



# KYSELYTESTIT SYNONYMIAN SELVITTÄMISESSÄ SANASTOTIETOUTTA KIELENPUHUILTA SÄHKÖISEEN SANAKIRJAAN

ULLA VANHATALO

VÄITÖKSENALKAJAISITELMÄ  
HELSINGIN YLIOPISTOSSA  
3. KESÄKUUTA 2005

Satuitteko seuraamaan huhtikuussa Helsingin Sanomien mielipidesivulla käytyä keskustelua siitä, ketä saa kutsua mummoksi tai vaariksi? Eräs lehden lukija oli tyrmistynyt siitä, että 60 vuotta täyttäneitä naisia mummitellaan tilanteissa, joissa lapsenlapsista ei ole tietoakaan. Toinen lukija vastasi hänelle Nykysuomen sanakirjaa lainaten, että mummohan tarkoittaa vanhaa naista. Keskustelu jatkui siitä, miten ovat latautuneet nimitykset *mummo*, *mummi*, *akka*, *eukko*, *muija*, *muori* ja *ämmä* tai *ukki*, *vaari*, *ukko* ja *äijä* tai *mukula*, *tenava*, *muksu*, *kakara* ja *pentu*.

Kirjoittajilla on sanojen sävyistä ja käytön sopivuudesta varsin varmoja näkemyk-

siä. Tietyt nimitykset ovat selvästi halventavia, sanottiin. Yhtä pontevasti väitettiin vastaan. Erään kirjoittajan mukaan suomen kieli muuttuu neljässäkymmenessä vuodessa niin paljon, että ihmisten olisi syytä tarkistaa sanatietoutta Nykysuomen sanakirjaa tuoreemmista opuksista. Keskustelu meni lopulta siihen, että esimerkiksi *mummo*-sanan käytön välttäminen nähtiin jopa huolestuttavana tarpeena silotella todellisuutta ja kieltää ihmisen vanheneminen.

Tämä pieni esimerkki avatkoon väitöstilaisuuteni. Puhumme synonyymeistä eli suunnilleen samaa tarkoittavista sanoista, niiden pienistä merkityseroista sekä ennen kaikkea siitä, miten näitä eroja voidaan tutkia ja miksi niitä kannattaa tutkia. Esimerkimmme *mummot* ja *vaarit* tuntuvat ehkä ensinäkemältä itsestään selviltä tai turhan pieniltä tutkimuksen aiheilta, mutta asian he-

rättämien tunteiden palo kertoo kuitenkin toista.

#### TEORIA VS. KÄYTÄNTÖ

Synonymiasta eli sanojen samantapaisista merkityksistä on kielitieteessä paljon teorioita, jotka ovat osin keskenäänkin ristiriitaisia. Nämä teoriat ovat kaukana tavallisesta ihmisestä, eivätkä teorioiden kirjoittajat ole useinkaan paneutuneet siihen kieleen, jota ihmiset arjessaan käyttävät. Samaan aikaan erityisesti sanakirjojen toimitustyössä, leksikografiassa, on tehty kovasti töitä synonymioiden keruun, lajittelun ja esittelyn kanssa. Tietoa on kerätty aikaisemmista sanakirjoista, ja työssä on käytetty paljon myös laajoja sanomalehtitekstikorpuksia. Jos varsinaiset synonymioiden teoriat ovatkin suurelle yleisölle tuntemattomia, on sanakirjoista tullut liki puolijumalia. Ihmisillä on tarve saada tietää, mikä kielessä on oikein, mitä sanat ihan oikeasti tarkoittavat, millaisia merkityksiä ja sävyjä ne mukanaan kantavat. Vuosikymmenien ja jopa vuosisatojen aikana sanakirjoihin kertynyt tieto edustaa kysyjälle usein lähes erehtymättömää ilmoitusta.

