

ÁLGU KÄYNNISTYI



Kotimaisten kielten tutkimuskeskuksessa ryhdyttiin vuonna 2000 suunnittelemaan saamelaiskielten sanojen alkuperää selvittävää hanketta. Asiasta todennäköisesti kiinnostuneille tahoille Suomeen, Ruotsiin, Norjaan, Venäjälle, Saksaan, Hollantiin ja Unkariin lähetettiin tällöin tiedustelu, jossa kyseltiin vastaanottajien halukkuutta ja mahdollisia valmiuksia tulla mukaan hankkeeseen. Viron Tartossa elokuussa 2000 olleen kansainvälisen fennougristikongressin yhteydessä pidettiin tutkimuskeskuksen aloitteesta hankkeeseen liittyvä kokous. Hanke käynnistyi vuoden 2001 alussa, ja tutkijoilta eri vaiheissa saadut kommentit otettiin huomioon sen suunnitelmassa. Esittelen hanketta käyttäen osittain hyväksi Kotimaisten kielten tutkimuskeskuksen johtokunnan vuonna 2001 hyväksymän hankesuunnitelman sanamuotoja.

Tämä uusi etymologinen ja varsin pitkäaikainen hanke aloitettiin tutkimuskeskuksessa Suomen sanojen alkuperä -teoksen eli SSA:n valmistuttua. Hankkeen nimeksi valittiin *Álgu*, joka on pohjoissaamea ja tarkoittaa alkua ja alkuperää. Hankkeessa työskentelee kolme tutkijaa: Kirsti Aapala, Eino Koponen ja Klaas Ruppel. Vuoden 2003 ajan työryhmään toi vahvistusta Ante Aikio, joka suoritti hankkeessa siviilipalvelustaan.

Saamelaiskielten etymologisen sanakirjan tarpeeseen on kiinnitetty toistuvasti huomiota niin Suomessa kuin kansainvälisestikin (ks. esim. Kulonen 1996: 48; Larsson 2001: 90). Saamelaiskielten sanaston alkuperän tutkijat ovat joutuneet tähän asti turvautumaan melko hajanaisiin lähteisiin ja etenkin lainasanojen sekä muiden uudempien sanojen etymologiaa selvitellessään poimimaan tiedot erillistutkimuksista. Tämä on vaikuttanut osaltaan siihen, että saamelaiskielten perussanastosta varsin suuri osa on edelleen vailla alkuperän selitystä. Saamelaiskielten järjestelmällinen etymologinen tutkimus tuo myös kaivattua lisävalaistusta saamelaisten historiaan ja kontakteihin, joita saamelaisilla on ollut muihin kansoihin ja kieliin. Tutkimushanke on tarpeellinen laajemminkin uralistiikan ja indoeurooppalaisten kielten tutkimuksen kannalta, sillä saamelaiskielten sanastoon sisältyy yhtä lailla vanhaa indoeurooppalaista ainesta kuin itämerensuomalaisiin kieliinkin.

Nykyisen tutkimustilanteen kartoittamisen lisäksi hanke on tarpeellinen myös siksi, että valmistuneet etymologiset sanakirjat yleensä elävöittävät sanojen alkuperän tutkimusta.

Suomalais-ugrilaisiin kieliin kuuluvia saamelaiskieliä puhutaan Norjan, Ruotsin, Suomen ja Venäjän pohjoisosissa, ja ne ovat jokaisen valtion alueella vähemmistökielisiä. Saamelaiskielten sanastoa koskeva tutkimushanke on koko uralistiikassa ensimmäinen askel pienempien kielten etymologisen tutkimuksen voimistamiseksi. On myös muistettava, että etymologinen sanakirja kuuluu jokaisen kielen perusteoksiin.

SAAMELAISKIELET

Hankkeessa on tarkoitus ottaa huomioon kaikki saamelaiskielet sen mukaan, miten niistä on saatavilla tutkimukseen soveltuvaa aineistoa. Aina ei ole suinkaan yksiselitteisesti osoitettavissa, milloin on kyse eri kielistä ja milloin saman kielen eri murteista. Vielä niinkin myöhään kuin 1980-luvulla puhuttiin varsin yleisesti yhdestä saamen kielestä, joka jakautuu eri murteisiin (esim. Korhonen 1981). Totta onkin, että saamen murteet muodostavat jatkumon, jossa vierekkäin puhuttavien kielimuotojen keskinäinen ymmärrettävyys on melko hyvä. Sama pätee kuitenkin myös esimerkiksi kiistatta kieli- eikä murreryhmänä pidettyihin itämerensuomalaisiin kieliin. Varsinkin toisistaan kaukana puhuttavien saamelaiskielten erot ovat niin suuret ja keskinäinen ymmärrettävyys niin vähäistä, että kyse on selvästi eri kielistä. Tätä näkökantaa puoltaa myös se, että usealle saamelaiskielille on kehitetty oma kirjakielensä. (Ks. esim. Aikio 1999; Hansegård 2000; Sammallahti 1998.)

