

# Koehenkilöiden taustan merkitys kokeellisessa leksikaalisessa semantiikassa ja sen sovelluksissa<sup>1</sup>

Kokeellinen semantiikka – kielen merkityksen ja käytön selvittäminen kielenpuhujakuntaa testaamalla – on viime aikoina herättänyt enenevää kiinnostusta. Leksikaalisessa tutkimuksessa koeasetelmat ovat tuottaneet uusia näkökulmia mm. polysemiaan ja synonymiaan. Aiheesta käydyssä keskustelussa on kuitenkin toistuvasti kiinnitetty huomiota koehenkilöryhmien homogeenisuuteen. Saatujen tulosten yleistettävyyden ja soveltamisen kannalta on merkityksellistä saada tietoa koehenkilöiden taustamuuttujien vaikutuksista tutkittuihin leksikaalisiin ominaisuuksiin.

Artikkelissa käsittelem seuraavia asioita:

- Mitä eroja kyselytestien vastauksissa ilmeni, kun koehenkilöt ryhmitettiin eri sosiodemografisten muuttujien suhteen?
- Mikä painoarvo koehenkilöiden sosiodemografisilla tiedoilla tulisi olla ja mitä esimerkiksi synonymiaa tutkivan testisarjan taustatietolomakkeen tulisi sisältää?
- Minkälainen on ihannekoehenkilö?
- Voiko kokeellisessa semantiikassa luottaa helpoiten saatavaan koehenkilöryhmään eli koululaisiin?

Tarkastelun esimerkkinä käytän lähisynonymiaa tutkivia koesarjoja, joissa sama kyselylomake esitettiin eri oppilaitosten eri-ikäisille opiskeluryhmille.

---

<sup>1</sup> Useat ihmiset ovat perehtyneet tässä esitellyn tutkimuksen eri vaiheisiin ja antaneet hyödyllisiä kommentteja. Kiitän erityisesti prof. Ulla-Maija Kulosta, prof. Jussi Niemeä, Antti Arpea, Reetta Kenttää, Jarno Raukkoa, Jussi Ylikoskea sekä kahta nimetöntä esilukijaa. Esittelin tämän artikkelin sisältöä myös Kielitieteen päivillä Tallinnassa 2004; yleisölle kiitos kommentteista.

## 1. Johdanto

### 1.1 Kokeellisesta semantiikasta ja siihen kohdistetusta kritiikistä

Karkeasti ottaen kokeellisen semantiikan alan voi jakaa kahtia: on ns. psykologiviestisesti suuntautuneita kovan linjan eksperimentalisteja, jotka käyttävät psykofysiologisia mittauslaitteita kuten aivosähkökäyriä tai silmänliikemittauksia. Toisena ryhmänä erottautuvat ns. pehmeän linjan edustajat, joiden menetelmänä ovat pääasiassa kyselylomakkeet. Suuntausten välillä ja myös sisällä käydään keskustelua mm. perusterminologian määritelmistä ja termien käytön oikeutuksesta: millaista tutkimusta saa kuvata sanalla *kokeellinen* tai millainen tutkimusasetelma on määriteltävissä *kokeeksi*? Nähdäkseni terminologiset ja myös metodiset ongelmat liittyvät ennen muuta tutkimussuunnan ja -suuntausten vakiintumattomuuteen, mikä toistaiseksi onkin paitsi ymmärrettävää myös suotavaa.

Tämä tutkimus edustaa pehmeää kokeellisuutta: tieto on kerätty kyselylomakkeilla äidinkielisiltä puhujilta. Käytän kyselytestiä, koska vastaavaa tietoa synonyimiasta on vaikeaa tai mahdotonta saada muilla menetelmillä<sup>2</sup>. Kyselymenetelmää on pidetty kiinnostavana, mutta se on saanut jonkin verran aiheellistakin kritiikkiä. On sanottu, että menetelmällä saadaan sitä, mitä etsitään, ja moitittu tilastomenelmien käyttöä tai käyttämättömyyttä. Lisäksi on kritisoitu mm. koeasetelman keinotekoisuutta, lähialojen tutkimustulosten sivuuttamista, koehenkilöiden homogeenisuutta sekä menetelmän työläyttä (esim. keskustelut työpajassa ”On the necessity of experimental methods in semantics” Skandinavian lingvistikonferenssissa Helsingissä 2004).

### 1.2 Miksi tämä tutkimus?

Yllä esitelty ja muukin kritiikki on antanut kokeellisen semantiikan tutkijoille aiheellista mietittävää. Menetelmää onkin tarkennettu ja kehitetty luotettavammaksi tai käytettyjä tutkimusasetelmia on ainakin pyritty perustelemaan paremmin. Tässä esityksessä keskityn yhteen kritiikin aiheeseen, koehenki-

<sup>2</sup> Suomalaisessa tutkimuksessa kyselytestejä ovat harrastaneet mm. Jarno Raukko (1994, 1999, 2002), Antti Arppe ja Juhani Järvikivi (2002), Jussi Ylikoski (2004) ja Ulla Vanhatalo (2002, 2003), Virossa mm. Renate Pajusalu (2004), Maria-Magdalena Jürvetson (2004), Ene Väinik (tulossa), Ilona Tragel ja Kaja Kährik (2004). Kyselytestimenetelmää arkitiedon keräämiseen on käytetty maailmalla laajemminkin: samantapaiseen asetelmaan perustuvat mm. MIT:n Open Mind Common Sense -projekti, jossa suurilta joukoilta ihmisiä kysytään tavallisia arkipäivän asioita ja niistä koostetaan ns. maallikkotietoa.

löiden sosiodemografisten taustatietojen huomiotta jättämiseen ja koehenkilöiden homogeenisuuteen. Vastaan kritiikkiin tekemällä saman koesarjan muutamalle toisistaan poikkeavalle koehenkilöryhmälle ja katson, löytyykö tulosten mahdollisille eroille selityksiä taustamuuttujista.

Miksi koehenkilöt ovat sitten olleet niin samanlaisia? Todennäköisesti merkittävin syy on koulu- ja opiskeluryhmien tutkimisen helppous – myötmieliset äidinkielenopettajat, yleensä rauhalliset luokkatilanteet sekä isot ja useimmiten innokkaat ryhmät. Usein tutkimuksissa todetaan, että tulokset pätevät vain tietyin ehdoin, ja varotaan yleistästä 17-vuotiaiden nuorten käsityksiä koko väestöä koskevaksi; aina ei yleistettävyysongelmaan kuitenkaan ole kiinnitetty huomiota (esim. Colston 2002). Menetelmän laajentamista aikuisväestöön on pidetty hankalana. Nettikyselyjä on suunniteltu ja pilottiluontoisesti kokeiltukin, mutta asia on vielä kansainvälisestikin kesken<sup>3</sup>. Koehenkilöiden taustojen vaikutusta kokeellisen leksikaalisen semantiikan tuloksiin on toistaiseksi tiettävästi vain arvuuteltu.

Kokeellisen leksikaalisen tutkimuksen yleistettävyydessä on pohjimmiltaan kyse siitä, onko sanasto – tämän tutkimuksen esimerkeissä synonyymit – koko kansan yhteistä, samalla tavalla ymmärrettyä ja käytettyä omaisuutta. Asialla on laajakantoista merkitystä kielen normittamisessa: kun dokumentoimme kieltä yleiskielen sanakirjaan, kenen kieltä sinne dokumentoimme? Miten varmistua, että sanakirjan tieto vastaa kieliyhteisön kieltä ja kielitajua mahdollisimman kattavasti? Kokeellisen leksikaalisen semantiikan tutkimuksessa sovellettavuus asettaa tarkentavia lisäehtoja jo lähtöasetelmille.

## *2. Tutkimuksen esittely*

### *2.1 Materiaali ja menetelmät*

Tässä esitettävän tutkimuksen aineisto on kerätty vuoden 2004 maalisi- ja huhtikuussa helsinkiläisten oppilaitosten äidinkielen tunneilla suomen äidinkielenlekkeseen ilmoittaneilta opiskelijoilta. Yhteensä koehenkilöitä oli 180, joista vajaa kolmasosa miehiä. Taulukossa 1 esitetään koehenkilöiden keski-ikä ja sukupuoli ryhmittäin sekä ryhmien suuruus.

---

<sup>3</sup> Ks. kuitenkin esim. Jack Chambersin murretopografiaprojekti <http://dialect.topography.chass.utoronto.ca/>.

