

MIKSI SUURET SANAKIRJAT PITÄÄ TEHDÄ?

Näinä aikoina, jolloin valtion niin sanottua tuottavuusohjelmaa toteutetaan eli resursseja systemaattisesti vähennetään, joudutaan kielentutkimuslaitoksosakin perimmäisten kysymysten äärelle. Mikä on tärkeää, mitä pitää tehdä? Mikä on vähemmän tärkeää, mitä voi vähentää tai mistä voi luopua kokonaan? Pitäisikö keskittyä nykyhetken tarpeisiin? Pitäisikö suurten, pitkäkestoisten — etenkin varhempia kielimuotoja kuvaavien — sanakirjojen tekeminen kyseenalaistaa?¹

Kaikkien kulttuurien perustana on kieli. Kielen keskeiset kuvausvälineet ovat kielioppi ja sanakirja. Kielioppi on rakenteiden kuvaus, sanakirja on sanaston kuvaus ja inventaario. Molemmat ovat yhtä välttämättömiä.

Euroopan sivistysvaltiot ovat uhranneet paljon aikaa ja varoja kieltensä sanaston kattavan inventaarion rakentamiseen. Svenska Akademiens ordbokin (1893–) valmistuminen vaatii kokonaisuudessaan ainakin 120 vuotta; tähän mennessä on ilmestynyt 34. osa hakusanaan *trivsel* asti. Jacob ja Wilhelm Grimmin *Deutsches Wörterbuch* ilmestyi 106 vuoden aikana (1854–1960). Oxford English Dictionary ilmestyi 10-osaisena 1884–1928, 20-osaisena 1989, ja uuden, entistä laajemman laitoksen on määrä ilmestyä 2010. Maailmanennätys on tietävästi hollannin kieltä kuvaavalla sanakirjalla *Woordenboek der Nederlandse Taal*, jota on tehty 150 vuotta (43 osaa 1863–1998; kolme täydennysosaa 2001).

Näiden kielten kirjallisella viljelyllä on pitkät perinteet. Niiden sanaston ään-

teellisestä ja semanttisesta kehityksestä on vanhojen kirjallisten lähteiden avulla luotavissa hyvä kuva. Suomen kielen monipuolinen kirjallinen viljely alkoi varsinaisesti 1800-luvun puolimaissa. Suomen sanaston kehitys ja historia on nähtävissä murteissa ja niiden sanastossa. Myös nykysuomen sanasto rakentuu yhä edelleen suurelta osin murteiden sanastolle ja sen edelleenkehitykselle. Lisäksi on huomattava, että kirjakielen sanasto on vain osa murteiden kokonaissanastosta.

Suomen kielen sanakirjatarpeet hahmotettiin viimeistään 1890-luvulla E. N. Setälän nimeä kantavassa sanakirjaohjelmassa. Setälä esitti Suomalaisen Kirjallisuuden Seuran kokouksessa 12.2.1896, että suomalaisen sivistyselämän tarpeisiin tulisi laatia kolme sanakirjaa: 1) murteiden sanavaraston mahdollisimman tyhjentävästi esittävä sanakirja, 2) tuonaikaisen kirjakielen sanakirja ja 3) Ruotsin vallan aikaisen kirjoitetun kielen sanakirja. Ohjelman suurin vajavuus on se, että siitä puuttui varhaisnykky-suomen, kirjakielen tärkeimmän kehitysvaiheen, sanakirja. Vasta tuona aikanahan vallattiin kulttuuri- ja sivistysala toisensa jälkeen suomen kielelle. Setälä ilmeisesti ajatteli, että tuo varhaisnykky-suomen sanasto tulisi esitellyksi kirjakielen sanakirjassa, kuten osittain tapahtuikin. Perusteellisempi 1800-luvun sanaston kartoitus on kuitenkin yhä tarpeen. Myöhemmin, 1910-luvulla, Setälä vielä käynnisti sanojemme alkupe-
rää kuvaavan eli etymologisen sanakirjan työstämisen.

