset eli leksikaaliset kartat kuvaavat eri kielten tunteman yhteisen käsitteen nimityksiä. Onomasiologinen karttatyypin on kartastossa selvästi yleisin.
- 2) Formatiivikartat kuvaavat kielenilmion, erityisesti sanan muodon levikkiä. Formatiivilla tarkoitetaan sanan äänteellisen hahmon yleistystä, "etymologista sanaa", esimerkiksi LÄMMIN-formatiivi on yleistys sanoista *lämmi(n)*, *lämmyn*, *lämi*, *lämmi* jne. Formatiivit on pääsääntöisesti kirjoitettu versaalikirjaimin. Pelkän formatiivin levikin lisäksi voidaan kuvata myös formatiivin kattamien sanojen äänne- ja muotopiirteiden levikkejä ja semantiikkaa. Usein karttoihin on yhdistetty kahden tai useamman, tavallisesti samaa merkitsevän, formatiivin tiedot.
- 3) Semanttiset kartat havainnollistavat sanan eri merkitysten levikkiä. Esimerkiksi formatiivilla HAKO (suomen *hako*, viron *haga*) on seuraavia merkityksiä: 'havupuun lehvä', 'maatunut tai veteen kaatunut puu', 'oksa', 'neulanen'. Jos kyseessä on semanttinen kartta, tämä on usein kerrottu kartan otsikkotiedoissa.
- 4) Motiivi- eli miellekartat kuvaavat käsitteen taustalla olevaa yhteistä motiivia. Esimerkiksi perunan nimitys voi perustua "päärynän" (*peruna* < muinaisruotsi, vrt. nykyruotsin *päron* 'päärynä', *jord|päron* 'maapäärynä = peruna'), "omenan" (*maaomena*), "nauriin" (*na.ggõrz*) tai "munan" (*maamuna*) nimitykseen. Jos kyseessä on motiivikartta, tämä on usein kerrottu kartan otsikkotiedoissa.

Edellä kuvattujen karttatyypin lisäksi painetun kartaston johdannossa mainitaan frekvenssi- eli tilastokartat sekä ideaalityypin eli rekonstruktio kartat, mutta näitä karttatyyppejä esiintyy vain vähän.

Kartasto kuvaa 1900-luvun alun kielimuotoa. Tiedot on kerätty 259 perushavaintopisteestä eli kartoituspisteestä, jotka jakautuvat kielittäin seuraavasti: suomi 185, karjala 29, vepsä 7, inkeroinen 3, vatja 3, viro 30 ja liivi 3. Tietoja on hankittu sekä arkistoista että kenttäkeruin. Suomen ja viron tiedot perustuvat yksinomaan arkistojen

kokoelmiin, ja karjalan ja vepsän tiedot ovat lähes kokonaan kenttäkeruun tulosta.

Sähköinen kartasto noudattaa painetun kartaston hierarkiaa ja kuvaustapaa, mistä aiheutuu muun muassa tietojen toisteisuutta. Käytännössä tämä näkyy esimerkiksi kartoissa, jotka kuvaavat 'muikun' (*Coregonus albula*) nimityksiä. Ensimmäisessä kartassa esitetään eri nimitysten levikit formatiivitasolla (*muikku, rääpys, maiva, silli*) ja seuraavissa kartoissa kuvataan *muikun* ja *rääpyksen* eri muotojen levikit, jolloin MUIKKU-levikki (256.1) on sama kuin *muikku* ja *muje* muotojen yhteislevikki (256.2) ja RÄÄPYS-levikki on sama kuin *rääpys, rääppö ~ rääpöi, rääbis, rääpus* ja *rääpyskä* -muotojen yhteislevikki (256.3).

Samassa palvelussa IMSK:in kanssa on tarjolla tietokantamuotoisena myös **Lauri Kettusen murrekartasto**, joka on ilmestynyt alun perin Suomalaisen Kirjallisuuden Seuran kustantaman *Suomen murteet* -teossarjan osana. I osassa (1930) Kettunen julkaisi keräämiään murrenäytteitä, II osassa (1930) hän kuvasi suomen murrealueet, ja III osa (1940) on kahteen niteeseen jakautuva murrekartasto. Osa III A sisältää 213 karttaa ja osa III B karttoihin kuuluvat yksityiskohtaiset selitykset.

