



Kansakunnan muistia elvyttävää synergiaa

Raimo Jussila

Keväällä 1998 ilmestyi Suomalaisen Kirjallisuuden Seuran kustantamana suomen vanhimman kirjakielen ensiesiintymien luettelo *Vanhat sanat*. Tämä sisällöltään sinänsä yksinkertainen teos (esittämiskaava: hakusana – ensiesiintymävuosi – lähde; esim. tietokirja 1738 Forseen) liittyy pieneenä palana laajaan kansainväliseen tutkimushankkeiden rypäeseen ja sen aineisto avaa uusia mahdollisuuksia tutkia kvantitatiivisesti suomen kielen sanastoa.

Synergia, yhteisvaikutus, on ilmiö, jossa kokonaisuus on enemmän kuin osiensa summa. Sana on alkuaan kreikkaa (*synergi'a* = myötävaikutus, apu). Teologiassa termi tunnetaan uskonpuhdistuksen jälkeisestä oppiriidasta, jossa luterilaiset tähdensivät kääntymisessä Jumalan armoaikutusta ja melanchtonilaiset synergistit ihmisen vapaan tahdon merkitystä. Tämä synergismi tuomittiin vuonna 1577.

Lääketieteessä synergia on samansuuntaisesti vaikuttavien lääkeaineiden yhteisvaikutus, joka voi olla tehokkaampi kuin kunkin aineen erillisvaikutuksien summa. Synergistejä ovat sellaiset lihakset, jotka toimivat samaan suuntaan tai pyrkivät vaikuttamaan samoin.

Kielitieteessä synergia-käsitettä on käytetty erityisesti saksalaisessa kvantitatiivisessa sanastontutkimuksessa. Professori G. Altmann käynnisti Bochumin yliopistossa 1980-luvun lopulla kansainvälisen tutkimushankkeen Language Synergetics. Sen tavoitteena on selvittää kvantitatiivisesti kielen leksikon sisäistä säännöstöä ja rakennetta sekä eri tekijöiden yhteisvaikutuksia. Erityyppisiä kieliä vertaamalla pyritään löytämään leksikon universaaleja piirteitä. Tätä nykyä tutkimusta johtaa professori Reinhard Köhler Trierin yliopistosta.

Tutkimushankkeessa ovat mukana seuraavat 12 kieltä: arabia, friisi, heprea, indonesia, korea, maori, puola, saame, saksa, suomi, unkar, vanha englanti. Suomen ja saamen osuuden yhteistyökumppaneina ovat olleet Kotimaisten kielten tutkimuskeskus ja Oulun yliopiston suomen ja saamen kielen laitos. Jokaisen mukana olevan kielen leksikosta koottiin ja kootaan osittain edelleenkin hakusanoittain mm. seuraavia tietoja: tavuluku, sanaluokka, taivutusparadigma, morfologinen status, funktio, merkitysten lukumäärä, kirjakielinen ikä, etymologia ja frekvenssi. Hankkeessa tutkitaan näiden eri tekijöiden riippuvuutta toisistaan. Tuloksia on julkaistu erityisesti sarjassa Quantitative Linguistics.


Suomen kielen osuus perustuu suureksi osaksi Tuomo Tuomen Suomen kielen käänteissanakirjaan. Sen noin 207 000 hakusanaa käsittävä perusaineisto oheistietoineen olivat jo olemassa ATK-muotoisina. Pauli Saukkosen ja kumppanien Suomen kielen taajuussanastosta saatiin sanojen frekvenssitiedot. Esimerkiksi sanojen funktioiden ja kirjallisen iän määrittämistä varten käynnistettiin tutkimusprojekti Oulun yliopistossa ja Kotimaisten kielten tutkimuskeskuksessa, koska valmiita aineistoja näiltä osin ei ollut.

Tämän kirjoituksen teema sisältyy leksikon synergiatutkimuksen suomen kielen sanojen kirjakielistä ikää koskevaan osaan. Tavoitteena on iätä jokainen suomen kirjakielen sana eli etsiä sanan vanhin painettuun kirjallisuuteen tai käsikirjoitukseen sisältyvä ensiesiintymä.