¹ Tämä kirjoitus on syntynyt vuorovaikutuksessa Kotimaisten kielten tutkimuskeskuksen, Kotuksen, sanakirjantekijöiden kanssa.



ERI KIELIMUODOISTA OMAT SANAKIRJANSA

Vain vajaat parikymmentä vuotta ennen sanakirjaohjelman esittämistä oli valmistunut Elias Lönnrotin suurtyö, Suomalais-ruotsalainen sanakirja (1866–1880). Se sisältää kirjakielen — käytössä olleen sanaston sekä uudissepelmien ja vanhentuneiden sanojen — lisäksi huomattavan määrän murteista peräisin olevaa mutta silloisen kirjakielen kuusiin mukautettua sanastoa. Lönnrotin sanakirja kuvaa suomen kieltä yhtenä kokonaisuutena, monoliittina. Setälän kolmen sanakirjan ohjelmassa korostuu puhutun ja kirjoitetun kielimuodon ero ja myös se, että kirjoitettuun kieleen pohjautuva yleiskieli oli jo eriytynyt omaksi kielimuodokseen, jota vakiinnutettiin ja jonka sanastoa voimaperäisesti kehitettiin.

Edellä mainitut neljä sanakirjaa — nykykielen sanakirja, vanhan kirjakielen sanakirja, murteiden sanakirja ja etymologinen sanakirja — ovat sivistyskielen välttämätön perusrustus. Ohjelmasta on kokonaan toteutettu jo osa. Nykysuomen sanakirja, joka sai vauhtia eduskunnan aloitteesta 1927, ilmestyi kuusiosaisena puoli vuosisataa sitten (1951–1961), ja se sisältää runsaat 200 000 hakusanaa. Jatkuvasti sanastoltaan uusiutuvan ja vähäiseltä osin myös normistoltaan muuttuvan kirjakielen tarpeisiin on sen jälkeen ilmestynyt kaksikin noin 100 000 hakusanaa sisältävää sanakirjaa: Suomen kielen perussanakirja (1991–1994, sähköinen versio 1997) ja Kielitoimiston sanakirja (sähköiset versiot 2004 ja 2008, painettu 2006). Kirjakielen luonteen vuoksi tarvitaan melko lyhyin väliajoin päivitetty versio, eli kirjakielen sanaston kehitystä tulee havainnoida jatkuvasti.

Kirjakielen sanakirjojen tapaan valmiiksi asti on edennyt myös etymologinen sanakirja. Ensimmäisen version, Suomen kielen etymologisen sanakirjan (1955–1981), laadinta aloitettiin Turun yliopiston yhteisen

perustetussa tutkimuslaitos Suomen suvussa (1930–). Uusi, hiukan toisin toimitettu ja tuoreimman tutkimuksen tulokset esittävä Suomen sanojen alkuperä ilmestyi viime vuosituhatlupussa (1992–2000). Myös etymologinen sanakirja vaatii alan tutkimuksen seuraamista ja ajoittain uusimmat tutkimustulokset noteeraavan päivityksen.

Ruotsin vallan aikaista kirjakieltä (1500-luvulta noin vuoteen 1810) kuvaavan Vanhan kirjasuomen sanakirjan varsinainen artikkelityö aloitettiin 1970-luvun alussa. Kuusiosaiseksi suunnitellusta sanakirjasta on valmistumassa kolmas osa (1. osa 1985).

Tämän maan todennäköisesti suurin humanistisen tieteenalan hanke on Suomen murteiden sanakirja. Juuri tätä hanketta toteuttamaan oli aikanaan (1924) perustettu Sanakirjasäätiö. Sanakirjasta on tähän mennessä ilmestynyt kahdeksan osaa (1. osa 1985). Kaikkiaan osia on tulossa 20, ja niiden on tarkoitus ilmestyä noin kolmen vuoden välein.

