

EI PELKÄSTÄÄN MIELIKUVITUKSEN PUUTTEEN VUOKSI — KIELIAINEISTOJEN SYSTEMAATTINEN KÄYTTÖ KIELENTUTKIMUKSESSA

Krista Ojutkangas kirjoittaa Virittäjän numerossa 2/2008 osuvasti aineistojen käytön merkityksestä kielentutkimuksessa. Ojutkangas perustelee aineistojen käyttöä mielikuviituksen puutteella eli sillä, että tutkija ei pysty mielikuviituksen ja kieli-intuitionsa avulla kuvaamaan kieltä riittävän tarkasti (mas. 277). Lopputulemana hän esittää, että erilaiset lähestymistavat ja näkökulmat — aineistoon pohjautuvat ja introspektiopohjaiset — eivät sulje toisiaan pois vaan pikemminkin täydentävät toisiaan. Katsauksessaan Ojutkangas perustelee omaa aineiston käyttöönsä tutkimustyössä mielikuviituksen puutteella, joskin hän esittelee myös muita syitä aineistojen käytölle kielentutkimuksessa; katsauksessa esitellään erityisesti »kognitiivishenkisessä nykyfennistiikassa» (mp.) aineistoja hyödyntäneitä tutkimuksia. Tarkastelen tässä puheenvuorossani edelleen aineistojen käyttöä kielentutkimuksessa, mutta pyrin käsittelemään aineistojen roolia Ojutkankaan katsausta systemaattisemmin ja tuomaan esiin erityisesti metodologisia eroja erilaisten lähestymistapojen välillä. Vaihtoehtoisista aineistonkäyttötavoista keskityn erityisesti aineistovetoisuuteen ja sen menetelmiin. Tarkastelunäkökulma on tässä esityksessäni korpuslingvistiikan ja korpuslingvistin, joiden tutkimukset ovat erittäin vahvasti aineistosisonnaisia; aineistolla tarkoitan tässä yhteydessä nimenomaan laajoja sähköisiä teksti-

aineistoja. Sivuan myös kielentutkijan intuitiota (sekä kieli-intuitiota että laajemmin käsitystä tutkimukseen liittyvistä valinnoista) ja introspektiometodia sekä näiden ja aineistoihin perustuvien menetelmien eroja ja yhdistämisenäkökulmia nimenomaan aineistotutkimuksen näkökulmasta (aiheesta tarkemmin ks. Featherston 2007 ja Sampson 2007 sekä näihin kirjoitettuja vastineita *Theoretical linguistics* 2007 ja *Corpus linguistics and linguistics theory* 2007 -teemanumeroissa).¹

ERILAISIA LÄHESTYMISTAPOJA AINEISTOON

Aluksi on tarpeen luoda katsaus siihen, millaisen roolin tutkija voi antaa aineistolle tutkimuksessaan ja sen eri vaiheissa. Jos jätämme käsittelemättä täysin tutkijan omaan kieli-intuitionsa (rationaalinen tieto kielestä ja sen normeista ja käytöstä) ja siihen pohjautuvaan introspektiomenetelmään perustuvan tarkastelun², kieliaineiston merkitys voidaan hahmottaa tutkimuksessa kolmijakoiseksi. Aineistoja hyödyntävät tutkimukset voidaan jakaa Tummersin, Heylenin ja Geeraertsin (2005: 234–238; ks. myös Geeraerts 2006) mukaan aineistoesimerkein tuettuihin (*corpus-illustrated*) ja aineistopohjaisiin (*corpus-based*) tutkimuksiin. Elena Tognini-Bonelli (2001) jakaa aineistoon perustuvat analyysit hienojakoisemmin

¹ Kiitän Virittäjän anonyymeja arvioijia erittäin hyödyllisistä kommentteista.

² Kielentutkijan on luultavasti kuitenkin mahdotonta tehdä tutkimusta **täysin** intuitiivisesti, sillä häntä, kuten kaikkia kielenkäyttäjiä, ympäröi jatkuvasti monenlaisten tekstien maailma, joka ainakin implisiittisesti toimii tutkijan »aineistona». Ilmausten hyväksyttävyyttä ja merkitystä pohditaan introspektiolingvistiikassa usein juuri kuvittelemalla ilmaukselle jokin käyttökonteksti (Wasow ja Arnold 2005: 1484–1485).



edelleen aineistopohjaisiin (*corpus-based*) ja aineistovetoisiin (*corpus-driven*).³ Kieliaineistoja käyttävät tutkimukset voidaan siis luokitella edelliset jaottelut yhdistelemällä aineistoesimerkein tuettuihin, aineistopohjaisiin ja aineistovetoisiin analyyseihin — jaottelu on osittain nähtävissä myös Ojutkankaan (2008) katsauksessa, joskaan ei näin systemaattisena. Tarkennan seuraavaksi näitä lähestymistapoja nimenomaan metodologisesta näkökulmasta ja tuon esiin lisäesimerkkejä tai poikkeavia näkemyksiä Ojutkankaan puheenvuoroon korpustutkimuksen näkökulmasta.

Edellä mainituista kolmesta lähestymistavasta **aineistoesimerkein tuettu** tutkimus antaa vähiten painoarvoa aineistolle. Aineiston roolina on toimia lähinnä materiaalipankkina, josta tutkija etsii teoriaan sopivia esimerkkejä kuvattavasta ilmiöstä — yleensä sellaisia esiintymiä, jotka tulevat tai täydentävät tutkijan kieli-intuitiota. Tällöin aineiston tehtävä on yleensä vahvistaa teoriaa (Wasow ja Arnold 2005: 1484). Korpus saattaa olla yhtenä esimerkivarantona muiden lähteiden rinnalla: Robin Turner (2006) esimerkiksi käsittelee turkinkielisten naista tarkoittavien *kiz* ja *kadin* -sanojen käyttöä ja semanttisia ominaisuuksia aineistonaan natiivien kielenuhujien tuottamat lausumat, tv-uutiset ja Ozturk Corpus. Aineistoja pidetään samanarvoisina, eikä niiden avulla pyritä tekemään kvantitatiivisia yleistyksiä sanojen käyttöaloista tai merkityksestä. Sanoja käsitellään merkityspiirreanalyysin avulla ja eri piirteitä kuvaavia esimerkkejä on haettu aineistoista, myös korpuksista.

