

SUOMEN MURTEIDEN MORFOLOGIAN TUTKIMUKSEN AINESLÄHTEISTÄ



HEIKKI HURTTA

VÄITÖKSENALKAJAISIESITELMÄ
TAMPEREEN YLIOPISTOSSA
18. MARRASKUUTA 2000

Murteiden tutkimus eli dialektologia on ollut suomen kielen tutkimuksen keskeisiä aloja jo yli sadan vuoden ajan. Suomen kielen tutkimusta on myös joskus kritisoitu dialektologian aseman vahvuuden ja tutkimuksen aineistokeskeisyyden takia, ja dialektologiaan on saatettu suhtautua aiheettoman nuivastikin.

Dialektologiassa on vanhastaan tutkittu ns. vanhaa murretta, kullekin seudulle ominaista kansanomaista puheenpartta. Suomen vanhoja kansanmurteita ei silti vieläkään tunneta perinpohjaisesti: useista laajoista aluemurteista puuttuu yhä systemaattinen kokonaisuus. Vain niukasti on harrastettu murteiden fonetiikan tutkimusta, vaikka laajat ääniteaineistot ja tietotekniikan analyysikeinot tarjoaisivat erinomaisen lähtökohdan. Siten maallikoillekin tutut käsitykset savolaismurteiden vokaaleista tai eri seutujen erilaisista puheennuoteista ovat vielä pääosin tutkimattomia.

Koska huomattava osa nykyisistä kielitieteen harjoittajista on indoeurooppalaisten kielten puhujia, voi varsinkin angloamerikkalaisen yleisen kielitieteen näkökulmasta jo suomi olla eksoottinen kieli; yksi suomen tällainen eksoottinen piirre

taas on muoto-opin eli morfologian runsaus.

Suomen murteiden muoto-opin tutkija on yleensä lähtökohdiltaan erilaisessa asemassa kuin vaikkapa tietyn aluemurteen äännehistorian tutkija: kun kiinnostuksen kohde määräytyy morfologisesti ja tutkittavana on suuri määrä erilaisia murteita, tutkija ei voi juuri harrastaa kenttätyötä vaan joutuu turvautumaan erilaisiin käytävissä oleviin aineslähteisiin.

Vanhan kansanmurteen tutkimuksen vuosikymmenten aikana on syntynyt useita eriluonteisia aineslähteitä, pääosin arkistoja. Murteentutkimuksen perinteen pituudesta johtuu, että eri aineskokoelmat ovat erilaisia: ne edustavat perustamisaikojensa ihanteita. Aineskokoelmia on myös alun perin kerätty hiukan erilaisin päämäärin, vaikka kaikissa onkin tähdätty vanhojen paikallismurteiden tallentamiseen. Siksi nykykäyttäjien olisikin tunnettava aineistot varsin hyvin, jotta he voisivat soveltaa ainekseen tarpeellista lähdekritiikkiä. Tarkastelen tässä murreaineskokoelmien käyttökelpoisuutta lähdekriittiseltä kannalta erityisesti muoto-opin eli morfologian näkökulmasta. Kuten aineistojen tähänastinen lähdekritiikki myös tämä katsaus painottuu siihen, missä määrin aineksen voi katsoa edustavan alueensa vanhaa paikallismurretta.

▷

Lähdekritiikin kannalta on huomattava myös se, että aineskokoelmat voivat tarjota runsaasti materiaalia myös muille kuin varsinaisille dialektologeille, mutta lähdekritiikin periaatteet eivät varsinkaan tälle tutkijakunnalle ole ilman muuta tuttuja. Siten aineistoja tuntematon käyttäjä saattaa pitää kaikkia tieteellisten arkistojen tietoja yhdenvertaisina, mikä voi jopa vääristää tutkimustuloksia. Aineistojen asema on myös nyttemmin eräässä suhteessa hyvin toisenlainen kuin aineistojen luojat aikoinaan olivat aavistaneet: erilaiset tietotekniset käsittelytavat merkitsevät sitä, että aineistoja voi haluttaessa käyttää hyvin monipuolisesti.

