

Avainrakenneanalyysillä uusia ikkunoita suomen oppimiseen

Ilmari Ivaska: *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti. Avainrakenneanalyysi.* Turun yliopiston julkaisuja C 409. Turku: Turun yliopisto 2015. 232 s. ISBN 978-951-29-6203-7.

Suomi toisena kielenä -alan tutkimus on pitkään keskittynyt oppimisen alku- ja keskivaiheisiin. Tämä on ymmärrettävää, sillä alkuvaiheen oppijoita on eniten – kaikkiaan aloittavat alusta – ja keskivaihe, Eurooppalaisen viitekehyksen (2014) taitotaso B, taas on tärkeä opiskelun, työllistymisen ja kansalaisuuden saamisen kannalta. Turun yliopisto on kuitenkin ottanut tehtäväkseen edistyneiden oppijoiden kirjoitusten keräämisen ja korpuksen (LAS2) rakentamisen. Tämä työ on yhteiskunnallisesti erittäin tärkeää, sillä koulutettujen maahanmuuttajien täytyy voida hankkia korkeatasoinen suomen kielen taito voidakseen hyödyntää aiempaa koulutustaan Suomessa ja toimia asiantuntijatehtävissä. Lisäksi ulkomaisien tutkinto-opiskelijoiden määrä kasvaa suomalaisissa yliopistoissa. Ilmari Ivaskan väitöskirja käsittelee juuri tämän ryhmän suomea. Ivaska on myös ollut keskeisesti mukana LAS2-korpuksen suunnittelussa ja kokoamisessa.

Ivaskan väitöskirja koostuu viidestä kotimaisissa ja kansainvälisissä tieteellisissä lehdissä ja kokoomateoksissa julkaistusta artikkelista ja niitä taustoittavasta laajasta johdantoluvusta. Tutkimuksella on kaksi päätavoitetta, joista ensimmäinen on metodinen: Tutkija kehittää työssään uuden menetelmällisen lähestymistavan ja siihen tarvittavia työkaluja. Toinen tavoite on valottaa edistyneen akateemisen kielitaidon ilmiöitä hyödyntäen näitä menetelmällisiä ratkaisuja. Var-

sinaiset tutkimuskysymykset ovat seuraavat (ks. johdanto-osan s. 5):

(1) Miten oppijankielen rakenteellisia ominaispiirteitä voidaan tunnistaa ja analysoida korpus- eli aineistovetoisesti – ilman ennalta asetettua hypoteesia mahdollisten erojen luonteesta?

(2) Mitkä kielen konstruktiot ovat tyypillisiä tai epätyypillisiä edistyneiden suomenoppijoiden kirjoituksille ensikielenään suomea käyttävien kirjoituksiin verrattaessa?

(3) Mitkä kielen konstruktiot erottavat eri ensikieliä puhuvia edistyneitä suomenoppijota toisistaan?

(4) Minkä kielen konstruktioiden käyttö muuttuu edistyneiden suomenoppijoiden kirjoituksissa ajan kuluessa ja miltä muutos näyttää ensikielenään suomea käyttävien kirjoituksiin verrattuna?

(5) Mistä syistä vielä edistyneidenkin suomenoppijoiden kieli poikkeaa rakenteellisesti joiltakin osin ensikielenään suomea käyttävien kielestä ja on edelleen muutoksessa?

Tutkimuskysymykset ovat sangen laajoja, eikä niihin ole mahdollista vastata täydellisesti suuremmassakaan tutkimushankkeessa. Kysymykset 2–5 koskevat kuitenkin laajaa toisen kielen oppimisen tutkimuksen aluetta: oppijankielen suhdetta äidinkielisten kieleen (2) ja näiden kielimuotojen erojen mahdollisia syitä (5), ensikielen vaikutusta (3) ja kielitaidon kehityskulkuja (4). Artikkeleissa ja joh-

danto-osassa kuhunkin kysymykseen kuitenkin saadaan vastaus ainakin joiltakin osin. Kehitettyä menetelmää on siis kokeiltu lähestymällä tutkimusaihetta useista keskeisistä toisen kielen oppimisen näkökulmista.

