

maan teoksen alusta loppuun, mutta sillä on myös tarjottavaa sieltä täältä tietoa poimivalle lukijalle.

Joitakin kirjan puutteita on toki myös mainittava. Teoksen alaotsikoiden hierarkia väliin uuvuttaa (*ja*-konjunktioita käsittelevässä toisessa luvussa viiden eri tason otsikkoja!), mutta toisaalta tarkkaa sisällysluetteloa voikin sitten käyttää muuten puuttuvan hakemiston korvikkeena. Toinen lukijaa paikoin häiritsevää puute on runsaan esimerkitön numeroimatta jättäminen. Tekijä joutuukin joskus viittaamaan tarkoittamiinsa esimerkkeihin mm. seuraavasti: »Kolmanneksi viimeisessä esimerkissä on joka tapauksessa kyse – ja toiseksi viimeisessäkin luultavasti –» (s. 36).

Rinnastusten normatiivinen luokittelu hyväksyttäviin ja virheellisiin tapauksiin ei mielestäni kytkeydy luontevasti kopulatiivikonjunktoiden historiaan, vaihteluun ja rinnastustyyppien rakennekuvauksiin. Kielenhuoltajan näkökulma kuitenkin pilkahtelee esiin silloin tällöin (vaikkakin melko har-

voin) Pulkkinen kirjan sivuilta, ehkä tuimimmin parikonjunktioilla *sekä – että* rinnastamista käsiteltäessä, vaikka tekijä muistaa toki myös itse korostaa normatiivisen lähestymistavan erillisyyttä (s. 14).

Tässä on tullut puheeksi vain joitakin yksityiskohtia Pulkkinen teoksesta. Olen pyrkinyt niiden avulla havainnollistamaan kirjan ominaisuuksia yleisemminkin sekä antamaan kuvan Pulkkinen tutkimusotteesta aiheeseensa. Toivottavasti on käynyt selville, että tuo ote on vahva. Rinnastuksesta, konjunktioista ja niiden käytöstä mutta myös laajemmin puhutun suomen variaation ja kirjoitetun kielen standardien yhteiselosta ynnä kirjakielemme historiasta kiinnostuneille Virittäjän lukijoille suosittelen tutustumista tähän kopulatiivisen rinnastuksen mainioon käsikirjaan. ■

JYRKI KALLIOKOSKI

Suomen kielen laitos, PL 3 (Fabianinkatu 33), 00014 Helsingin yliopisto

TILASTOTIETOJA SUOMEN MURTEIDEN SANASTOSTA

Raimo Jussila – Erja Nikunen – Sirkka Rautoja *Suomen murteiden taajuussanasto. A frequency dictionary of Finnish dialects.* Kotimaisten kielten tutkimuskeskuksen julkaisuja 66. Kotimaisten kielten tutkimuskeskus, Helsinki 1992. XXIX + 319 s. ISBN 951-37-0850-0.

Suomen kielen taajuussanasto (1979) on saanut rinnalleen merkittävän tilastollisen sanaston: Suomen murteiden taajuussanaston. Suomen murteiden taajuussanasto (jäljempänä MTS) selvittää suomen murteiden sanojen yleisyyttä. Aineistona on 16 Suomen kielen näytteitä -sarjan sanastoa. Näyt-

teistä kahdeksan edustaa länsimurteita ja kahdeksan itämurteita. Näytteiden valinta on pyritty tekemään tasaisesti eri murteet huomioon ottaen. Suomen murteiden taajuussanaston aineisto koostuu 207 636 saneesta ja 11 129 eri sanasta. Sanaston esikuvana on käytetty Suomen kielen taajuussanastoa

(STS), ja tuloksia on verrattu STS:n tuloksiin.

Suomen murteiden taajuussanasto ja kaantuu kahteen osaan: yleisyysjärjestysluettelo ja aakkosjärjestysluettelo. Yleisyysjärjestysluettelo sisältää kaikki ne sanat, joita aineistossa on ollut vähintään kaksi. Aakkosellinen luettelo sisältää normaalissa aakkosjärjestyksessä kaikki yleisyysluettelossa esiintyvät sanat. Yleisyysluettelossa on järjestystä ilmaisevan luvun ja hakusanan lisäksi ilmoitettu sanan absoluuttinen kokonaisfrekvenssi, haje, suhteellinen frekvenssi, summaprosentti ja pitäjäkohtaiset yleisyystiedot. Hakusanakentässä on merkittynä myös sanaluokka ja tarvittaessa homonymianumero. Aakkosluettelo puolestaan sisältää hakusanan lisäksi yleisyysjärjestystä osoittavan luvun, absoluuttisen kokonaisfrekvenssin ja pitäjäkohtaiset absoluuttiset frekvenssit. Aakkosluettelosta puuttuvat tiedot suhteellisista frekvensseistä, jotka ehkä olisivat mahdollistaneet jonkinasteisen pitäjittäisen vertailun. Pitäjätiedot on luetteloissa esitetty siten, että ensin on länsimurteiset aineistot etelästä pohjoiseen (Hin, Vel, ATa, Ika, Lop, Hol, Kal, ATo) ja sitten itämurteiset pohjoisesta etelään (Suo, Sot, Rau, Jou, Ker, Lip, Kur, Mol).