Jos aineistosta on havaittavissa tapauksia, joita ei teorian tai introspektion perus-

teella tiedetty olevan olemassa, aineistoesimerkkien avulla voidaan muuttaa teoriaa tai falsifioida se (Tummers ym. 2005: 234). W. Detmar Meurers (2005: 1621) pitää tärkeänä juuri teorian korjaamista aineiston avulla, sillä todelliseen kielenkäyttöön perustavan aineiston avulla voidaan päästä kiinni sellaiseen tietoon kielessä esiintyvistä monipuolisesta variaatiosta, jota ei yksistään introspektion käytöllä saavuteta. Aineistoesimerkkejä sisältävässä tutkimuksessa kieliaineiston käyttöä voidaan perustella sillä, että se tuo introspektion ja kieli-intuition rinnalle tietoa todellisesta kielenkäytöstä ja sen monimuotoisuudesta esimerkkien avulla eikä tutkija ole tutkimuksessaan pelkästään niin sanottujen laboratoriolauseiden varassa (ks. aiheesta Huumo 2007). Lähestymistavassa aineistoesiintymiä käsitellään edelleen kieli-intuitioon pohjautuen eikä yleensä oteta kantaa siihen, ovatko havainnot kielenkäytössä marginaalisia tai frekventtejä; aineistosta saatavan kvantitatiivisen tiedon ja tilastollisten menetelmien avulla voitaisiin kuitenkin saada informaatiota myös ilmiön käyttötaajuudesta. (Tummers ym. 2005: 234–235.)

Aineistopohjaisesta tutkimuksesta on tullut laajojen tekstiaineistojen (korpusten) kokoamisen myötä yhä suosittumpaa. Lähestymistapa eroaa edellisestä selvästi siten, että todelliseen kielenkäyttöön perustuva materiaali saa tutkimuksessa selvästi suuremman roolin: aineisto ei ole vain esimerkivaranto vaan tutkimus muodostuu laajamittaisesta aineiston analysoimisesta. Aineistopohjaisessa tutkimuksessa lähtökohtana ovat yleensä tutkijan tekemä havainto kielestä, hänen intuitionsa tai aiempiin tutkimuksiin perustuva teoria ja hypoteesi. Näiden poh-

³ Tarkasti ottaen aineistopohjainen/-vetoinen ja korpuspohjainen/-vetoinen eivät ole käsitteinä samaa merkittäviä, sillä aineisto voi olla periaatteessa mikä tahansa tutkimuksessa käytettävä kielimateriaali, mutta korpus on kielentutkimusta varten systemaattisesti koottu sähköinen tekstimateriaali. Koska käsiteltävät teemat koskevat kuitenkin yleisesti tutkimusaineistojen käyttöä eivätkä vain korpuksia, käytän katsauksessani aineisto-termiä korpuksen sijasta.

jalta syntyvää tutkimuskysymystä tarkastellaan aineiston avulla, ja aineiston roolina on pääasiassa teorian ja aiempien havaintojen selittäminen tai testaaminen. Aineistopohjaisessa tutkimuksessa kieltä analysoidaan olemassa olevien kategorioiden avulla, ja aineiston tärkein merkitys on sen voimassa osoittaa aiempien teorioiden epäkohdat ja muutostarpeet sekä mahdollisuudessa tarjota kvantitatiivista tietoa. (Tognini-Bonelli 2001: 65–66.) Biber ja Jones (2005: 152) nostavat aineistopohjaisen tutkimuksen hyödyiksi lisäksi erityisesti tutkimuksen toistettavuuden ja havaintojen yleistettävyyden, joista viimeksi mainittu on mahdollista, jos tutkimus perustuu tutkimuskysymyksen kannalta tarpeeksi isoon ja edustavaan aineistoon. Lähestymistavalle on tyypillistä leksikon ja kieliopin erillinen tarkastelu, ja analyyseissa on perinteisesti painottunut kieliopin tutkimus (kuten aineistoesimerkein tutussakin tutkimuksessa), vaikka laajat kieliaineistot loisivat tilaisuuden tarkastella leksikkoa ja kielioppia toisiinsa vahvasti kytkeytyvinä ilmiöinä (Tognini-Bonelli 2001: 81; 2002: 74–75). Lähestymistapa on yleistynyt esimerkiksi kognitiivisessa kielitieteessä, jossa kieliaineistojen analyysistä on tullut yhä merkittävämpi työkalu tutkijalle (ks. Ojutkangas 2008), ja erityisesti niin sanotussa kognitiivisessa korpuslingvistiikassa (ks. esim. Grondelaers, Geeraerts ja Speelman 2007) kvalitatiivisen analyysin rinnalle ovat nousseet kvantitatiiviset havainnot tutkittavasta ilmiöstä.

Edellä käsiteltyihin metodologisiin lähestymistapoihin verrattuna **aineistovetoisen** metodi antaa tutkimusaineistolle entistä suuremman merkityksen, ja sen avulla saatava näkökulma kieleen on myös osittain erilainen kuin aineistoa vähemmän painottavilla lähestymistavoilla. Myös Ojutkangas (2008: 281) viittaa tähän menetelmään ja antaa tutkimusesimerkkejä, jotka edustaisivat tätä lähestymistapaa. Mainitut Visapään

(2005) ja Päiviön (2007) tutkimukset eivät kuitenkaan tarkkaan ottaen edusta aineistovetoista tutkimusta, sillä molemmissa tutkimuskysymykset ja tarkasteltavat kieli-ilmiöt ovat valikoituneet ensisijaisesti aiemman tutkimuksen sekä tutkijan mielenkiinnon ja havaintojen perusteella. Ne lukeutuvatkin aineistopohjaisiksi tutkimuksiksi huolimatta siitä, että paikoin tutkimusaineisto on tuonut esiin mielenkiintoisia ja tutkimuksellisesti merkittäviäkin havaintoja muun muassa uusien hypoteesien pohjaksi. Aineistovetoisen metodin perusajatuksena on tarkastella tutkimusmateriaalia kokonaisuutena ja antaa aineistolle suuri painoarvo niin tutkimushypoteesin luomisessa, teorian kehittämisessä kuin kielen ilmiöiden kategorisoinnissakin. Ilmiön toistuvuudella ja frekvenssillä (eli kvantiteetilla) on tässä lähestymistavassa merkittävä rooli, ja usein niitä pidetään kielellisten kategorioiden perustana. Myös se, että jotain kielellistä ilmiötä ei aineistosta löydy, on teorian kannalta mahdollisesti merkittävää. Aineistovetoisen tutkimuksen tavoitteena ei ole pelkästään todistaa aiempia havaintoja ja teorioita oikeiksi, uudelleenmuokkausta vaativiksi tai vääriksi, vaan se pyrkii nostamaan esiin myös sellaisia potentiaalisesti tärkeitä kielen ilmiöitä, joita ei ole mahdollista havaita ilman laajan aineiston todistusvoimaa ja kvantitatiivista, tilastollista tietoa. (Tognini-Bonelli 2001: 84; 2002: 75.)