Kettusen karttojen sisältö on siirretty tietokantamuotoon 1990-luvun lopussa Kanadassa Yorkin yliopistossa Sheila Embletonin ja Eric S. Wheelerin johtamassa Finnish Dialect Atlas Project -hankkeessa. Kotus sai aineiston vuonna 2000. Aineisto tarkastettiin ja muokattiin nykyiseen asuunsa vuodesta 2009 alkaen Urho Määttäin johtamassa monitieteisessä BEDLAN-tutkimushankkeessa (Biological Evolution and the Diversification of Languages Project) yhteistyössä Kotuksen kanssa.

Kettusen kartasto kuvaa suomen murteiden äänne- ja muotoopillisia piirteitä. Kartat on jaoteltu äänneopillisin perustein. Pääjako on konsonantismi (kartat 1–140) ja vokalismi (kartat 141–213). Karttojen otsikointi on Kettusen kartaston sisällysluettelon mukainen. Verkkosovelluksesta on ladattavissa myös painetun kartaston kartat pdf-tiedostoina.

Kotus on ollut mukana myös *Atlas Linguarum Europaen* (ALE) toimitustyössä. Aiemmin Kotuksessa säilytteillä olleet materiaalit on siirretty Turun yliopistoon lukuun ottamatta toimitustyön tueksi kootua **Euroopan kielikartaston sanalipustoa**.

## 6. Äänitteitä kaikista itämerensuomalaisista kielistä

**Suomen kielen nauhoitearkiston** kokoelmissa on suomen murteiden ja puhekielen lisäksi paljon myös suomen sukukielten äänitteitä. Parhaiten on edustettuna karjala (n. 1 500 tuntia). Runsaasti on kerätty myös eri saamen kieliä, viroa ja unkaria. Aivan vähäisiä eivät ole muidenkaan sukukielten kokoelmat.

Valtaosa nauhoitearkiston äänitteistä sisältää haastatteluja ja kerrontaa, mutta erityisesti sukukielisissä äänitteissä on myös kansanmusiikkia. Tiedot kielenoppaasta, haastattelijasta ja tallennustilanteesta on tallennettu kokoelmatietokantaan, mutta tietoja tallenteiden sisällöistä ei juuri ole. Nauhoitearkiston kokoelmatietokanta on käytettävissä erillisen liittymän kautta Kotuksen verkkosivuilla osoitteessa <http://kaino.kotus.fi/naark/>. Tarkoituksenmukaista olisi yhdistää tiedot Kotuksen yleiseen aineistotietokantaan, minkä jälkeen myös av-aineistojen tiedot näkyvät Finna-hakupalvelun kautta. Henkilösuoja tosin asettaa rajoituksia sille, missä laajuudessa tietoja voidaan avoimessa verkossa näyttää.

Litterointeja on tehty suomen murteiden äänitteistä noin tuhannesta tunnista, mutta sukukielten äänitteistä litterointeja on varsin vähän. Usein aineiston valinnassa onkin tukeuduttava äänitteiden kuuntelemiseen. Tutkijoiden julkaisuista, keruu- ja matkakertomuksista ja muistiinpanoista voi myös löytyä arvokasta tietoa aineiston käytön tueksi. Esimerkiksi Pertti Virtarannan julkaisuissa ja systemaattisissa muistiinpanoissa on paljon taustatietoa informanteista ja äänitystilanteista.

Nauhoitearkiston perustaja Pertti Virtaranta itse oli arkiston ahkerimpia kerääjiä. Hän oli erityisen kiinnostunut suomesta ja karjalasta, mutta nauhoitti entisen Neuvostoliiton alueelle suuntautuneilla lukuisilla retkillään myös lyydiä, viroa, vepsää, inkeroista, vatjaa ja liiviä (keruumääräjärjestyksessä). Karjalaa Virtaranta tallensi 1950-luvun loppupuolelta alkaen lähes vuosittain aina 1980-luvun lopulle saakka. Usein hänen mukanaan keruuretkillä kulki puoliso Helmi Virtaranta ja lisäksi suomalainen kollega tai paikallinen asiantuntija, kuten esimerkiksi itäkarjalainen kielentutkija Grigori Makarov. Pertti Virtarannan ja Väinö Kaukosen kesällä 1968 äänittämistä karjalankielisistä



Kuva 2. Pertti Virtaranta haastattelemassa karjalaisnaista. Kuva Museoviraston kokoelmista.

nauhoista (n. 90 tuntia) on laadittu sisällysluettelot, jotka tutkija voi saada digitaalisina käyttöönsä. Kerronnan lisäksi aineisto sisältää hää- ja itkuvirsiä, joikuja, kehto- ja muita lauluja.