Taulukko 1. Tutkimukseen osallistuneiden koehenkilöiden ryhmäkohtainen keski-ikä, sukupuoli ja ryhmien koko.

oppilaitos	keski-ikä	naisia	miehiä	yht.	huomioitavaa
Kallion lukio	16 v.	40	10	50	ilmaispainotteinen, korkea sisään-pääsykeskiarvo
Aikuislukiot	37 v.	16	5	21	Mäkelänrinne ja Ressu
Hesote-nuoret	18 v.	45	3	48	Malmi, ammattiin valmistava, terveydenhuoltoalan oppilaitos
Hesote-aikuiset	43 v.	27	1	28	Malmi, täydennyskoulutusryhmä, terveydenhuoltoalan oppilaitos
Heltech	18 v.	0	33	33	Vallila, ammattiin valmistava, teknisen alan oppilaitos, matala sisään-pääsykeskiarvo
		<b>128</b>	<b>52</b>	<b>180</b>	

Aineisto on lähtökohtaisesti ryhmitelty ja analysoitu oppilaitosten mukaan, koska ryhmien erityispiirteet, kuten sukupuoli ja ikä, ovat selviä. Ryhmiä verrattaessa voidaan kiinnittää huomiota mm.

- samassa oppilaitoksessa opiskelevien samaa sukupuolta olevien koehenkilöiden ikäeroon (Hesote-N ja Hesote-A)
- samanikäisten samaa sukupuolta olevien koulutustasoeroon (Kallio ja Hesote-N)
- samanikäisten samalla koulutustasolla olevien sukupuolieroon (Hesote-N ja Heltech)

Olisi ollut suotavaa saada koehenkilöiksi vielä ryhmä aikuisia miehiä, mikä ei kuitenkaan järjestynyt.

Jokaiselle koehenkilölle jaettiin samanlainen 13 sivua eli osatestiä käsittävä kyselylomake, johon kuului sekä monivalinta- että tuottotestejä (ks. Liite 1). Äidinkielen, iän ja sukupuolen selvittäminen oli suhteellisen yksinkertaista (lomakkeen kysymykset 1–3), kun taas murretaustan (kysymykset 4, 5, 6, 9, 11, 12), vieraiden kielten vaikutuksen (kysymykset 4, 5, 6, 7, 8, 10, 11, 12), sosioekonomisen luokan (kysymykset 13, 14, 18, 19, 20), sosiaalisen aktiivisuuden (kysymykset 35, 36, 37, 38) ja kielellisen aktiivisuuden (kysymykset 15, 16, 17, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 39, 40; lisäksi 21 ja 22, joita ei laskettu mukaan myöhemmin esiteltävään indeksiin) selvittämiseen käytettiin runsaasti erilaisia kysymyksiä. Sosioekonomisen taustan kysymyksissä keskityttiin opiskelu- ja työtaustaan, ei tuloihin (vrt. Tilastokeskuksen ohjeet, [www.stat.fi/tk/tt/luokitukset/index\\_henkilö\\_keh.html](http://www.stat.fi/tk/tt/luokitukset/index_henkilö_keh.html)). Koehenkilöiden kuuluminen tietyn oppilaitoksen piiriin voidaan varauksin tulkita myös sosioekonomiseksi taustamuuttujaksi. Vieraiden kielten vaikutuksen selvittämi-

sen tarkoitus oli ottaa huomioon mahdolliset äidinkielen kielitajua selvästi sekoittavat tekijät. Sosiaalisen aktiivisuuden selvittämisen tavoite oli tarkistaa, ovatko koehenkilöt suurin piirtein normaalisti tekemisissä kieliyhteisön kanssa<sup>4</sup>.

Tulosten esittelyn yhteydessä on käytetty kielellisen aktiivisuuden indeksiä. Se laskettiin siten, että jokaisesta kielellisen aktiivisuuden mittaamiseen kuuluvasta myöntävästä vastauksesta tai a/b-vastauksesta sai yhden pisteen, kieltävästä vastauksesta tai c/d/e/f-vastauksesta nolla pistettä (kysymykset 15, 16, 17, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 39, 40). Poikkeuksena oli kieliaineiden yliopisto-opiskelua koskenut kysymys 17, jossa myöntävästä vastauksesta sai kolme pistettä. Täysi pistemäärä oli 18, josta seurasi indeksiluku 10. Koehenkilöiden indeksit laskettiin kaavalla

$$\frac{\text{pisteluku} \times 10}{18} = \text{indeksi}$$

Indeksiluvut pyöristettiin kokonaisluvuiksi.

Sisältökysymysten kohdalla kullakin sivulla esitettiin kysymys tai kysymyksiä yhdestä synonyymistä tai synonyymiryhmästä. Testilomake oli nidottu reunoistaan siten, että sivut pystyi täyttämään vain yhden kerrallaan. Testitilannetta valvottiin ja paperit kerättiin pois sitä mukaa kuin ne oli täytetty. Koehenkilöillä kului lomakkeen täyttämiseen aikaa noin 20–25 minuuttia.

Analyysivaiheessa hylättiin lomakkeet, joiden täyttäjät ei ollut ilmoittanut äidinkielekseen suomea. Varauduttiin hylkäämään myös tapaukset, joissa koehenkilö olisi palauttanut täysin tyhjän tai epäasiallisesti täytetyn testilomakkeen. Yksittäiset tyhjat aukot muuten täytetyissä testilomakkeissa otettiin huomioon tyhjinä vastauksina.

## 2.2. Tulokset

Kiinnitän tulosten käsittelyssä erityistä huomiota seuraaviin kysymyksiin:

- Eroavatko eri koehenkilöryhmien vastaukset keskenään?
- Eroavatko koehenkilöiden vastaukset sanakirjamääritelmästä?
- Eroavatko vastaukset perusluokittelusta poikkeavan tai sitä tarkentavan taustamuuttujan tai -muuttujien mukaan analysoituina (kohdassa 2.2.3 äänen voimakkuus)?

<sup>4</sup> Tutkimusta suunniteltaessa tarkoitus oli saada mukaan enemmän ns. kadunmiehiä. Koehenkilöiden poiminta kadulta osoittautui kuitenkin hitaaksi ja hankalaksi, joten päädyin käyttämään oppilaitoksia, mikä oli jo eräänlaista esiseulontaa. Tästä seurasi, että päädyin ainakin toistaiseksi olemaan analysoimatta tuloksia sosiaalisen aktiivisuuden indeksin mukaan, onhan sosiaalista aktiivisuutta oltava ainakin jonkin verran opiskelutilanteessa.





Synonyymivalintojen tulokset esitetään Kuviossa 2. Eri-ikäisten vastaukset ovat suhteellisen samankaltaisia. Kiinnostavia yksityiskohtia silti löytyy: *voitollista* ei juuri mainita aikuisryhmissä (Aikuislukiot, Hesote-A), kun se nuorten ryhmissä on edustettuna (Kallio, Hesote-N, Heltech). Sukupuolen vaikutuksesta voisi mainita *taloudellisen* runsaamman esiintymisen naisten kuin miesten vastauksissa (Hesote-N vs. Heltech). Koulutustaustaa tarkastellessa kiinnittää huomiota *halvan* väistyminen erityisesti *suotuisan* ja *hyödyllisen* tieltä nuorten ilmaisutaidonlukiolaisten (Kallio) vastauksissa. Vastauksissa edelleen ylivoimaisesti useimmin esiintyvät merkitykset 'halpa, vähän mak-sava' ja muut (rahalliseen) hintaan tai arvoon liittyvät adjektiivit. Muut mainitut synonyymit näkyvät kaavion yläosassa. Jos otetaan huomioon koehenkilöiden rastiittamat tärkeimmäksi valitut synonyymit (ei näkyvissä Kuviossa 2), *halpa* on ylivoimaisesti laajimmin edustettuna.

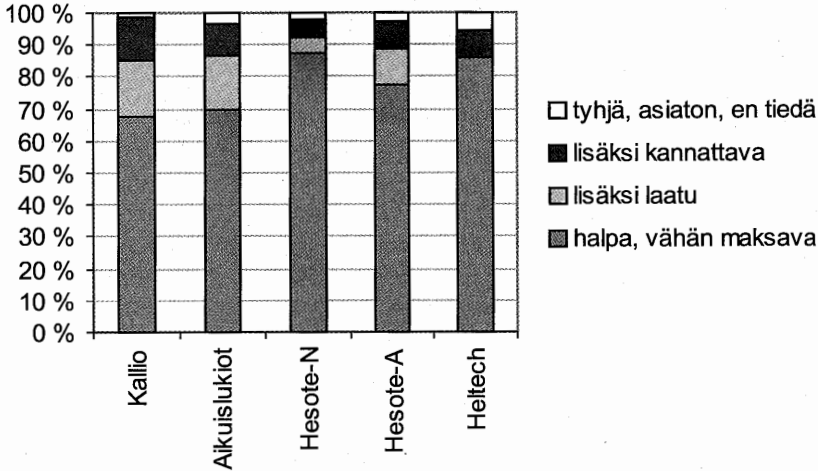
Kiinnostava esimerkki mahdollisesta sukupolvien välisestä sanastoerosta on *huokea*, joka on varsin voimakkaasti edustettuna aikuisryhmissä (Aikuislukiot ja Hesote-A), mutta jonka edustus nuorten ryhmissä (Kallio, Hesote-N, Heltech) on huomattavasti suppeampi. Edustukseen saattaa vaikuttaa myös *huokean* 'edullinen' merkityksen itäisehkö murrelevikki (SMS:n mukaan KarE, KarK-P, Sav, Kai Ink<sup>5</sup>). Aikuisryhmissä neljällätoista kolmestakymmenestä *huokean* maininneella on taustassaan itämurteita. Kiinnostava on myös *etuisan* valitseminen jokaisessa ryhmässä. *etuisan* nykyisestä marginaaliasemasta kertoo mm. se, ettei Microsoft Wordin sanakirja sitä tunne. *etuisan* samoin kuin *edukkaan* valintaan erityisesti nuorten ryhmissä on mahdollisesti vaikuttanut sama kantasana *edullisen* kanssa. Synonyymien hahmottumista voisi selvittää perusteellisemmin pyytämällä uusia koehenkilöitä luettelemaan avoimessa tuottotestissä spontaanisti *edullisen* synonyymeja.