Aineistovetoisessa lähestymistavassa hypoteesin ja tutkimuskysymyksen luominen voidaan jättää jopa aineiston motivoimaksi, vaikka laaja tutkimuskehys ja -kysymys ovat luonnollisesti tutkijan etukäteen valitsemia (usein intuition pohjalta, kuten Ojutkangaskin huomauttaa). Näin laajamittainen aineiston varaan heittäytyminen ei ole kuitenkaan kovin tavallista — ensiksi luultavasti siksi, että tutkimuskohteeksi voi nousta hyvinkin toisenlaisia asioita kuin tutkija on aiemmin tarkoittanut, ja toiseksi

▷

siksi, että asetelma vaatii aineiston tutkimiseen tarkoitettujen korpus- ja tilastollisten menetelmien hallintaa. Usein tilanne lienee myös sellainen, että tutkijan mielenkiinto on herännyt jotakin kielen ilmiötä kohtaan jo ennen kuin saatavilla on ollut sopiva kieliaineisto, jolloin aineisto tulee osaksi tutkimusta vasta tutkimuskysymyksen syntymisen jälkeen. Aineistovetoisen tutkimuksen määrä alkaa kuitenkin lisääntyä, ja Suomesakin tehtävä lingvistinen tutkimus on sitä jo hieman hyödyntänyt, kuten seuraavista esimerkeistä käy ilmi.

SYSTEMAATTINEN AINEISTOVETOISUUS: AVAINSANA-ANALYYSI

Jos tutkimuskysymystä halutaan lähteä muovaamaan vahvasti aineiston ohjaamana heti alusta alkaen, tutkija voi käyttää apukeinona muun muassa aineistosta tehtäviä frekvenssilistoja. Niitä hyödyntää myös niin sanottu avainsana-analyysi. Siinä käytetään hyväksi vähintään kahden tekstiaineistoa — tutkimusaineistoa ja vertailuaineistoa — ja analyysiin tarvittavaa tilastopohjaista tietokoneohjelmaa. Tekstin (tai aineiston) avainsanaisuuden — eli avainsanojen (*keywords*) — avulla paljastuu pohjimmiltaan se, mistä tekstissä on kysymys. Jos esimerkiksi suomalaisista kielitieteellisistä tutkimusteksteistä muodostettua aineistoa verrattaisiin yleensä laaja-alaisesti tieteellisiä tekstejä sisältävään aineistoon, saattaisivat sen avainsanoiksi nousta esimerkiksi *objekti*, *koheesio*, *etuvokaali* tai *pääte*, koska niitä luultavasti käsitellään enemmän lingvistisissä teksteissä kuin yleensä tieteellisessä tutkimuksessa. Tilastollisesta näkökulmasta katsottuna

avainsanoja ovat sellaiset sanat, jotka esiintyvät tutkimusaineistossa poikkeuksellisen taajaan vertailuaineistoon verrattuna (ks. mm. Scott ja Tribble 2006: 58–59).

Analyysi voi myös paljastaa kulttuurisia avainsanoja, kuten käy ilmi Hannu Kempasen (2008) suomennosten ideologisuutta tarkastelevasta tutkimuksesta: tutkimuksen aluksi tehtiin avainsana-analyysi, jossa suomennettujen historiatekstien avainsanoiksi, ja siten suomennokset ei-käännöksistä erottaviksi sanoiksi, nousivat monet kulttuuriset ja yhteiskunnalliset sanat, kuten *hallitus*, *yhteistyö*, *sopimus* ja *ystävyyys* (mts. 77–78, 94). Tämän tutkimuksen avainsana-analyysissä esiin tulleet sanat ja niitä laajemminkin ilmaukset valikoituivat myöhemmin tutkimuksessa tehdyn aktanttianalyysin perustaksi, eli aineistovetoisuus ja avainsana-analyysi ohjasivat työn kulkua alusta asti ja toimivat sen punaisena lankana ja pontimena käännöstekstien ideologisuutta tarkasteltaessa.

Avainsana-analyysia voidaan käyttää myös tietyn kielivariantin ominaispiirteiden kartoituksen ensiaskeleena. Omassa tutkimuksessani olen käyttänyt aineistovetoista lähestymistapaa suomea vieraana kielenä opiskelevien tuottamien tekstien analysoimisessa. Avainsana-analyysi nostaa esiin muun muassa oppijankielelle tyypillisiä sanoja (tarkemmin sananmuotoja), siis sanoja, jotka esiintyvät oppijansuomessa huomattavasti useammin kuin natiivisuomessa.⁴ Tällaisia ovat muun muassa *koska*, *paljon* ja *kello*. Koska esimerkiksi *kello* on erityisen frekventti oppijansuomessa, on luonnollisesti tarpeen tarkastella korkean esiintymisfrekvenssin syytä. Syvempi aineiston analyysi paljastaa, että sanaa käytetään epätyypillisissä kollokaatorakenteissa

⁴ Oppijansuomen aineistona on tutkimuksessani ollut Kansainvälinen oppijansuomen korpus (ks. Jantunen 2007; <http://www.oulu.fi/hutk/sutvi/oppijankieli/index.html>) ja vertailuaineistona Käännössuomen korpuksen (Mauranen 2000) alkuperäissuomen osakorpus.

ja että sen paradigmaattinen morfologinen preferenssi eli taivutusmuotojen variaatio (ks. Jantunen 2004: 27–29) on natiivisuomesta poikkeava. Tällainen leksikaalis-kieliopillisen rakenteen tarkastelu paljastaa, että suomenoppija näyttäisi käyttävän *kello*-sanaa välttääkseen ajanilmauksissa esiintyvien lukusanojen taivuttamisen: *Minä härrän kello kymmenen vrt. Minä herään kymmeneltä*. Oppijansuomesta näyttää puuttuvan myös natiivisuomessa tavallinen leksikaalis-semanttinen rakenne *KELLO <KATSOA, VILKAISTA, TUIJOTTAA>*, jossa siis *kello*-sanankontekstista löytyy 'katsomista' merkitseviä verbejä.

Tällainen kielenoppijan leksikkoa ja kontekstuaalisia valintoja koskeva analyysi on mahdollista juuri aineistovetoisen lähestymistavan avulla; intuitiivisesti ei kielenopettaja, eikä luultavasti tutkijakaan, huomaa natiivisuomesta poikkeavaa käyttöä, koska rakenne on täysin kieliopinmukainen. (Analyysistä tarkemmin Jantunen 2008; oppijankielen aineistovetoisesta analyysistä ks. myös Jantunen 2007.) Avainsana-analyysi on erityisesti leksikkoa ja leksikaalis-kieliopillisia suhteita tarkastelevaan tutkimukseen sopiva työkalu: soveltuvuus puhtaasti kieliopin tutkimukseen lienee vielä syvemmin tarkastelematta (ks. kuitenkin Scott ja Tribble 2006).

