



KYSELYTESTIT VS. KORPUSLINGVISTIIKKA LÄHISYNONYMIEN SEMANTTISTEN SISÄLTÖJEN ARVIOINNISSA — MITÄ VIELÄ KESKEISESTÄ JA TÄRKEÄSTÄ?

Vertaan tässä artikkelissa korpuslingvististä analyysiä ja kyselytestejä lähisynonymian tutkimusmenetelminä¹. Artikkelin pohjana olevassa tutkimuksessa käytämäni kyselytestisarja testaa aiemman korpuslingvistisen tutkimuksen tuloksia adjektiiveista *keskeinen* ja *tärkeä* (Jantunen 2001). Artikkelin analyysiosuudessa keskityn siihen, miten yksimielisiä koehenkilöt olivat sanojen sopivuudesta annettuihin lauseyhteyksiin, kuinka koehenkilöiden vastaukset suhteutuivat korpustutkimuksen tuloksiin ja mistä kyselytestin ja korpustutkimuksen tulosten erot mahdollisesti johtuivat. Kyselytesteillä saatiin tutkitusta sanastosta sellaista tietoa, jota korpuslingvistinen menetelmä ei tavoita.

JOHDANTO

KORPUSLINGVISTIIKKA SANASTON SEMANTTISEN SISÄLLÖN ARVIOINNISSA

Korpuslingvistiikka on nykyisen sanastontutkimuksen tärkeimpiä työkaluja. Erityisesti leksikografiassa korpuksista on etsitty esimerkkilauseita, kollokaatioita, fraaseja, sanojen esiintymistajuuksia ja taustatietojen kertomia tyylillisiä huomioita (esim. Biber ym.

¹ Lämpimät kiitokseni tutkija Tarja Riitta Heinoselle ja tutkija Jarmo H. Jantuselle sekä kahdelle nimettömälle esilukijalle arvokkaista kommentaiteista.



1998). Nykyiset tehokkaat korpustyökalut tuottavat hetkessä valtavan määrän toivottuja esiintymiä, kuten tietyn sanan kollokaatteja tai tietyn kieliopillisen lausekkeen edustuksia. Lähisynonyymien kollokaatiota on nykyisin korpuslingvistisin menetelmin tutkittu englannin kielessä esimerkiksi sanoista *big*, *great*, *large* (Biber ym. 1998), suomen kielessä tietävästi vain adjektiiveista *keskeinen* ja *tärkeä* (Jantunen 2001) sekä verbeistä *mieltii* ja *pohtii* (Arppe ja Järvikivi 2002 ja Arppe 2002).

Korpuslingvistisen analyysin periaatteissa korostuu kokonaisuuden hahmottamisen tärkeys: yksittäisen sanan merkitys määrittyy paitsi sitä ympäröivien sanojen merkitysten myös koko tekstin kautta; lisäksi tekstiä on käsiteltävä osana sen käyttöyhteyttä erottamatta merkitystä ja muotoa (esim. Stubbs 1993). Kiistattomista eduista ja mahdollisuuksista huolimatta korpuslingvistiseen sanastontutkimukseen liittyy myös paljon ongelmia ja rajoituksia. Semanttisessa kuvauksessa rajoituksia tulee siitä, onko aineisto riittävän laaja ja ovatko tulokset yleistettävissä. Käytännön ongelmia ovat muun muassa korpusten ja analyysiohjelmien saatavuus. Suomalaisessa kielitieteellisessä keskustelussa on kiinnitetty huomiota muun muassa kirjoitetun tekstin (erityisesti sanomalehtitekstin) yliedustukseen, tekstien saatavuuteen liittyviin tekijänoikeuskysymyksiin, korpustyökalujen tarkoituksenmukaisuuteen ja käytettävyyteen, Internet-hakukoneiden hakukielten kómpe-lyyteen, suurten korpusten automaattisessa koodaamisessa tapahtuviin virheisiin, hakujen ohessa tullesiin runsaisiin ei-toivotuttuihin osumiin ja tekstityyppien luokitteluun (Lahdenmäki 2002 Helsingin kielitieteen päivien korpustyöpajoista 2002). Muun muassa nämä ongelmat nostavat esiin ilmeisen tarpeen kehittää myös muita sanastontutkimuksen menetelmiä erityisesti silloin, kun tarkastelun kohteena ovat sanaston semanttiset tai pragmaattiset piirteet.

KYSELYTESTIT SANASTON SEMANTTISEN SISÄLLÖN ARVIOINNISSA

Kyselytestit² ovat suhteellisen uusi menetelmä sanastontutkimuksessa. Testejä on toistaiseksi käytetty lähinnä toisen kielen oppimisen tutkimiseen (esim. Paribakht ja Wesche 1999; Henriksen 1999; Newton 1995), jolloin tutkimuskohde on kuitenkin ollut sanaston sijaan oppimisprosessi. Varsinaisessa sanastontutkimuksessa menetelmää on sovellettu ainakin polysemian (muiden muassa Raukko 1999; 2002) ja lähisynonymian (Vanhatalo 2002) tutkimuksessa.

Kyselytestien käyttö perustuu käsitykseen tiedon intersubjektiivisuudesta (*intersubjectivity*; termin määrittelystä kielitieteessä ks. esim. Raukko 1999): mahdollisimman objektiivinen tieto sanan yhteisöllisestä merkityksestä saadaan tutkimalla kvantitatiivisesti monen yksilön subjektiivista ymmärrystä. Itse asiassa tällainen empiirinen tutkimusstrategia on vakiintunut useimmilla muilla tieteenaloilla. Kyselytestimenetelmässä tutkimus voidaan kohdentaa hyvin tehostetusti haluttuun osaan sanastoa. Kohdesanoista ja niiden käytöstä voidaan tarpeen mukaan saada esille hyvin monenlaista tietoa. Informanttijoukkoa voi tarvittaessa kasvattaa esimerkiksi Internetin välityksellä satoihin tai tuhansiin. Testimenetelmä mahdollistaa myös prosessimaisen jatkotestauksen sekä on joustavasti

² Kyselytesteistä käytetään myös nimitystä populaatiotestit. Termien englanninkielisiä vastineita ovat *population tests*, *survey tests* ja *experimental method*.

yhdistettävissä muihin sanastontutkimusmenetelmiin.

Kyselytestimetodin teoreettisista ongelmista huomio on keskittynyt lähinnä siihen, miten todenmukaisina ja luonnollisina testituloksia voidaan pitää. On totta, että tuloksia voi ohjata haluttuun suuntaan tarkoitushakuisesti testaamalla. Kyselytestisarja tulee rakentaa siten, että se nauttii kielentutkijoiden luottamusta. Testisarjaa koottaessa on otettava tarkkaan huomioon se, mitä sanastoa tutkitaan, miksi tutkitaan ja ketkä ovat informantteja. Myös käytännön olosuhteet asettavat vaatimuksia testisarjalle. On otettava huomioon, tehdäänkö testi paperilomakkeille vai Internetiin. Kuinka paljon aikaa testisarjaa varten voidaan käyttää? Millä informantit saadaan motivoitua luotettaviin vastauksiin? Hyvän testisarjan on oltava toistettavissa. Nykyisellään useat näistä kysymyksistä odottavat käytännön vakiintumista.

MENETELMIEN VERTAILU

Korpus- ja kyselytestimenetelmien tavoite on sama: saada lisää tietoa sanojen merkityksestä ja käytöstä. Merkittävimmät erot löytyvät sekä tutkimusaineiston keruutavasta että kerätyn aineiston olemuksesta. Korpusten kielen tuottaja ei tiedä, että aineistoa tullaan analysoimaan, kun taas kyselytestin koehenkilö tietää tuottavansa analysoitavaa kielenainesta. Korpustutkimuksessa enemmän tai vähemmän spontaanisti tuotettua materiaalia on seulottavana erittäin runsaasti; kyselytesteissä materiaalia tulee kohdennetusti ja suhteellisen vähän. Korpustutkimuksissa alkuperäisistä kielentuotostilanteista on analyysivaiheeseen mennessä kadonnut paljon kontekstiin liittyvää tietoa, vaikka sitä olisi eri vaiheissa tallennettu koodauksin ja selityksin. Kyselytesteillä voidaan taas pureutua koehenkilöiden kielentuntemukseen ja kognitiivisiin käsityksiin.