Nämä neljä sanakirjaa ja joukko muita tärkeitä sanakirjahankkeita ovat sitemmin kuuluneet ja kuuluvat edelleen Kotimaisten kielten tutkimuskeskuksen perustehtäviin. Sanakirjat mainitaan ensimmäisinä myös sen lain perusteluissa, jolla tutkimuskeskus 1976 perustettiin.

Vanhan kirjasuomen sanakirjan ja Suomen murteiden sanakirjan käyttöikä tulee olemaan pitkä, ja nämä sanakirjat tehdään vain kerran. Syynä on se, että niiden kuvaama kielimuoto on määritelty ajallisesti ja aineistoltaan rajalliseksi. Molempia on luonnehdittu historiallisiksi, vaikka Suomen murteiden sanakirjan kuvaamat murteet ovat edelleen eri määrin aivan elävää kieltä.

YKSIKIELISET JA KAKSIKIELISET SANAKIRJAT

Tutkimuskeskuksen toimialaan kuuluvat suomi, ruotsi, saamen kielet, romanikieli ja

viittomakieli; laissa mainitaan vielä erikseen suomen sukukielten tutkimus. Kirjakieltä kuvaavan laajan yksikielisen ruotsin kielen sanakirjan laadinta on Ruotsin tehtävä. Sen sijaan Suomen ruotsalaismurteita kuvaava Ordbok över Finlands svenska folkhäl kuuluu tutkimuskeskuksen tehtäväkenttään. Seitsenosaisesta sanakirjasta on tähän mennessä ilmestynyt neljä osaa (1. vihko 1976). Suomen sukukielten sanakirjoista ovat valmiina kuusiosainen Karjalan kielen sanakirja (1968–2005), viisiosainen Mordwinisches Wörterbuch (1990–1998) ja Tschermisssches Wörterbuch (2008). Wogulisches Wörterbuch on valmistumassa. Mordvan, tšeremissin ja vogulin murteita kuvaavien sanakirjojen toimitustyö aloitettiin muualla (Suomalais-Ugrilaisen Seuran stipendiaattivoimin), ja tutkimuskeskus otti vastatakseen näiden lähinnä tutkimuksen käyttöön tarkoitettujen sanakirjojen loppuun saattamisesta. Sukukielten sanakirjoista on vuonna 2010 käynnistymässä pohjoissaamen yleistajuinen etymologinen sanakirja. Se perustuu Kotuksessa tehtyyn ja edelleen täydennettävään tietokantaan, johon sisältyy kaikkien saamelaiskielten aineistot.

Vähemmistökielistä on laadittu yhteistyössä Kuurojen Liiton kanssa Suomalaisen viittomakielen perussanakirja (1998) sekä kaikille avoin ja ilmainen verkkosanakirja Suvu (2003). Romanikielestä on julkaistu Viljo Koiviston laatimat Romano–finitiko–angliko laavesko liin (1994) ja Suomi–romani-sanakirja (2002).

Ruotsi on Suomen toinen virallinen kieli, minkä vuoksi velvoitteemme sitä kohtaan ovat suuremmat kuin muita kieliä kohtaan: tarvitsemme Suomessa aina laadukkaan, kattavan ja ajantasaisen suomalais-ruotsalaisen ja ruotsalais-suomalaisen sanakirjan. Siksi Kotuksessa on laadittu Suomen kielen perussanakirjan pohjalta yhteistyössä WSOY:n kanssa Suuri suomi–ruotsi-sanakirja (1997), josta tehdään uusia

ajantasaisia painoksia. Saman tietokannan pohjalta on yhteistyössä Eesti Keele Instituutin kanssa toimitettu Suomi–viro-suur-sanakirja (2003). Molemmista sanakirjoista on myös sähköiset versiot. Laajan ajantasaisen virolais-suomalaisen sanakirjan pohjaa valmistellaan Eesti Keele Instituutissa. Työ vaatii ja tulee vaatimaan yhteiskunnan tukea sekä Virossa että Suomessa.