Otan tästä pari esimerkkiä. Esimerkit eivät liity synonymiaan, mutta kuvaavat tilannetta, jossa sanat elävät yhteiskunnassa. Toukokuussa 2005 Helsingin Sanomissa keskusteltiin *suometumisesta*, siis tästä lähihistorian ilmiöstä, jossa Suomi mukautti politiikaansa Neuvostoliiton vaikutuksen takia. Keskustelu kääntyi suometumisen merkityksen vähättelyyn tai tunnustamiseen, ja yhdeksi todistusaineistoksi otettiin — ei enempää eikä vähempää kuin sanakirja. Sanottiin suunnilleen niin, että *suomettumista* ei ole ollutkaan, koska sitä ei löydy Nykysuomen sanakirjasta. Sanakirjalle annettiin aikamoinen auktoriteetti-asema. Samoina päivinä saman lehden mielipidesivulla otettiin voimakkaasti kan-

taa toiseenkin sanastokysymykseen. Erään kotimaisen matkatoimiston konkurssin myötä alan liitto vaati, että käsitettä *valmismatka* on täsmennettävä. Matkatoimistojen, viranomaisten ja asiakkaiden näkökulmissa kun on ollut toistuvia erimielisyyksiä siitä, mitä *valmismatkaan* sisältyy ja mitä korvataan, kun jo maksettu valmismatka jostain syystä peruuntuu. Kirjoituksessa vaadittiin, että käsite määritellään kauppa- ja teollisuusministeriön määräämässä valmismatka-asiain neuvottelutoimikunnassa.

Elämme siis ajassa, jossa monet asiat perustuvat sanojen käyttöön. Sanakirjojen sisältämä tieto kohtaa ympäröivän yhteiskunnan monenlaisia odotuksia ja vaatimuksia. Sanakirjasta haetaan vastauksia hyvin erilaisiin kysymyksiin. Mutta mihin sanakirjan sisältämä tieto perustuu? Onko tieto riittävää ja ajantasaista? Jos jokin ilmiö sanakirjasta puuttuu, tarkoittaako tämä, ettei kyseistä ilmiötä ole olemassa?

#### KOLLEKTIIVINEN TIETO JA SEN KERÄÄMINEN TESTEILLÄ

Väitöskirjassani osoitan, että meillä on olemassa sanojen — näin myös synonymioiden — merkityksistä runsaasti yhteistä tietoa, jota sanakirjoista ei löydy mutta jota voidaan tavoittaa kyselytesteillä. Kyselytestillä tässä tutkimuksessa tarkoitetaan sitä, että tavallisille ihmisille annetaan kynä ja paperilomake, jossa on erinäisiä kysymyksiä äidinkielen sanoista. Palatkaamme nyt alussa esitettyihin esimerkkeihin, tarkemmin *lapsi*-sanan synonyymeihin.

Pyytäisin nyt arvon yleisöä osallistumaan pieneen testinäytteeseen. Alkuvalmistelujen perusteella oletan, että osa sanoista *mukula*, *tenava*, *muksu*, *kakara*, *pentu* ja *kersa* on ihmislapsen viitattaessa jollain tavalla latautuneita, joko positiivisesti tai negatiivisesti. Pyydän yleisöä nosta-

maan käden ylös, jos kohta sanomallani sanalla on kielteinen eli negatiivinen lataus.

Testitilanne on tällainen: tapaat vanhan ystäväsi kadulla ja haluat kysyä, mitä hänen lapselleen kuuluu. Jos seuraavissa kysymyksissä on negatiivinen lataus, nosta käsi ylös.

- Mites sun mukulas jaksaa?
- Mites sun kakaras jaksaa?
- Mites sun tenavas jaksaa?
- Mites sun muksus jaksaa?
- Mites sun pentus jaksaa?
- Mites sun kersas jaksaa?

Ylös nousseiden käsien määrä osoittaa, että ainakin sanoilla *kersa* ja *kakara* näyttää olevan negatiivinen lataus. Sanojen *tenava* ja *pentu* tuloksissa on jonkin verran hajontaa. Tähän hajontaan voi olla monia syitä. Kyseisten sanojen lataus on ehkä positiivinen tai lataus puuttuu, tai kenties sanat olivat koehenkilöille tuntemattomia tai tehtävänanto epäselvä. Tai ehkä koehenkilöt nukkuivat tai heidän kättään pakotti juuri testin aikaan niin, ettei se millään noussut ylös — tai he olivat yksinkertaisesti kyllästyneitä testiin ja miettivät, että loppuisi jo koko lektio ja saisipa täytekakkaa. Tämänkaltaiset epävarmuustekijät vastauksien suhteen ovat mukana myös väitöstutkimuksessani.

Voinemme kuitenkin todeta, että tässä salissa olevilla koehenkilöillä on aika yksimielinen käsitys siitä, että sanat *kersa* ja *kakara* ovat negatiivisemmin latautuneita kuin sanat *mukula* ja *muksu*. Meillä on todennäköisesti yhteinen käsitys myös siitä, että nämä sanat eroavat toisistaan vielä monissa muissakin suhteissa, joiden selvittämiseen tarvittaisiin hieman tarkempia tutkimuksia.