Tehtävässä 3 koehenkilöiltä kysyttiin, mitä eroa on *edullisella* ja *halvalla*. Vastaus tuli nyt nuotoilla itse.

<sup>5</sup> (Etelä-Karjala, Keski- ja Pohjois-Karjala, Savo, Kainuu, Inkeri)



Kuvio 3. Koehenkilöiden vastaukset avoimeen tuottotestiin, jossa kysyttiin, mitä eroa on *edullisella* ja *halvalla*. Noin kolmasosa joka ryhmästä vastasi, että *halpa* maksaa vähemmän kuin *edullinen*. Rahallisen arvon lisäksi toinen dimensio oli laatu: *edullisen* katsottiin olevan parempilaatuinen kuin *halvan*. Silmiinpistävää on, että lähes puolet teknillisen oppilaitoksen nuorista miehistä ei osaa tai halua määritellä *edullisen* ja *halvan* eroa.



Vertailun tulokset esitetään Kuviossa 3. Koulutustaustassa huomio kiinnittyi nuorten ilmaisutaidonlukiolaisten (Kallio) runsaisiin mainintoihin muusta kuin rahallisesta tai arvollisesta merkityksestä. Nuoret ammattiopiskelijat eivät näitä juuri maininneet (Hesote-N ja Heltech)<sup>6</sup>. Vähiten koulutetuilla oli myös eniten vastauksia *ei eroa* tai *en tiedä*. Nuorten hyvin koulutettujen (Kallio) ja aikuisten terveydenhuolto-opiskelijoiden (Hesote-A) vastauksissa korostui *edullisen* ja *halvan* laatuero.

Yhteenvetona koehenkilöiden vastauksista *edullisesta* ja sen synonyymeista voisi todeta seuraavaa: yleisesti ottaen vastauksissa ei ollut suuria eroja tutkittujen taustamuuttujien suhteen. Tiettyihin poikkeamiin merkityksen määrittelyssä saattaa selitys kuitenkin löytyä taustamuuttujista. Suhteessa leksikografiaan voidaan todeta, että kaikkien koehenkilöiden mukaan *edullisella* on hyvin tärkeä merkitys 'vähän maksava, halpa', joka yleiskielen sanakirjoista – hämmästyttävää kyllä – puuttuu. Tässä esitettyjen tulosten valossa ko. merkityksen olemassaolo sanakirjamääritelmässä olisi kuitenkin perusteltua. *edullisen* ja 'vähän maksava' perinteisesti edustavan *halvan* eroja kuvatessaan koehenkilöt mainitsivat lähinnä hinnan (*halpa* maksaa vähemmän kuin *edullinen*) ja laadun (*edullinen* on laadultaan parempi kuin *halpa*).

<sup>6</sup> *edullisen* merkitys 'olla alennuksessa' saattaa tarjota kiinnostavan tulkinnan: kertoo merkityksen vähäinen esiintyminen tai puuttuminen (Kallio, Heltech) siitä, etteivät kyseiset ryhmät kuulu tiettyjen kalliiden tuotteiden alennusmyyntitajuihin?

### 2.2.2 erehdys, virhe

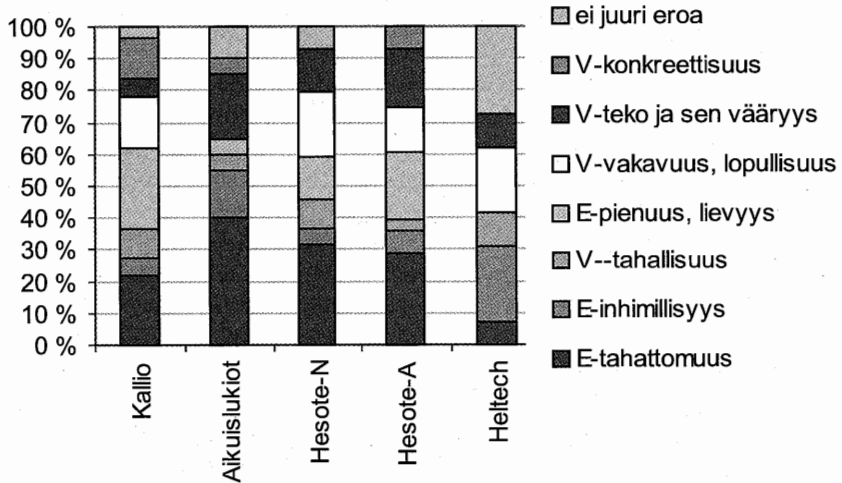
**erehdy[s]**<sup>39</sup> epähuomiosta, vahingosta, ajattelemattomuudesta tms. johtuva virhe, hairahdus, haksahdus, lapsus; väärä käsitys, harhaluulo. *Suuri, pieni, paha, karkea e. Tehdä e. Hänelle sattui, tuli e. laskelmissa. Korjata e:ksensä. Oli e. uskoa, että --. E:ksessä epähuomiossa, vahingossa.*

**virhe**<sup>48</sup> hyväksyttävästä t. normaalista poikkeava seikka, vika. 1. erheellinen, väärä, viallinen suoritus, erehdys, hairahdus, lapsus; sellaisen tulos. *Pieni, paha, vähäinen v. Arviointi-, muistiv. Mittausv. Kirjoitusv-, käännösv-, ääntämisv. Virkav. Tehdä v., v:itä. Teki sen v:en, ettei kysynyt, mihin lapset tarvitsevat moottorisahaa. Oli paha v. vaihtaa työpaikkaa. Hoidossa tapahtunut v. Selostukseen pujahti asiav:itä. Korjata, oikaista v. Ark. Pilkun pois jättämisestä otettu [arvosanaan vaikuttavaksi merkitty] v. Pelaaja otti [teki tahallaan] v:en. 2. vajavuutta, vajaalaatuisuutta, -kuntoisuutta tms. aiheuttava seikka, vika. *Maku-, pintav. V. tuotteen värissä. Koiran rakenteen v:et. Kaikilla meillä on omat v:emme.**

(PS)

Tehtävässä 4 koehenkilöiltä kysyttiin, mitä eroa on *erehdyksellä* ja *virheellä*. Kuviossa 4 esitetään koehenkilöiden vastauksia *erehdyksen* ja *virheen* erois-

*Kuvio 4.* Koehenkilöiden vastauksia *erehdyksen* ja *virheen* eroista. Teknillisen oppilaitoksen nuorten miesten tuloksissa esiintyy runsaasti 'ei eroa' -vastauksia (vrt. Kuvio 3).



ta. Kaikissa ryhmissä tärkeimpänä *erehdyksen* ja *virheen* erona mainitaan, että *erehdys* on tahaton, inhimillinen ja lievä verrattuna *virheeseen*. *Virheen* ja *erehdyksen* vakavuuseron mainitseminen on vähäistä niin aikuislukiolaisilla kuin ammattopiskelijapojillakin (Heltech). Kyseiset ryhmät poikkeavat huomattavasti ilmaisutaidonlukiolaisten ryhmästä (Kallio), jossa suurin osa *erehdyk-*

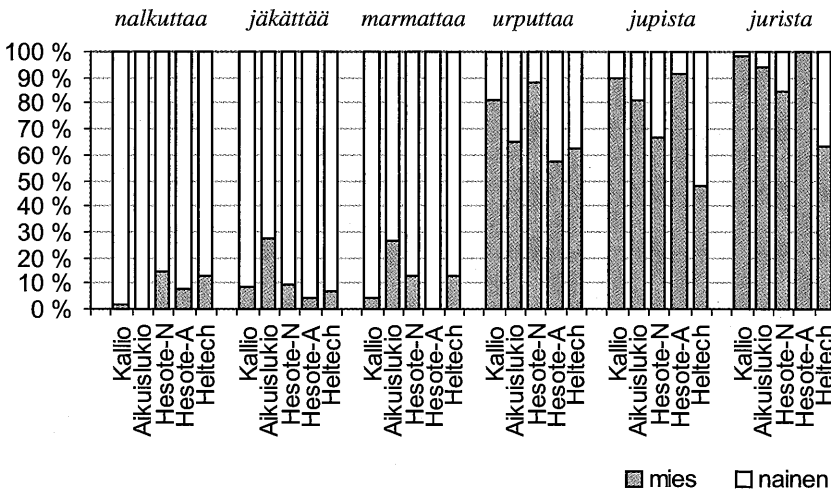
sen ja virheen erojen kuvauksesta keskittyi nimenomaan vakavuuden, suuruuden ja lopullisuuden pohdintaan. Kiinnostava on virheen konkreettisuuden mainitseminen aikuisryhmissä (Aikuislukiot, Hesote-A) sekä nuorten lukiolaisten (Kallio) ryhmässä.