Toistaiseksi suomen kielestä on tarjolla hyvin vähän korpuslingvistiikkaa ja muita menetelmiä vertailevaa lähisyonymiaa koskevaa tutkimusta. Eri tutkimusmenetelmien vertailu on kuitenkin tieteenalan arvioinnin ja kehittämisen kannalta erittäin tärkeää. Tämä artikkeli pyrkii osaltaan vastaamaan tuohon tarpeeseen. Vertailututkimukseni pohjaksi valitsin yhden olemassa olevan korpustutkimuksen (Jantunen 2001 sanoista *keskeinen* ja *tärkeä*), jonka vertailupariksi tein vastaavanlaisen tutkimusasetelman kyselytestimenetelmällä. Tutkimuksessani vertasin *keskeisen* ja *tärkeän* kollokaatioita selvittäneen korpustutkimuksen tuloksia tekemääni kyselytestiin.

TUTKIMUSASETELMA

TUTKITTAVA SANASTO

Ennen varsinaista koeasetelmaa on syytä katsoa, miten suomen kielen tärkeimmät sanakirjat, Nykysuomen sanakirja (NS) ja Suomen kielen perussanakirja (PS), määrittelevät *keskeinen* ja *tärkeä* -sanojen merkitykset.

keskeinen *a.* -sesti *adv.* -syys *omin.* vrt. keskinen, keskinäinen. **1.** keskellä, keskustassa sijaitseva, vars. tämän aseman tärkeyttä t. sopivuutta ajatellen. | **K:** set hämäläisalueet. Kirkolle on valittava k. paikka. — *Yhd.* itse-, minä-, omak.; ihmis-,

▷

jumala-, kristusk. **2.** ed:een liittyen: huomattava, olennainen, tärkeä. | Seuran, romanin k:sin henkilö. Varsin k. kysymys. Saavuttaa k. asema jssak. Olla k:sesti mukana jssak toiminnassa. **3.** par. keskinen | Jäsenten k. kilpailu. Meidän k:semme asia. (NS)

keskeinen 1. keskellä, keskustassa sijaitseva; huomattava, olennainen, tärkeä. *Hottellin k. sijainti. Olla k:sessä asemassa jssak. Romanin k:sin henkilö. K. kysymys.* **2.** yhdys sanojen jälkiosana: jhk keskittyvä, kohdistuva, jtk painottava, korostava. *Itse-, minäk. Ihmis-, yksilö-. Perhe-, kuluttajak. Helsinki-k. aluepolitiikka.* **3.** *Ystävien k. [paremmin: keskinen, välinen] naljailu. keskeisesti Ratkaisuun k. vaikuttanut seikka. keskeisliike fys. kappaleen liike keskeisvoiman alaisena. Ympyräliike ym. k:keet. keskeislyriikka = keskuslyriikka. keskeisvoima fys. voima jonka vaikutussuora kulkee aina saman pisteen kautta. keskeisyys Sijainnin k. Ihmis-, jumalak. (PS)*

tärkeä a. -ästi *adv.* -ys *omin.* **1.** tarpeellisuutensa, arvokkuutensa, yleisyytensä tms. puolesta merkityksellinen, merkittävä, huomattava, tähdellinen. **a.** asioista, esineistä. | T:ät yhteiskunnalliset kysymykset. Taloudellisesti t. ongelma. Ratkaisevan t:itä tietoja. Valanteon t:ys. Itseluottamuksen t:ys. Jllak on t. merkitys jkllle. Jk on jkllle t:ää. Pitää jtk t:änä. Asian t:yden oivaltaen. Ongelma, joka t:ältä osaltaan on kansainvälinen. On minulla t:ämpääkin tekemistä kuin kuunnella juttujasi. On t:ätä, että tuomio tulee oikeudenmukainen. Lienen jo esittänyt t:immän sanottavani. Tarvitsen t:ästi lisätietoja. Minä otan t:än askeleen [= menen avioliittoon] KIVI. Hattu ei — — näyttele yhtä t:ää osaa kuin ennen BAGHEERA. — — seikka, jota ei kyllin t:ästi voi nuorten mieliin teroittaa Kianto. — Teknillisesti t:ät raaka-aineet. Suomen t:ämmät puulajit. Vehnä maan t:impänä viljalajina. Tekstiilit t:änä kauppatavarana. Taiteilijan kehitykselle t:ät ulkomaanmatkat. Kaupungin t:immät kadut. Gibraltarin t:ys Englannille. T. rautateiden risteusasema. Savuke on monelle ruokakin t:ämpi. Kirveen t:ys aiheutti sen jatkuvan täydellistämisen A.M. TALLGREN. — *Yhd. elin-, kuoleman-, nyky-, päivät.* **b.** ihmisistä. | T:immät kansanrunouden kerääjät. Johtajalleen t. apulainen. N.N. on ammattiyhdistysliikkeen t. johtohenkilö. — Vrt. 2. | Hysterinen t:ltä näyttämisen tarve K.MARJANEN. **2.** ihmisistä: itseään tärkeänä pitävä, omahyväinen, koppava. | Olla t:än näköinen, t:ätä poikaa. Otsa t:ässä rypyssä. »Olen saanut ylennyksen», hän kehaisi t:ästi. He ovat kovin t:än ja painavan näköisiä kaikki, sillä jokainen luulee, että juuri hänestä riippuu koko kokouksen menestys K.KAJANDER. — — ihan nenänpää t:yttään punoitti NUOLIV. — — heidän ääneensä on tullut t:yttä, joka kertoo heidän ihailevan itseään VALENTIN. — — ettei kannata suhtautua omaan itseensä lainkaan t:ästi A.KIVIMAA. — *Yhd. itse-, pikku-, turhan-, virant.* **3.** *kans. taipuvainen, perso, hanakka. | Oli vähän liian t. naisten perään KARHUM. Hän oli näet t. viinaan AK. (NS)*

tärkeä 1. Tarpeellisuutensa, arvonsa tms. takia merkittävä, merkityksellinen, huomattava, tähdellinen. *Elint. Ratkaisevan t. tieto. Sillä on t. merkitys meille kaikille. Kiitos on lapsille t:ää. Pitää jtk t:änä. On t:ää, että tiedät asiasta. On t:ää hoitaa sairaus mahdollisimman pian. Onko raha t:ämpi, t:ämpää kuin rakkaus? Eikö sinulla ole mitään t:ämpää tekemistä? Maan t:in satama. T:immät vientituotteemme.* **2.** omahyväinen, koppava, tärkeilevä; lapsista: (innostuneen) ylpeä. *Turhant. T. ja mahtipontinen mies. Lapset kertoivat t:inä seikkailustaan. tärkeästi Nenä t. pystyssä. (PS)*

Lyhyesti voidaan todeta, että yhteisten merkitysten lisäksi *keskeisellä* on sellaisia tilaan, paikkaan ja suhteisiin liittyviä merkityksiä, joita *tärkeällä* ei ole.

KORPUSTUTKIMUKSEN ASETELMA

Jantunen selvitti korpustutkimuksessaan kahden akateemisissa teksteissä taajaan esiintyvän lähisynonymien — *tärkeän* ja *keskeisen* — vieruskollokaatteja niin kutsutun tilastol-

lisen merkitsevyysindeksin avulla (Jantunen 2001: 176)³. Korpusaineisto oli peräisin Joensuun yliopiston kansainvälisen viestinnän laitoksella kootusta sähköisestä Käännös-suomen korpuksesta. Jantunen esittää, että korpuslingvistisin keinoin voidaan osoittaa *keskeisellä* ja *tärkeällä* olevan sekä yhteisiä (esiintyvät molempien sanojen yhteydessä) että omia (esiintyvät vain toisen sanan yhteydessä) vieruskollokaatteja. *Keskeisen* omia kollokaatteja ovat Jantunen mukaan muiden muassa *teema*, *idea*, *osuus* ja *ajatus*, kun taas *tärkeän* omia kollokaatteja ovat muiden muassa *seikka*, *asia*, *suhde* ja *tieto* (Jantunen 2001: 180). Jantunen artikkelissa on poimittu esimerkkilauseiksi lähinnä sellaisia lauseita, joiden katsottiin edustavan yksiselitteisesti jommankumman tutkitun sanan kollokaatioesiintymää. Minun tekemässäni kyselytestin analyysissä taas katsottiin, miten yksimielisiä koehenkilöt olivat suhteessa korpuksista uutettuun tietoon.