Meillä on siis yhteistä tietoa näistä merkityseroista. Tätä tietoa kutsutaan kollektiiviseksi tai intersubjektiviseksi tiedoksi. Se

tarkoittaa sitä, että me suomalaiset yhdessä tiedämme parhaiten, mitä sanat tarkoittavat. Äitini tietää paljon muttei kaikkea suomen kielen sanojen merkityksistä, arvoisa kustos tietää ehkä enemmän kuin äitini muttei kuitenkaan kaikkea, eikä edes paras sanakirjankirjoittaja tiedä läheskään kaikkea sitä, mitä asiasta voitaisiin tietää. Jos koko tämä sali kertoo näkemyksensä jostakin sanasta, väitän kuitenkin, että tiedämme siitä jo aika paljon. Paikalla on miehiä ja naisia, nuoria ja vanhoja ja eri paikkakunnilta kotoisin olevia ihmisiä, joilla on erilainen koulutus- ja työelämätausta ja jotka ovat kielellisesti eri tavoin aktiivisia. Me yhdessä muodostaisimme aika hyvän koehenkilöjoukon. Rohkenen tosin epäillä, että tästä salista löytyy tänään enemmän kielentutkijoita kuin normaali-väestöstä keskimäärin, ja tämä mahdollinen vääristymä olisi tietenkin otettava huomioon vastauksia tulkittaessa.

Tutkimuksessani olen hyödyntänyt sitä, että me kaikki tiedämme. Teistä, äidinkielisistä suomen kielen puhujista, ikään kuin uutetaan tätä tietoa sanojen merkityksistä ja merkitysten eroista. Olen keskittynyt siihen, miten tällainen menetelmä toimii, millaisia hyötyjä ja millaisia ongelmia sen käytössä on. Seuraavaksi tarkastelen sitä, mitä iloa tuon tiedon saavuttamisesta on. Miksi ihmeessä tällaista tutkimusta kannattaa tehdä?

#### TIEDON SOVELLUS

Meille noiden äskeisten *lapsi*-sanan synonyymien merkityserot ovat aika selviä asioita, mutta sellaiselle, joka opettelee suomea vieraana kielenä, erot eivät aukeine ilman selityksiä. Samalla tavalla olemme itse pulassa, kun meidän pitäisi valita vieraassa kielessä yksi sana annetusta synonyymijoukosta. Aina emme voi pyytää jotakuta kyseisen kielen puhujaa selittämään

▷

meille sanan merkitystä. Voisimmeko sitä paitsi edes luottaa siihen, että yksi ihminen osaa kertoa kaiken sen, mitä meidän tarvitsee tietää?

Sanakirjakaan ei tässä usein auta, näennäisesti pieniä merkityseroja kun ei ole sinne kirjattu. Yksi syy tähän kirjaamattomuuteen on se, ettei merkityseroja ole juurikaan tutkittu. Kun kerran kyseessä on näin selvä asia, miksi ihmeessä merkityserojen tarkkaa kuvaamista ei ole aikaisemmin suunniteltu ja toteutettu?

Ajatelkaamme nyt vaikka Nykysuomen sanakirjaa. Siinä on monta paksua osaa pientä tiheää tekstiä. Jos sinne ryhdyttäisiin lisäämään yhtä ja toista lisätietoa, tulisi kirjasta sataniteinen eikä sen käyttäminen olisi enää tarkoituksenmukaista. Entä miten toimittaisiin sitten, kun jonkin sanan merkitys muuttuisikin käytössä? Entä jos käyttäjä haluaisikin vertailla vaikkapa synonyymien käyttöyhteyksiä sanakirjan eri osista? On selvää, että tällainen kyselyteillä saatava varsin moniulotteinen tieto vaatii sähköisen esitysmuodon. Perinteinen paperisanakirja ei millään voi sulattaa tarvittavaa tietomäärää.