ITSENÄISET, TOISIAAN TÄYDENTÄVÄT SANAKIRJAT

Kotuksen suuret sanakirjat ovat itsenäisiä, mutta niillä on yhteys toisiinsa ja ne ovat toisiaan täydentävä kokonaisuus. Vanhan kirjasuomen sanakirja ja Kielitoimiston sanakirja (sekä sen edeltäjät) kuvaavat eri aikojen kirjakieltä; 1800-luvun sanastoa kattavasti kuvaava sanakirja täydentäisi sarjan ajalliseksi jatkumoksi. Aikajänne ulottuu kehittyvästä ja kehitettävästä, vapaasti varioivasta kielimuodosta nykyiseen normikieleen. Historia liittyy kirjakielen sanakirjat Suomen murteiden sanakirjaan: kirjakielemme perustuu murteisiin, ja sitä on eri vaiheissa rakennettu eri murteiden pohjalta. Nämä sanakirjat kuvastavat suomen kielen nykyisyyttä ja menneisyyttä, luontaista vaihtelua ja eri aikoina tietoisesti tehtyjä ratkaisuja. Historialliset sanakirjat ovat tärkeitä nykykielen ja nyky-yhteiskunnan sekä niiden taustan ja kehityksen ymmärtämiseksi.

Murteet ovat olleet lukemattomien sukupolvien ainoa ajattelun, ilmaisun ja yhteydenpidon kieli. Suurten murre-sanakirjojen sivuilta välittyvät kansanhimisen aineellinen ja henkinen perintö, hänen kokemuksensa, tunteensa, ajatuksensa, asenteensa, arvomaailmansa ja suhteensa ympäristöönsä. Vanhan kirjasuomen sanakirja taas kuvaa 1500–1700-lukujen kirjallista kielenkäyttöä, suomenkielisten tekstilajien alkua ja alkuvaiheita. Tuo kielimuoto syntyi sekä

▷

hengellisen että maallisen vaikuttamisen ja vallanpidon välineeksi. Vanhan kirjasuomen sanakirjan kuvaama uskonnon ja lainkäytön kieli on paljolti käännöskieltä, mikä yhdistää vanhan kirjakielen muihin eurooppalaisiin kulttuurikieliin ja länsimaiseen sivistykseen. Tässä suhteessa myös ruotsalaismurteidemme sanakirja täydentää suomen kielen sanakirjoja: se on kaksikielisessä maassa tärkeä apuväline molemmipuolisten kulttuurivai- kutteiden ymmärtämiseen.

Kulttuurisesti ja kansallisen identiteetin kannalta tärkeät suuret sanakirjahankkeet perustuvat valtavaan aineistotyöhön: toisaalta poimintoihin painetuista lähteistä sekä toisaalta puhutun kielen kenttäkeruusiin ja laajoihin systemaattisiin kyselyihin. Nykysuomen sanakirjan aineistona oli 4,5 miljoonan sanalipun arkisto. Sen perillisen Kielitoimiston sanakirjan aineistoa kartutetaan päivittäin lehdistä ja kirjallisuudesta tehtävin poiminnoin sekä käyttämällä hyväksi tekstikorpuksia ja hakukoneita. Vanhan kirjasuomen sanakirjan pohjaksi koottiin aikoinaan puolimiljoonainen sanalippuaineisto. Nykyään on käytössä myös varsin kattava sähköinen tekstikorpus, josta sana-artikkelien aineisto etsitään hakuohjelmien avulla. Korpuksen koostamistyö on tehty Vanhan kirjasuomen sanakirjan toimitustyön ohessa. Suomen murteiden sanakirjaa tehdään noin 8 miljoonan sanalipun uniikista aineistosta, jollaista nykyoloissa olisi enää mahdotonta tavoittaa. Toisen puolen tästä aineistosta ovat vuosikymmenten aikana koonneet koulutetut ja toisen puolen harrastajakerääjät. Noin 7 000 harrastajakerääjän talkoohengessä tekemä työ on sekä kulttuuritekona että tulosten kannalta ainutlaatuinen maailmassa.