Kuten Ojutkangaskin (2008: 281) havainnollistaa katsauksessaan, aineistoa hyväksi käyttävä tutkimus (aineistoesimerkein tuettu tai aineistopohjainen tutkimus) voi aineiston syvällisen tarkastelun myötä muuttua osittain aineiston johtamaksi »aineistovetoiseksi» tutkimukseksi: aineisto itsessään nostaa esiin mielenkiintoisia ilmiöitä, jotka kaipaavat selityksiä esimerkiksi mainituissa Visapään (2005) ja Päiviön (2007) tutkimuksissa. Erilaiset lähestymistavat aineistoon eivät siis useinkaan toteudu »puhtaina», vaan yksi tutkimus voi hyödyntää erilaisia lähestymistapoja — jopa

ilman että tutkija on tehnyt aineiston käytön suhteen periaatteellisia etukäteisvalintoja.

Käsitteinä aineistopohjainen ja -vetoisen tutkimus eivät ole vakiintuneet selväräjaisiksi, mistä kertovat muun muassa seuraavat esimerkit. Scott ja Tribble (2006) puhuvat aineistopohjaisesta (*corpus-based*) tutkimuksesta, vaikka heidän esittelemänsä avainsanamenetelmä on nimenomaan aineistovetoiseen tutkimukseen soveltuva. Ute Römerin (2004) analyysi englannin modaalisisista apuverbeistä (mm. *may, can*) on vuorostaan edellä esitettyä jaottelua noudattaen pohjimmiltaan aineistopohjainen tutkimus, sillä lähtökohtana ovat aiemmat kieliopit ja tutkijan omat havainnot, joiden perusteella tutkimuskohteet on valittu aineistoanalyysia varten — tutkimus on otsikoitu kuitenkin aineistovetoiseksi tutkimukseksi. Tognini-Bonellin (2001) esittelemä käsite aineisto- tai korpusvetoisuus (ja korpusvetoinen lingvistiikka, *corpus-driven linguistics*, mts. 177–186) onkin suhteellisen tuore ja vakiintumaton, ja aineistopohjainen lähestymistapa voidaan jossain mielessä ajatella aineistovetoisen yläkäsitteeksi.

Siinä missä aineistoesimerkein tuettua tutkimusta pidetään ongelmallisena frekvenssi- ja tilastotiedon puutteen vuoksi, aineistoa laajemmin hyödyntävää (korpus-)tutkimusta pidetään usein liikaa ilmiöiden esiintymistiheyteen ja tilastolliseen merkitsevyyteen keskittyvänä. Esimerkiksi Hoffmann (2004) kritisoi pelkästään frekvenssipohjaista, taajakäyttöisiin ilmiöihin keskittyvää aineistotutkimusta ja tähdentää, että myös vähäfrekvenssiset, mutta aineistoon perustuvassa tutkimuksessa helposti ulkopuolelle jäävät ilmiöt ovat kielen ja tutkimuksen näkökulmasta oleellisia (sanastontutkimuksen osalta ks. myös Vanhatalo 2003: 362). Toisaalta Hoffmann (2004: 188) huomauttaa, että kvantitatiivistakin tietoa tarvitaan: esimerkiksi kieliopillistumisen analysoimisessa ilmiön yleisyyteen liittyvä

tieto on oleellista, sillä leksikaalisen aineksen kieliopillistuminen ja sen frekvenssin muuttuminen ovat usein kytköksissä toisiinsa. Kielentutkimuksessa tarvitaankin yhtä lailla sekä kvalitatiivista että kvantitatiivista analyysia. Hieman yksinkertaistaen Penke ja Rosenbach (2004: 486–487) korostavat, että kvalitatiivisen tarkastelun tarkoituksena on osoittaa, onko tietty ilmiö mahdollinen jossakin kontekstissa tai esiintyykö se kielessä. Kvantitatiivisen analyysin tehtävä on heidän mukaansa puolestaan kvantifioida ilmauksen yleisyyttä. Lisäksi on todettava, että kvalitatiivinen analyysi saavuttaa paremmin vähäfrekvenssiset ilmaukset ja pysyy erityisesti selittämään kielenaineksen käyttötaajuuden taustalla olevia syitä. Biber, Conrad ja Reppen (1998: 5) huomauttavat, että ilmausten frekvenssit eivät saa jäädä

yksinomaiseksi tiedoksi aineistotutkimuksessa, vaan määrällistä tietoa on selitettävä laadullisella. Empiirisessä tutkimuksessa materiaalin edustavuus ja materiaalin edustavuuden rajoitusten tunteminen on yksi tärkeimmistä periaatteista (ks. mm. Biber 1993) sekä kvantitatiivisen että kvalitatiivisen tiedon saatavuuden ja luotettavuuden näkökulmasta. Hoffmann (2004: 197–198) painottaa lisäksi, että aineiston antama tieto ilmiön frekvenssistä voi kovastikin poiketa aineiston ulkopuolisesta todellisuudesta muun muassa siinä, että korpuukset sisältävät toistaiseksi huomattavan vähän puhutun kielen aineistoa kirjoitettuun aineistoon verrattuna. Edellisissä kappaleissa kuvatut luonnehdinnat kolmesta lähestymistavasta on koottu pääpiirteittäin ja yksinkertaistusti taulukkoon 1.

Taulukko 1. Aineistoesimerkein tuettu, aineistopohjainen ja aineistovetoinen tutkimus.

	aineistoesimerkein tuettu tutkimus	aineistopohjainen tutkimus	aineistovetoinen tutkimus
aineiston rooli	esimerkkireservi, aineiston käyttö epäsystemaattista	aineiston avulla tutkitaan aiemman tutkimuksen tai havaintojen avulla tehtyä hypoteesia tai kiinnostavaksi osoittautunutta ilmiötä	aineisto ohjaa tutkimusta alusta alkaen systemaattisesti ja toimii hypoteesien ja teorian kehittämisen perustana
intuitio / introspektio	tutkimus perustuu voimakkaasti intuitioon ja introspektioon	ohjaa tutkimusta ja on merkittävä aineiston valinnassa	ohjaa aineistosta nousevien tulosten tulkintaa
kvalitatiivisuus / kvantitatiivisuus	kvalitatiivista	pohjimmiltaan kvalitatiivista, kvantitatiivisilla havainnoilla tuetaan kvalitatiivisia havaintoja	pohjimmiltaan kvantitatiivista, kvalitatiivisilla havainnoilla selitetään kvantitatiivisia havaintoja
kielioppi / leksikko	kielioppi tutkimuskohteena merkittävä	kielioppi ja leksikko erillisiä tutkimuskohteita, kielioppipainotteista	yhdistää leksikon ja kieliopin tarkastelun, leksikaalis-kieliopilliset yksiköt