KYSELYTESTIEN TUTKIMUSASETELMA

Vertailukelpoisen aineiston saamiseksi tämän tutkimuksen monivalintatestit⁴ laadittiin vastaamaan mahdollisimman tarkasti korpustutkimuksen asetelmaa. Kyselytestien koehenkilöt eivät olleet tietoisia kollokaatiotutkimuksen teksteistä, vaan arvioivat annettuja (eli Jantunen artikkelissa mainittuja) lauseita itsenäisinä kokonaisuuksina.

Kyselytestitutkimuksen materiaali kerättiin syyskuussa 2002 suomen äidinkieleeseen ilmoittaneilta lukiolaisilta (n=73)⁵. Testit tehtiin Helsingissä Lauttasaaren yhteiskoulussa neljässä äidinkielen opetusryhmässä. Koehenkilöiden taustatiedoista kysyttiin ikää (16–19 vuotta), sukupuolta (naisia 38 %), syntymäpaikkaa ja entisiä asuinpaikkoja (88 % vastaajista oli asunut ainoastaan Suur-Helsingin tai Etelä-Suomen alueella).

Jokaiselle koehenkilölle jaettiin neljä sivua käsittävä testisarja, joka sisälsi yhden tuotto-testin ja 19 monivalintakysymystä (ks. seuraava luku ja liite 1 s. 365)⁶. Kaikki koehenkilöt saivat täytettäväkseen samat monivalintatehtävät, kun taas tuottotesti vaihteli opetusryhmittäin. Tuottotesti sijoitettiin testisarjan alkuun ennen monivalintatestejä, jotta monivalintatestin valmiit esimerkkilauseet eivät olisi vaikuttaneet sen täyttämiseen. Testilomake oli nidottu reunoistaan siten, että siitä pystyi täyttämään vain yhden sivun kerrallaan. Testitilannetta valvottiin ja paperit kerättiin pois sitä mukaa kuin ne oli täytetty. Testien täyttämiseen koehenkilöt käyttivät noin 6–20 minuuttia.

Monivalintatestin (n=73) 15 lauseparia perustuivat Jantunen korpustutkimuksen esimerkkilauseisiin (Jantunen 2001: 178–183; katso liite 1). Jokaiselle koehenkilölle esitettiin myös yksi rakenteeltaan seuraavanlainen tuottotestikysymys:

Ulkomaalainen ystäväsi osaa jo erittäin hyvin suomea. Nyt hän haluaisi tietää, mitä eroa on »keskeisellä» ja »tärkeällä». Miten selität hänelle?

³ Korpustutkimuksessa selvitettiin myös kolligatioita, jotka morfosyntaktisina ilmiöinä jäävät tämän artikkelin ulkopuolelle. — Tutkimusasetelmassa otettiin huomioon sanojen harvinaisuus, mikä on erinomainen asia. Tämän tapaisia tilastollista lähestymistapaa vaativaa sanastoa suomen kielessä riittää runsaasti esimerkiksi ekspressiivisessä eli omatopoeettis-deskriptiivisessä sanastossa.

⁴ Kyselytestiasetelmasta vrt. Vanhatalo 2002.

⁵ n viittaa koehenkilöiden lukumäärään kyseisessä tehtävässä.

⁶ Tässä artikkelissa esitettyjen ja analysoidujen kysymysten lisäksi koehenkilöt vastasivat vielä neljään rakenteeltaan samanlaiseen monivalintakysymykseen (Jantunen 2001: esimerkkilauseet 16–19).

Tuottotestissä selitettävänä adjektiivipareina olivat eri koehenkilöryhmissä joko *keskeinen ja tärkeä* (n=17), *olennainen ja tärkeä* (n=35) tai *merkittävä ja tärkeä* (n=21). Tuottotestikysymyksen lisäadjektiivit (*olennainen ja merkittävä*) oli valittu esiintymistiheyden perusteella korpustutkimuksessa esitetystä frekvenssitaulukosta (Jantunen 2001: 175). Analyysivaiheessa varauhin hylkäämään täysin tyhjä tai epäasiallisesti täytetyt testisivut, mutta yksikään vastauspaperi ei osoittautunut hylättäväksi.

KYSELYTESTIN MONIVALINTAOSUUDEN TULOSTEN ESITTELY JA ANALYYSI

Tulosten analyysissä etsin vastauksia kolmeen pääkysymykseen:

- Kuinka yksimielisiä koehenkilöt keskenään olivat vastauksissaan? (Tulokset on esitetty koehenkilöiden yksimielisyysasteen mukaisessa järjestyksessä.)
- Kuinka koehenkilöiden vastaukset vastaavat korpustutkimuksen tuloksia? (Vastaaavat a-kohdan järjestystä lukuunottamatta lukuunottamatta lauseita (5) ja (7).)
- Mistä kyselytestin ja korpustutkimuksen tulosten erot mahdollisesti johtuivat?

Tämän analyysistrategian tarkoitus oli koetella korpuslingvistiksi saadun tiedon pätevyyttä. Tarkoitus oli myös etsiä korpusmenetelmän mahdollisiin heikkoihin kohtiin täydennysmahdollisuuksia empiirisestä kyselytestimenetelmästä.

Analyysia varten jaoin kyselytestin 15 lausetta ryhmiin sen mukaan, kuinka yksimielisiä koehenkilöt keskenään olivat siitä, kumpi tutkituista sanoista (*keskeinen* vai *tärkeä*) sopii paremmin kyseiseen lauseeseen (ks. taulukko alla). Mitä suurempi prosenttiluku on, sitä yksimielisempiä koehenkilöt keskenään olivat eli sitä vahvempi on intersubjektiviinen näyttö tietyistä sananvalinnasta. Useimmissa tapauksissa kyseessä on korpustutkimuksen tuottama niin sanottu oma kollokaatti (lukuunottamatta lauseita (5) ja (7)).

Kuinka samaa mieltä koehenkilöt olivat keskenään sanan sopivuudesta	Lauseet
Täysin samaa mieltä: 96–100 %	6, 14
Hyvin samaa mieltä: 81–95 %	3, 7, 9, 12, 13
Melko samaa mieltä: 66–80 %	2, 4, 5
Varovasti samaa mieltä: 51–65 %	8, 10
Varovasti eri mieltä: 36–50 %	11, 15
Melko eri mieltä: 21–35 %	1
Hyvin eri mieltä: 6–20 %	–
Täysin eri mieltä: 0–5 %	–

Lauseissa (1)–(8) oli 1L-kollokaatteja (1 left, so. tutkimuskohteeseen kuuluu yksi sana tutkittavan sanan vasemmalta puolelta). Lauseet (9)–(15) sisälsivät puolestaan 1R-kollo-

kaatteja (1 right, so. tutkimuskohteeseen kuuluu yksi sana tutkittavan sanan oikealta puolelta). Seuraavassa esitetyissä lauseissa kursiivi osoittaa, mitkä sanat olivat korpustutkimuksessa tarkastelun kohteena. Kyselytestissä lauseet esitettiin ilman minkäänlaisia kursivointeja. Lauseiden numerointi on alkuperäisen korpustutkimusartikkelin (Jantunen 2001) mukainen.