Suuret sanakirjahankkeet vaativat pitkäaikaisen resursoinnin, eivätkä ne ole kultakaivoksia, ei edes hyvin kaupaksi käyvät Kielitoimiston sanakirja. Siksi Suomen kaltaisessa pienessä maassa niiden tekeminen

on mahdollista vain valtion rahoituksella — käytännössä Kotimaisten kielten tutkimuskeskuksessa.

KORVAAKO AINEISTO SANAKIRJAN?

Pitkäkestoinen työ on vuosien mittaan silloin tällöin virittänyt kysymyksen, onko lainkaan tarpeen tehdä Vanhan kirjasuomen sanakirjaa ja Suomen murteiden sanakirjaa. Onhan olemassa aineistot, jotka ovat tutkijoiden ja muiden kiinnostuneiden käytettävissä. Aineisto ei kuitenkaan ole sanakirja. Sanakirja-artikkeli syntyy ammattitaitoisen leksikografin työnä aineiston semanttisen ja moninaisen muun analyysin perusteella, ja esimerkiksi Suomen murteiden sanakirjan arkiston lähdekritiikki on vaativaa. Jos sanakirjan vaihtoehtona olisi aineisto, pitäisikö tutkimustenkin sijaan tarjota pelkkä aineisto?

Vielä edessä oleva pitkäaikainen työ on herättänyt semmoisenkin ajatuksen, että sanakirjatyön keskeyttämistä tulisi harkita. Tämä ajatus on Kotuksen taannoisessa arvioinnissa (Kotuksen arviointi s. 46) puettu hiukan tulkinnanvaraiseen muotoon: »Sanakirjatyö sitoo resursseja siinä määrin, että uusien sanakirjojen käynnistäminen ja erityisesti historiasta kumpuavan pitkäkestoinen sanakirjatyön 'katkaiseminen' tulee tehdä yhteisymmärryksessä opetusministeriön kanssa.» Mikä olisi sanakirja, joka käsittäisi vain aakkosvälin *a–n*? Se olisi torso, haaskatun rahan ja tärvätyn työn muistomerkki.

Toimitustyön pituus asettaa työille, varsinkin kuvattavan kielimuodon ja hankkeen keston kannalta, monenlaisia haasteita — tietysti jo siksi, että kuvattava kielimuoto ja etenkin sen edustama maailma käyvät sanakirjan toimittajille aina vieraammiksi. Toisaalta tietotekniikan kehitys tarjoaa aivan uusia mahdollisuuksia: esimerkiksi

suomen kielen sanakirjat tehdään nykyisin Kotuksessa rakenteisessa muodossa. Työtapoja ja sähköisiä apuneuvoja on kehitetty ja kehitetään kaiken aikaa toimituksissa, yhteistyössä tietotekniikan asiantuntijoiden kanssa, keskeisenä tavoitteena myös synergiaedut. Eräiltä osin työ nopeutuu, mutta aineiston tulkinta — semanttinen analyysi ja sanaston muu kuvaus — vaatii saman ihmisäivoin tehtävän työn kuin ennenkin.