Suomen murteiden suuria arkistoja on kolme: Suomen murteiden sana-arkisto, muoto-opin arkisto ja suomen kielen nauhoitearkisto, ja merkittäviä ovat myös lauseopin arkisto ja nimiarkisto. Nämä aineistot sijaitsevat pääosin Helsingissä Kotimaisten kielten tutkimuskeskuksessa; muoto-opin arkisto on Helsingin ja lauseopin arkisto Turun yliopistossa, mutta aineistoista on kopiokokoelmia muissakin korkeakoulukaupungeissa.

Primaariaineistoa sisältävät myös kielenäytejulkaisut ja tutkijan tallentamaa omaa murreaineistoa sisältävät tutkimukset.

Kaikki nämä aineslähteet ovat tärkeitä, ja niiden lähdekritiikissä tarkastelussa olisi vältettävä jälki viisautta. Lähdekritiikki ei myöskään voi olla yksioikoista, vaan tietojen luotettavuus ja käyttökelpoisuus riippuvat kulloisestakin näkökulmasta. Samakin tieto voi yhden tutkimuksen kannalta olla luotettava, toisen tutkimuksen kannalta epäluotettava.

Murreaineistojen tausta ja aineiden käytön lähdekritiikin periaatteet ovat pitkään todennäköisesti olleet käyttäjäjoukolleen selvät. On kuitenkin ilmeistä, että vuosikymmenten aikana kertynyt aineistoja koskeva lähdekriittinen perimätieto ei ole enää

kaikkien nykyisin arkistoja käyttävien tiedossa. Tilanne on siis muuttunut: 1960-luvulle, ehkä vielä 1970-luvulle asti julkista lähdekritiikkiä käytännön tutkimusenteon kannalta tuskin sanottavasti tarvittiin. Suomen kielen tutkimus oli keskittynyt Helsinkiin, ja murrearkisto ja yliopisto olivat läheisessä yhteydessä, joten traditionaalinen lähdekritiikki oli paremmin aineistojen käyttäjien tiedossa; myös ajallinen välimatka vanhimpiinkin aineistoihin oli lyhempi. Toisaalta ajallisen välimatkan lyhemmyydestä seurasi se, että monet sisäistä kritiikkiä osakseen saaneet kerääjät vielä olivat elossa, joten hienotunteisuusyistäkään ei voinut esittää yksittäisiin kerääjiin asti ulottuvaa julkista lähdekritiikkiä.

Nykytilanteessa, jossa tutkimusala on heterogeeninen ja maantieteellisesti laajalle hajautunut, on lähdekritiikkien esiintuominen tutkimuksessa aiheellista niin itse tutkimuksen kannalta kuin pedagogisestakin näkökulmasta. Ei voine katsoa, että lähdekritiikki on vain osa tutkimusprosessia, jonka esiintuominen ei enää kuuluisi valmiiseen, siloiteltuun julkaisuun.

VIIME VUOSIKYMMENTEN LÄHDEKRITIIKKIÄ

Murreaineiston julkisen lähdekritiikin historiassa merkittävä on Päivi Rintalan väitöskirja (1972): Rintala pohtii laajahkosti Suomen murteiden sana-arkiston eli silloisen Sanakirjasäätiön arkiston käyttökelpoisuutta erityisesti johto-opin tutkimuksen kannalta. Hän päätyy perustellen pitämään eräitä kerääjiä epäluotettavina, ja hänen tutkimuksensa liitteenä on myös nimilistä näistä varottavista kerääjistä.

Useissa muissakin suomen murteiden aineistoon perustuvissa laajahkoissa tutkimuksissa on esitetty tärkeää lähdekritiikkiä. Tutkijoista mainittakoon Heikki Leskinen, Ilkka Savijärvi, Matti Punttila, Maija

Länsimäki, Jaakko Sivula, Jouko Koivisto, Lea Laitinen, Matti Mäkelä, Pentti Suihkosen ja Aila Mielikäinen.