Vanhoiden sanojen osuus nykysuomessa pieni mutta keskeinen

Urakka oli etukäteenkin arvioiden ylivoimaisen laaja, mutta se liittyi luontevasti tutkimuskeskuksen perustehtäviin: tutkimuksen perusaineistojen hankkimiseen, sanakirjatyöhön ja kielen kehittämiseen. Liikkeelle päätettiin lähteä siten, että ajoitetaan ensin sanaston vanhimmat kerrostumat ja tyydytään nuorien sanojen osalta likimääräiseen ikäykseen. Toisin sanoen kun on saatu dateeratuksi sanat, jotka ovat tulleet kirjakieleen ennen vuotta 1900, jäljellä olevat sanat voidaan merkitä 1900-luvun sanoiksi. Urakka jaettiin kolmeen osaan. Omana työskäntäni on ollut ns. vanha kirjasuomi eli Ruotsin vallan aika 1543–1810. Eri sarkana on ollut 1800-luvun suomi. 1900-luku on jätetty ainakin tässä vaiheessa tarkemman selvitystyön ulkopuolelle. Ensiesiintymien pioneerityö oli tehty jo aiemmin. Akateemikko Martti Rapolan (1891–1972) keskeisin tutkimusala oli vanha kirjakieli ja tutkimusote filologinen. Hän käynnisti Vanhan





kirjasuomen sanakirjatyön ja sen vaatiman aineistojen keruun. Hän oli erityisen kiinnostunut sanaston historiasta. Rapola metsästi vuosikymmenten mittaan myös sanojen vanhimpia kirjallisia esiintymiä, jotka ilmestyivät valikoimana *Sanojemme ensiesiintymiä Agricolasta Yrjö Koskiseen* (SKS 1960). Hänen luettelossaan on tiedot noin 4 500 sanasta alkaen Agricolan aapisesta (1543) ja päättyen 1900-luvun alkuun. Rapolan teos on ollut Vanhat sanat -luettelon esikuvana. Uutta ensiesiintymäaineistoa ei ole ennätetty vielä käyttää varsinaisiin synergiatutkimuksiin. Joitakin havaintoja on kuitenkin aineiston kokoamisen ja julkaisun toimitustyön kestäessä tehty, ja selostan niitä seuraavassa lyhyesti. Siinä kirjakielessä, jota *Nyky-suomen sanakirja* (1951–1961) edustaa, on 25 700 sellaista sanaa, jotka on kirjattu kirjakielestämme ns. vanhan kirjasuomen aikana eli Ruotsin vallan aikana vuosina 1543–1810. Vanhassa kirjasuomessa on eri arvioiden mukaan 52 000–60 000 sanaa, joten 1900-luvulle asti eläneitä sanoja on 43–49 %. Nyky-suomen sanastosta näiden vanhojen sanojen osuus on vain noin 12 %, mutta suurelta osalta se on kieleemme keskeisintä leksikkkoa.

Yleisin sanasto vanhinta

Yleisin sanasto on samalla vanhinta niin etymologisesti (eli vuosituhaten taakse) kuin kirjallisestikin (suomen kielen osalta vuoteen 1543 saakka). Esimerkiksi tuhannesta yleisimmästä sanasta vain 1 % on peräisin 1900-luvulta, 15 % 1800-luvulta ja 84 % vanhan kirjasuomen ajalta. Nyky-suomen sadasta yleisimmästä sanasta peräti 93 on kirjattu jo 1500-luvun lähteistä. Laskelmat on tosin tehty 1960-luvun kielestä, jossa tuhannen yleisimmän sanan joukkoon kuuluvia 1900-luvun sanoja ovat vain todeta, auto, tilanne, taulukko, Neuvostoliitto, osallistua, menetelmä, tavoite, Tšekkoslovakia, radio ja elokuva. Tästä luettelosta ovat romahtaneet Neuvostoliitto ja Tšekkoslovakia, ja tuhannen yleisimmän joukkoon ovat tunkeuneet ainakin televisio ja tietokone, mutta vanhojen sanakerrostumien osuus lienee edelleen yhtä vahva kuin 1900-luvun alkupuoliskon suomessa.