### 2.2.3 Ns. valitusverbien pesye

Kolmantena sanaryhmänä käsiteltiin ns. valitusverbien pesyettä eli verbejä *jupista, jurista, jurnuttaa, jäkättää, marista, marmattaa, motkottaa, mukista, nalkuttaa, napista, nurista, purnata, ruikuttaa, urputtaa, valittaa, voihkia ja voivotella*. Peruskoululaisilla ja lukiolaisilla aikaisemmin tehdyissä tutkimuksessa (Vanhatalo 2002) on tarkasteltu näihin verbeihin liittyviä agenteja ja mm. verbien ilmaisevan äänen voimakkuutta. Tuolloisten tulosten mukaan esim. verbien *nalkuttaa* agentiksi miellettiin yksinomaan keski-ikäinen nainen. Nyt selvityksen kohteena oli, vastaavatko eri koehenkilöryhmät eri tavoin sukupuolisesti ja iällisesti latautuneisiin merkityspiirteisiin. Vertailun vuoksi mukana oli myös lataukseltaan neutraali(mpi) piirre, äänen voimakkuus.

Tehtävässä 6 koehenkilöitä pyydettiin vastaamaan, kenen puhetta verbi kuvaa.

Kuvio 6. Koehenkilöiden vastauksia kysymykseen verbien agenteista. Kuviossa esitetään vastausten mukaan selvivimmän nais- tai miesagenttiin liittyvät verbit. Vastauksissa ei ole huomattavia poikkeamia miesvaltaisten (Heltech) ja naisvaltaisten (Kallio, Aikuislukiot, Hesote-N ja Hesote-A) ryhmien kesken. Eniten poikkeamaa esiintyy miesten ryhmässä *jupisemisen* ja *jurisemisen* yhteydessä: nuorten ammattikoulua käyvien miesten mukaan *jupisija* tai *jurisija* on lähes yhtä usein nainen kuin mies.

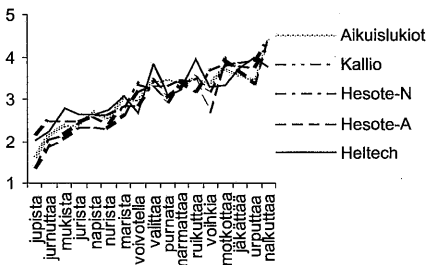


Kuviossa 6 esitetään, mitä koehenkilöt vastasivat, kun kysyttiin, kenen puhetta verbi kuvaa. Koska verbiaineistoa oli varsin runsaasti (yhteensä 17 verbiä), kuvioon 6 on otettu mukaan vain voimakkaimmin mies- tai naisagenttiin yhdistetyt verbit. Huomattavimmat poikkeamat keskiarvosta näkyvät nuorten ammattikoulua käyvien miesten ryhmässä (Heltech; *jupista* ja *jurista* liittyvät myös naisagenttiin) sekä keski-ikäisten aikuislukiossa opiskelevien naisten ryhmässä (neljäsosan mukaan *jäkättäjät* ja *marmattajat* ovat miehiä; kolme neljäsosaa on silti enemmistön kanssa samaa mieltä). Jälkimmäisessä kyse ei kuitenkaan näytä olevan vastaajien iän tai sukupuolen vaikutuksesta, vrt. keski-ikäisten naisopiskelijoiden lukiolaisryhmä (Aikuislukiot). Yhteenvetona voidaan todeta, että vastaukset noudattavat hyvinkin tarkasti aikaisemmin peruskoululaisilta ja lukiolaisilta saatuja tuloksia agentin sukupuolesta (Vanhatalo 2002).

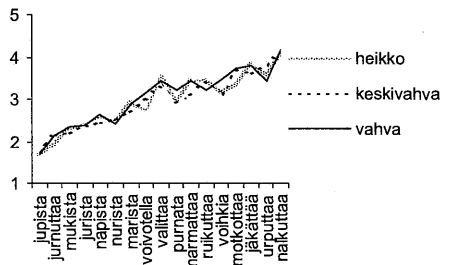
Tehtävässä 7 koehenkilöiltä kysyttiin, miten voimakas ääni verbeihin liittyy. Verbiä äänen voimakkuudesta saadut tulokset on analysoitu ensin perusryhmien, sitten kielellisen aktiivisuuden, iän ja sukupuolen mukaan (kuvio 7).

*Kuvio 7.* Koehenkilöt vastasivat kysymykseen, miten voimakas ääni verbeihin liittyy. Vastaukset esitetty koehenkilöryhmittäin, koehenkilöiden kielellisen aktiivisuuden indeksin, iän sekä sukupuolen mukaan. Erot vastauksissa ovat pieniä lukuun ottamatta ikäryhmittäistä jakoa, jossa ikäryhmä 55–64 poikkesi eniten vastausten keskiarvosta (koehenkilöitä oli tosin myös vähiten tässä ryhmässä). Hienoista eroa oli sukupuolen mukaisessa jaottelussa siten, että miesten vastauksissa verbiä äänet ovat keskimäärin hieman voimakkaampia kuin naisten vastauksissa.

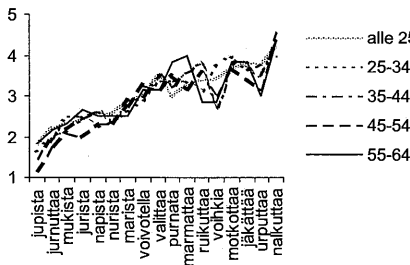
Koehenkilöt oppilaitosten mukaan ryhmiteltyinä



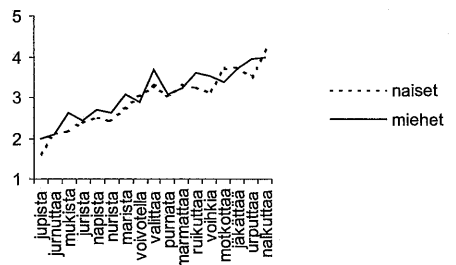
Koehenkilöt kielellisen aktiivisuuden mukaan ryhmiteltyinä



Koehenkilöt iän mukaan ryhmiteltyinä



Koehenkilöt sukupuolen mukaan ryhmiteltyinä



Tulokset ovat jälleen hyvinkin yksimielisiä; voimakkainta hajontaa tuloksissa näkyy ikäryhmittäin. Tutkituista verbeistä hiljaisin on *jupista*, äänekkäin *nal-kuttaa*. Yksittäiset piikit tarjoavat kiinnostavia huomioita. *Valittaa*-verbin äänen voimakkuus Heltechin opiskelijoiden keskuudessa saattaa johtua ko. verbin ilmeisestä yläkäsitteisyydestä nuorten kielessä (Vanhatalo 2002, 342). Tulokset tukevat Vanhatalon aiempien löydösten yleistettävyyttä.

#### 2.2.4 Yhteenvedo tuloksista ja analyysista

Huolimatta yksittäisistä eroista koehenkilöryhmien tulkinnat sanojen merkityksistä ovat huomattavan yhdenmukaiset.

Tulosten yleistettävyyttä heikentävät jonkin verran aineistoon liittyvät varaukset, joista merkittävimpiä ovat miesten osuuden vähyys ja koehenkilöiden maantieteellinen sijoittuminen Helsinkiin. Miehiä oli kuitenkin noin kolmasosa vastaajista, ja suuri osa vastaajista ilmoitti omaksi tai vanhempiensa murretaustaksi muun kuin pääkaupunkiseudun tai oli asunut osan lapsuuttaan ja nuoruuttaan pääkaupunkiseudun ulkopuolella. Koehenkilöiden ikäryhmien kattavuus voidaan myös kyseenalaistaa; olihan mukana vain nuorten ja keski-ikäisten ikäryhmät. Lasten kielen kehitysvaiheet huomioon ottaen alle 13-vuotiaita ei kuitenkaan yleensä lasketa mukaan yleiskielen tutkimuksissa. Ikään-tyneiden kielestä tehtyjen tutkimusten mukaan kieli ei enää juurikaan vaihtelee myöhäisessä aikuisiässä.