TÄYSIN SAMAA MIELTÄ

- (6) *Tutkimuksen keskeinen* idea maallikoiden ja asiantuntijoiden riskin määrittelyjen eroista jätetään mainitsematta – –

Lauseessa (6) *keskeisen* 1L-kollokaattina oli *tutkimus*; kontekstia tarkastellessa ei tosin voi välttyä ajatukselta, että koehenkilöt olisivat tässä tapauksessa tehneet valintansa pikemmin 1R-kollokaatin eli *idean* mukaan (vrt. samanlainen tilanne lauseessa (4)). Korpustutkimuksessa pidetäänkin 1R-kollokaatteja yleisesti vahvempina ja merkittävimpinä kuin 1L-kollokaatteja, näin erityisesti *keskeisen* osalta (Jantunen 2001: 179, 181). Tämä vahvojen substantiivisten 1R-kollokaattien asema näkyy mitä ilmeisimmin myös testituloksissa.

- (14) Yksi puheaktiteorian *keskeinen ongelma* on epäsuoruus, jossa Searlen (1975, 60) mukaan puhuja esittää – –

Lauseen (14) osalta koehenkilöt ja korpustulokset ovat täysin yksimielisiä: *ongelma* siis näyttäisi olevan *keskeisen* omia 1R-kollokaatteja (Jantunen 2001: 179).

HYVIN SAMAA MIELTÄ

- (3) Ajallisen jatkuvuuden (samuuden) ohella myös erottautuminen on *hyvin tärkeää*.

84 % kyselytestiin vastanneista oli sitä mieltä, että *hyvin tärkeää* on sopivampi kuin *hyvin keskeistä* (lause (3)). *Tärkeän* omien 1L-kollokaattien osalta suurin yksimielisyys korpustutkimuksen ja kyselytestin tuloksissa saavutettiin astemääritteen *hyvin* kohdalla.

- (7) Mielikuvat ovat *tärkeitä*, mutta eivät merkitse koko asiaa.

Korpustutkimuksen mukaan *olla*-verbi on sekä *keskeisen* että *tärkeän* taajimmin esiintyvä 1L-kollokaatti (Jantunen 2001: 178); tässä on muistettava myös *olla*-verbin erityisasema usein esiintyvissä peruslauseissa. Lauseessa (7) 86 % kyselytestiin vastanneista oli *tärkeän* kannalla. Tässä tapauksessa huomio kiinnittyy kuitenkin 2L-kollokaattiin eli *mielikuviin*, joka todennäköisesti on vaikuttanut vahvasti sanavalintaan.

- (9) Semioottisina olioina tekstejä yhdistää muuan tuiki *tärkeä seikka* — nimitään se, että ne kaikki ovat representaatioita.

Korpustutkimuksessa *tärkeän* omista 1R-kollokaateista merkittävimpiä oli substantiivi *seikka*. Koehenkilöistä 82 % valitsi lauseeseen (9) sopivammaksi adjektiivin *tärkeä*

▷

kuin *keskeinen*. Voimakkaasti vaikuttavana tekijänä on mitä ilmeisimmin ollut myös L1-kollokaatin paikkaa pitävä adverb *tuiki*, joka enemmistön suomea äidinkielenään puhuvien kielitajun mukaan ei sovi *keskeisen* määritteeksi⁶. On mielenkiintoista, että 18 % vastaajista kuitenkin hyväksyi *tuiki keskeisen*. Onko taustalla nuorten kielitajun muuttuminen vai esimerkiksi huolimattomuus testintekotilanteessa, jää toistaiseksi arvoitukseksi.

- (12) Kirjoittajien *keskeinen teesi* on, että sosiaaliset ja pragmaattiset tekijät vaikuttavat luokittelujärjestelmiin – –
- (13) Myös Lehtosen romaanissa paloviinan salapoltto ja salakauppa olivat mitä *keskeisin teema*.

79 % (lause (12)) ja 84 % (lause (13)) kyselytestiin vastanneista oli korpustulosten kannalla, joiden mukaan *teesi* ja *teema* ovat *keskeisen* omia kollokaatteja.

MELKO SAMAA MIELTÄ

- (2) Samastumisen ulottuvuus ei ylipäänsä ole *yhtä tärkeä* kuin Suomessa.

Korpustutkimuksen mukaan astemäärite *yhtä* on yksinomaan *tärkeän* määrite. Kyselytestin koehenkilöistä 67 % oli samaa mieltä.

- (4) Näiden *kahden keskeisen* jännitteen strukturoima feminististen tieto-oppien kolmijako on Sandra Hardingin käsittein – –

Keskeisen omista L1-kollokaateista korpustutkimuksessa esitettiin lukusanakvanttori *kaksi*. 77 % koehenkilöistä oli korpustutkimuksen kanssa yhtä mieltä. Lauseen (4) kohdalla herää taas kysymys 1R-kollokaatin vaikutuksesta (vrt. lause (6)), uskoisin koehenkilöiden tehneen valintansa pikemmin 1R- kuin 1L-kollokaatin mukaan. Lukusanakvanttorin omistautuminen tietyille adjektiiville tuntuu muutenkin arveluttavalta ja johtunee ehkä pikemminkin metodisista tekijöistä (esimerkiksi suhteellisen pienestä korpusaineistosta).

- (5) Oppikirjat tarjoavat tähän monimutkaiseen tarinaan *yhden keskeisen* kertomuksen.

Tilastollisesti merkittävin *keskeisen* kollokaatti on korpustutkimuksen mukaan *yksi*; sanalla on tosin kollokaatiosuhde myös *tärkeän* kanssa. *Yhden* ja *keskeisen* suhde on korpustutkimuksessa kuitenkin huomattavasti tiiviimpi kuin *yhden* ja *tärkeän* suhde (Jantunen 2001: 177). Kyselytestin koehenkilöt olivat samoilla linjoilla: lauseessa (5) 68 % vastanneista piti *keskeistä* sopivampana sananvalintana. Toki taas herää epäily substantiivisen 1R-kollokaatin merkityksestä sananvalintaan.

⁶ Asia oli päinvastoin (julkaisemattomassa) rinnakaistutkimuksessani, jossa erittäin hyvin suomea taitavat englantia äidinkielenään puhuvat aikuiset (n=7) vastasivat samoihin kysymyksiin. Suurin osa suomea toisena kielenä puhuvista oli *tuiki keskeisen* kannalla.

VAROVASTI SAMAA MIELTÄ

- (8) Kuten myöhemmin käy ilmi, varsinaisen metaforan käytön ymmärtämiseksi Väyrysen kampanja *on keskeinen* osa laajempaa kontekstia.
- (10) – eri puolilla Lappia havaitut yksittäiset metsävauriot kasvoivat *tärkeäksi yhteiskunnalliseksi* ilmiöksi nimenomaan siksi, että –

Lauseessa (8) koehenkilöiden valinnat *tärkeän* hyväksi saattavat johtua vahvasta 1R-kollokaatista (substantiivi *osa*). *olla*-verbi on korpustutkimuksen mukaan molempien adjektiivien frekventein 1L-kollokaatti; katso myös lause (7). Lauseessa (10) vain niukka enemmistö koehenkilöistä oli samaa mieltä kuin korpustutkimuksen tulokset, joiden mukaan *yhteiskunnallinen* oli *tärkeän* niin sanottu oma kollokaatti. Onko mukana taas 2R-vaikutusta?

VAROVASTI ERI MIELTÄ

- (11) Jos joku näistä seikoista olisi kulttuurin kannalta yhtä *tärkeä kuin* meille sattuu olemaan homous, juoppous tai aggressiivisuus –

Niukka vähemmistö (47 %) kyselytestiin vastanneista oli samaa mieltä kuin korpustutkimus, jonka mukaan *kuin on tärkeän* oma kollokaatti (lause (11)). Laajempaa kontekstia tarkastellessa paljastuu 5L-asetmassa oleva substantiivi *seikka*, joka on mahdollisesti vaikuttanut sanavalintaan. Tässä on erinomainen esimerkki siitä, miten suomen kaltaisessa vapaan sanajärjestyksen kielessä olisi erittäin suotavaa tarkastella myös sanan välitöntä ympäristöä kauemmaksi ulottuvia kollokaatteja.

- (15) – esimerkiksi sellainen, jossa informaation erittelyn asemasta *keskeiseksi nousse* merkityksen tulkinta.

Kyselytestin tulosten mukaan lauseessa (15) ei ollut eroa *keskeisen* ja *tärkeän* välillä (48 % vastauksista oli *keskeisen* puolesta).