Kirjakielemme varhaisvaiheille on ominaista, että uutta sanastoa ilmestyy sysäyksittäin. Jotkin vuosikymmenet nousevat vahvasti esiin. Toisaalta on on pitkiä pysähdyksen kausia, jolloin julkaistaan vähän tai pelkästään uusia painoksia. Kirjakielemme alku 1543–1552 nousee esiin mahtavana huippuna Mikael Agricolan yli kaksituhatsivuisen tuotannon takia. Sata vuotta myöhemmin 1640-luvulla kirjakielemme sai vahvan sanastoruiskeen erityisesti ensimmäisen kokoraamatun ansiosta (Biblia 1642). Daniel Jusleniuksen sanakirja (1745) taas sata vuotta myöhemmin on sanastomme suuria lähde-oksia, samoin kuin 1780-luvulla valmistunut Christfrid Gananderin *Nytt Finskt Lexiconin* käsikirjoitus. Erityisen hitaasti uusien sanojen määrä on karttunut Agricolan jälkeen 1500-luvun loppupuolella, isonvihan aikoina 1710–1720-luvuilla ja Ruotsin vallan loppuajana 1790–1810.

Sanastonkehityksen kymmenen ensimmäistä ovat ensiesiintymien mitattuna seuraavat (sulkeissa ensiesiintymien määrä): 1) Mikael Agricola 1543–1552 (5 228), 2) Christfrid Ganander 1763–1789 (4 102), 3) Daniel Juslenius 1745 (2 676), 4) Biblia 1642 (1 062), 5) Henrik Gabriel Porthan 1770– (Jusleniuksen sanakirjan välilehtiin ja sivuihin tehdyt lisäykset; 1 017), 6) Henrik Florinus 1678–1733 (667), 7) Ericus Schroderus 1637 (548), 8) kuningas Kristofferin maanlain käänös 1570– (464), 9) *Variarum Rerum Vocabula* 1644 (377) ja 10) Andreas Aspegreen 1745 (lisäyksiä Jusleniuksen sanakirjaan; 361). Tämän kärkekkymmenikön osuus ensiesiintymien määrästä on peräti 64 %. Tässä suhteessa 1900-luku poikkeaa suuresti aiemmista ajoista, yksittäisen kirjoittajan kieltä uudistava tuotanto hukkuu julkaisujen paljouteen, vaikka sanasto kasvaakin huimasti varsinkin tieteen ja tekniikan erityisaloilla.

Ensiesiintymäaineistoa koottaessa on käytetty 1 227:ää lähdeä, jotka kattavat 50–60 % tuonaikaisen suomenkielisen kirjallisuuden nimekemäärästä. Valtaosa niistä on pienpainatteita, asetuksia, kuulutuksia, juhlarunoja tms., joissa uusia sanoja ei ole lainkaan tai on vain muutama.

Kansakunta ilman muistia on dementikko

Edellä olen esittänyt muutamia hajahuomioita, joita aineistosta on ollut varsin vähällä vaivalla tehtävissä. Kvantitatiivisin menetelmin on tarkoitus selvittää sanojen kirjallisen iän riippuvuutta ja synergisia ominaisuuksia esimerkiksi frekvenssin, sanan pituuden, taivutustyyppin, sanaluokan tai merkitysten lukumäärän suhteen. Aineiston käänneiselutellon



avulla voidaan selvittää erilaisten johdostyyppien kehittymistä. Myös yhdys sanojen kehityshistorian tutkimukseen on nyt entistä parempia mahdollisuuksia.

Kirjakielten historiat poikkeavat toisistaan suuresti. Hyvä esimerkki ovat kolme samaan kielikuntaan kuuluvaa kieltä, unkari, suomi ja saame. Unkarin ensimmäiset kielenmuistomerkit ovat jo 1200-luvun alusta. Suomen kirjakielen komea alku on 1500-luvun puolivälissä. Ja Euroopan pohjoisilla ääriillä saamen kieltä sovitettiin kirjalliseen asuun 1600-luvulla. Synergeettikatutkimusten tavoitteisiin ei kuulu eri kirjakielten sanaston historioiden vertailu keskenään, vaan tältäkin alueelta pyritään löytämään leksikkoo koskevia yleisiä lainalaisuuksia. Aineistot antavat kuitenkin mahdollisuuden myös monipuoliseen ja -kieliseen filologiseen sanastontutkimukseen.