Esittelemani tutkimusmenetelmän varauksia ja riskejä arvioidessa merkittävimpanä pidän tutkijan subjektiivista analysointivaltaa avoimia vastauksia tulkittaessa. Suurin osa vastauksista on kuitenkin tulkittavissa yksiselitteisesti: *Edullinen on sama kuin halpa. Eli ei maksa paljoa; et se on halpa; jokin tavara ei ole kallis, vaan sen saa (ostettua) halvalla.* Toisaalta on myös vähemmän yksiselitteisiä, kuten *edullinen on halvempaa verrattuna kalliimpaan. edullisessa voi säästää enemmän rahaa.* Menetelmän soveltaminen laajemmassi edellyttää, että erityisesti avointen tuottotestien analyysissä käytetään useita analyysoijia. Myös tulosten yleistys kategorioihin on mahdollista tehdä monella tavalla. Merkityspiirteet voidaan yhdistää dimensioittain siten, että esim. *erehdyksen* ja *virheen* kohdalla määritellään asian koko tai merkittävyys sen sijaan, että mainitaan *erehdyksen* kohdalla pienuus ja lievyys ja *virheen* kohdalla suuruus ja vakavuus. Harkinnanvaraista on myös tilastomenetelmien käyttö: mitkä ovat mielekkäät ja löytämisen arvoiset tilastolliset riippuvuudet?

#### 2.2.5 Yhteenvedo teknisestä onnistumisesta

Koetilanteet sujuivat hyvin kaikissa ryhmissä lukuun ottamatta kahta Heltechin ryhmää ja yhtä Hesoten nuorten ryhmää, joissa oli jonkin verran hälinää. Näissä nousi myös eniten kysymyksiä ja kommentteja (*Mä en tajuu mitä tässä pitää*

*tehdä; Ei vittu kiinnosta; Mitä tää tarkoittaa?*). Myös kyselylomakkeeseen jääneisiin painovirheisiin tartuttiin eniten näissä ryhmissä. Lomakkeen tehtävä 5 aiheutti jonkin verran kysymyksiä; vastauksista oli nähtävissä, etteivät kaikki koehenkilöt olleet ymmärtäneet tehtävänantoa samalla tavoin. Aiempia tutkimuksia kirjavampi testipopulaatio antoi kuvaa myös siitä, miten kyselyn tehtävät voidaan ymmärtää väärin. Nuorilla lukiolaisilla (Kallio) näytti olevan keskimääräistä suurempi taipumus ymmärtää tehtävät niin kuin testin laatija oli tarkoittanut. Tehtävänantojen tulisi kuitenkin olla mahdollisimman yksiselitteisiä ja yksinkertaistettuja, niiden tulisi myös mitata kielitajua eikä koehenkilön viitseliäisyyttä. Kyselytutkimuksessa on otettava huomioon, että tutkimusasetelma saattaa vaikuttaa eri koehenkilöihin eri tavoin. Osalle kirjoittaminen on tuttua ja jokapäiväistä, toisille harvinaista, osalla saattaa olla keskittymisvaikeuksia noin puoli tuntia kestävässä kyselykaavakkeen täyttämässä.

Toisinaan on kysytty, oppivatko koehenkilöt testin kuluessa ja kuinka merkityksellistä oppiminen on. Parhaina koehenkilöinä on yleensä pidetty sellaisia, jotka eivät ole tehneet samantapaisia testejä aikaisemmin. Tutkimuksissa on myös pyritty minimoimaan eri osien vaikutus toisiinsa. Jos kyselylomakkeen ensimmäinen tehtävä aktivoi tiettyjä piirteitä, vaikkapa akselilla positiivisuus–negatiivisuus, vaikuttaako tämä seuraaviin vastauksiin? Toistaiseksi tiedossani ei ole tutkimuksia, joissa asiaa olisi natiivipuhujien ollessa kyseessä perusteellisesti selvitetty; joitakin tutkimuksia on tosin tehty englannin verbeillä *get* ja *date* (Raukko 1999). On myös muistettava, että toisen ja vieraan kielen opetuksen varsin runsasta testausentutkimusta voisi hyödyntää myös natiivien kyselytesteissä (mm. Huhta jm. 1993). Testien lainaaminen kielenopetuksesta ei kuitenkaan ole aivan ongelmatonta johtuen mm. testauksen erilaisista tavoitteista sekä vieraan ja oman kielen hahmotuksen erilaisuudesta.

### 3. Pohdintaa

#### 3.1 Mitä annettavaa sosiolingvistiikalla olisi kokeelliselle sanastontutkimukselle?

Sosiolingvistisessä tutkimuksessa ja variaationtutkimuksessa on osoitettu, että kielenpuhujien taustat vaikuttavat kielen käyttöön ja käsitykseen kielestä. Kielen vaihtelussa on suomalaisessa tutkimusperinteessä kiinnitetty huomiota ennen kaikkea dialektologisiin eroihin. Labovilaiset muuttajat tulivat sosiolingvistiikan nousun myötä (ks. esim. Paunonen 1982). Iän, sukupuolen ja sosiaalisen taustan lisäksi on nykyisin tutkittu entistä enemmän tilanteista variaatiota (esim. Sorjonen jm. 2001; Lappalainen 2003). Toistaiseksi sosiolingvistinen näkökulma on näkynyt kohtalaisen heikosti (kokeellisessa) sa-

nastontutkimuksessa ja edelleen sanakirjoissa ja leksikografisessa tutkimuksessa.

Luokka- ja rotuerojen ohittamisesta suomalaisessa tutkimuksessa voidaan päätellä, ettei niiden osuutta ole pidetty kaikkialla tarpeellisena selvittää<sup>7</sup>. Kielellistä vaihtelua aiheuttavien taustamuuttujien merkitys vaihtelee: jokin varioiva ilmiö selittyy mahdollisesti vain yhden muuttujan vaihtelulla muiden muuttujien pysyessä merkityksettöminä. Jos asia on näin, tästä seuraa myös se, että jossain on otettava huomioon sellaista vaihtelua, mitä yleensä ei ole huomattu. Haasteellista onkin selvittää, mikä kaikki aiheuttaa vaihtelua kielenkäytössä ja miten aiheuttajiin päästään käsiksi. Kysymys on varsin monitahoinen erityisesti sanastontutkimuksessa, jossa yksilön sanastoon vaikuttavat tekijät voivat olla varsin kirjavia ja kohdistua tiettyihin sanaston alueisiin, esim. ammattiin tai harrastuksiin liittyvään sanastoon. Kyseessä ei tarvitse välttämättä olla varsinainen ammattikieli, vaan pelkkä harjaantuneen huomiointikyvyn tuottama kielellisen kuvauksen taso, joka poikkeaa tavallisesta keskiwerrosta. Kartuttaessaan mahdollisten taustamuuttujien kirjoja kokeellisen sanastontutkimuksen tulisi olla sosiolingvistikäsitteitä kuin sosiolingvistiikka itse.

Millaisella otannalla kokeellisen leksikaalisen semantiikan koehenkilöt tulisi valita ja millaisia taustamuuttujia heiltä pitäisi kerätä, että tulokset olisivat sosiolingvistisesti luotettavia? Koehenkilöiden eli informanttien taustoihin on otettu kantaa niin kielentutkimuksessa kuin laajemmassakin yhteiskunnallisessa mittakaavassa, esim. Tilastokeskuksen yleisohjeissa. On ilmeistä, että ainakin osa taustatiedoista on usein kysytty ”varmuuden vuoksi” ilman raportoinnissa esiteltyä analyysiä (esim. Mielikäinen 1986; Suojanen 1985); näin tässäkin esitellyssä tutkimuksessa. Taustatietojen keruu ja analysointi on suhteellisen helppoa, ne toimivat samalla alkulämmittelyinä itse sisältökysymyksille ja niiden perusteella on mahdollista seuloa joukosta poikkeustapauksia kuten leksikografit tai vieraskieliset. Standardoitua taustatietolomaketta on mahdollista hyödyntää myös muissa tutkimuksissa. Taustatietoja ei aineistonkeruuvaiheessa kannata eikä pidä ohittaa perusteita.

<sup>7</sup> Labovilaisen tutkimuksen tärkein muuttuja – luokka- ja rotuerot – on suomalaisessa tutkimuksessa usein ohitettu. Tähän on mitä ilmeisemmin vaikuttanut yhteiskuntien erilaisuus: Suomessa rotueroja ei juuri ole eikä luokkaerojen mahdollista olemassaoloa ole pidetty merkityksellisenä. Useissa suomalaisissa sosiolingvistisissä tutkimuksissa on painotettu sitä, että amerikkalaisten tutkimusten asetelmia ei sellaisenaan voida toteuttaa suomalaisessa ympäristössä. Kiinnostavaa olisi kuitenkin tietää, onko mahdollista, että kaikessa tasa-arvoisuudessaan (tai ainakin siihen pyrkimisessään) myös suomalainen ympäristö ja siinä puhuttu kieli sisältäisi laajemman sosiaalisen variaation kuin tutkimukset antavat ymmärtää. Luokkaerot eivät tule tutkimuksissa näkyviin, jos koko tutkittu ryhmä kuuluu samaan luokkaan tai jos tuloksia ei ole analysoitu taustamuuttujien mukaan (niiden keräämisestä huolimatta).