MELKO ERI MIELTÄ

- (1) Tämän seurauksena kristinuskon rooli muuttui *vähemmän tärkeäksi*, vaikka se edelleen toimi sekä normatiivisena rauhoittajana –

Korpustutkimuksen mukaan kvantiteettia ilmoittava astemääräte *vähemmän* on yksi merkitevimmistä *tärkeän* omista kollokaateista. Korpustutkimuksessa *vähemmän* ei esiintynyt kertaakaan *keskeisen* tilastollisesti merkittävänä 1L-kollokaattina. On erittäin kiintoisaa, että kyselytestin koehenkilöt olivat asiasta lähestulkoon päinvastaisella kannalla. Vastaajista vain 33 % piti lauseessa (1) sopivampana vaihtoehtona adjektiivia *tärkeä*. Hajaannuksen syy saattaa tässä tapauksesta löytyä 3L-kollokaatista: *rooli* on suhteellisen merkittävä sekä *keskeisen* että *tärkeän* R-kollokaatti (Jantunen 2001: 179). Tämä on mainio esimerkki vapaan sanajärjestyksen mahdollistamasta kollokaatiosuhteiden sekoittumisesta.

▷

On ilmeistä, että koehenkilöiden kielenkäyttö ja kyky kertoa käsityksistään vaihtelevat testitutkimustilanteen mukaan. Tekemäni tutkimus perustui korpustutkimuksessa annettuihin esimerkkeihin ja oli siten luonteeltaan hyvin tarkasti rajattu monivalintatesti: koehenkilöt pakotettiin valitsemaan yksi annetuista vaihtoehdoista. Luonnollisen kielen semantiikka pohjautuu kuitenkin huomattavasti laajempiin kokonaisuuksiin kuin 1L-/1R-kollokaatit. Siksi monipuolisemman kognitiivisen tiedon saamiseksi koehenkilöitä pyydettiin vastaamaan myös yhteen vapaamuotoiseen tuottokysymykseen, joka esitettiin ennen monivalintatestiä. Tuottotestillä saadaan selvitettyä, mitä koehenkilöt spontaanisti kertovat kyseisistä sanoista.

Tuottotestien tuloksista käy ilmi (liite 2 s. 367), että saatujen vastausten perusteella *keskeinen* ja *tärkeä* (samoin kuin lähisyronyymit *olennainen* ja *merkittävä*) saavat todella runsaasti sellaista luonnehdintaa, jota kollokaatteihin painottuvasta korpustutkimuksesta (tai nykyisistä sanakirjoista) ei saada selville. Liitteessä 2 esitetyn kaltainen raakatieta vaatii toki jatkokäsittelyä: saatujen vastausten perusteella tulee rakentaa monivalintapohjainen kyselytestisarja (vrt. Vanhatalo 2002), jonka uudet koehenkilöt täyttävät arvioiden näin ensimmäisen kierroksen koehenkilöiden vastauksia. Mahdollisia tutkittavia piirteitä voisivat tuottotestien perusteella olla esimerkiksi henkilökohtaisuus, yleisyys, kiireellisyys, välttämättömyys, pakollisuus, itsestänselvyys, yleinen tunnettuus ja tunnustettuus. Tällaisessa tutkimusasetelmassa kannattaisi tutkia *keskeisen*, *tärkeän*, *olennaisen* ja *merkittävän* lisäksi myös muita merkitykseltään läheisiä sanoja kuten *huomattava*, *oleellinen*, *arvokas*, *merkityksellinen* ja niin edelleen.

Yhtä tuottotesteissä ilmennyttä seikkaa on syytä kommentoida: *tärkeää* määriteltäessä vastaukset sisälsivät myös jonkin verran kehämääritelmiä, esimerkiksi *tärkein on tärkein asia* tai *tärkeä on asia joka on vaan hyvin tärkeä*. Tämä viittaa siihen, että *tärkeää* pidettiin eräänlaisena yleiskäsitteenä muiden tutkittujen sanojen joukossa. Kehämääritelmien aiheuttaja voi tosin olla myös tutkitun sanaston tuttuus. Kyseessä on varsin abstraktit adjektiivit; ainakin *merkittävän*, *keskeisen* ja *olennaisen* käyttö on harvinaisempaa puhekielessä kuin kirjakielessä⁷.

MITÄ TULOISTA VOIDAAN LUKEA?

Tuloksia tarkasteltaessa on muistettava sekä korpusaineiston että kyselytestin koehenkilöryhmän laatua ja määrää koskevat rajoitukset. Korpusaineisto sisälsi vain akateemista tekstiä ja oli suhteellisen suppea. Kyselytestin koehenkilöjoukko oli puolestaan rajattu sekä kooltaan (n=73) että taustaltaan (ikä, koulutus, asuinpaikka). Valmista korpustekstiä tarkastellessa on vaikea tietää, millaisiin kokonaisuuksiin liittyväksi ja mistä lähtökohdista teksti on alun perin kirjoitettu. Kielenkäyttäjän suhde kieleen on hyvin erilainen hänen kirjoitettaessaan korpustekstiä kuin kyselytestivastausta. Merkittävä rajoitus testiasetel-

⁷ Huomattakoon, että erittäin hyvin suomea taitavat englantia äidinkielenään puhuvat aikuiset (n=7) vastasivat samoihin kysymyksiin perusteellisesti ja runsain esimerkein (julkaisematon rinnakkaistutkimus).

massa oli se, että suurin osa esimerkkilauseista oli korpusten perusteella mustavalkoisia (vain jommankumman tutkitun sanan omia kollokaatteja), mikä saattoi siis ohjata tuloksia korpustutkimuksen tulosten suuntaan. Tässä artikkelissa esitettyjen tulosten tarkoitus onkin ensisijaisesti toimia välineenä kahden tutkimusmenetelmän vertailussa.

Testivastaukset ovat keskimäärin hyvin samaa tai melko samaa mieltä kuin korpustutkimuksen tulokset. Hyvin eri mieltä olevia vastauksia ei ollut lainkaan. Näin ollen korpuslingvistinen menetelmä ja kyselytestimenetelmä tuottivat suhteellisen yhdenmukaisia tuloksia *keskeisen* ja *tärkeän* vieruskollokaattien osalta. Tulosten yhdenmukaisuus ei tosin aina välttämättä johtunut samoista syistä (ks. lukua Kyselytestin monivalintaosuuden tulosten esittely ja analyysi s. 356).

Testituloksista on pääteltävissä, että lähisynonyymien valintaan mitä ilmeisimmin vaikutti laajempi konteksti kuin vain vierussanat tai mahdollisesti jokin kollokaatiota laajempi seikka. Kun tutkimuksen kohteena on suhteellisen vapaan sanajärjestyksen kieli kuten suomi, olisi tärkeää tutkia kollokaatteja laajemmalla säteellä (vrt. Jantunen 2001: 177 alaviite). Kotekstin rajaaminen yhden sanan säteeseen voi aiheuttaa hankaluuksia aineiston tulkinnassa ja jatkohyödyntämisessä: mukaan lipsahtaa helposti teknisiä vääristymiä⁸. Vielä hankalampaa voi olla jäljittää harvasta seulasta läpi menneitä muutaman sanan yli hyppiviä kollokaatioita tai kollokaatioiden yhdistelmiä. Adjektiiviveja tutkittaessa olisi erityisen tärkeää kiinnittää huomiota tavanomaista voimakkaampiin substantiivisiin kollokaatteihin. Oikeastaan voi olla jopa harhaanjohtavaa tarttua 1R- tai 1L-asemaan, koska kollokaatiosuhteen aiheuttajana on usein pääsanasubstantiivi, jonka paikka lauseessa voi olla hyvinkin liikkuva (esimerkiksi lauseet (1) tai (11)).