Nauhoitearkiston vanhimmat äänitteet ovat vahalieriötallenteita 1900-luvun alkuvuosikymmeniltä. Kokoelmassa on noin 950 lieriötä, muun muassa K. F. Karjalaisen (hantia 1900–1902), G. J. Ramstedtin (kalmukkaa 1903), Artturi Kanniston (mansia 1905–1906), Yrjö Wichmannin (maria 1906), Kai Donnerin (samojedia 1912–1914), T. I. Itkosen (koltansaamea 1913), Lauri Kettusen (suomen murteita, vepsää,

vatjaa, liiviä ja unkaria 1914–1932) ja A. O. Väisäsen (mordvaa 1914, viroa 1922) äänityksiä. Osa kokoelmista on vielä tarkemmin käymättä läpi ja tunnistamatta.

Monet sukukielten lieriökokoelmat siirtyivät kerääjiltä Suomalais-Ugrilaiselle Seuralle ja sitä kautta vuonna 1959 perustetulle Suomen kielen nauhoitearkistolle, missä osa lieriöistä kopioitiin avokelanauhoille 1960-luvulla. Kaikki lieriöt jäivät Vironkadulle Helsingin yliopiston fonetiikan laitoksen ullakolle, kun nauhoitearkisto muutti vuonna 1988 Sörnäisiin. Lieriöt ”löydettiin” reilut kymmenen vuotta sitten, minkä jälkeen ne järjestettiin, luettelointiin ja pakattiin vanerisiin reikäpehmustein varustettuihin laatikkoihin. Vuonna 2008, kun Kotus muutti Kaisaniemeen, lieriöt pääsivät muiden tallenteiden kanssa asianmukaisiin ja vakioilmastoituihin arkistotiloihin. Kotuksen muuttaessa Hakaniemeen vuonna 2016 av-aineistojen säilytys ulkoistettiin, koska nykyisessä kiinteistössä ei ole tarvittavia erikoistiloja.

Vuonna 2009 kaikki toistokelpoiset lieriöt digitoitiin nauhoitearkiston 50. toimintavuoden kunniaksi. Arvokas lisä kokoelmiin saatiin samana vuonna, kun Kansallisarkisto luovutti E. N. Setälän lieriöt (liiviä 1912, vepsää 1916) Kotukselle. Myös nämä lieriöt on digitoitu.

E. N. Setälä tallensi vepsää vuonna 1916 yhdessä A. O. Väisäsen kanssa. Tältä retkeltä on säilynyt eri arkistoissa paljon materiaalia. Väisäsen valokuvat ovat Museoviraston kuva-arkistossa. Lisäksi Väisänen on luettellonut aineiston ja litteroinut sen sisältämät laulut. Muistiinpanoista käyvät ilmi pitäjät, esiintyjät ja heidän ikänsä. Käsien kirjoitettu luettelo ja joitakin litterointeja sisältävä muistikirja ovat Kansallisarkistossa (kopio Kotuksessa). Konekirjoitettu luettelo sekä puhtaaksikirjoitetut ja tarkastetut laulujen tekstilitteraatit ovat puolestaan Suomalaisen Kirjallisuuden Seuran arkistossa Väisäsen aineiston osana (näistäkin kopiot Kotuksessa). Lieriöt sisältävät lauluja, runoja ja itkuvirsiä, paimensoittoa, eläinten kutsuntaa, huhuilua sekä kanteleen- ja harmonikansoittoa. Nauhoitearkistossa on myös Väisäsen haastattelu, jossa hän ikämiehenä muistelee muun muassa vepsän maille suuntautunutta keruuretkeä.