Tekniikan kehitys on johtanut siihen, että perinteisten painettujen sanakirjojen lisäksi on laadittu sähköisiä sanakirjoja. Pitäisikö entistä enemmän tai jopa kokonaan siirtyäkin sähköiseen julkaisemiseen? Kielitoimiston sanakirja on hyvä esimerkki siitä, että molemmille on kysyntää ja molempia tarvitaan: painettu ja sähköinen versio eivät ole toisensa pois sulkevia vaihtoehtoja, vaan niillä on erilaisia ominaisuuksia, ja ne täydentävät toisiaan. Kotuksen sisällä on käytössä myös Suomen murteiden sanakirjan kaikkien ilmestyneiden kahdeksan osan sähköinen korpus, jossa on jo nyt runsaasti hakuominaisuuksia. Sähköinen versio on suunnitteilla myös Vanhan kirjasuomen sanakirjasta.

Kielitoimiston sanakirjan sähköinen versio on kustannussopimuksen mukaisesti maksullinen. Murre-sanakirjan sähköisen version julkiseen käyttöön saattaminen edellyttää sopimista painetun kirjan kustantajan kanssa. Kotuksen intressissä tietenkin olisi, että sen kaikki sanakirjat leviäisivät mahdollisimman laajaan käyttöön. Toiveen vastikkeettomasta käytöstä toteuttaa aivan lähivuosina Karjalan kielen sanakirja, josta paraikaa valmistetaan vapaasti käytettävää verkkoversiota.

SANAKIRJOJEN UUDET YHDISTÄMISMAHDOLLISUUDET

Leksikografisen työn päämääränä on valmis sanakirja. Kotuksen suuret sana-

kirjahankkeet tarjoavat mahdollisuuksia myös erilaiseen hyödyntämiseen ja jatkopalostamiseen. Kuten Suomen kielen perussanakirjan tietokannasta muokattiin suomalais-ruotsalaisen ja suomalais-virolaisen sanakirjan pohja, Kielitoimiston sanakirjasta voidaan muokata esimerkiksi kaksikielisten sanakirjojen pohjia maahanmuuttajia varten. Työ vaatii tietysti paitsi resursseja myös yhteistyötä näiden kielten asiantuntijoiden kanssa.

Sähköiset sanakirjat tarjoavat jo nyt mutta varsinkin tulevaisuudessa monenlaisia liittämismahdollisuuksia. Sanakirjat on mahdollista linkittää paitsi toisiinsa myös vaikkapa korpuksiin, tutkimusjulkaisuihin ja kielenhuollon aineistoihin; jossain vaiheessa mukaan voitaisiin kytkeä myös käsittehierarkiat. Kiehtovia näköaloja tarjoaa sanakirjan tai sanakirjojen liittäminen kielioppiin eli kielen rakenteen ja sanaston kuvausten yhdistäminen. Suomen murteiden sanakirjaan voidaan lisätä esimerkiksi levikkikartat (joita on jo nyt sähköisessä muodossa yli 32 000), kuvia ja piirroksia sekä ääntä (murreäänitteitä). Kotus on jo alustavasti suunnitellut yhdessä Suomalais-Ugrilaisen Seuran kanssa sanakirjaportaalina, jota silmällä pitäen Mordvan sanakirjaa saatetaan parhailaan seuran työnä rakenteeseen muotoon ja johon myös Karjalan kielen sanakirjan sähköinen versio on tarkoitus sijoittaa.

Tulevaisuudessa siintelevän sanakirjojen verkkotaivaan rakentamiseksi vaaditaan paljon työtä. Haaveen toteuttamiseksi on suuret sanakirjat tehtävä valmiiksi, aivan loppuun asti. Se on kansallinen kulttuuri-tehtävä. ■

MAIJA LÄNSIMÄKI

Kotimaisten kielten tutkimuskeskuksen
Sanakirjaosaston johtaja

Sähköposti: maija.lansimaki@kotus.fi

▷

LÄHTEET

Kotuksen arviointi = *Kotimaisten kielten tutkimuskeskuksen (Kotus) arviointi*. Opetusministeriön työryhmä-

muistioita ja selvityksiä 2008:6.
http://www.minedu.fi/OPM/Julkaisut/2008/kotimaisten_kielten.html?lang=fi&extra_locale=fi.