Toisaalta tämän tutkimuksen pohjana ollut Jantusen alkuperäinen kollokaatiotutkimus on osa laajempaa kokonaisuutta; tässä vertailuaineistona käytetty osa tutkimuksesta oli keskittynyt juuri vieruskollokaatteihin ja antoi niistä paljon kontekstuaalista tietoa⁹. Korpustutkimuksessa oli alun pitäen hakualueena L2–R2, joka antoi Jantusen mukaan yllättävän samanlaisen kuvan adjektiivien valinnoista kuin L1–R1. On myös muistettava, että korpustutkimuksen tarkoituksena ei ollut antaa kattavaa kuvaa kyseisten adjektiivien kollokationaalisesta profiilista vaan kuvata joitain synonymian tutkimiseen soveltuvia korpusten menetelmiä käyttämällä esimerkkisanoina *keskeistä* ja *tärkeää*. Näin ollen hakualueen rajaus on tässä tapauksessa ymmärrettävä. Kotekstin laajentamisessakin tulee vastaan monenlaisia ongelmia: rajautuako vieruskollokaatteihin, koko teokseen, tekstilajiin vai tietyn tyyppiseen diskurssiin? Hakualueen laajetessa tutkittavien muuttujien määrä kasvaa nopeasti hallitsemattomaksi.

Mikä sitten on kollokaatiosuhteiden merkitys laajemmassa mittakaavassa lähisynonyymien eroja tarkasteltaessa? Vaikka kollokaatiosuhteiden selvittäminen on oleellista (ks.

⁸ Laajempi konteksti on erityisen tärkeä silloin, kun kuvataan sanojen kollokatiivisia rajoitteita (Cruse 1986: 279). Jantusen tulosten mukaan *tärkeällä* ja *keskeisellä* on yhteisten kollokaattien lisäksi runsaasti sellaisia kollokaatteja, joita toisella ei ole (2001: 180). Tämä pätee vieruskollokaatteihin, ja olisi kiinnostavaa, jos se pätsisi myös laajempaan hakualueeseen. Sitä kannattaisi selvittää sekä laajemmalla kontekstilla että kyselytestimenetelmällä.

⁹ Kappaleen tiedot hakualueesta sekä näkemykset kotekstin laajentamisesta ja merkityksestä perustuvat Jantusen lähettämään yksityiseen sähköpostiin 12.2.2003.

esim. Inkpen ja Hirst 2002 sekä Edmonds ja Hirst 2002), ei muuta sanaan ja sen käyttöön liittyvää kielellistä tietoa tulisi unohtaa (esim. Cruse 1986 tai Jones 1986: 42–). Sanaston semanttisia piirteitä ja sanojen keskinäisiä suhteita tutkittaessa kollokaatiot ovat vain yksi semanttisten kytkösten laji. Vaikka kollokaatioiden poiminta sähköisistä korpuksista on nykyisin teknisesti helppoa ja suhteellisen yksiselitteistä, ne eivät ole kielen käytön, tutkimisen tai ymmärtämisen kannalta sen merkittävämpiä kuin aiemminkaan. Toisin sanoen tiedonsaannin helppous ei sanele tiedon tarpeellisuutta tai hyödyllisyyttä. Siksi on suhtauduttava tosissaan muihinkin sanastontutkimuksen menetelmiin. Luonnollista kielenymmärrystä tutkiva empiirinen metodi, kyselytestit, on mitä varteenotettavin menetelmä lähestyä sanaa ja sen merkityskomponentteja laajemmin, leksikologisen semantiikan perusteihin nojauten (esim. Nida 1975: 32–67, 232).

Sanastontutkimusmenetelmien tulisi ottaa huomioon paitsi kielen variaatio myös sanaston osien erityisominaisuudet; tutkimusmenetelmän valinnan tulisi perustua tutkittavan sanastoryhmän erityispiirteisiin. Leksikaalis-semanttisen tutkimuksen rajaaminen korpuslingvistiikkaan ei tee oikeutta esimerkiksi harvafrekvenssiselle sanastolle. Tutkimusvoimavarojen puitteissa on myös tärkeää tarkastella tiettyä sanaryhmää eri tutkimusmenetelmin, esimerkiksi sekä korpusten että kyselytestien avulla. Tutkimusasetelman tulisi olla tutkijasta riippumaton ja avoin: on hyvin hankala perustella tutkimusasetelmaa, jossa tutkijan intuitio asettaa rajat tutkittavalle ilmiölle. Saadun tiedon paikkansapitävyys, yksityiskohtaisuus ja hyödyllisyys ovat tärkeitä pohdinnan aiheita. Äidinkieltään tutkivan on viisasta pidättäytyä tuomitsemasta joidenkin merkitys- tai käyttöpiirteiden tutkimista turhaksi vain omaan intuitionsa vedoten. Oman näkemykseni mukaan kaikkeen sanan merkityksestä esitettyyn tietoon tulisi suhtautua ensisijaisesti kiinnostuneesti. Kirjoihin päätyvä lingvistinen määritelmä ei ole lopullinen totuus, vaan tieteellinen hypoteesi, jota tulee testata, jotta sitä voitaisiin parantaa, testata uudelleen tai hylätä (Wierzbicka 1996: 239).

LOPUKSI

Esimerkkiaineistoon pohjautuva menetelmävertailu ja pohdinta osoittavat, että kyselytesteillä voidaan sekä tukea korpuslingvistiikan tuloksia että hankkia sanojen merkityksistä sellaista tietoa, jota korpuslingvistiikan keinoin ei saada selville. Korpustutkimuksen ja kyselytestitutkimuksen filosofisia eroja ei tarvitse paisutella. Molemmissa tehdään empiiristä kvantitatiivista tutkimusta: korpuslingvistiikka pureutuu jo tuotettuun, kyselytestimenetelmä elävään aineistoon. Tässä artikkelissa esitetyn kyselytestin monivalintaosuuden analyysi perustui muutamien virkkeiden testaamiseen, kun taas korpustutkimus perustui tilastollisesti merkitseviin esiintymiin. Kyselytestit mahdollistavat hyvin monenlaisten rakenteiden käytön ja hyvinkin harvafrekvenssisten sanojen valikoidun tutkimisen.

Yksityiskohtaisen sanastontutkimuksen lisäksi on keskusteltava siitä, miten ja mihin sanastontutkimusten tuloksia voidaan hyödyntää. Motiivien ja tavoitteiden selvittely on tärkeää ainakin kahdesta syystä: ensinnäkin sanastontutkimuksen kenttä on laaja ja työtä on paljon, joten tutkimusresurssit tulee käyttää tehokkaasti hyväksi, ja toiseksi työ ei saa jäädä vain tieteellisten artikkelien tasolle, vaan tuloksille on voitava nähdä käytännön tason sovelluksia.

Niin teoriankehittämissä kuin myös yksittäisten sanojen merkitysten määrittelyssä olisi pidettävä mielessä se, miksi tutkimusta tehdään. Teoreettisesti hyvät tutkimustulokset jäävät ilmaan, ellei tieto päädy kuluttajille eli sanakirjoihin. Siksi tarvitaan avointa ja luovaa yhteistyötä kokeellisen semantiikan, korpuslingvistiikan, kielenoppimisen tutkimuksen, leksikografian ja kielenkäyttäjien välillä.

Voimakkaasti kehittyvä sähköinen leksikografia on luonnollinen kokeellisen sanastontutkimuksen tulosten hyödyntäjä. Määritellössään tehtäviään kokeellisen sanastontutkimuksen olisi palattava ruohonjuuritasolle tarkkailemaan sanakirjan käyttäjän ongelmia. On kysyttävä, millaisessa tilanteessa sanakirjaa tarvitaan, miten se jättää käyttäjänsä puolaan ja mitä ongelmia käyttäjä huomaa. On syytä perusteellisesti ja jatkuvasti pohtia sitä, millainen tieto on sanakirjan käyttäjän kannalta tärkeää. Samalla on myös määriteltävä se, mikä on asianmukaista ja todenmukaista tietoa. Onko sanakirjoissa käyttäjän kannalta yhdentekevää tietoa? Sähköisten sanakirjojen aikakaudella tekniset mahdollisuudet ovat muuttuneet ja monet perustavaa laatua olevat kysymykset tulevat uudelleen arvioitaviksi. Lähisynonymian tutkiminen on yksi haastavimmista nykykysymyksistä: mitä syvemmin kielen — oman tai vieraan — rikkaus tajutaan, sitä hienojakoisemmista merkitys- ja käyttöeroista ollaan kiinnostuneita.