Käsitykset siitä, kuinka monta eri saamen kieltä on, vaihtelevat. Melko vakiintuneen näkemyksen mukaan kieliä voidaan laskea nykyisin olevan kymmenen. Läntiseen ryhmään niistä kuuluvat etelä-, uumajan-, piitimen-, luulajan- ja pohjoissaame ja itäiseen ryhmään inarin-, koltan-, akkalan-, kildinin- ja turjansaame. (Korhonen 1981: 15–17; Sammallahti 1998: 6.)

Oma kirjakieli saamelaiskielistä on kuudella: etelä-, luulajan-, pohjois-, inarin-, koltan- ja kildininsaamalla. Puhujamäärältään suurin on Norjassa, Ruotsissa ja Suomessa puhuttava pohjoissaame. Arviot sen puhujamäärästä vaihtelevat 17 000:n ja 30 000:n välillä. Luulajansaamea käyttää muutama tuhat henkeä, kildinin-, inarin-, koltan- ja eteläsaamea kutakin 300–500. Kirjakieltä vailla olevia uumajan-, piitimen- ja turjansaamea puhuu korkeintaan muutama kymmen, akkalanansaamea ilmeisesti enää yksi.

Saamen kielistä tutkituina on pohjoissaame, ja sen nykyistä ortografiaa voidaan pitää vakiintuneimpana, vaikka se on ollut sellaisenaan käytössä vasta vuodesta 1978 alkaen. Tätä ennen Norjassa, Ruotsissa ja Suomessa pohjoissaamea kirjoitettiin eri tavoin, pääkielistä omaksuttujen kirjoitustottumusten mukaan.

KOLME OSAHANKETTA

Álgu-hanke jakautuu kolmeen osahankkeeseen. Ensimmäisenä tavoitteena on luoda etymologinen tietokanta, joka palvelee ensin hankkeensisäisenä apuneuvona ja toimii myöhemmin muillekin tutkijoille avoimena etymologisena arkistona. Tietokannan rakenteen

▷

suunnittelu ja siihen liittyvä koekäyttö ovat olleet mittava ja vaativa urakka. Tietokannan ensimmäinen versio on valmistunut, ja tietojen tallennus on voitu aloittaa kesäkuussa 2003.

Álgu-hankkeen tietokannalle ei käsittääkseni löydy juuri esikuvaa, vaan kyse on uraa uurtavasta työstä. Mainitsematta ei kuitenkaan saa jäädä Koblenzin yliopistossa meneillään oleva hanke Uralische etymologische Datenbank (UEDb), jossa pyritään Uralisches etymologisches Wörterbuch -teoksen (UEW) esittämiseen sähköisessä muodossa. Lisäksi hankkeessa pyritään tietävästi kehittämään sääntöjä, joiden avulla tietokone voi löytää automaattisesti sanoille vastineita toisista kielistä. UEDb lähtee liikkeelle olemassa olevasta painetusta etymologisesta sanakirjasta; Álgu-hankkeessa sen sijaan on tarkoitus rakentaa ensin arkisto ja laatia sen perusteella sanakirjoja.

ÁLGU-TIETOKANTA

Tietokanta on toteutettu SQL-ohjelmointikielellä. Tämän teknisen ratkaisun ansiosta tietotekniikan huimasta kehitysvauhdista joskus mahdollisesti aiheutuva tarve tietokantaohjelman vaihtoon ei tuota suuria vaikeuksia, ja tallennettuja tietoja voidaan edelleen käyttää. Tietokannan käyttöä varten on kehitetty käyttöliittymä, joka avautuu selainohjelman ikkunaan. Tallennettava kieliaines sisältää runsaasti tarkkeita ja erikoisia kirjaimia, minkä takia tietojen tallennus tietokantaan ja tietojen esittäminen käyttöliittymässä pohjautuu Unicode-koodaukseen (UTF-8).