Lauri Kettunen suuntasi vepsän puhuma-alueille pian Setälän jälkeen vuonna 1917–1918. Vuonna 1918 hän äänitti etelävepsäläisissä kylissä muun muassa satuja, tarinoita ja lauluja. Kuurinmaan liiviläiskylissä Kettunen vieraili toistakymmentä kertaa. Vuosina 1920



Kuva 3. Sakari Pietarila digitoimassa Kotuksen vahalieriöitä vuonna 2009. Valokuvaaaja Toni Suutari. Kuva Kotuksen kokoelmista.

ja 1923 hän äänitti satuja ja tarinoita. Vuodelta 1927 Kettuselta on yksi Helsingissä tehty liivinkielinen äänitys, ja lisäksi on muutama tallennusajan ja -paikan suhteen tunnistamattomaksi jäänyt äänitys. Vätjä Kettunen tallensi vuosina 1915, 1923 ja 1932, satuja, lauluja, muistelmia ja muuta vastaavaa. Kaikkiaan Kettusen eri kielistä äänittämiä lieriöitä on säilynyt lähes kaksisataa. Kettunen on julkaissut osan äänittämästään aineistosta teksteinä, ja lisäksi hänen kirjoina ilmestyneet matkakertomuksensa sisältävät paljon tietoa kielenoppaista ja keruutilanteista. Kotuksen verkkosivuilla osoitteessa [http://www.kotus.fi/aineistot/puhutun\\_kielen\\_aineistot/sukukielten\\_vahalieriokokoelmat](http://www.kotus.fi/aineistot/puhutun_kielen_aineistot/sukukielten_vahalieriokokoelmat) on näytteitä ja kattavat kirjallisuusviittein varustetut luettelot Kettusen tallentamista vepsän, liivin ja vätjän vahalieriöäänitteistä.



Kuva 4. E. N. Setälä vuonna 1916 Korvoilan kylässä vepsäläisten luona haastattele-  
massa 11-vuotiasta Mark-poikaa. Taustalla näkyy äänittämisessä käytetty parlografi  
ja vahalieriötä. Valokuvaaja A. O. Väisänen. Kuva Museoviraston kokoelmista.

## 7. Etymologiset aineistot

**Etymologinen arkisto** sisältää suomen sanaston historiaa sekä kieli-  
tieteellisesti että asia- ja kulttuurihistoriallisesti selvittävää aineistoa.  
Aineistoa on käytetty *Suomen kielen etymologisen sanakirjan* ja *Suo-  
men sanojen alkuperä* -teoksen toimitustyössä. Teosten ilmestyttyä  
etymologista arkistoa ei ole enää kartutettu. Erillisenä kokoelmana  
on suomen etäsukukielten sellaisten sanojen etymologioita, joille ei  
ole vastineita itämerensuomessa. Sähköisenä aineistona on käytös-  
sä **Etymologinen viitetietokanta**, joka nimensä mukaisesti sisäl-  
tää viittauksia sanojen etymologioita käsittelevään kirjallisuuteen  
(<http://kaino.kotus.fi/sanat/evita/>).

**Saamelaiskielten etymologinen tietokanta Älgu** osoitteessa <http://kaino.kotus.fi/algu/> sisältää n. 88 000 saamelaiskielten sanaa etymologisine tietoineen. Eniten tietoja on tutkituimmasta ja puhutuimmasta kielestä pohjoissaamesta (n. 35 000 sanaa). Tietokantaan on koottu suurin osa tähän asti ilmestyneestä saamelaiskieliä koskevasta etymologisesta tiedosta.

Etymologiset tiedot näkyvät tietokannassa sanojen välisinä suhteina, jotka kertovat esimerkiksi sanavastaavuuksista tai lainautumisesta. Tällä hetkellä tietokannassa on lähes 0,5 miljoonaa suhdetietoa. Tietokanta antaa tietoa myös etymologisesta kirjallisuudesta, sillä suhteen lähdetieto on aina näkyvissä.

Kotuksessa on toimitettu etymologinen sanakirja *Suomen sanojen alkuperä*, joka ilmestyi vuosina 1992, 1995 ja 2000. Sanakirjasta tehty sähköinen versio on ollut tutkijoiden käytettävissä, ja tavoitteena on, että jossakin vaiheessa sanakirja voitaisiin avata laajemmin käytettäväksi.