LÄHTEET

- ARPPE, ANTTI 2002: The usage patterns and selectional preferences of synonyms in a morphologically rich language. – Annie Morin & Pascale Sébillot (toim.), *JADT 2002. 6th international conference on textual data statistical analysis 13.–15.3.2002*, vol. 1 s. 21–32. Rennes: INRIA.
- ARPPE, ANTTI – JÄRVIKIVI, JUHANI 2002: Verbal synonymy in practice. Combining corpus-based and psycholinguistic evidence. – <http://www.cogsci.uni-osnabrueck.de/qitl/QITL-Dateien/Abstracts/arppejaervikivi.rtf> 21.2.2003.
- BIBER, DOUGLAS – CONRAD, SUSAN – REPPEN, RANDI 1998: *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- CRUSE, D. A. 1986: *Lexical semantics*. Cambridge: Cambridge University Press.
- EDMONDS, PHILIP – HIRST, GRAEME 2002: Near-synonymy and lexical choice. – *Computational linguistics* 1:1 s. 1–40.
- HENRIKSEN, BIRGIT 1999: Three dimensions of vocabulary development. – *Studies in second language acquisition* 21 s. 303–317.
- INKPEN, DIANA ZAIU – HIRST, GRAEME 2002: Acquiring collocations for lexical choice between near-synonyms – SIGLEX Workshop on unsupervised lexical acquisition, 40th meeting of the Association for Computational Linguistics, Philadelphia, June 2002. <http://ftp.cs.toronto.edu/pub/gh/Inkpen+Hirst-2002.pdf> 27.6.2003.
- JANTUNEN, JARMO H. 2001: »Tärkeä seikka» ja »keskeinen kysymys»: Mitä korpuslingvistinen analyysi paljastaa lähisynonyymeista? – *Virittäjä* 105 s. 170–192.
- JONES, K. SPARK 1986: *Synonymy and semantic classification*. Information technology series. Edinburgh: Edinburgh University Press.
- LAHDENMÄKI, PEKKA 2002: sähköposti 8. toukokuuta 2002 korpuslingvistiikka-listalla. – Liittymisosoite: korpuslingvistiikka-lista-request@kotus.fi.

▷

- NEWTON, JONATHAN 1995: Task-based interaction and incidental vocabulary learning: a case study. – *Second language research* 1995 volume 11, number 2 s. 159–177.
- NIDA, EUGENE A. 1975: *Componential analysis of meaning: An introduction to semantic structures*. Paris: Mouton.
- NS = *Nykysuomen sanakirja*. Helsinki: WSOY 1951–1961.
- PARIBAKHT, T. SIMA – WESCHE, MARJORIE 1999: Reading and »incidental» L2 vocabulary acquisition: An introspective study of lexical inferencing. – *Studies in second language acquisition* 21. s. 195–224.
- PS = *Suomen kielen perussanakirja*. Helsinki: Kotimaisten kielten tutkimuskeskus 1990–1994.
- RAUKKO, JARNO 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* s. 87–105. *Current Issues in Linguistic Theory*, 152. Amsterdam: John Benjamins.
- 2002: Pitämisen polysemia: miten koehenkilöt hahmottavat *pitää*-verbin merkitystyyppensä. – *Virittäjä* 106 s. 354–374.
- STUBBS, MICHAEL 1993: British tradition in text analysis. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology: In honour of John Sinclair* s. 1–33. Amsterdam: John Benjamins.
- VANHATALO, ULLA 2002: »Naiset motkottaa aiheesta ja nalkuttaa syyttä»: Kyselytestit verbien semanttisten sisältöjen arvioinnissa. – *Virittäjä* 106 s. 330–353.
- WIERZBICKA, ANNA 1996: *Semantics: Primes and universals*. Oxford: Oxford University Press.

LIITE 1

MONIVALINTATESTI	TULOKSET		
Valitse annetuista vaihtoehtoista mielestäsi sopivampi.	keskeinen	tärkeä	Jantunen 2001
(1) Tämän seurauksena kristinuskon rooli muuttui vähemmän <u>tärkeäksi</u> , vaikka se edelleen toimi sekä normatiivisena rauhoittajana – – Tämän seurauksena kristinuskon rooli muuttui vähemmän <u>keskeiseksi</u> , vaikka se edelleen toimi sekä normatiivisena rauhoittajana – –	67 %	33 %	tärkeä
(2) Samastumisen ulottuvuus ei ylipäänsä ole yhtä <u>tärkeä</u> kuin Suomessa. Samastumisen ulottuvuus ei ylipäänsä ole yhtä <u>keskeinen</u> kuin Suomessa.	33 %	67 %	tärkeä
(3) Ajallisen jatkuvuuden (samuuden) ohella myös erottautuminen on hyvin <u>tärkeää</u> . Ajallisen jatkuvuuden (samuuden) ohella myös erottautuminen on hyvin <u>keskeistä</u> .	16 %	84 %	tärkeä
(4) Näiden kahden <u>keskeisen</u> jännitteen strukturoima feminististen tieto-oppien kolmijako on Sandra Hardingin käsittein – – Näiden kahden <u>tärkeän</u> jännitteen strukturoima feminististen tieto-oppien kolmijako on Sandra Hardingin käsittein – –	77 %	23 %	keskeinen
(5) Oppikirjat tarjoavat tähän monimutkaiseen tarinaan yhden <u>keskeisen</u> kertomuksen. Oppikirjat tarjoavat tähän monimutkaiseen tarinaan yhden <u>tärkeän</u> kertomuksen.	68 %	32 %	keskeinen
(6) Tutkimuksen <u>keskeinen</u> idea maallikoiden ja asiantuntijoiden riskin määritelmien eroista jätetään mainitsematta – – Tutkimuksen <u>tärkeä</u> idea maallikoiden ja asiantuntijoiden riskin määritelmien eroista jätetään mainitsematta – –	93 %	7 %	keskeinen
(7) Mielikuvat ovat <u>keskeisiä</u> , mutta eivät merkitse koko asiaa. Mielikuvat ovat <u>tärkeitä</u> , mutta eivät merkitse koko asiaa.	14 %	86 %	tärkeä
(8) Kuten myöhemmin käy ilmi, varsinaisen metaforan käytön ymmärtämiseksi Väyrysen kampanja on <u>tärkeä</u> osa laajempaa kontekstia. Kuten myöhemmin käy ilmi, varsinaisen metaforan käytön ymmärtämiseksi Väyrysen kampanja on <u>keskeinen</u> osa laajempaa kontekstia.	56 %	44 %	keskeinen

▷

<p>(9) Semioottisina olioina tekstejä yhdistää muuan tuiki <u>keskeinen</u> seikka — nimittäin se, että ne kaikki ovat representaatioita. Semioottisina olioina tekstejä yhdistää muuan tuiki <u>tärkeä</u> seikka — nimittäin se, että ne kaikki ovat representaatioita.</p>	18 %	82 %	tärkeä
<p>(10) -- eri puolilla Lappia havaitut yksittäiset metsävauriot kasvoivat <u>keskeiseksi</u> yhteiskunnalliseksi ilmiöksi nimenomaan siksi, että -- -- eri puolilla Lappia havaitut yksittäiset metsävauriot kasvoivat <u>tärkeäksi</u> yhteiskunnalliseksi ilmiöksi nimenomaan siksi, että --</p>	59 %	41 %	tärkeä
<p>(11) Jos joku näistä seikoista olisi kulttuurin kannalta yhtä <u>keskeinen</u> kuin meille sattuu olemaan homous, juoppous tai aggressiivisuus -- Jos joku näistä seikoista olisi kulttuurin kannalta yhtä <u>tärkeä</u> kuin meille sattuu olemaan homous, juoppous tai aggressiivisuus --</p>	53 %	47 %	tärkeä
<p>(12) Kirjoittajien <u>tärkeä</u> teesi on, että sosiaaliset ja pragmaattiset tekijät vaikuttavat luokittelujärjestelmiin -- Kirjoittajien <u>keskeinen</u> teesi on, että sosiaaliset ja pragmaattiset tekijät vaikuttavat luokittelujärjestelmiin --</p>	79 %	21 %	keskeinen
<p>(13) Myös Lehtosen romaanissa paloviinan salapoltto ja salakauppa olivat mitä <u>keskeisin</u> teema. Myös Lehtosen romaanissa paloviinan salapoltto ja salakauppa olivat mitä <u>tärkein</u> teema.</p>	84 %	16 %	keskeinen
<p>(14) Yksi puheaktiteorian <u>keskeinen</u> ongelma on epäsuoruus, jossa Searlen (1975, 60) mukaan puhuja esittää -- Yksi puheaktiteorian <u>tärkeä</u> ongelma on epäsuoruus, jossa Searlen (1975, 60) mukaan puhuja esittää --</p>	97 %	3 %	keskeinen
<p>(15) -- esimerkiksi sellainen, jossa informaation erittelyn asemasta <u>keskeiseksi</u> nousee merkityksen tulkinta. -- esimerkiksi sellainen, jossa informaation erittelyn asemasta <u>tärkeäksi</u> nousee merkityksen tulkinta.</p>	48 %	52 %	keskeinen