NOJATUOLISTA OHJELMISTOJEN ÄÄREEN JA TOISINPÄIN

Altenberg ja Granger (2002: 6) väittävät, että kielitieteessä empiirinen, todellisen vuorovaikutuksen tutkimus on vallannut alaa natiivin kielenpuhujan kieli-intuition käytöltä pääasiallisena tiedonlähteenä. Luonnollisen diskurssin tutkiminen ja empiirisen aineiston, erityisesti korpusaineistojen käyttö on yhä suosituempaa: Sampsonin (2005: 16, 33) havaintojen mukaan korpus-tutkimusten määrä on lisääntynyt voimakkaasti 1990-luvulla, joskin kasvu on hieman tasaantunut. Aineistoperustainen tutkimus on saanut enemmän sijaa myös sellaisilla kielentutkimuksen aloilla, joilla introspektiivinen tutkimus on ollut pääasiallinen menetelmä: esimerkiksi jo edellä mainitussa kognitiivisessa lingvistiikassa on nähtävissä paineita laaja-alaisemmalle kieliaineistojen käytölle (ks. Huumo 2007). Kädenvääntö siitä, kumpi lähestymistapa — introspektio-metodi vai aineistometodi — on ylipäättään kielentutkimuksessa parempi, on mielestäni jokseenkin turhaa (ks. myös Ojutkangas 2008): molemmat lähestymistavat ovat tarpeellisia ja ne täydentävät toisiaan.

Todistelu lähestymistapojen paremmuudesta tiedonkeruun kannalta jatkuu edelleen kuitenkin vilkkaana. Tästä on osoituksena esimerkiksi Huumon (mt.) raporttoima vastakkainasettelu kognitiivisen lingvistiikan piirissä, ja keskustelua käydään myös korpuslingvistiikassa. Muun muassa John Sinclairin (2004: 9–10) mielestä kielentutkimuksessa on tähän asti luotettu liikaa spekulatioon (introspektioon), ja hän painottaa isojen tutkimusaineistojen laaja-alaisempaa hyödyntämistä. Toisaalta taas esimerkiksi Cardey ja Greenfield (2002: 231, 246) kärjistävät, että korpus-tutkimuksessakin nimenomaan tutkijan intuitio ja kyky tarkastella aineistoa analyttisesti ovat ratkaisevassa asemassa

tietokoneen tarjotessa lähinnä apukeinoja siihen, mitä tutkija tarvitsee analyysissaan tai mitä hän jo kieli-intuitionsa perusteella tietää. Nämä ensisilmäyksellä vastakkaisilta tuntuvat kannanotot mahtuvat kuitenkin hyvin yhden tieteenalan piiriin, eikä niitä pitäisi nähdä polaarisisina vaan pikemminkin toisiaan täydentävinä korpuslingvistiikasakaan.

Aineistotutkimusta on turha ja väärin linjata intuitiottomaksi. Tutkijan intuitio ja kokemus on läsnä silloin, kun tehdään ratkaisuja siitä, mitä tutkimusaineistoksi (korpukseksi) valitaan ja mitä aineisto tulee sisältämään, miten tutkimus tehdään ja miten tuloksia tulkitaan (Stubbs 1995: 48; 2002: 9; ks. myös Ojutkangas 2008: 277). Korpusten kokoaminen on viime kädessä kiinni kokoamisesta vastaavien tekemistä ennakkopäätöksistä; vastaavasti yksittäinen tutkija tekee päätöksensä siitä, mitä hän laajasta aineistosta käyttää oman tutkimuksensa materiaalina. Edelleen esimerkiksi korpus-tutkijan on tehtävä päätöksiä siitä, millaista tietoa hän aineistosta etsii (ellei lähestymistapa ole selvästi aineistovetoinen).

Itse aineiston käsittelyssäkin on intuitiolla paikkansa. Borsley (2002: 240) korostaa kriittisessä korpus-tutkimusta koskevassa kirja-arviossaan, että esimerkiksi epätyypillisten aineistoesiintymien havaitsemiseen tarvitaan nimenomaan tutkijan kielitietoa; kannanottoon on korpuslingvistinkin helppo yhtyä, joskin esiintymäfrekvenssit ja tilastolliset menetelmät voivat tässä auttaa. Myöskään sanaleikkeihin, metaforiin ja kekeilevaan kielenkäyttöön, neologismeihin ja kielenvastaisiin ilmauksiin ei automaattinen analyysi vielä kovin hyvin pääse käsiksi. Aineiston analyysissäkin on jatkuvasti turvaututtava kieli-intuitionsa, vaikka aineistoa käsiteltäisiinkin tilastollisin menetelmin. Esimerkiksi oman havaintoni mukaan leksikaalisten yksiköiden kotekstuaalisia

▷

valintoja koskevassa analyysissäni (Jantunen 2004) tutkijan intuitiivisten päätelmien määrä lisääntyy sitä mukaa kun tutkimuksessa siirrytään leksikaalisia (kollokaatioita) ja kieliopillisia valintoja abstraktimmille kontekstuaalisille (semanttisille) tasoille; kieli-intuitiota korpusanalyysiakaan ei pysty tekemään (ks. mts. 24, 142, 221). Itse tulosten tulkinta ja niiden merkitys kielen ja sen tutkimuksen näkökulmasta on luonnollisesti analysoijan intuition varassa. Edustavinkaan aineisto ja paraskaan analyysiohjelma ei voi tehdä tulkintoja ja päätelmiä tulosten merkityksestä ja arvosta, siihen tarvitaan intuitiivista näkemystä tieteenalasta (aiheesta ks. myös Arppe 2008: 1–7).

Aineiston arvo suhteessa introspektion käyttöön on pitkälti siinä, että aineistosta pystyy (aineiston rajoitukset huomioon ottaen) havaitsemaan, miten kieltä käytetään. Kieli-intuitioon perustuva introspektio puolestaan kertoo kielenkäytön mahdollisuuksista ja normeista ja siitä, mitä kielestä tiedetään ja millaiseksi kieli hahmotetaan, mutta se on suhteellisen huono antamaan tietoa siitä, miten kieltä todellisuudessa käytetään. Esimerkiksi kielenpuhujan kuvaus ja oletus omasta kielenkäytöstä on usein ristiriidassa todellisen kielenkäytön kanssa. (Sinclair 1991: 4; Francis 1993: 139; Tsui 2004: 39.) Natiivien kielenpuhujien kieli-intuitioissa on lisäksi runsaasti variaatiota muun muassa ilmausten hyväksyttävyyden ja merkityksen tulkittamisen suhteen, ja etenkin rajatapausien kohdalla kielenpuhujat kohtaavat vaikeuksia tulkittaessaan ilmauksia intuition pohjalta (Wasow ja Arnold 2005: 1483).