Tutkimuksissa esitetyssä lähdekriittisessä keskeisellä sijalla on ollut spontaanista puheesta tai lause-esimerkistöstä peräisin olevien tietojen tärkeys: sitä korostavat ainakin Punttila (1985), Koivisto (1990), Laitinen (1992), Mäkelä (1993) ja Suihkonen (1994). Mäkelä korostaa spontaanien esimerkkien tärkeyttä varianttien yleisyys-suhteiden selvittelyn kannalta juuri siksi, että »kokoelmien kerääjillä ja ilmeisesti äännehistorioitsijoillakin on taipumus kirjata harvinaisehkoja tyyppejä». Toinen keskeinen teema on ollut kyselyiden avulla saadun aineksen käyttökelpoisuuden pohdinta kriittisessäkin sävyssä: kyselyitä ovat käsitelleet Länsimäki (1987), Laitinen, Punttila ja Sivula (1989).

MUOTO-OPIN ARKISTO

Suomen murteiden morfologisen aineksen kannalta keskeisin arkisto on Muoto-opin arkisto (MA). Sen aineisto on signumisysteemin mukaan morfologisesti ryhmitelty, mikä tekee käytön helpoksi. Arkistoa ovat 1960-luvulta alkaen kartuttaneet kielitieteellisen koulutuksen saaneet kerääjät. Paradoksaalisesti siis suomen vanhojen kansanmurteiden muoto-opin kannalta jäsentynein aineisto on varsin nuorta.

Muoto-opin arkiston aineksen voi jakaa lähdekriittikin kannalta kolmeen ryhmään: – ensinnäkin spontaanista puheesta, monesti nauhastakin poimitut useimmiten kontekstilliset sananmuodot, joiden yhteydessä usein on mainittu myös oppaan nimikirjaimet, – toiseksi kontekstittomat muodot, joita ei erikseen ole mainittu kysymällä tai provoimalla saaduiksi; tällaisia muotoja on varsinkin teemamuotojen sarjoissa, – kolmanneksi muodot, jotka on yksi-

selitteisesti ilmoitettu kysytyiksi tai provoimoiduiksi.

Lähdekriittisesti epävarmin on toinen ryhmä, jonka edustajia on varsinkin vanhimmassa aineistossa. Keruun alkuaikoina kerääjät ilmeisesti ovat pyrkineet saamaan talteen täysiä paradigmoja, mutta aina ei ilmene, ovatko ne kyselyjä vai spontaanista puheesta poimittuja.

Tutkijoiden esittämät lähdekriittiset huomiot liittyvätkin provosoituun aineistoon: Pentti Suihkosen tutkimuksessa (1994) mukana on ensi sijassa spontaani aineisto, ja teemoja on vain primaarimateriaalin tueksi. Provosoinnin ongelmiin kiinnittää huomiota myös Länsimäki (1987). Aila Mielikäinen on kartastossaan (1991) ilmoittanutkin käyttäneensä ensisijaisesti spontaanin esimerkkistön muotoja; provosoituja tai kyselemällä saatuja hän ei ole käyttänyt lainkaan. Punttila (1985) taas huomauttaa MA:n aineksen nuoruudesta. — Toisaalta ainakin tapauksittain — esimerkiksi *askel*-tyyppisten nominien kannalta — taas morfologian monimuotoisuuden kattava tavoittaminen on vaikeaa pelkkien lause-esimerkkien turvin.

Aineiston käytön näkökulmasta kiintoisaa on, että Muoto-opin arkisto on ongelmattomampi lauseopin tutkijan kuin morfologin kannalta: syntaktikko saa arkistosta käyttöönsä hyvin järjestettyjä esimerkkilauseita eikä voi muuta käyttääkään, morfologi taas joutuu pohtimaan erilaisen aineksen asemaa arkiston materiaalissa.

SUOMEN MURTEIDEN SANA-ARKISTO

Murrearkistoista laajin on Suomen murteiden sana-arkisto, aiempi Sanakirjasäätiön arkisto, joka on Suomen murteiden sanakirjan keskeinen lähde. Se käsittää noin 8 miljoonaa lippua, ja sitä on alettu kartuttaa E. N. Setälän alkuunpanemaa sanakirja-

▷

hanketta varten jo 1800-luvulla. Koska aineisto on kerätty nimenomaan sanakirjaa varten, on ollut tarpeen ottaa huomioon monenlaisia kielellisiä näkökulmia: tarkasteltavan sanan kuvaukseen kuuluu taivutusmorfologiankin tarkastelu, mutta murteiden leksikografiassa on otettava huomioon myös monia muita tekijöitä: äännevariantit, levikki ja ennen kaikkea merkitys. Lisäksi arkiston kehityksessä on vaikuttanut Wörter und Sachen -tutkimusperinne, niin että eräät liput ovat etupäässä kansatieteellisiä, jopa kokonaan vailla kielenaineiksia. Tällaisetkin liput ovat käyttökelpoisia sanan levikkiä ja merkitystä selvitellessä.