## 8. Variaa

Erilaisia pieniä itämerensuomalaisten kielten kokoelmia, kenttämuistiinpanoja, käsikirjoituksia ja vastaavia dokumentteja sisältyy moniin Kotuksessa säilytettäviin tutkijoiden henkilöarkistoihin. Suomen kieleen liittyviä tietoja on eniten, mutta muihin itämerensuomalaisiin kieliin liittyviä tietoja löytyy ainakin (kaikkia ei ole yksityiskohtaisesti luetteloitu) seuraavista henkilöarkistoista:

- **Elvi Erämetsän henkilöarkistossa** on kasvinnimitietoja vatjasta, karjalasta, liivistä, virosta sekä muistiinpanoja Eesti Rahva Muuseumin kasvivaritiedoista, myös Vilho Setälän muistiinpanoja vatjasta. Erämetsän käsikirjoitus kasvivarjäyksestä on digitoitu ja saatavissa arkistojen asiakaspalvelun kautta.
- **Helmi Helmisen henkilöarkistossa** on Karjalaa, Inkeriä ja Viroa koskevaa aineistoa, muistiinpanoja ja kuvia Karjalasta sekä tietoja inkerikoista ja vatjalaisista.
- **Jalo Kaliman henkilöarkistossa** on karjalaisia paikannimiä sekä sana- ja nimitietoja suomalais-ugrilaisten kansojen asuma-alueilta Venäjältä.



- **Eino Leskisen henkilöarkistossa** on karjalan kieltä koskevia lehtiartikkeleita, luetteloita, sanastokartunnan luetteloita sekä *Karjalan kielen sanakirjan* kokouspöytäkirjoja.
- **Pertti Virtarannan henkilöarkisto** on hyvin laaja ja sisältää muun muassa hänen yksityiskohtaisia muistiinpanojaan keruumatkoilta. Aineistoa on paljon monesta itämerensuomalaisesta kielestä.

Lähisukukansoihin ja -kieliin liittyvää aineistoa on myös muissa erilliskokoelmissa. Vaikkapa **Suomen murteiden sana-arkiston käsikirjoitusten** joukosta löytyy Laina Syrjäsen kirjoitus *Vatjalaisten leivän valmistuksesta* ja Paavo Siron *Vatjalaisten teurastustapoja*. Kirjallisen aineiston lisäksi löytyy valokuviakin, muiden muassa Reino Peltolan kuvia Karjalan Šungusta vuodelta 1942. Erityisesti nauhoite-arkiston aineistoihin liittyviä valokuvia on myös Museoviraston kuva-arkistossa, esimerkiksi kuvia Setälän Liivinmaan matkalta vuodelta 1912 sekä Setälän ja A. O. Väisäsen vepsäläisalueen matkalta vuodelta 1916. Myös Pertti Virtarannan laaja kenttäkeruiden yhteydessä karttunut valokuva-arkisto on luovutettu Museovirastolle.

## 9. Lopuksi

Edellä esillä olleiden aineistojen lisäksi Kotuksessa on paljon erilaisia suomen kielen aineistoja, esimerkiksi Nykysuomen sana-arkisto, Vanhan kirjasuomen arkisto, 1800-luvun suomen kokoelmat ja monia sähköisiä tekstikorpuksia. Kotuksen aineistoihin voi perehtyä pääpiirteissään Kotuksen verkkosivuilla ja Finna-hakupalvelussa (<http://www.kotus.fi/aineistot>, <https://www.finna.fi/>), mutta kiinnostuneen kannattaa vierailla arkistossa, tutustua kokoelmiin ja keskustella henkilökunnan kanssa, sillä luettelointitiedot eivät ole kattavia. Henkilökunnalla on hiljaista tietoa, joka voi auttaa tiedon lähteille, ja varsinkin moniin sisällöltään epäyhtenäisiin erilliskokoelmiin ja henkilöarkistoihin voi sisältyä materiaalia, joka on aineistoa läpikäymällä tutkijan itse löydettävä. Tutkijoilta ja muilta käyttäjiltä saatava tieto kokoelmien sisällöstä otetaan mielellään vastaan ja se voidaan liittää täydentämään ja tarkentamaan luettelointia. Myös digitointiehtotuksia kannattaa esittää. Vuorovaikutus arkiston ja asiakkaiden välillä on kaikille hyödyksi.