Se, miten koulutettujen lingvistien tulkinnat samasta ilmiöstä ovat erilaisia (tai samanlaisia), olisi tärkeä tarkastelunaihe nimenomaan introspektiolingvistiikan ja

tutkimustulosten toistettavuuden näkökulmasta. Griesin (2002) havainnot kielentutkijoiden käsityksistä englannin genetiivirakenteiden käytöstä viittaavat alustavasti siihen, että kokeneetkin kielentutkijat päätyvät ratkaisuihinsa toisistaan poikkeaviin tulkintoihin. Lisäksi aineistojen valossa on usein helposti huomattavissa, että esimerkiksi leksikografit eivät pysty kieli-intuitionsa perusteella kuvaamaan merkityksiä, niiden yleisyyksiä ja semanttisia rakenteita tyhjentävästi (Stubbs 2002: 9).⁵ Haasteellisia introspektiomethodille ovat muun muassa lekseemien monitasoiset kontekstuaaliset rakenteet, rakenteiden esiintymisfrekvenssit ja vaikkapa yhden lekseemin eri taivutusmuotojen käyttöympäristöön kohdistuvat valinnat, joskin toisaalta Moon (2007: 166) huomauttaa, että puristinen korpusten käyttö esimerkiksi leksikografian apuna ei riitä, vaan tutkijoiden on käytettävä myös introspektiota, sillä laajoistakaan aineistoista ei saa irti kaikkia merkityksiä ja käyttötapoja, jotka voidaan kuitenkin kuvata kielitiedon avulla. Charles Fillmore (1992: 35) toteaa, ettei luultavasti ole olemassa niin laajaa tekstiaineistoa (korpusta), että se sisältäisi informaatiota kaikesta siitä leksikaalisesta ja kieliopillisesta tiedosta, josta hän nojatuolilingvistinä (Fillmoren oma luonnehdinta, mp.) on kiinnostunut. Kieliaineistot ovat väistämättä rajallisia, mutta iso ja hyvin suunniteltu tekstiaineisto on varmasti kokonaisuudessaan luotettavampi ja edustavampi kuin yhden tutkijan kielitieto.

Intuition ja introspektion käytön riittämättömyys ei korostu pelkästään kielen käytön kuvauksessa, vaan sen rajallisuus — kuten luonnollisesti myös aineistojen rajallisuus — on nähtävissä myös kielen **opetuksessa**: pelkästään kieli-intuitionsa

⁵ Arppen (2008: 163–169, 217–219, 253) tulosten mukaan kielenkäytön ja sanakirjojen esimerkkien vastaavuudessa on suuresti eroja: sana-artikkelit sisältävät mm. harvinaisiksi tai epätyypilliseksi luokiteltavia esimerkkejä tai niistä puuttuu kokonaan joitakin tavallisia käyttöyhteyksiä.

varassa kielenopettajan on hyvin usein vaikea selittää monia kielen ilmiöitä. Muun muassa synonyymisten ilmausten käytön rajat samoin kuin niiden keskinäinen suhdekin sekä lekseemien kollokationaaliset suhteet ovat ilmiöitä, joita koskeviin kysymyksiin kokeneetkaan opettajat eivät kovin helposti pysty kielitietonsa perusteella antamaan vastauksia (Tsui 2004: 43–44). Mauranen (2004: 90, 94) väittää puolestaan, että se mitä kielenopetuksessa opetetaan, ei edes ole välttämättä sen mukaista, mikä on kielessä taajaan esiintyvää tai mitä on kielessä olemassa. Tämä koskee erityisesti puhekielen opetusta. Tällaisiin ongelmiin voivat kieliaineistot tarjota apua introspektion rinnalle.

Introspektiotutkimuksen ja aineistotutkimuksen vuoropuhelun merkityksen voisi yrittää tiivistää kahden eri lähestymistavan omaavan kielentutkijan, Charles Fillmoren ja Geoffrey Leechin, ajatuksin: Fillmore (1992: 35, 45) toteaa, että olihan hän introspektiomenetelmän rinnalla testannut kuinka pientä kieliaineistoa tahansa, aina on ollut löydettävissä sellaista tietoa, jota hän ei olisi saavuttanut muuten kuin aineistoon syventymällä (ks. myös Ojutkankaan huomautus 2008: 282). Leech (1991: 74) taas painottaa voimakkaasti, ettei aineistoon (korpukseen) perustuvassa tutkimuksessa pidä yrittää hylätä intuitiota: »corpus use is – a question of corpus *plus* intuition, rather than corpus *or* intuition» (korostus GL). Kun tarkoituksena on kuvata kieltä laaja-alaisesti, tarvitaan siis väistämättä välillä pistäytymistä toisaalta nojatuolissa ja toisaalta aineiston ja ohjelmiston äärellä.

**LOPUKSI:
ETUA ERI MENETELMIEN
YHDISTÄMISESTÄ**

Melko hedelmättömäksi jäävän joko–tai vastakkainasettelun sijaan tärkeämpää on

tarkastella erilaisten metodologisten lähestymistapojen etuja ja rajoituksia objektiivisesti Huumon (2007; ks. myös siinä mainittu lähteet) ja Ojutkankaan (2008) tavoin. Tällöin ollaan kuitenkin vasta puolitiessä: erilaisten metodologisten ratkaisujen yhdistäminen, sen avulla saatujen tulosten analyysi sekä yhdistämisen etujen ja ongelmien arviointi ovat askel pitemmälle. Biber, Conrad ja Reppen (Biber ym. 1998: 9–10) sekä Stubbs (2002: 10) painottavat, että aineistojen käyttö pitäisikin nähdä muuta menetelmiä täydentävänä metodina, eikä korpustutkimuksessakaan ainoana oikeana lähtökohtana; syiksi mainitaan aineistopohjaisen tutkimuksen tunnuspiirre eli se, että tutkimuskysymykset nousevat usein aiemmasta tutkimuksesta, muiden menetelmien käyttämisestä tai puhtaasti intuitiosta, mikä siis jo pohjimmiltaan on menetelmien yhdistämistä. Myös aineistosta havaittavien ilmiöiden taustalla vaikuttavia syitä on tarkasteltava eri menetelmin kuin millä analyysi on tehty (Fillmore 1992: 59). Eri tutkimuslähtökohdista aineiston merkitystä tarkastelevat painottavat voimakkaasti siis nimenomaan monimetodisuutta.