Nykyisin laajaksi ja kirjavaksi kasvanut sosiolingvistiikka – sen enempää kuin muukaan kielentutkimus – ei varsinkaan Suomessa ole tavoitteiltaan välttämättä lainkaan yhteiskunnallisesti suuntautunutta (vrt. esim. Labov 1997). Usein on jopa päinvastoin: korostetaan lingvistisen tutkimuksen autonomisuutta ja ympäröivän yhteiskunnan asettamiin toiveisiin tai tavoitteisiin suhtaudutaan hyvin varauksellisesti. Tutkimuksen motiivia ja relevanssia tulisi miettiä. Tämä tutkimus on alkuperäisen labovilaisen tutkimuksen kanssa samalla linjalla siinä, että se ainakin pyrkii hyödyttämään muitakin kuin tekijäänsä. Labov etsi epästandardia englantia vakuuttaakseen sen olemassaolon amerikanenglannin valkoisen murteen puhujille; tämän kirjoittajan motiivatio on löytää ”ihan normaalia suomea” ja sen perusteita sanakirjoja ja niiden käyttäjiä varten.

### 3.2 Sosiolingvistisesti suuntautuneen kokeellisen tutkimuksen huomioiminen leksikografiassa<sup>8</sup>

Ehkäpä leksikologia ja leksikografia saavat tulevaisuudessa määritteekseen sosiolingvistisen. Jos sosiolingvistisessä tai kokeellisessa tutkimuksessa selvitetään, miksi kielessä valitaan niin kuin valitaan, tulisi tiedon olla leksikografian näkökulmasta äärettömän kiinnostavaa.

Kielellisen tiedon tuottajan tai tallentajan taustatietojen merkinnässä on leksikografiassa ollut vaihtelua. Suppeimmillaan tieto on ns. tavallisessa yleiskielen sanakirjassa, jossa taustatietoihin viittaaminen puuttuu täysin. Murre-sanakirjassa on tiedossa ilmauksen tuottajan maantieteellinen tausta sekä yleensä myös ikä (aineistoa kerätty pääosin vanhoilta ihmisiltä); sukupuolesta ja sosioekonomisesta taustasta ei tavallisesti ole mainintaa.

Kaikesta variaatiosta tärkeimmälle tulisi olla tilaa ns. tavallisessa sanakirjassa. Sanaston sosiolingvistiseen tutkimukseen liittyy kuitenkin ongelmia (esim. Lappalainen 2003, 15–16). Erityisiä riskejä otetaan, kun aletaan määrittellä sanojen semantiikkaa ja pragmatiikkaa, käytön sosiaalista variaatiota jne. Toisaalta tähän asti on hyvin vapaasti tehty luokittelua esim. dialektologisin

<sup>8</sup> Polysemiatutkimuksessa on osoitettu (mm. Raukko 1994, 2002), että kokeellisella semantiikalla ja kyselytestimenetelmällä olisi annettavaa sanakirjojen polyseemisten sanojen artikkelien kokoamisessa. Niinkin yksinkertainen asia kuin merkitysten järjestys sanakirja-artikkelissa on kaikkea muuta kuin selvä (vrt. esim. Raukko 1997). Sanakirjan käyttäjälle – varsinkin vieraskieliselle – merkitysten järjestyksellä on kuitenkin väliä. Leksikografian pohdissa merkitysten keskinäistä järjestystä sanakirja-artikkeliä kirjoittaessa voisi olla hyvinkin kiinnostavaa tietää, miten äidinkieliiset puhujat tällä hetkellä merkitysten tärkeysjärjestyksen käsittävät. Puhumattakaan siitä, että vanhoille sanoille väistämättä tulee uusia merkityksiä, joiden ehdottomasti tulisi näkyä sanakirjassa. Valaisevana esimerkkinä mainittakoon tutkimukset *pitää*-verbin polysemiasista (Raukko 2002), joissa koehenkilöiden mukaan verbin ensimmäinen merkitys oli ’tykätä’; Perussanakirjan mukaan ’tykätä’ on merkitysjärjestyksessä vasta 15. sijalla.



perustein. Jos sosiolingvistiset mittarit ovat kunnossa, miksi ei uskallettaisi käyttää leksikografiassa myös niitä? Ns. normaali yleiskieli ei ole murreta vapaampaa sosiodemografisten taustamuuttujien vaikutuksesta – valkoinen amerikanenglanti on lingvistisesti samanarvoinen variantti kuin musta amerikanenglanti, vaikka paineet ensimmäisen variantin pitämiseen normina ovatkin suuret. Sosiolingvistien keskustelussa tulee usein esiin individualismin kunnioittaminen ja yleistysten varominen. Stereotyyppien luomista tulee varoa, mutta pitäisi löytyä rohkeutta tehdä välttämättömiä yleistyksiä. Mikään ei ole ikuista eikä muuttumatonta, ei edes sanakirjamääritelmä. Sen on saatava elää. Sanakirjamääritelmän elävyys ja ajantasaisuus on erityisen tärkeää toisen tai vieraan kielen puhujalle. On kannattavampaa varustaa sanakirjamääritelmä varauksilla kuin rajata kokonaan pois mahdollisesti liioittelevat yleistykset.

Sanakirjan tavoitteena tulisi olla käyttäjän mahdollisimman hyvä palvelu, ei niinkään jokaisen artikkelin toimituksellinen täydellisyys. Sisällön olenaisuutta ja tarkoituksenmukaisuutta tulisi korostaa, vaikka muodollinen johdonmukaisuus siitä kärsisi. Sanakirjan käyttäjä ja käyttäjän odotukset tulisi ottaa huomioon ensimmäisenä. Ehkäpä loppukäyttäjä olisi hyvinkin kiinnostunut interaktion järjestäytymisestä sosiaalisessa tilanteessa, jos nettisanakirjassa olisi hakusanasta linkkiesimerkki kotivideokatkelmaan – elävän kielen korpukseen. Suomea toisena kielenä puhuvilta tulisi selvittää, kuinka ongelmallisina eri synonyymiryhmät koetaan, ja tutkimussuunnitelma tulisi sopeuttaa siihen. Ei-äidinkielisten panos tutkimuksen suunnitteluun ja kulkuun saattaisi olla hyvinkin merkittävä. Tulisi pohtia uudelleen myös sanakirjan ja tietosanakirjan tehtäviä ja suhdetta.

Näistä näkökulmista sosiolingvistisellä tutkimuksella voisi olla runsaasti annettavaa leksikografialle myös ilman yhteyttä kokeelliseen sanastontutkimukseen. Liikkeelle voisi lähteä esim. asioimiskielen tutkimusten tuloksista ja vaikkapa miesten ja naisten eroista tervehdyksissä tai puheenvuorojen aloituksissa.

Kokeellisen semantiikan – ja muunkin kielitieteellisen tutkimuksen soveltamisen leksikografiaan tulisi perustua joustavaan suhtautumiseen kieleen. Perinteistä sanakirjaa vaalivien ensimmäisiä vastalauseita on, ettei koko sanastoa voitaisi kuitenkaan käsitellä yhtä perusteellisesti. Miksi pitäisikään? Koko sanaston pusertaminen saman tutkimuskoneiston läpi ei tee oikeutta kielelle. Joustavassa ja tarkoituksenmukaisessa leksikografian sovelluksessa hyödynnetään tarjolla olevaa tietoa ja pyritään saamaan lisätietoa ongelmallisista tai muuten kiinnostavista ilmiöistä. Tulisi huolehtia siitä, että suuret sanakirjaprojektit pysyisivät asiantuntevissa käsissä ja lopputulokset olisivat riittävän eläviä tullakseen hankituiksi ja käytetyiksi.

Kaikkeen kokeelliseen tutkimukseen liittyy menetelmän monentasoista hiomista kokemuksen ja uusien ideoiden pohjalta. Tutkimuskenttää voi halu-

nessa laajentaa esim. joukkoliikennevälineisiin tai www-pohjaisiin kyselylomakkeisiin, joiden käyttö mahdollistaisi moninkertaisesti suuremmat koehenkilöjoukot. Leksikografisesti kiinnostava on kysymys tutkittavan sanaston laadusta ja määrästä. Synonymiatutkimuksessa tulisi käydä systemaattisesti läpi sanakirjat (PS, NS, synonyymisanastot) ja luoda yleiskatsaus lisätutkimusta tarvitsevasta sanastosta.

### Lopuksi

Kokeellisen sanastontutkimuksen avainkysymyksiä on, onko mahdollista löytää peruskieltä ja sitä puhuvia ihannekoehenkilöitä. Nyt saamani tulosten perusteella totean, että sopiva koehenkilö on kuka tahansa aikuinen äidinkielenä tutkimuksen kohdekieltä puhuva kirjoitustaitoinen, joka saadaan rauhalliseen tilaan puoleksi tunniksi osallistumaan tutkimukseen. Koehenkilöjoukon suuruus ja heterogeenisyys parantavat tulosten yleistettävyyttä, mutta eivät näytä suuresti vaikuttavan tulosten keskeisiin kvalitatiivisiin ominaisuuksiin.