Kuudestatoista pitäjänsanastosta kootun aineiston hakusanointus on yhdenmukaistettu muuttamalla äänne-, muoto- ja johto-opillisesti varioiva aineisto tutkimuksen kannalta tarkoituksenmukaisiksi abstrakteiksi perusmuodoiksi. Murre on karkeistettu Suomen murteiden sanakirjan toimitusohjeiden mukaisesti. Koska tavoitteena on ollut luoda puhekielinen ja murteellinen aineisto STS:n frekvenssitietoihin vertailtavaksi, on hakusanointuksessa noudatettu paljolti STS:ssa käytettyä tapaa, jota taas ohjasivat Nykysuomen sanakirjan hakusanointus- ja oikeinkirjoitusperiaatteet. Aineiston erilaisuuden vuoksi MTS:n periaatteet kuitenkin eroavat joistakin STS:n laadinnassa käytetyistä. Esimerkiksi johto-opin osalta tehtäisiin murteille

väkivaltaa, jos kaikki johdostyytit (esim. *kokoontua* ~ *kokointua* ~ *kokoutua* ~ *kokou-da*) yhdistettäisiin yhdeksi yleiskieliseksi variantiksi. Samoin vartalon ja johdinaineksen liittymän äänteellinen variaatio (esim. *avanto* ~ *avento*) on mielekästä ottaa hakusanointuksessa huomioon.

Murteiden taajuussanaston johdannossa (s. xi) todetaan, että 16 pitäjän otos on liian pieni taajuuserojen murrealueelliseen tutkimiseen. Korkeintaan itä- ja länsimurteita koskevia sanastopäätelmiä voi tehdä. Myöskään alueelliselta edustavuudeltaan aineisto ei ole paras mahdollinen, koska omaleimaisista eteläpohjalaismurteista ei ole yhtään näytettä aineistossa. Murreaineisto on myös sanamäärältään heterogeeninen: sanoja Alatornion näytteessä on vain kaksi kolmasosaa Kalajoen näytteen sanoista, ja Alatornion näytteessä saneita on vain puolet Moloskovitsan saneista. Herää kysymys, miksi Alatornion näyte on otettu edustamaan peräpohjalaisia murteita, kun vielä lisäksi näytteessä mainitaan olevan 3 183 hakusanointuksen ulkopuolelle jäänyttä tapailua ja epäselvää esiintymää. Laajuudeltaan MTS:n aineisto on noin puolet STS:n aineistosta. Kahden taajuussanaston vertailussa tulee kuitenkin ongelmia, vaikka sanastojen koostamisen peruslähtökohdat ovatkin samat. Tulosten tulkittavuuteen vaikuttaa nimittäin erilainen otostekniikka: murteiden otosyksikkönä on noin 6 488 saneen laajuinen litteroitu näyte, kun taas yleiskielen otosyksikkönä on viiden virkkeen ja noin 60 saneen tekstikatkelma. STS:n otostekniikka antaa mahdollisuudet aihepiiriltään moninaisten näytteiden mukaan ottamiseen, kun taas MTS:n aineisto rajoittuu aihepiiriltään vanhakantaiseen maalaisyhteisöön. Jälkimmäinen seikka johtuu puolestaan murteiden nauhoituksessa käytetyistä yleisistä periaatteista: kysellään vanhaan maalaiskulttuuriin liittyvistä asioista.

Taajuussanaston johdannon mielenkiintoisin luku on Tuloksia, jossa selväsanaisesti ja selkeiden taulukoiden avulla vertaillaan

yleis- ja kirjakielen (STS) ja murteiden (MTS) taajuussanastoja. Kirjoitetun kielen suurempi tiiviys ja toisaalta myös otosyksikkö koko vaikuttavat siihen, että yleiskielessä keskitaajuusluku (= saneiden lukumäärä yhtä sanaa kohti) on 7,03, kun taas murteissa keskitaajuusluku on 18,7. Sanaluokkajakaumat kuvastavat myös puhekielen ja kirjallisperäisen kielen eroja: puheessa pronominit, adverbit ja konjunktioit ovat huomattavasti yleisempiä (esim. pronomineja murteissa 16,43 %, yleiskielessä 8,42 %). Substantiiveja ja adjektiiveja yleiskielessä (S 35,83 %; A 9,22 %) on kaksi kertaa niin paljon kuin murteisissa (S 16,31 %; A 4,62 %). Verbejä käytetään kummassakin yhtä paljon.