Kuva 5. Karjalaisnainen Reino Peltolan kuvaamana Šungun kylässä vuonna 1942.  
Kuva Kotuksen kokoelmista.

## Lähteet

- Koonnut Sofia Björklöf (Helsingin yliopisto & Kotimaisten kielten keskus)
- Ahtia, E. V. 1938: *Karjalan kielioppi. Äänne- ja sanaoppi*. Suojärvi: Karjalan kansalaisseura. Karjalan Kielen Seura julkaissut näköispainoksena 2009 ja antanut pdf-version julkaistavaksi Kotuksen aineistopalvelu Kainossa 2014. Saatavissa: <[http://kaino.kotus.fi/digitointi/pdf/Ahtian\\_sana-\\_ja\\_aanneoppi.pdf](http://kaino.kotus.fi/digitointi/pdf/Ahtian_sana-_ja_aanneoppi.pdf)>
- Ahtia, Edvard V. 2014a: *Karjalan kielioppi II. Johto-oppi*. Konehkirjutti Sirpa Hentula, skanniruičči, toimitti i lad'd'ai Martti Penttonen, ilmahpiästai Karjalan Kielen Seura r.y. 2014. Saatavissa: <[http://kaino.kotus.fi/digitointi/pdf/Ahtian\\_johto-oppi.pdf](http://kaino.kotus.fi/digitointi/pdf/Ahtian_johto-oppi.pdf)>
- 2014b: *Karjalan kielioppi III. Lauseoppi*. Skanniruičči Kotus, konehkirjutti Tuomo Salonen, toimitti i lad'd'ai Martti Penttonen, ilmahpiästai Karjalan Kielen Seura r.y. 2014. Saatavissa: <[http://kaino.kotus.fi/digitointi/pdf/Ahtian\\_lauseoppi.pdf](http://kaino.kotus.fi/digitointi/pdf/Ahtian_lauseoppi.pdf)>
- Karjalan kielen lauseoppi. Käsikirjoitus. Säilytteillä Kotimaisten kielten keskuksessa. Saatavissa: <[http://kaino.kotus.fi/digitointi/pdf/Ahtia\\_Karjalan\\_kielen\\_lauseoppi.pdf](http://kaino.kotus.fi/digitointi/pdf/Ahtia_Karjalan_kielen_lauseoppi.pdf)>
- ALE = *Atlas Linguarum Europae*. 1975–2007. Assen: Van Gorcum & Roma: Poligrafico.
- ALFE = *Atlas Linguarum Fennicarum* 1–3. *Itämerensuomalainen kielikartasto. Läänemeresoome keeleatlas. Ostseefinnischer Sprachatlas. Lingvističeskij atlas pribaltijsko-finskix jazыkov*. 2004–2010. Helsinki: SKS & Kotimaisten kielten tutkimuskeskus. Saatavissa: <<https://avaa.tdata.fi/web/avaa/-/kotus-kielikartastot>>
- Álgu. Saamelaiskielten etymologinen tietokanta. Helsinki: Kotimaisten kielten keskus. Saatavissa: <<http://kaino.kotus.fi/algu/>>
- Digitaalinen Nimiarkisto. Helsinki: Kotimaisten kielten keskus. Saatavissa: <<https://nimiarkisto.fi/>>
- Eesti-soome sõnaraamat. Viro-suomi-sanakirja* 2019. Verkkosanakirja. Päätoim. Heikki Hurttu & Margit Langemets. Tallinn–Helsinki: Eesti Keele Instituut & Kotimaisten kielten keskus. Saatavissa: <<http://www.eki.ee/dict/efi/>>
- Elomaa, Jarmo & Johanna Laakso (toim.) 1986: *Lyydiläismurteiden käänteissanasto*. LSFU IX:2, KKTKJ 44. Helsinki: Suomalais-Ugrilainen Seura & KKTK.
- Etymologinen arkisto. Helsinki: Kotimaisten kielten keskus.