### LÄHTEET

- ARPE, ANTTI – JÄRVIKIVI, JUHANI 2002: Verbal synonymy in practice. Combining corpus-based and psycholinguistic evidence. <http://www.cogsci.uni-osnabrueck.de/qitl/QITL-Da-teien/Abstracts/arppejaervikivi.rtf>
- COLSTON, HERBERT 2002: Pragmatic Justifications for Nonliteral Gratitude Acknowledgments: "Oh Sure, Anytime". *Metaphor and Symbol*, 17(3), 205–226.
- HUHTA, ARI – SAJAVAARA, KARI – TAKALA, SAULI (TOIM.) 1993: Language Testing: New Openings. Institute for Educational Research. University of Jyväskylä.
- JÜRVTSON, MARIA-MAGDALENA 2004: Joidenkin naurua kuvaavien ekspressiivisten verbien äännevariaatioiden merkityseroista. 31. Keeleteaduse päevad. 31. Kielitieteen päivät. *The annual Finnish and Estonian conference of linguistics. Teesid. Esitelmäkirja. Abstract book*. Tallinn, 6.–7.5.2004, 31. Tallinn.
- LABOV, WILLIAM 1972: *Sociolinguistic Patterns*. Basil Blackwell. Oxford.
- 1997: How I got into linguistics, and what I go out of it. <http://www.ling.upenn.edu/~wlabov/Papers/HowIgot.html>
- LAPPALAINEN, HANNA 2003: *Variaatio ja sen funktiot. Erään sosiaalisen verkoston jäsenten kielellisen variaation ja vuorovaikutuksen tarkastelua*. SKST 964.
- MIELIKÄINEN, AILA 1986: *Nyky-suomalaisen puhekielen murros*. Jyväskylän osatutkimus. Raportti 4. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 32.
- NS = *Nyky-suomen sanakirja*. Helsinki: WSOY 1951–1961.
- PAJUSALU, RENATE 2001: Kas moos ja buss seisavad sarnaselst ehk väike katse verbiga *seisma*. R. KASIK (toim.): *Keele kannul. Pühendusteos Mati Ereli 60. sünnipäevaks*. TÜ eesti keele õppetooli toimetised 17, 250–274.
- PAUNONEN, HEIKKI 1982: Suomen kielen sosiolingvistinen vaihtelu tutkimuskohteena. M. K. SUOJANEN – PÄIVIKKI SUOJANEN (toim.): *Sosiolingvistiikan näkymiä*. Gaudeamus. Helsinki.
- PS = *Suomen kielen perussanakirja*. Helsinki: Kotimaisten kielten tutkimuskeskus 1990–1994.

- RAUKKO, JARNO 1994: Polysemia: Merkitysten verkosto – merkityksen verkko. PENTTI LEINO – TINA ONIKKI (toim.): *Näkökulmia polysemiaan*. Kieli 8. Helsingin yliopiston suomen kielen laitos.
- 1997: The Status on Polysemy in Linguistics: From Discrete Meanings to Default Flexibility. *The 1997 Yearbook of the Linguistic Association of Finland*, 145–170.
- 1999: An "intersubjective" method for cognitive-semantic research on polysemy: The case of *GET*. MASAKO K. HIRAGA – CHRIS SINHA – SHERMAN WILCOX (toim.): *Cultural, psychological and typological issues in cognitive linguistics. Selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995*, 87–105. [Current Issues in Linguistic Theory, 152.] Amsterdam/Philadelphia: John Benjamins.
- 2002: Pitämisen polysemia: miten koehenkilöt hahmottavat *pitää*-verbin merkitystyyppettä. *Virittäjä* 106, 354–374.
- SMS = *Suomen murteiden sanakirja*. 1985–. Kotimaisten kielten tutkimuskeskus. SKS.
- SORJONEN, MARJA-LEENA – PERÄKYLÄ, ANSSI – ESKOLA, KARI (toim.) 2001: *Keskustelu lääkärin vastaanotolla*. Vastapaino.
- SUOJANEN, M. K. 1985: *Mitä Turussa puhutaan? Raportti Turun puhekielen tutkimuksesta*. Turun puhekielen projektin julkaisuja 3. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 23.
- TRAGEL, ILONA – KÄHRİK, KAJA 2004: Experimental methods in analysing construction meaning. *20<sup>th</sup> Scandinavian Conference of Linguistics, January 7–9, 2004. University of Helsinki*, 129. Yleisjäljennös.
- VAINIK, ENE (tulossa): Intracultural Variation of Estonian Emotion Vocabulary: The Effect of Age and Gender of the Results of a List Task.
- VANHATALO, ULLA 2002: "Naiset motkottaa aiheesta ja nalkuttaa syyttä": Kyselytestit verbien semanttisten sisältöjen arvioinnissa. *Virittäjä* 106, 330–353.
- 2003: Kyselytestit vs. korpuslingvistiikka lähisyronymien semanttisten sisältöjen arvioinnissa – mitä vielä *keskeisestä* ja *tärkeästä*? *Virittäjä* 107, 351–368.
- YLIKOSKI, JUSSI 2004: Hilja käi kahtoo katiskat – käykääs kattoos mun kotisivut! Suomen "VI infinitiivin inessiivi" ennen ja nyt. *31. Keeleteaduse päevad. 31. Kielitieteen päivät. The annual Finnish and Estonian conference of linguistics. Teesid. Esitelmäkirja. Abstract book*. Tallinn, 6.–7.5.2004, 79. Tallinn.

## LIITE 1.

Ulla Vanhatalo/16.3.2004  
Helsingin yliopisto

Koodi  
Päivämäärä  
Paikka  
Aika

Jättäkää tämän viivan yläpuolinen osa täyttämättä.

## OHJEITA

- Tämä kysely tehdään nimettömänä.
- Täyttäkää yksi sivu kerrallaan katsomatta seuraavaa sivua. Kun olette täyttäneet sivun, irrottakaa se nipusta ja jättäkää pöydälle. Älkää tehkö täytettyyn sivuun enää korjauksia.
- Vastatkaa testin kysymyksiin juuri niin kuin Teistä itsestänne tuntuu. Kysymyksiin ei ole olemassa "oikeita" tai "väärä" vastauksia.
- Jos ette osaa rengastaa jonkin kysymyksen vaihtoehtoista yhtäkään, kirjoittakaa kysymyksen kohdalle "en tiedä".
- KIITOS vastauksistanne!

## 1. Äidinkielenne:

- a) suomi  
b) ruotsi  
c) muu, mikä? \_\_\_\_\_

## 2. Sukupuolenne:

- a) nainen  
b) mies

## 3. Ikänne:

\_\_\_ vuotta

## 4. Rengastakaa kaikki alueet, joilla olette asunut elämänne aikana yhtäjaksoisesti yhden vuoden tai kauemmin.

- a) Pohjois-Suomessa  
b) Keski-Suomessa  
c) Itä-Suomessa  
d) Länsi-Suomessa  
e) muualla Etelä-Suomessa kuin  
pk-seudulla  
f) pääkaupunkiseudulla  
g) ulkomailla

## 5. Rengastakaa paikka, joissa olette asunut suurimman osan lapsuuttanne ja nuoruuttanne.

- a) Pohjois-Suomessa  
b) Keski-Suomessa  
c) Itä-Suomessa  
d) Länsi-Suomessa  
e) muualla Etelä-Suomessa kuin

- pk-seudulla  
f) pääkaupunkiseudulla  
g) ulkomailla

## 6. Rengastakaa alue, joissa olette asunut suurimman osan aikuisikäänne.

- a) Pohjois-Suomessa  
b) Keski-Suomessa  
c) Itä-Suomessa  
d) Länsi-Suomessa  
e) Etelä-Suomessa  
f) pääkaupunkiseudulla  
g) ulkomailla

## 7. Jos olette asunut ulkomailla yli vuoden, mitä kieltä siellä puhuitte?

- a) vain omaa äidinkieltä  
b) vain paikallista kieltä  
c) sekä omaa äidinkieltä että paikallista kieltä

## 8. Jos vastasitte edelliseen (b) tai (c), mitä kieltä / kieliä puhuitte?

## 9. Mitä suomen murretta tai minkä alueen puhekieltä puhutte?

## 10. Jos osaatte äidinkielenne lisäksi vieraita kieliä, mitä ne ovat ja kuinka hyvin niitä osaatte? Merkitkää rastilla.

	osaan hyvin	tulen tarvit- taessa toimeen	osaan vain hiukan
--	----------------	---------------------------------------	-------------------------

ruotsi	___	___	___
englanti	___	___	___
saksa	___	___	___
ranska	___	___	___
viro	___	___	___
venäjä	___	___	___
espanja	___	___	___
italia	___	___	___
suomi	___	___	___
_____	___	___	___
_____	___	___	___

11. Mikä on/oli isännen äidinkieli?

- a) suomi  
b) ruotsi  
c) muu, mikä? \_\_\_\_\_

Jos isännen äidinkieli on/oli suomi, osaatteko sanoa, mitä murrettä hän puhuu/puhui?  
\_\_\_\_\_