SMSA:n aineisto jakautuu kahteen pääryhmään, yhtäältä kielitieteellisen koulutuksen saaneiden stipendiaattikerääjien, toisaalta vapaaehtoisten vastaajien ainekseen.

Stipendiaattikerääjät keräsivät keruun alkuaikoina pitäjänsanastoja aakkosellisen Keräilysanaston avulla. Sanojen kokoaminen tällaisten listojen avulla osoittautui ymmärrettävästi epäonnistuneeksi ratkaisuksi. Tunnetusti juuri näin kerättyjä stipendiaattikokoelmia pidetään suurelta osin epäluotettavina; sellaisia on enin osa siitä aineistosta, jota Päivi Rintala ilmoittaa käyttäneensä vain erityisestä syystä. Myöhemmät stipendiaatit ovatkin keränneet aineistojaan toisella tavoin. Tuoreimmat stipendiaattikokoelmat ovat 1970-luvulta. Uusien stipendiaattikokoelmien aines on varsin luotettavaa, joskin ne joiltain osin saattavat herättää morfologian tutkijassa samantapaisia kysymyksiä kuin muoto-opin kokoelmat: niissäkin on ilmoitettu pitkiä ja monimutkaisiakin paradigmoja, joihin sisältyvien muotojen spontaanisuus jää avoimeksi.

Vapaaehtoiset vastaajat ovat enimmäkseen kielestä kiinnostuneita maallikoita, jotka ovat vastanneet Sanastaja-lehden kysymyksiin ja lähettäneet myös spontaaneja tiedonantaja.

Kyselyin kerätyn aineksen yhtenä ongelmana on se, että kyselyiden sanamuoto on voinut vaikuttaa vastaajiin eri tavoin. Vastauksessa voi olla juuri kyselyssä ollut määräsanan asu, vaikka murteessa käytettäisiinkin yleensä erilaista asua. Kyselyssä voi olla myös syöttinä (termi Länsimäeltä 1987) esimerkkilauseita, joiden asu voi vaikuttaa niin esimerkkien morfologiaan, syntaksiin kuin sanastoonkin. — Vapaaehtoisten vastaajien aines voi lisäksi vertautua murre- ja slangikirjallisuuteen siinä mielessä, että se voi olla äänne-, muoto- ja sanastopiirteiltään murretta, tihennetytkin murteellista, mutta syntaktisesti yleiskieltä. Morfologin näkökulmasta vapaaehtoisten vastaajien itsenäisesti sepittämiä esimerkkilauseita voinee kuitenkin pitää luotettavina, ellei murteen merkintä joissain tapauksissa jää liian karkeaksi.

Vapaaehtoisten vastaajien tiedoissa ei myöskään aina ole selvää, mikä on murretta. Selvää on, että jos lippu sisältää sanan merkityksen yleiskielisen luonnehdinnan, ei morfologian tutkijalle lipusta ole hyötyä, vaikka sen yleiskieliseen tekstiin sisältyisikin tarkasteltavan sanan muotoja. Lipussa annettu sellainen kontekstion nominatiivi, joka voisi olla myös yleiskielessä mahdollinen, ei sekään välttämättä ole murretta. Ongelmallisia ovat tapaukset, joissa osa sanalipun sisällöstä on selvästi tunnistettavaa yleiskieltä, osa taas yhtä selvää murretta. Yhden pulmaryhmän muodostavat liput, jossa merkityksenselite on käännetty yleiskielestä murteelle: voiko taivutusopin tutkija käyttää tällaisiin murremuotoisiin selitteisiin sisältyviä sanamuotoja?