Frekvenssien mukaan tarkasteltuna sekä STS:n että MTS:n aineistossa on yhden keran esiintyviä sanoja noin puolet. Yli 20-esiintymäisten sanojen osuus on ilmeisesti otostekniikan vuoksi murteissa jonkin verran suurempi kuin yleiskielessä. Murteissa saattaa olla taajafrekvenssistenkin sanojen käytössä alue-eroja (esim. *että* on lännessä sijaluvuilla 5–11, idässä 6–687, kun taas *jotta* lännessä sijaluvuilla 0–1 223, idässä 0–36). Sanaston kattavuudessa murteet ja yleiskieli eroavat toisistaan huomattavasti: murteisissa jo 25 yleisintä sanaa kattaa puolet tekstistä, mutta yleiskielessä samaan yltää vasta 350 sanaa. Murteissa tarvitaan vain 400 yleisintä sanaa kattamaan 80 % tekstistä, yleiskielessä yli 3 000 sanaa. Kirjoitettu kieli näyttää olevan tekstisanastoltaan rikkaampaa kuin puhuttu kieli, jos rikkaudella tarkoitetaan eri sanojen suurta määrää.

Kiinnostavaa on verrata, minkälaisia ovat sata yleisintä sanaa STS:ssa ja MTS:ssa. Kuten aiemmin totesin, sanaluokkajakauman mukaan aineistot eroavat siinä, että puhekieli käyttää enemmän pronomineja, adverbeja ja konjunktioita, yleiskieli taas substantiiveja ja adjektiiveja. Sadan yleisimmän sanan osalta sanaluokkaero näkyy selvästi vain substantiivien (STS 15, MTS 10) ja adverbien (STS 13, MTS 29) määrässä. Jos tarkastellaan sa-

nojen merkityksiä, murteet ja yleiskieli eroavat selvästi sanastoltaan, yhteisistä, runsasfrekvenssisistä sanoista huolimatta. Vertailen seuraavassa STS:n ja MTS:n sataa yleisintä sanaa. Näiden perusteella voi päätellä jotakin yleistä paitsi suomalaisen maailmankuvasta ennen kaikkea myös sanastojen pohjana olevista aineistoista ja niiden erilaisista keruutavoista.

Murteiden yleisimmät substantiivit (kummankin sanaston yhteisten, ihmisille tärkeiden sanojen *aika, mies, päivä* ja *vuosi* lisäksi) ovat *hevonen, talo, kerta, vesi, lehmä* ja *poika*. Yleiskielessä suomalaisen yleisimmät substantiivit ovat yhteisten sanojen lisäksi *asia, maa, ihminen, osa, Suomi, työ, tapa, Jumala, kysymys, lapsi* ja *mieli*. Murteita tallennettaessa on nimenomaan kerätty vanhaa talonpitoon liittyvää perinnettä, joten maaseudun elämässä ja kielessä tärkeät *hevonen, lehmä* ja *talo* ovat päässeet frekvenssilistan kärkeen, samoin elämää eteenpäin vievät *vesi* ja *poika*.

Yhteisten adjektiivien (*suuri, hyvä, sellainen*) lisäksi murteissa sadan ensimmäisen sanan joukkoon ovat päässeet maalaiselämän kuvaukseen liittyvät adjektiivit *semmoinen, vanha, iso* ja *pitkä*. Yleiskielen tärkeitä adjektiiveja taas ovat *oma, uusi* ja *pieni* (sekä *eri* ja *koko*). Pronomineista yhteisten pronomien (*se, ne, minä, me, toinen, mikä, joka, hän, tämä, muu*) lisäksi murteissa käytetään runsaasti pronomineja *tu* ja *nuo*. Yleiskielessä runsaasti esiintyvä *eräs* on murteissa todella harvinainen. Konjunktioista selvimmän yleiskielisiä ovat *tai, sekä* ja *koska*, murteellinen puolestaan *jotta*. Deiktiset adverbit (*siinä, siitä, siihen, sinne, tuolla, tuossa*) puheyhteyteen kuuluvina ovat murteelle ominaisia, yleiskielisiä puolestaan ovat *myös, kuitenkin* ja *siis*.