12. Mikä on/oli äitinnen äidinkieli?

- a) suomi  
b) ruotsi  
c) muu, mikä? \_\_\_\_\_

Jos äitinnen äidinkieli on/oli suomi, osaatteko sanoa, mitä murrettä hän puhuu/puhui?  
\_\_\_\_\_

13. Minkä peruskoulutustason olette suorittanut?

(rengastakaa vain yksi vaihtoehto)

- a) en mitään  
b) kansakoulu, kansalaiskoulu  
c) keskikoulu tai peruskoulu  
b) ylioppilastutkinto

Koulun kieli oli

- a) suomi  
b) ruotsi  
c) muu, mikä? \_\_\_\_\_

14. Mikä on ammattikoulutuksenne?

- (rengastakaa vain yksi vaihtoehto)  
a) ei ammatillista koulutusta  
b) ammattikurssi (väh. 4 viikkoa)  
c) ammattikoulu  
d) opistotason tutkinto (esim. tekn.opisto, kauppaopisto, sair.hoit.opisto)  
e) yliopisto- tai korkeakoulututkinto  
f) aloitetut opinnot jossain mainituissa  
g) muu, mikä? \_\_\_\_\_

Oppilaitoksen kieli oli

- d) suomi  
e) ruotsi  
f) muu, mikä? \_\_\_\_\_

15. Oletteko opiskellut koulussa vieraita kieliä?

- a) olen  
b) en ole

16. Oletteko opiskellut kansalais- tai työväenopistossa vieraita kieliä?

- a) olen  
b) en ole

17. Oletteko opiskellut kieliaineita yliopistossa (esim. suomen kieli, jokin vieras kieli, yleinen kielitiede)?

- a) olen  
b) en ole

18. Millä alalla teette / olette tehnyt elämäntyönne?

- a) yrittäjänä
- b) maatalousyrittäjänä
- c) palvelualalla
- d) teollisuudessa
- e) muulla tuotantoalalla
- f) toimihenkilöalalla
- g) opetus- tai kasvatusalalla
- h) hoitoalalla
- i) tutkimusalalla
- j) vapaan ammatin harjoittajana
- k) kotona
- l) muualla

19. Millä alalla vanhempanne tekevät / ovat tehneet elämäntyönsä?

- a) yrittäjänä
- b) maatalousyrittäjänä
- c) palvelualalla
- d) teollisuudessa
- e) muulla tuotantoalalla
- f) toimihenkilöalalla
- g) opetus- tai kasvatusalalla
- h) hoitoalalla
- i) tutkimusalalla
- j) vapaan ammatin harjoittajana
- k) kotona
- l) muualla

20. Mikä seuraavista kuvaa tämänhetkistä elämäntilannettanne parhaiten?

- a) Käyn koulua.
- b) Opiskelen.

- c) Käyn töissä.
- d) Etsin töitä.
- e) Hoidan kotona lapsia.
- f) Olen eläkkeellä.
- g) Ei mikään edellisistä.

21. Mitä mieltä olette suomen kielen tilasta nykyaikana?

- a) suomen kieltä puhutaan huomattavasti paremmin nyt kuin ennen
- b) suomen kieltä puhutaan paremmin kuin ennen
- c) suomen kieltä puhutaan samalla tavoin kuin ennenkin
- d) suomen kieltä puhutaan huonommin kuin ennen
- e) suomen kieltä puhutaan nykyisin todella huonosti

22. Millaiset seikat suomen kielessä ovat teidän mielestänne vaikeita vieraskielisille?

- a) en ole asiaa erityisemmin miettinyt
- b) ääntäminen
- c) oikean sanan valitseminen
- d) yleinen sujuvuus
- e) puhuminen
- f) ymmärtäminen
- g) kirjoittaminen
- h) lukeminen
- i) aksentti
- j) sanojen taivutus
- k) sanajärjestys
- l) muu, mikä? \_\_\_\_\_

1) Ulkomaalainen tuttavanne osaa jo oikein hyvin suomea. Aivan kaikkea hän ei vielä kuitenkaan osaa. Nyt hän kysyy Teiltä, mitä tarkoittaa sana *edullinen*. Kuinka selittäisitte hänelle? \_\_\_\_\_

2) Mikä seuraavista on mielestänne tarkoittaa samaa kuin *edullinen* (eli on sanan *edullinen* synonyymi)?  
Rengastakaa yksi tai useampi. Jos rengastatte useamman kuin yhden, merkitkää vielä rasti sen sanan kohdalle, joka on Teidän mielestänne tämän päivän suomen kielessä kaikkein lähimpänä *edullista*.  
sopiva kannattava edukas etuisa hyödyllinen otollinen positiivinen myönteinen miellyttävä huokea taloudellinen halpa mainio suopea hyvä onnellinen tuottava voitollinen suotuisa \_\_\_\_\_

3) Kuinka selittäisitte ulkomaalaiselle tuttavallenne, mitä eroa on sanoilla *halpa* ja *edullinen*? \_\_\_\_\_

4) Mitä eroa on mielestänne sanoilla *erehdys* ja *virhe*? \_\_\_\_\_

5) Rastittakaa seuraavista asioista ne, jotka Teidän mielestänne kuvaavat *erehdystä* ja *virhettä*.

	<i>erehdys</i>	<i>virhe</i>
suuri	___	___
vakava	___	___
inhimillinen	___	___
ymmärrettävä	___	___
negatiivinen	___	___
vahinko	___	___
johtuu	___	___
huolimattomuudesta	___	___
väärin tehty asia	___	___
voi sattua helposti	___	___
pahat seuraamukset	___	___
siihen liittyy	___	___
riskinotto	___	___

6) KENEN PUHETTA SANA MIELESTÄNNE KUVAA?

Merkitkää mielestänne paras vaihtoehto rastilla. Vain yksi rasti kummallekin puolelle!

	mies	nainen	vaihtoehto			
			lapsi	nuori	keski- ikäinen	vanhus
jurnuttaa	___	___	___	___	___	___
jäkättää	___	___	___	___	___	___
valittaa	___	___	___	___	___	___
marissta	___	___	___	___	___	___
jupista	___	___	___	___	___	___
voivotella	___	___	___	___	___	___
jurista	___	___	___	___	___	___
marmattaa	___	___	___	___	___	___
napista	___	___	___	___	___	___
nurista	___	___	___	___	___	___
motkottaa	___	___	___	___	___	___

urputtaa	—	—	—	—	—	—
mukista	—	—	—	—	—	—
ruikuttaa	—	—	—	—	—	—
nalkuttaa	—	—	—	—	—	—
purnata	—	—	—	—	—	—
voihkia	—	—	—	—	—	—

7) KUINKA VOIMAKAS ÄÄNI  
SEURAAVIIN SANOIHIN MIELESTÄNNE LIITTYY?  
Merkitkää rastilla.

	hiljainen ääni		voimakas ääni		
	1	2	3	4	5
jurnuttaa	—	—	—	—	—
jäkättää	—	—	—	—	—
valittaa	—	—	—	—	—
marista	—	—	—	—	—
jupista	—	—	—	—	—
voivotella	—	—	—	—	—
jurista	—	—	—	—	—
marmattaa	—	—	—	—	—
napista	—	—	—	—	—
nurista	—	—	—	—	—
motkottaa	—	—	—	—	—
urputtaa	—	—	—	—	—
mukista	—	—	—	—	—
ruikuttaa	—	—	—	—	—
nalkuttaa	—	—	—	—	—
purnata	—	—	—	—	—
voihkia	—	—	—	—	—

8) Nyt Teidän pitäisi määritellä seuraavien sanojen merkitykset. Miten merkitykset mahdollisesti eroavat toisistaan?

ihmeellinen \_\_\_\_\_

outo \_\_\_\_\_

kummallinen \_\_\_\_\_

omituinen \_\_\_\_\_

KIITOS VASTAUKSISTANNE!

**Huomautus Liitteeseen 1. Kyselylomake esitetään tässä asetuksiltaan tiivistettynä. Alkuperäisessä lomakkeessa sisältökysymykset 1–8 esitettiin kukin omalla sivullaan. Artikkelissa ei käsitellä tehtäviä 5 ja 8.**



ULLA VANHATALO: *The impact of the backgrounds of the informants in experimental lexical semantics and in its applications*

Experimental semantics – which refers to exploring the semantics through testing the live speakers of the target language – has recently brought up several new perspectives to, for instance, polysemy and synonymy. In its criticism, there has been a raising concern about the potential impact of the homogeneity of the informants to the test results. The present study, belonging to so-called soft experimentalism, was set out to demonstrate whether results from population test series may vary according to the common background variables of the informants. To test this possibility, I have chosen to present the same questionnaire (“population test”) to informant groups that differ in their age, gender or educational status. Altogether 180 native Finnish speakers participated in this study. The results demonstrate that the backgrounds of the speakers do not have as much qualitative impact on the test results as commonly assumed. The size and heterogeneity of the study group would thus mainly improve the possibility to quantitatively generalize the results to the whole population.