Vanhat sanat -teos soveltuu nimenomaan yksittäisten sanojen kehitysvaiheiden tutkimiseen. Siitä voi nopeasti selvittää, milloin esimerkiksi sanat *tiiede* ja *tapahtua* ovat tulleet kirjakieleemme. Koska luettelo ei tunne *tiedetä*, se on myöhemmiltä ajoilta; muista lähteistä selviää, että se on Volmari Kilpisen uudissana vuodelta 1842. Sen sijaan verbi *tapahtua* esiintyy jo Agricolaalla 1544 ja muissakin vanhoissa lähteissä. Etymologista sanakirjaa vilkaisemalla tarkentuu, että se on *tavata*-verbin johdos, joka esiintyy ainakin karjalan kielessä. Se juontaa siis alkunsa Agricolaan edeltävistä puhutuista kielimuodoista.

Sanojen ikäysprojektiin ryhdyttiin kymmenisen vuotta sitten puhtaasti kielitieteellisin perustein. Hankkeen aikana ruvettu yhä enemmän keskustelemaan käsitteestä kansankunnan muisti. Kansakunta muistaa kielen avulla myös kirjallisesti dokumentoidun historiansa taakse. Tietoa menneisyydestä on säilynyt myös kieleen itseensä, varsinkin sanastoon. Sanojen juurten tutkimus on yhteisömuistin tietoista virkistämistä.

Tutkimuslaitokseni – Kotimaisten kielten tutkimuskeskus – on tärkeä kansankunnan muistelin, sen kielikeskus, laajoine arkistoineen ja sanakirjatöineen. Etymologinen sanakirja – *Suomen sanojen alkuperä* (valmiina osat 1–2, A–P) Ä tunkeutuu tieteellisin menetelmin tuhansia vuosia kirjoitetun kielen taakse ja luo huikaisevia näkyjä menneisyyteemme. *Suomen murteiden sanakirja* (valmiina osat 1–5, A–kakuttaa) valottaa kansankielen perusteellisena inventaariona viime vuosisadan ja tämän vuosisadan alun kieltä ja elämänmuotoa. *Vanhan kirjasuomen sanakirja* (valmiina osat 1–2, A–K) esittää tyhjentävästi kirjakielemme sanastohistorian Agricolaasta alkaen. Ajantasainen *Suomen kielen perussanakirja* (1990–1994) sulautuu sekini aikanaan yhteiseen kansankunnan muistiin.

Vanhat sanat -julkaisu sisältää eräänlaisia syvyysluotaustuloksia, jotka ilmoittavat kuinka kauas sanan kirjallinen menneisyys ulottuu. Sanan todellista ikää se ei tietenkään aina osoita, ts. puhuttua kieltä on ollut tietysti ennen Agricolaakin, mutta kirjoitettua kieltä vanhempia sanoja voidaan ajoittaa vain summittaisesti etymologisin menetelmin. Sanastonluotauksissa ei voi siis päästä Agricolaan syvemmälle kuin ehkä joidenkin asiakirjakatkelmien kautta, mutta mitä lähemmäs nykyaikaa tulemme, sitä enemmän riittää ajoitettavaa. Nykyluotauksen kannalta keskeinen 1800-luku on vielä monelta osin kartoittamatta, eikä sen kauden sanakirjasta ole kuin kaukaisia haaveita. Vaikka 1900-luvulta onkin olemassa laajoja arkistoaineistoja, tietomme oman vuosisatamme uudissanoista ovat osin varsin ylimalkaisia. Lähimuisti tuntuu unohtavan erityisen nopeasti, ellei sitä virkistetä tai sen sisältöä vasiten tallenneta.

Vanhat sanat -luetteloon on koottu vuosisatojen taakse viittä tiedonmuruja noin 25 000 sanasta. Näistä muruista voidaan rakentaa menneisyyden muistikuvia ja kehityskaaria, jotka auttavat meitä entistä syvemmin ymmärtämään myös nykyistä kieltämme ja yhteiskuntaamme. Tällaista tutkimusta ei tehdä pelkästä tiedonhalusta ja uteliaisuudesta, vaan siihen pakottaa myös jokaista yksilöä pelottava uhkakuva, sillä myös kansakunta ilman muistia on dementikko.

Kirjoittaja on Kotimaisten kielten tutkimuskeskuksen erikoistutkija ja Vanhat sanat -teoksen tekijä. Kirjoitus perustuu osittain Suomalaisen Kirjallisuuden Seurassa 3.9.1998 pidettyyn teoksen esittelyyn.