Tietokanta on relaatiotietokanta, jossa yksiköiden eli lekseemien väliin luodaan relaatioita eli suhteita. Kahden eri kielen vastineiden välillä voi esimerkiksi olla vastine-suhde, jos kyseessä on etymologisesti sama sana (pohjoissaamen *guolli* ~ suomen *kala*) tai lainasuhde (saP *ássat* < sm. *asua*), tai kahden saman kielen sanan välillä voi olla johtosuhde (saP *álgu* < saP *álgit*) tai muuta sellaista. Aivan aluksi suhteet syötetään tietokantaan olemassa olevan kirjallisuuden perusteella, ja syöttöihin liitetään lähdetiedot. Työn edetessä työryhmän jäsenet täydentävät suhdeverkostoa oman harkintansa mukaan. Sitä mukaa kuin lekseemejä ja niiden suhteita lisätään, tietokantaan muodostuu vähitellen etymologinen verkosto. Tietokanta pystyy myös varovaiseen päättelyyn ja ehdottaa löytämiään suhteita implisiittisinä. Jos esimerkiksi tietokantaan on syötetty tiedot saP *geaidnu* ~ sm. *keino* ja saP *geaidnu* ~ eteläsaamen *geajnoe*, niin tietokanta esittää, että myös eteläsaamen ja suomen sanat olisivat etymologisia vastineita.

Tietokannan ytimen muodostavat saamelaiskielten lekseemit ja niiden väliset suhteet. Lekseemi- ja suhdetietojen tallentaminen aloitetaan parhaiten dokumentoiduista saamelaiskielistä. Helsingin yliopiston suomalais-ugrilaisessa laitoksessa on meneillään Konrad Nielsenin pohjoissaamen sanakirjan (Nielsen ja Nesheim 1932–1962) uudistamistyö, jonka yhteydessä sanakirja on tallennettu sähköiseen muotoon. Tämä aineisto on saatu Álgu-hankkeen käyttöön, ja se muodostaa tietokannan rungon. Álgu-hankkeessa on kehitetty ohjelma, joka muuntaa Nielsenin sanakirjassa ja etymologisessa tutkimuksessa-kin yleisimmin käytetyn kirjoitustavan automaattisesti nykyortografian asuun. Muunnoksen tuloksia hyödynnetään sekä Nielsenin sanakirjan uudistamistyössä että Álgu-hankkeessa.

Álgu-tietokanta on rakennettu niin, että sanan tallennus ja haku eivät ole sidoksissa yhteen ortografiaan. Ratkaisu on seurausta siitä, että saamelaiskielten ortografiat ovat

vaihdelleet ja juuri aineistoltaan rikkaimpien isojen sanakirjojen kirjoitustapa ei ole sama kuin nykyortografia. Sama koskee myös esimerkiksi Venäjällä puhuttavia etäsukukieliä, joita kirjoitetaan nykyisin kyrillisin kirjaimin, mutta joista etenkin etymologisissa yhteyksissä on käytetty suomalais-ugrilaista transkriptiota. Myös kantakielten lekseemien rekonstruktiot saattavat eri lähteissä poiketa toisistaan.

Tietokantaan voidaan myös syöttää epävarmoja tietoja, joihin liitetään kysymysmerkki. Myös negatiivisia tietoja, siis esimerkiksi tieto, että jonkin lähteen mukaan tietyt sanat eivät ole toistensa vastineita, voidaan sisällyttää tietokantaan. Jos eri lähteissä esitetään keskenään ristiriitaisia käsityksiä, nämä tiedot voidaan niin ikään syöttää tietokantaan. Näin tietokannan avulla voidaan kuvata tutkimuksen tilannetta juuri niin selkeän yksimielisenä, eriytyneenä tai ristiriitaisena kuin se todellisuudessa on.

ÁLGU-SANAKIRJAT

Toinen ja kolmas Álgu-hankkeen osahanke ovat julkaisuja: ensin ilmestyväksi on suunniteltu yleistajuinen pohjoissaamen etymologinen sanakirja, ja sitä seuraa tieteellinen kaikkien saamelaiskielten etymologinen sanakirja. Sanakirjojen päätoimittajaksi on lupautunut Pekka Sammallahhti.

Tietokannan käyttöliittymän kieli on suomi. Sanakirjojen metakielestä päätetään myöhemmin. Kaikkia saamelaiskieliä koskevan sanakirjan hakusanojen kielestä tai kielistä täytyy myös päättää. Sähköisessä muodossa olevia sanakirjatietoja voidaan hakea jouhevammin eri kielillä, eikä hakusanojen kieli ole niin ratkaiseva kuin perinteisessä sanakirjassa.