Verbit, joita kummassakin taajuussanastossa on lähes yhtä paljon, ovat merkitykseltään erityyppisiä murteissa ja yleiskielessä. Yhteisiä yleisiä verbejä ovat *olla, ei, tulla, mennä, antaa, saada, sanoa, pitää, ottaa,*

tehdä, käydä, nähdä, jäädä ja tietää. Yleiskielessä kolmanneksi yleisin verbi *voida* on murteiden sanastossa vasta sijalla 246. Yleiskielen puhuja *saa* kaksi kertaa enemmän kuin *antaa*, murteiden puhuja *saa* melkein neljä kertaa enemmän kuin *antaa* — mikä onkin itsekeskeiselle ihmiselle ymmärrettävää. Näiden verbien perusteella näyttäisi siltä, että minään päin suuntautuva liike olisi ensisijainen. Tätä käsitystä vahvistaa vielä se, että tuleminen (*tulla*) on yleiskielessä kolme kertaa yleisempää kuin meneminen (*mennä*). Sen sijaan murteissa taas minästä pois päin suuntautuva liike näyttää yhtä yleiseltä kuin minään päin suuntautuva, jos tarkastellaan *tulla*- ja *mennä*-verbejä. Jos kuitenkin laskeetaan mukaan murteissa hyvin yleinen *lähteä*-verbi, liike minästä pois päin voittaa selvästi minään päin suuntautuvan liikkeen. Inhimilliseltä tuntuu kuitenkin murteenpuhujia, joka kaksi kertaa mieluummin *ottaa* kuin *antaa*. Yleiskielen puhuja sen sijaan mieluummin *antaa* kuin *ottaa*. Yleiskielen *tehdä, käyttää* ja *suorittaa* saavat murteissa yleisimmin vastineen *tehdä*. Yleiskielessä *aletaan*, murteissa *ruvetaan*. *Sanoa*-verbin lisäksi yleiskielessä yleisiä ovat *mainita* ja *esittää*, murteissa käytetään yleisimmin *sanoa*-verbiä. *Saattaa, tapahtua* ja *joutua* kuuluvat yleiskielen, *panna, viedä, syödä* ja *päästä* murteen yleisimpään sanastoon.

Suomen kielen taajuussanasto ja sen perusaineistona oleva Oulun korpus ovat olleet pohjana mm. A student's Glossary of Finnish -sanakirjassa, Nykysuomen sanakirjan Etymologisessa sanakirjassa, Pauli Saukkosen tyylintutkimuksissa ja Antero Niemikorven sanastotutkimuksissa. Kun Suomen murteiden taajuussanastoa verrataan STS:oon, voidaan tehdä monenlaisia päätelmiä puhutun ja kirjoitetun kielen sekä kansankielen ja yleiskielen välisistä eroista ja yhtäläisyyksistä. Vaikka sanastojen otostekniikka eroaakin ja MTS:n aineisto on laajuudeltaan puolet STS:n aineistosta, yleisimpien erojen ja yhtäläisyyksien tutkimiseen MTS:n aineisto on

riittävä. Samoin yleisimpien sanojen osalta sanojen vertailu on relevanttia. Murteiden taajuussanasto on heti ilmestyessään osoittanut tarpeellisuutensa: se on niin murteen ja sen sanaston tutkijalle kuin myös puhutun ja kirjoitetun kielen tutkijalle oiva lähde vastaisiin tutkimuksiin. Uskon, että tällainen suuren työn vaatinut sanasto saa runsaasti käyttäjiä ja on tulevina vuosina aineistona lukuisissa tutkimuksissa. Toivon kuitenkin, että murteista tehtäisiin vielä alueellisesti kattavampi taajuussanasto, jossa kaikki pienemmätkin murrealueet olisivat edustettuina. Tällainen lienee kuitenkin mahdollista vasta sitten, kun Suomen kielen näytteitä -sarja on valmis. ■

HELENA SULKALA

Oulun yliopisto, Suomen ja saamen kielen laitos, PL III, 90571 Oulu

LÄHTEET

- BRANCH, MICHAEL – NIEMIKORPI, ANTERO – SAUKKONEN, PAULI 1979: A student's glossary of Finnish. WSOY, Porvoo.
- NIEMIKORPI, ANTERO 1990: Suomen kielen sanaston frekvenssianalyysia. Vaasan Korkeakoulun julkaisuja. Tutkimuksia No 150. Kielitiede 18. Vaasa.
- 1991: Suomen kielen sanaston dynamiikka. Acta Wasaensia No 26. Kielitiede 2. Vaasa.
- SAUKKONEN, PAULI – HAIPUS, MARJATTA – NIEMIKORPI, ANTERO – SULKALA, HELENA 1979: Suomen kielen taajuussanasto. A frequency dictionary of Finnish. WSOY, Porvoo.