- Etymologinen viitetietokanta. Viitteitä vuodesta 1966 2000-luvun alkuun mennessä ilmestyneeseen kirjallisuuteen. Helsinki: Kotimaisten kielten keskus. Saatavissa: <http://kaino.kotus.fi/sanat/evita/>
- Euroopan kielikartaston sanalipusto. Helsinki: Kotimaisten kielten keskus.
- Henkilöarkistot. Helsinki: Kotimaisten kielten keskus. Tarkempia tietoja henkilöarkistojen sisällöistä saatavissa: [http://www.kotus.fi/aineistot/tietoa\\_aineistoista/henkiloarkistot](http://www.kotus.fi/aineistot/tietoa_aineistoista/henkiloarkistot)
- Karjalan kielen sana-arkisto. Digitoidut sanaliput (alkuperäinen aineisto siirretty helmikuussa 2016 Joensuuhun Itä-Suomen yliopiston yhteyteen). Helsinki: Kotimaisten kielten keskus. [https://www.kotus.fi/aineistot/sana-aineistot/karjalan\\_kielen\\_sana-arkisto](https://www.kotus.fi/aineistot/sana-aineistot/karjalan_kielen_sana-arkisto)
- Karjalan kielen sanakirja* 1–6. 1968–2005. Toim. Pertti Virtaranta (1–3), Raija Koponen (4–6). LSFU XVI, 3–6 myös KKTKJ 25. Helsinki: Suomalais-Ugrilainen Seura & Kotimaisten kielten tutkimuskeskus.
- Karjalan kielen verkkosanakirja*. Toim. Marja Torikka, vastaava toimittaja 8.9.2010 lähtien Leena Joki. Verkkosovellus Jari Vihtari. Kotimaisten kielten tutkimuskeskuksen verkkojulkaisuja 18. Helsinki: Kotimaisten kielten tutkimuskeskus. Julkaistu 15.10.2009, päivitetty 22.2.2019. <http://kaino.kotus.fi/kks>
- Kettunen, Lauri 1930a: *Suomen murteet I. Murrenäytteitä*. SKST 188. Helsinki: SKS.
- 1930b: *Suomen murteet II. Murrealueet*. SKST 188. Helsinki: SKS.
- 1940: *Suomen murteet III. A. Murrekartasto. B. Selityksiä murrekartastoon*. SKST 188. Helsinki: SKS. Saatavissa: <http://avaa.tdata.fi/web/avaa/-/kotus-kielikartastot>. Kartat pdf-tiedostoina saatavissa: <http://kettunen.fnhost.org/?i=1>
- Laakso, Johanna (toim.) 1988: *Rückläufiges Wörterbuch des Livischen*. Anhand des Livischen Wörterbuches von Lauri Kettunen, herausgegeben von Johanna Laakso. LSFU V:2, KKTKJ 47. Helsinki: Suomalais-Ugrilainen Seura & Kotimaisten kielten tutkimuskeskus.
- 1989: *Vatjan käänteissanasto*. LSFU XXII, KKTKJ 49. Helsinki: Suomalais-Ugrilainen Seura & Kotimaisten kielten tutkimuskeskus.
- Nimiarkisto. Helsinki: Kotimaisten kielten keskus. Kokoelmätietokanta saatavissa: <http://kaino.kotus.fi/nimikokoelmat/?a=etusivu> (Katso: <https://www.kotus.fi/aineistot/nimiaineistot>)
- Sananparsikokoelmat. 2016–. Kansallisarkiston digitaaliarkisto. Saatavissa: <http://digi.narc.fi/digi/dosearch.ka?sartun=385077.KA>. 25 kunnan sananparret korpusmuodossa saatavissa: [http://kaino.kotus.fi/korpus/sp/meta/sp\\_coll\\_rdf.xml](http://kaino.kotus.fi/korpus/sp/meta/sp_coll_rdf.xml)