YHTEISTYÖ

Álgu-hankkeelle perustettiin tukiryhmä, johon kuuluu tutkimuskeskuksesta hankkeen työntekijöiden lisäksi sanakirjaosaston johtaja Maija Länsimäki. Tukiryhmässä ovat lisäksi edustettuina ne laitokset, joiden kanssa on sovittu yhteistyöstä: Helsingin yliopiston suomalais-ugrilainen laitos (Ulla-Maija Kulonen) sekä Oulun yliopiston suomen ja saamen kielen ja logopedian laitoksen suomen ja saamen kielen jaos (Pekka Sammallahhti).

Hankkeen työntekijät ovat laatineet noin 50 etymologista artikkelia saamelaiskulttuurin ensyklopediaan (SENC), joka tätä kirjoitettaessa on vielä viimeisteltävänä. Verkkojulkaisuna ensyklopedia ilmestyy vielä tänä vuonna. Álgu-hankkeella on lisäksi yhteyksiä Uppsalan yliopiston suomalais-ugrilaiseen laitokseen, Uumajan yliopiston arkeologian ja saamen tutkimuksen laitokseen ja Uumajan murre-, paikannimi- ja kansanperinnearkistoon (DAUM). Muita yhteistyömahdollisuuksia saattaa avautua Pohjoismaisen tutkijankoulutusakatemian (Nordisk Forskerutdanningsakademi, NorFA) rahoittaman Saamen kielen leksikografian ja kieliteknologian verkoston kautta (Saletek). Verkoston tavoitteena on edistää yhteistyötä Pohjoismaiden välillä saamen kieliteknologian alalla. Tutkimusohjelman suomalainen osapuoli on Helsingin yliopiston yleisen kielitieteen laitoksen ylläpitämä Suomen kieliteknologian dokumentaatiokeskus (FILT). ■

▷

LÄHTEET

- AIKIO, SAMULI 1999: Saamelaiskielistä ja niiden huollosta. – *Kielikello* 2 s. 26–30.
- HANSEGÅRD, NILS-ERIK 2000: Dialekt eller språk? Om de västsamiska och norrbottensfinska skriftspråken. – URSUS 7. Skriftserie utgiven av Finsk-ugriska institutionen i Uppsala.
- KORHONEN, MIKKO 1981: *Johdatus lapin kielen historiaan*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- KULONEN, ULLA-MAIJA 1996: *Sanojen alkuperä ja sen selittäminen. Etymologista leksikografiaa*. Suomi 181. Helsinki: Suomalaisen Kirjallisuuden Seura.
- LARSSON, LARS-GUNNAR 2001: Lappische Sprachforschung an der Schwelle zum 21. Jahrhundert. – *Finnisch-ugrische Forschungen* 56:79–94. Helsinki.
- NIELSEN, KONRAD – NESHEIM, ASBJØRN 1932–1962: *Lappisk ordbok. Grunnet på dialektene i Polmak, Karsjok og Kautokeino* 1–5. Instituttet for sammenlignende kulturforskning, serie B: skrifter XVII:1–5. Oslo: Universitetsforlaget.
- SAMMALLAHTI, PEKKA 1998: *The Saami Languages. An Introduction*. Kárášjohka: Davvi Girji.
- UEW = *Uralisches etymologisches Wörterbuch* 1–3. Wiesbaden: Harrassowitz 1988–1991.

Álgu: www.kotus.fi/tutkimus/saame/algus.shtml

DAUM: www.sofi.se/daum/

FILT: www.ling.helsinki.fi/filt/index-fi.shtml

Saletek: www.ling.helsinki.fi/uhlcs/saletek

SENC: www-db.helsinki.fi/saami/home.html (koartikkeleita)

UEDb: www.uni-koblenz.de/~uedb/uedb_aktuell/index.html

Sähköposti: klaas.ruppel@kotus.fi

OIKAISU

Virittäjässä 4/2003 ilmestyneessä Boglárka Straszerin kirjoituksessa Havaintoja suomen kielen prosodisten piirteiden oppimisesta tietokoneavusteisella alkeiskurssilla on toimituksessa syntynyt virhe. Kuviodien pysty akselin selityksessä hertsin lyhenne on sijoitettu väärin; kohdan tulisi kuulua »Perustaajuus (Hz) ja intensiteetti».