Uutta ei eri metodisten lähestymistapojen yhdistäminen luonnollisestikaan ole, mutta erityisesti 2000-luvulla julkaistujen tutkimusten myötä on selvästi havaittavissa halu tarkastella ilmiöitä usean menetelmän ja aineiston avulla (aiheesta tarkemmin Arppe ja Järvikivi 2007). Esimerkiksi Griesin ja Stefanowitschin (2006) toimittama kokoelma kognitiivisen kielentutkimuksen alalla tehdystä aineistopohjaisesta tutkimuksesta tuo hyvin esiin yhdistämismahdollisuuksia (ks. myös Tummersin ym. (2005) koonti aineistopohjaisesta kognitiivisesta tutkimuksesta). Vastaavaa menetelmien onnistunutta yhdistämistä on harjoitettu myös muun muassa diskurssianalyyseissä (esim. Biber ja Jones 2005; ks. myös Hoey ym. 2007) ja erityisen runsaasti ja pitkään leksikografi-

assa aineistopohjaisten sanakirjojen laadinnassa (ks. kuitenkin korpusten käytön rajoituksista leksikografiassa esim. Cardey ja Greenfield 2002). Suomalaisessa tutkimuksessa korpusten käyttöä ovat yhdistäneet muihin menetelmiin muun muassa Arppe ja Järvikivi (2007) ja Haddington, Sivonen ja Jantunen (käsikirjoitus). Edellisessä tutkimuksessa korpusmenetelmät on yhdistetty kahteen kokeelliseen testaukseen tutkittaessa suomen kielen synonyymeja *mieltiä* ja *pohtia* ja jälkimmäisessä vuorovaikutus- ja kognitiivisen lingvistiikan menetelmiin selvitettäessä *tulla sanomaan* ja *mennä sanomaan* -tyyppisten konstruktoiden käyttöä. Kummassakin tutkimuksessa paljastuu, että monimetodisella tutkimuksella ja usean aineiston käytöllä saadaan selvästi kattavampi kuva ilmiöstä kuin tyydyttäessä yhteen menetelmään; lähestymistapoja yhdistämällä löydetään myös selityksiä yhden metodin avulla saaduille tuloksille. Aineiston käytössä ei kuitenkaan kannata minkään tutkimustradition puitteissa jäädä pelkästään sattumanvaraisen »aineistovetoisuuden» varaan, jolloin tutkija yllättäen huomaa jotain mielenkiintoista aineistosta muuta menetelmää käyttäessään, vaan aineistolle voidaan antaa tutkimuksen alusta lähtien metodologisesti perustelu asema (erityisesti aineistovetoisessa lähestymistavassa) ja aineiston suhteen voidaan tehdä järjestelmällisiä, harkittuja, teoriaan ja tavoitteisiin soveltuvia ratkaisuja. Luonnollisestikaan tämä ei estä mielenkiintoisten havaintojen tekemistä ja niiden esiin nousemista myös sattumalta tutkimuksen kuluessa. ■

JARMO HARRI JANTUNEN
Sähköposti: jarmo.jantunen@oulu.fi

LÄHTEET

- ALTENBERG, BENGT – GRANGER, SYLVIANE 2002: Recent trends in cross-linguistic lexical studies. – Bengt Altenberg & Sylviane Granger (toim.), *Lexis in contrast: Corpus-based approaches* s. 3–48. Amsterdam: Benjamins.
- ARPPE, ANTTI 2008: *Univariate, bivariate, and multivariate methods in corpus-based lexicography: A study of synonymy*. Publications of the department of general linguistics no. 44. University of Helsinki. – http://www.ling.helsinki.fi/~aarppe/Publications/Arppe_Dissertation_Final_Print.pdf 26.1.2009.
- ARPPE, ANTTI – JÄRVIKIVI, JUHANI 2007: Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. – *Corpus linguistics and linguistic theory* 3 (2) s. 131–159.
- BIBER, DOUGLAS 1993: Representativeness in corpus design. – *Literary and linguistic computing* 8 (4) s. 243–257.
- BIBER, DOUGLAS – CONRAD, SUSAN – REPPEN, RANDI 1998: *Corpus linguistics: Investigating language structure and use*. Cambridge approaches to linguistics. Cambridge University Press.
- BIBER, DOUGLAS – JONES, JAMES K. 2005: Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. – *Corpus linguistics and linguistic theory* 1 (2) s. 151–182.
- BORSLEY, ROBERT 2002: Book review on S. Hunston and G. Francis: *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. – *Lingua* 112 (3) s. 235–241.
- CARDEY, SYLVIANE – GREENFIELD, PETER 2002: Computerised set expression dictionaries: Analysis and design. – Bengt Altenberg & Sylviane Granger (toim.), *Lexis in contrast: Corpus-based approaches* s. 231–248.

- Amsterdam: Benjamins.
- FEATHERSTON, SAM 2007: Data in generative grammar: The stick and the carrot. – *Theoretical linguistics* 33 (3) s. 269–318.
- FILLMORE, CHARLES J. 1992: »Corpus linguistics» or »Computer-aided armchair linguistics». – Jan Svartvik (toim.), *Directions in corpus linguistics. Proceedings of Nobel symposium 82, Stockholm, 4–8 August 1991* s. 35–60. Berlin: Mouton de Gruyter.
- FRANCIS, GILL 1993: A corpus-driven approach to grammar: Principles, methods and examples. – Mona Baker, Gill Francis & Elena Tognini-Bonelli (toim.), *Text and technology: In honour of John Sinclair* s. 137–156. Amsterdam: Benjamins.
- GEERAERTS, DIRK 2006: Methodology in cognitive linguistics. – Gitte Kristiansen, Michel Achard, René Dirven & Francisco J. Ruiz de Mendoza Ibáñez (toim.), *Cognitive linguistics: Current applications and future perspectives* s. 21–50. Berlin: Mouton de Gruyter.
- GRIES, STEFAN TH. 2002: Evidence in linguistics: Three approaches to genitives in English. – http://www.linguistics.ucsb.edu/faculty/stgries/research/EvidenceLinguistics_LACUS2002.pdf 26.1.2009.
- GRIES, STEFAN TH. – STEFANOWITSCH, ANATOL (toim.) 2006: *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Trends in linguistics. Studies and monographs 172. Berlin: Mouton de Gruyter.
- GRONDELAERS, STEF – GEERAERTS, DIRK – SPEELMAN, DIRK 2007: A case for a cognitive corpus linguistics. – Monica Gonzalez-Marquez, Irene Mittelberg, Seana Coulson & Michael J. Spivey (toim.), *Methods in Cognitive Linguistics. Human Cognitive Processing* 18. Amsterdam: Benjamins.
- HADDINGTON, PENTTI – SIVONEN, JARI – JANTUNEN, JARMO HARRI (käsikirjoitus): A Corpus linguistic, cognitive semantic and interactional linguistic view to *go-say* and *come-say* constructions in Finnish.
- HOFFMANN, SEBASTIAN 2004: Are low-frequency complex prepositions grammaticalized? On the limits of corpus data — and the importance of intuition. – Hans Lindquist & Christian Mair (toim.), *Corpus Approaches to grammaticalization in English* s. 171–210. Studies in corpus linguistics 13. Amsterdam: Benjamins.
- HOEY, MICHAEL – MAHLBERG, MICHAELA – STUBBS, MICHAEL – TEUBERT, WOLFGANG 2007: *Text, discourse and corpora: Theory and analysis*. Studies in corpus and discourse. London: Continuum.
- HUUMO, TUOMAS 2007: Joko lingvistin nojatuoli joutaisi kaatopaikalle? Introspektioliingvistiikan asemasta kognitiivisessa kielitieteessä. – *Emakeele Seltsi aastaraamat* 53 s. 163–180.
- JANTUNEN, JARMO H. 2004: *Synonymia ja käännessuomi: Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännskielen leksiikaalisiin erityispiirteisiin*. Joensuun yliopiston humanistisia julkaisuja 35. Joensuu: Joensuun yliopisto.
- 2007: Oppijansuomen piirteitä korpusvetoisesti. – Pirkko Muikk-Werner, Ossi Kokko & Hannu Remes (toim.), *Virsu III: Suomalais-ugrilaisia kohdekieliä ja kontakteja* s. 67–83. Studies in languages 42. Joensuu: Joensuun yliopisto.
- 2008: On a blind date with the data: Corpus-driven analysis as a method for studying learner language. Plena-