Julkista lähdekritiikkiä harjoittaneista tutkijoista Lea Laitinen (1992) katsoo, että SMSA:n tiedot — ainakin kyselyin kootut — ovat morfosyntaksin tutkimuksen kannalta usein epäluotettavia. Heikki Leskinen (1970) suhtautuu SMSA:n aineistoon morfologian tutkimuksen näkökulmasta varsin

kriittisesti: »Kokonaisen sana-arkiston seurlonta on kyllä kunnioitettava suoritus, mutta morfologille se tuottaa tavallisesti uskomattoman laihan saaliin.» Suhtautuminen SMSA:n ainekseen näyttää kuitenkin riippuvan tutkijan alasta. Leksikologien asenne on myönteisempi: Sivula (1989) pitää kyselyin saatua materiaalia virhelähteistä huolimatta hyvänä, ja samoin tuntuu katsovan Länsimäki (1987). Huomiota on kiinnitetty myös kokoelmien eri osien eroihin.

Koska SMSA on karttunut hyvin pitkän ajan kuluessa ja hyvinkin erilaisten kerääjien voimin, on lähdekritiikin asema sen käytössä hyvin keskeinen. Sisäinen lähdekritiikki on murreanakirjan toimittamista varten ollut jo varhain välttämätöntä, ja arkistosta on nykyisin käytettävissä myös 45-sivuinen opas, joka on välttämätön kaikille arkiston käyttäjille, eikä kenenkään pitäisi mennä itsekseen käyttämään arkistoa. Toisaalta on otettava huomioon, että vaikka aineiston käyttäjien on hyvä tietää ns. varottavista kerääjistä, on tällaisten kerääjien luetteloinnissa kuitenkin ongelmansa. Voi käydä niin, että varottavista kerääjistä tuleekin kartettavia kerääjiä ja näiden kerääjien lähettämä hyväkin aineisto jää käyttämättä.

SMSA on oikein käytettynä korvaamaton. Leksikologisiin tutkimuksiin ja myös sellaisiin muita kielen tasoja käsitteleviin tutkimuksiin, joiden aineisto(a) voidaan kerätä leksikaalisin perustein, se sopii mainiosti. Jos siis morfologisen ilmiön aineisto on käsiteltävissä tavalla tai toisella sanalähtöisesti (taajoja mallisanoja, verrattain harvoihin sanoihin rajoittuva ilmiö), on Suomen murteiden sana-arkisto varsin käyttökelpoinen.

LAUSEOPIN ARKISTO

Turun yliopistossa luotu lauseopin arkisto on useissa tapauksissa morfologian tutkijalle hyödyllinen, koska siinä voidaan

tehdä sanaesiintymittäisiä hakuja murreaastattelujen spontaanista puheesta. Pulmana on kuitenkin aineiston pienuus: jos tutkittavana on pientaajuinen ilmiö, ei sen esiintymiä Lauseopin arkiston aineistoon kovin paljon satu.

PAINATTEET JA NAUHOITTEET

Murrenäytejulkaisuista erityisen hyödyllisiä leksikaalisehtosten ilmiöiden tutkijalle ovat ne, joihin sisältyy sanahakemisto. Vanhemmissa julkaisuissa ei tosin sanahakemistoa aina ole ollenkaan, ja systemaattisia, kaikki tekstin sanaesiintymät käsittäviä sanahakemistoja on vain uusimmissa julkaisuissa.

Joitain murrekirjoja on luettavuuden parantamiseksi kaunisteltu poistamalla toistoja ja takelteluja. Morfologian tutkijaa tämä ei kuitenkaan ratkaisevasti haitanne. Enemmän voi haitata se, että joissakin kirjoissa on jätetty merkittävää sandhi-ilmiötä, jopa loppukahdennuskin.

Hyödyllisiä ovat myös primaariaineistoon perustuvat murremonografiat. Niissäkin on samoja ongelmia kuin SMSA:n ja MA:n kokoelmissa: esimerkkilauseet sisältänevät spontaania aineistoa, mutta yksittäiset sananmuodot jäävät luotettavuudeltaan avoimiksi.