- SMSA = Suomen murteiden sana-arkisto. Helsinki: Kotimaisten kielten keskus. (Katso: <[http://www.kotus.fi/aineistot/sana-aineistot/suomen\\_murteiden\\_sana-arkisto](http://www.kotus.fi/aineistot/sana-aineistot/suomen_murteiden_sana-arkisto)>)
- Suomen kielen etymologinen sanakirja* I–VII. Toivonen, Y. H., Erkki Itkonen, Aulis J. Joki & Reino Peltola 1955–1981. LSFU XII. Sana-hakemiston (VII) koostaneet Satu Tanner ja Marita Cronstedt. Helsinki: Suomalais-Ugrilainen Seura.
- Suomen kielen nauhoitearkisto. Helsinki: Kotimaisten kielten keskus. Kokoelmatietokanta saatavissa: <<http://kaino.kotus.fi/naark/>>
- Suomen murteiden sanakirja* 1–8 (*a–kurvottaa*). 1985–2008. KKTKJ 36. Helsinki: Kotimaisten kielten tutkimuskeskus.
- Suomen murteiden sanakirja (a–lysmä)*. 2012–. Kotimaisten kielten keskuksen verkkojulkaisuja 30. Helsinki: Kotimaisten kielten keskus. Päivitetty 22.5.2019. Saatavissa: <<http://kaino.kotus.fi/sms>>
- Suomen sanojen alkuperä. Etymologinen sanakirja* 1–3. 1992–2000. Päätoimittajat Erkki Itkonen & Ulla-Maija Kulonen. SKST 556, KKTKJ 62. Helsinki: SKS & Kotimaisten kielten tutkimuskeskus.
- Suomi–viro-suursanakirja. Soome-eesti suursõnaraamat* 2003. Päätoim. Valdek Pall. Tallinn–Helsinki: Eesti Keele Instituut & Kotimaisten kielten keskus. Saatavissa: <<http://www.eki.ee/dict/fie/>>
- Vepsän verkkosanasto*. 2007. Helsinki: Kotimaisten kielten tutkimuskeskus. Saatavissa: <<http://kaino.kotus.fi/sanat/vepsa/>>

## Finnic materials at the Institute for the Languages of Finland

*Toni Suutari & Ulriikka Puura*

The Institute for the Languages of Finland is a central expert organization specializing in languages. Its core areas are language planning and dictionary compilation, together with related research. The Institute's tasks also include maintaining and developing linguistic archives and collections. These dictionaries are based on large scientific collections accumulated over decades, and the collections of the Institute are also widely used by researchers, students, and the general public.

The archives of the Institute cover approximately 2 200 shelf meters. There are millions of card-files and other documents and about 24 000 hours of audiovisual recordings. Some of the collections have been digitized, and corpora have been created directly in digital form. The digital collections include hundreds of millions of words and approximately 17 000 hours of audiovisual recordings that have been digitized or collected directly in a digital format.

The archives and collections of the Institute contain a great deal of lexical and onomastic data, but there are also many smaller collections containing, for example, field notes of researchers and photographs. The main collection of the Lexical Archive and the Dictionary of Finnish dialects comprises more than 8 million items of data. The Names Archive currently contains about 2.7 million original or comparable place name card-files covering Finland and its neighboring territories, all itemized by parish. The cartographic collection contains about 10 000 maps connected to the place name collections. The place name collections of the Names Archive have been converted into a digital archive. The digitization project (started in 2014) was one of the initiatives celebrating Finland's centenary in 2017.

The Audio Recording Archive holds almost 24 000 hours of audio and video recordings (about 70% is digitized), mainly Finnish dialects and contemporary Spoken Finnish, but also a significant number of languages related to Finnish (especially the Saami languages and Karelian). Recordings have been systematically made since 1959, and

the oldest recordings – wax cylinders containing, for example, Estonian, Veps, Votic, and Livonian – are from the beginning of the 20<sup>th</sup> century.

There are over 550 000 card-files in the Karelian Language Lexical Archive. All the card-files are digitized and can be found on the Institute's websites. The Dictionary of the Karelian language (1968–2005) is also available online. Other dictionaries compiled by the Institute are also now being published on the Internet, e.g., the Dictionary of Finnish Dialects, and digital versions are being created of books originally published in paper format.

Other works available online include, e.g., the Online Vocabulary of Veps, based on the manuscript of Lauri Kettunen, the Finnic Language Atlas (ALFE = *Atlas Linguarum Fennicarum*), the Etymological Reference Database, and the Etymological Database of the Saami Languages (Álgu). There are also many lexical collections available in paper format, which cover the Finnic languages, e.g., the card-files of Veps vocabulary collected by Reino Peltola.

The focus of the Institute's data services is on providing archives, collections, dictionaries, and other databases for public use as freely as possible. Users will be involved in collecting and developing data by increasing interaction and by finding modern ways of crowdsourcing.