- riesitelmä, 7th EAAL Conference linguistic resources and database use, 24–25 April 2008, Tallinn. – <http://www oulu.fi/hutk/sutvi/oppijankieli/tutkimus> 25.7.2008.
- KEMPPANEN, HANNU 2008: *Avainsanoja ja ideologiaa: Käännettyjen ja ei-käännettyjen historiatekstien korpuslingvistinen analyysi*. Joensuun yliopiston humanistisia julkaisuja 51. Joensuu: Joensuun yliopisto.
- LEECH, GEOFFREY 1991: Corpora. – Kirsten Malmkjaer (toim.), *The linguistics encyclopedia* s. 73–80. London: Routledge.
- MAURANEN, ANNA 2000: Strange strings in translated language: A study on corpora. – Maeve Olohan (toim.), *Intercultural faultline: Research models in translation studies I, textual and cognitive aspects* s. 119–141. Manchester: St. Jerome.
- MAURANEN, ANNA 2004: Spoken corpus for an ordinary learner. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 89–105. Studies on corpus linguistics 12. Amsterdam: Benjamins.
- MEURERS, W. DETMAR 2005: On the use of electronic corpora for theoretical linguistics: Case studies from the syntax of German. – *Lingua* 115 s. 1619–1639.
- MOON, ROSAMUND 2007: Sinclair, lexicography, and the Cobuild Project: The application of theory. – *International journal of corpus linguistics* 12 (2) s. 159–181.
- OJUTKANGAS, KRISTA 2008: Kielentutkijan mielikuvituksen puute eli mihin todelliseen kielenkäyttöön perustuvaa aineistoa tarvitaan. – *Virittäjä* 112 s. 276–284.
- PENKE, MARTINA – ROSENBAACH, ANETTE 2004: What counts as evidence in linguistics? An introduction. – *Studies in language* 28 (3) s. 480–526.
- PÄIVIÖ, PIA 2007: *Suomen kielen asti ja saakka: Terminatiivisten partikkelien synonymia, merkitys, käyttö ja kehitys sekä asema kieliopissa*. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 75. Turku: Turun yliopisto.
- RÖMER, UTE 2004: A corpus-driven approach to modal auxiliaries and their didactics. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 185–199. Studies on corpus linguistics 12. Amsterdam: Benjamins.
- SAMPSON, GEOFFREY 2005: Quantifying the shift towards empirical methods. – *International journal of corpus linguistics* 10 (1) s. 15–36.
- 2007: Grammar without grammaticality. – *Corpus linguistics and linguistic theory* 3 (1) s. 1–32.
- SCOTT, MIKE – TRIBBLE, CHRISTOPHER 2006: *Textual patterns: Key words and corpus analysis in language education*. Studies in corpus linguistics 22. Amsterdam: Benjamins.
- SINCLAIR, JOHN 1991: *Corpus, concordance, collocation*. Describing English language. Oxford: Oxford university press.
- 2004: *Trust the text: Language, corpus and discourse*. London: Routledge.
- STUBBS, MICHAEL 1995: Collocations and semantic profiles: On the cause of the trouble with quantitative studies. – *Functions of language* 2 (1) s. 23–55.
- 2002: On text and corpus analysis: a reply to Borsley and Ingham. – *Lingua* 112 s. 7–11.
- TOGNINI-BONELLI, ELENA 2001: *Corpus linguistics at work*. Studies in corpus linguistics 6. Amsterdam: Benjamins.

- 2002: Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. – Bengt Altenberg & Sylviane Granger (toim.), *Lexis in Contrast: Corpus-based approaches* s. 73–95. Amsterdam: Benjamins.
- TSUI, AMY B. M. 2004: What teachers have always wanted to know – and how corpora can help. – John McH. Sinclair (toim.), *How to use corpora in language teaching* s. 39–61. *Studies on corpus linguistics* 12. Amsterdam: Benjamins.
- TUMMERS, JOSE – HEYLEN, KRIS – GEERAERTS, DIRK 2005: Usage-based approaches in cognitive linguistics: A technical state of the art. – *Corpus linguistics and linguistic theory* 1 (2) s. 225–261.
- TURNER, ROBIN 2006: »How do you know she’s a woman?» Features, proto-types and category stress in Turkish. – June Luchjenbroers (toim.), *Cognitive linguistics investigations: Across languages, fields and philosophical boundaries* s. 219–234. Philadelphia: Benjamins.
- VANHATALO, ULLA 2003: Kyselytestit vs. korpuslingvistiikka lähisynonyymien semanttisten sisältöjen arvioinnissa – Mitä vielä *keskeisestä* ja *tärkeästä*? – *Virittäjä* 107 s. 351–369.
- VISAPÄÄ, LAURA 2005: Alastomat, inhimilliset infinitiivit: Kahden infinitiivikonstruktion pragmaattinen ja morfosyntaktinen analyysi. – Ilona Herlin & Laura Visapäälä (toim.), *Elävä kielioppi: Suomen infiniittisten rakenteiden dynamiikkaa* s. 72–98. Helsinki: Suomalaisen Kirjallisuuden Seura.
- WASOW, THOMAS – ARNOLD, JENNIFER 2005: Intuitions in linguistic argumentation. – *Lingua* 115 s. 1481–1496.