Suomen kielen nauhoitearkiston aineisto on erityisen arvokasta siksi, että käyttäjä pääsee itse kuulemaan murteenpuhujan tuotoksen. Morfologian tutkijan on kuitenkin hankala turvautua nauhoiteaineistoon, jos hän on tutkimassa pientaajuisia ilmiöitä kaikkien suomen murteiden kannalta: nauhoista poimiminen vaatisi hyötyynsä nähden kohtuuttoman työmäärän.

LOPUKSI

Arkistojen murreainesta on yleensä lähdetty keräämään tiettyyn päätarkoitukseen: mur-

▷

resanakirjan laadintaan tai morfologian taikka syntaksin tutkimukseen. Käytännössä kuitenkin SMSA, MA ja LA sopivat monenlaiseen tutkimukseen, vaikka alkuperäinen tarkoitus varsinkin SMSA:n ja MA:n vanhemmassa aineistossa voi rajoittaa käyttöä muihin tarkoituksiin: tarkoitan tässä SMSA:n kansatieteellisesti painottunutta aineistoa ja MA:n kontekstittomia teemoja.

Väitöskirjani aineisto antaa pääpiirteisen kuvan siitä, miten morfologinen materiaali jakautuu eri aineslähteiden kesken. Lähdekritiikin jälkeen käyväksi katsomani aineisto käsittää tietokannassani runsaat 49 000 tietuetta. Niistä vähän yli puolet on Suomen murteiden sana-arkiston ja runsas kolmannes muoto-opin arkiston aineistoa. Kielennäytteistä ja tutkimuksista tehtyjen poimintojen osuus on n. kuusi prosenttia, ja muita tietoja on niukasti. Aineistosta vähän alle 20 000 on kontekstillisiä esiintymiä. Tässä lause-esimerkkien aineistossa SMSA:n aineksen osuus on pienempi (vähän alle puolet) ja poimintojen osuus suurempi (runsaat 10 prosenttia).

Määräsuhteet ovat arkistojen kokoon katsoen odotuksenmukaiset. Samalla on todettava, että minkään aineiston pois jättäminen ei olisi hevin mahdollista, jos murteiden ilmiöistä halutaan saada edustava kuva: Suomen murteiden sana-arkisto ja Muoto-opin arkisto ovat niin mittavia, et-

tei niitä käy ohittaminen; toisaalta poimintot ja pienistä ainesryhmistä myös Lause-opin arkisto tarjoavat runsaasti juuri erikoisen tärkeää spontaanin puheen aineistoa, niin että niitäkin täytyy pitää tutkimuksen kannalta välttämättöminä.

Morfologin näkökulmasta eri arkistojen aineksen käyttökelpoisuus riippuukin tutkimusaiheesta. Jos aihe käsittelee selviä sanaryhmiä tai on käsiteltävissä mallisanojen avulla, on helppo käyttää Muoto-opin arkiston lisäksi muitakin aineslähteitä; muissa tapauksissa taas Muoto-opin arkiston merkitys korostuu entisestään.

Runsas aineisto kontrolloi itse itseään, ja tämä onkin ollut vanhastaan paljon käytetty lähdekritiikin muoto. Aineistoon liittyvä yleinen lähdekriittinen tieto lienee kuitenkin tutkijalle hyödyksi, vaikka jokaisen tiedon luotettavuuden hän joutuukin viime kädessä punnitsemaan itse. Aineslähteiden taustan ja historian tunteminen on ainakin välttämätöntä.

Jos murrearkistoja kerättäisiin nyt, olisi tähdennettävä mahdollisimman tarkkaa lähteiden ilmaisemista, niin että kaikista koulutettujen kerääjien aineksista kävisi ilmi, mikä on spontaania, mikä kyselyä. Tähän liittyy huomattava osa validiteetin ja reliabiliteetin ongelmista ja myös virhelähteistä. Vastaisuudessa tärkeää on murteiden korpusten kehittäminen. ■

HEIKKI HURTTA *Suomen murteiden askel-tyyppisten nominien morfologiaa ja murremaantiedettä*. Opera Fennistica & Linguistica 9. Tampereen yliopiston suomen kielen ja yleisen kielitieteen laitos.

Kotimaisten kielten tutkimuskeskus, Sörnäisten rantatie 25, 00500 Helsinki
Sähköposti: *heikki.hurta@kotus.fi*