Asiat ovat vaikuttaneet toisiinsa: tietotekniikan ja erityisesti verkkotekniikan kehitys on viidentoista vuoden aikana ollut räjähdysmäistä. Vasta tämä on tehnyt mahdolliseksi moniulotteisten merkitysten kaappaamisen sanakirjoihin, nimenomaan sähköisiin sanakirjoihin. Jos sanastontutkijoilla onkin ollut kiinnostusta yksityiskoh- taiseen merkitysvahteiden selvittämiseen, on tämä kuitenkin jäänyt varsin suppeaksi tutkimukseksi ilmeisesti juuri siksi, ettei tiedolle ole nähty tarvetta eikä sovellusmahdollisuuksia. Nykyisessä tilanteessa on valtava epäsuhta: teknisesti olisi mahdollista hyödyntää tietokoneiden ohjelmia niin pitkälle kuin vain hurjimmat mielikuvat kantavat ja rakentaa uusi uljas sanakirja. Sisältö kuitenkin puuttuu, tämä hienovireinen tie-

to sanastosta, vaikkapa juuri synonyymien pienistä merkityseroista. Teknisesti näppäristä mutta sisällöllisesti jälkeenjääneistä sanakirjoista on Internetissä jo olemassa kiinnostavia esimerkkejä.

Ainut ratkaisu tilanteeseen — jos tilanteen korjaaminen nähdään yhteiskunnassa ja kielentutkijoiden kesken mielekkääksi — on voimakas panostus nykykielen sanaston tutkimukseen ja tutkimusmenetelmien kehittämiseen. Kokeellisen sanastontutkimuksen edustajana nostan tietenkin mieleläni esiin juuri tämän menetelmän merkityksen. Vakavasti ottaen tarvitsemme tällä hetkellä laajaa yhteistoimintaa kaikkien sanastoa eri muodoissa käsittelevien tahojen ja tutkijoiden kanssa. Tarvitaan ennakkoluulotonta, luovaa otetta. Asiaa täytyy uskaltaa katsoa eri näkökulmista kuin mistä sitä on tähän asti on katsottu. On mietittävä perusteellisesti esimerkiksi seuraavia kysymyksiä: Ketä varten sanaston tutkimusta tehdään? Mikä oikeasti aiheuttaa ongelmia sanakirjan käyttäjille? Mitä sanoja tutkia? Mitä jättää suosiolla tutkimatta?

#### LOPUKSI

Tutkija joutuu vastaamaan muille ja erityisesti itselleen siihen, miksi hän työtänsä tekee. Mikä on tutkimuksen motiivi? Kun kuuntelemme ympäristöämme, voimme todeta, että tutkimustyön motiivit ovat varsin kirjavia, usein hyvin arkisia: tutkimustyötä tekemällä hankitaan elantoa, pysytään pois pahanteosta, kokeillaan, mihin pystytään, tuotetaan uutta tietoa tai tyydytetään uteliaisuutta tai jopa kunnianhimoa. Kielen- tutkijat harvoin mainostavat tekevänsä työtänsä siksi, että yhteiskunnassa olisi sille tilausta. Joskus voitaisiin ehkä keskustella myös siitä, että varsinkin niin kauan kuin liki kaikkea humanistista tutkimusta tehdään yhteiskunnan varoilla, voisi tutkimuk- sellä ehkä olla vastuu yhteiskunnan autta-

misesta niissä tilanteissa, joissa se asiantuntija-apua tarvitsee. Kun yhteiskunnassa pohditaan, mikä on korrekti ilmaus 'vanhalle naiselle' tai mitä tarkoittaa *valmismatka*, pitäisi sanastontutkijain herätä. Ei vain aukaisemaan sanakirjaa, vaan niin sanotusti painumaan turpeeseen tutkimaan, mitä sanat tällä hetkellä tässä yhteiskunnassa elävien suomalaisten kielessä sisällään pitävät. Näin tulisi tehdä, jottemme päätyisi tilan-

teeseen, jossa sanojen merkityksestä päätetään sanomalehtien mielipidesivuilla tai kauppa- ja teollisuusministeriön määräämässä valmismatka-asia neuvottelutoimikunnassa.

Jos sanastontutkimusta tehdään näin, yhteiskuntalähtöisesti, ei tutkijan enää tarvitse miettiä sitä, mitä hän haluaa tutkia, vaan hänen täytyy löytää se, mitä ehdottomasti *pitää* tutkia. ■

ULLA VANHATALO *Kyselytestit synonymian selvittämisessä — sanastotietoutta kielenpuhujilta sähköiseen sanakirjaan*. Elektroninen väitöskirja on luettavissa osoitteessa <http://ethesis.helsinki.fi/>.

*Suomalais-ugrilainen laitos, PL 25 (Franzeninkatu 13), 00014 Helsingin yliopisto*  
Sähköposti: [ulla.vanhatalo@helsinki.fi](mailto:ulla.vanhatalo@helsinki.fi)