LIITE 2

Koehenkilöt vastasivat 'tärkeän' ja 'keskeisen' eroista (n=17), 'tärkeän' ja 'olennaisen' eroista (n=35) ja 'tärkeän' ja 'merkittävän' eroista (n=21).

tärkeä	keskeinen
henkilökohtaisuus nimenomainen, tietty juttu yleisemmin käytetty muut oleelliset asiat ihan mikä vain asia vaatii kiireellistä huomiota jota ilman ei tule toimeen tärkeä on tärkeä ei eroa keskeiseen tärkeä on tärkein jotain todella tärkeää	ajankohtaisuus maailmanlaajuisuus asian tai tekstin ydin tai pääidea (6) jokin tietty asia avoimempi kuin tärkeä ei niin tärkeä paikka tai juttu tärkeä osa asiaa monesta tärkeästä tärkein olennainen paljon puhuttu myös tärkeää ei eroa tärkeään pääoleellisin asia asiassa
esim. <i>kielioppi ja ääntäminen erityisen rakas päivä</i>	<i>puhekieli tarinan ratkaiseva henkilö</i>
tärkeä	olennainen
erittäin tärkeä asia asia, jota ilman on vaikea tulla toimeen (2) jokin merkittävä seikka asia, joka on merkittävä, muttei täysin välttämätön erittäin erikoinen asia jotain, mikä pitää tietää ei ole suurta eroa olennaiseen (2) jotain, mitä täytyy tehdä (2) tärkeä on tärkeää merkittävä asiaan kuuluva juttu merkittävämpi kuin olennainen asia voi käyttää yksin ilman että tarttee selittää tärkeää oman itsensä takia (2) jotain, mitä ei saa unohtaa voi jättää kertomatta, mutta asia tulee selväksi pakollinen tai välttämätön asia (2) laajempi käsite kuin olennainen (2) vahva merkitys, voidaan käyttää eri tilanteissa jotain todella merkityksellistä ei välttämättä välttämättömiä mitä vain tärkeää ikään kuin sääntö en selitä	enemmän itsestäänselvyys asia, joka tiedetään seikka, jota ilman ei käytännössä pärjää (2) huomioitava asia (2) jotain, mitä pitäisi tietää tilanteesta riippuen ei ole suurta eroa tärkeään (2) odotetusti jo hyödylliseksi todettua kuuluu kiinteästi asiaan, josta on kyse (5) tarpeellinen kokonaisuuden osana (4) asia erottuu muista asioista ei välttämättä ollenkaan tärkeä asia tärkeä jonkin tietyn asian takia (4) asiaa tarvitaan »selkeyttämään» koko kuvaa välttämättömät pääkohdat tai tekstin pääsanoma (3) suuren asian suurin osa suppeampi käsite kuin tärkeä perimmäinen syy jostain »tärkeästä» ei yhtä tärkeää kuin tärkeä keskeistä ja suositeltavaa, muttei pakollista en selitä

▷

- esim. *maailmanhistorian tietäminen on tärkeää ihmisen sydän perheenjäsenet tai ystävät tai joku ihminen* (3) *maailmanhistorian tietäminen ei ole olennaista asiaa, joita kysytään kokeissa ihmisen jalka jalkapallossa pallo on olennainen asia peliä*

tärkeä

yleisempi ilmaus (2)
 henkilökohtaisuus (4)
 normaali adjektiivi niin kuin muissakin kielissä
 ei suurta eroa merkittävään (2)
 asia lähes välttämätön
 se, mikä vaikuttaa elämäämme
 ei vastausta (2)

kaikkien tulisi olla tietoisia
 jotain, mikä pitäisi ottaa vakavasti

asia tai tehtävä, jolla on arvoa
 asia joka on vaan hyvin tärkeä
 sama kuin merkittävä, mutta pitää tehdä nopeasti
 tärkeä asia ihan erikseen

merkittävä

suuret asiat
 asia, joka vaikuttaa johonkin
 ei suurta eroa tärkeään (2)

asia, joka on myös tärkeä
 suuri yleinen merkitys tai vaikutus (3)
 joku tehnyt jotakin, mikä jää muistiin
 asiayhteyden kannalta oleellinen ja merkityksellinen
 ei vastausta (2)
 joku kuuluisuus, joka ei suuresti vaikuta elämäämme
 ei tarvitse ottaa niin vakavasti kuin tärkeää
 suuri, merkityksellinen asia jollekin (2)
 aika lähellä tärkeää

tärkeä merkitys tulevaisuuteen ja asioihin

- esim. *hektiset pikkujutut, jotka olisi hyvä muistaa Tarja Halonen jokin esine* (2) *koe on tärkeä kurssi-arvosanan kannalta jokin asia*
- perheen menetys Matti Nykänen luonnonmullistus henkilö juhlapäivä tapahtuma paikka ajanjakso*

EVALUATING THE SEMANTIC CONTENT OF NEAR-SYNONYMS: POPULATION TESTS VERSUS CORPUS LINGUISTICS

The article compares two methods for analysing the meanings of near-synonyms: one using population tests and the other based on corpus linguistics. The population tests used in the study were designed on the basis of a previous corpus-linguistic study (Jantunen 2001), which examined collocations of two near-synonym adjectives: *tärkeä* (~ ‘important’) and *keskeinen* (~ ‘central’). The writer here examines the extent to which the tested individuals agreed with each other and with the results of the corpus study. Multiple-choice tests were devised to correspond as closely as possible to the format of the corpus study. Data for the population test study was col-

lected in September 2002 from upper comprehensive school pupils with Finnish as their mother tongue (n=73; age 16–19 yrs; 38% female; 88% always resided in Helsinki region). The participants were given a four-page series of tests comprising one free-response test and 19 multiple-choice questions based on example sentences. The free-response question asked participants to outline the differences in meaning between the near-synonyms *keskeinen*, *tärkeä*, *olennainen* (~ ‘essential’) and *merkittävä* (~ ‘notable’).

In general, the corpus-linguistic method and the population test method produced fairly similar results for the contiguous collocations of *tärkeä* and *keskeinen*. However, the similarity of these results could, the writer notes, have been attributable to a variety of factors. Nevertheless, it was evident that the choice of near-synonyms was most probably affected by a broader context than simply two contiguous words. From the differences in the results it was apparent that the lexical choices of the participants were influenced by a context wider than the contiguous collocations, a consideration that had not been taken into account in the corpus study. In the analysis it was concluded that the study of adjectives should focus attention particularly on unusually strong nominal collocations. The writer also concludes that in studying semantic features of the lexicon and mutual relationships between words, it should be remembered that the collocations sought in corpus linguistics represent only one form of semantic coupling. Development of new methods is still needed in lexical research and especially in dictionary compilation, and the empirical population test method does offer a potentially useful tool for analysing lexical meanings and usage. ■

Kirjoittajan osoite (address):
Suomalais-ugrilainen laitos
PL 25 (Franzéninkatu 13)
00014 Helsingin yliopisto
 Sähköposti: *ulla.vanhatalo@helsinki.fi*