Aineiston esittelystä syvälliseen analyysiin

Ensimmäinen artikkeli ”The Corpus of Advanced Learner Finnish (LAS2): Database and toolkit to study academic learner Finnish” (2014, *Apples – Journal of Applied Language Studies* 8 s. 21–38) on siinä mielessä poikkeuksellinen väitöskirja-artikkeli, että siinä esitellään työn aineisto ja päämenetelmä, mutta ei lainkaan näiden avulla saavutettuja tuloksia. Sen sisältämä tieto on siis oikeastaan osa artikkeleita yhdistävää johdantoa. Artikkelin julkaiseminen itsenäisenä työnä on kuitenkin ollut järkevää, koska näin tietoa korpuksesta on voitu levittää tutkijoiden käyttöön jo ennen koko väitöskirjan valmistumista. On tietenkin myös käytännöllistä sijoittaa tämä artikkeli omaksi osakseen väitöskirjaa – näin johdantoon tarvitaan vain lyhyt aineiston esittely. Artikkelin 1 esittelee huolellisesti aineiston taustamuuttujat, työprosessin, annotoinnin esimerkkeineen sekä aineiston käyttöön kehitetyt työkalut. Siinä käydään läpi myös LAS2-korpuksen genret, koonti- ja annotointiperiaatteet sekä aineiston määrä kirjoitushetkellä.

Virittäjässä ilmestynyt artikkeli 2 ”Edistyneen oppijansuomen avainrakenteita: Korpustenäkökulma kahden kieli- muodon tyypillisiin rakenteellisiin eroihin” (2014, *Virittäjä* 118 s. 161–193) esittelee aluksi taustaksi korpustutkimusta ja konstruktionäkökulmaa, joihin paneudutaan laajemmin johdanto-osassa. Tässä artikkelissa esitellään myös muita artikkeleita perusteellisemmin koko väitöskirjan keskeinen menetelmä, avainrakenteenanalyysi. Tulokseksi saadaan

joukko n-grammeja, potentiaalisia avainrakenteita, jotka erottavat oppijansuomen ja äidinkielen suomen korpuksia. Tulokset ovat kiinnostavia, sillä ne määrittelevät aiempia tutkimuksia tarkemmin, mitä morfosyntaktinen kompleksisuus tällä kielitaidon tasolla tarkoittaa ja millaisia ovat aikamuotojen ja modaali-ilmausten erot oppijankielen ja syntyperäisen kielen välillä. Tuloksia asetetaan laajempaan kontekstiin jo tämän artikkelin puitteissa, mutta etenkin yhdessä muiden artikkeleiden tulosten kanssa johdanto-osassa.

Kolmas artikkeli ”Tracing crosslinguistic influences in structural sequences: What does key-structure analysis have to offer?” (2015, *Bergen Language and Linguistics Studies* 6. *Learner Corpus Research: LCR2013 Conference Proceedings* s. 23–44) suuntaa huomion paljon tutkituun, mutta edelleen erittäin ajankohtaiseen kysymykseen kieltenvälisestä vaikutuksesta kielenoppijan tuotoksessa. Siinä tarkastellaan viittä eri ensikieltä puhuvien opiskelijoiden tenttivastauksista löytyviä eroja ottamatta kantaa erojen kieliteiseen tai myönteiseen vaikutukseen oppimisen kannalta. Artikkelin on menetelmällisesti perusteellinen. Joitakin eroja ryhmien väliltä löytyy, vaikka aineiston pienuus estää tilastollisesti riittävän suurten erojen löytämisen. Tämän artikkelin tuoman kokemuksen pohjalta voisi kirjoittaa johonkin kansainväliseen alan julkaisuun avainrakenteenanalyysin mahdollisuuksista selvittää positiivista siirtovaikutusta, jota on vaikea tutkia muilla keinoin.

Kielitaidon pitkittäiskehitys on teemana neljännessä artikkelissa ”Longitudinal changes in academic learner Finnish: A key structure analysis” (2015, *International Journal of Learner Corpus Research* 1 s. 210–241). Jälleen menetelmien esittelyllä on artikkelissa keskeinen osuus, mutta esiin nousee myös kiinnostava ja opetuksessakin hyödyllinen tulos aikamuotojen käytön muutoksesta. Kyseessä on juuri sellainen tulos, joka olisi toden-

näköisesti jäänyt löytämättä ilman väitte-
lijän kehittämää menetelmää.

Myös viidennessä artikkelissa ”Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä” (2014, *Lähivertailuja. Lähivördlusi* 24 s. 47–80) löydetään sekä sanastollisia että rakenteellisia oppijankielen piirteitä, jotka voi myös nostaa esille edistyneiden kielenoppijoiden opetuksessa. Menetelmälle uskollisena artikkeli nostaa esiin frekventeimpiä eroja, vaikka ne tässä tapauksessa eivät ehkä ole aihepiirin kannalta kiinnostavimpia. Perusteellisimmin käsitellään *voida*-verbin eri muotojen käytön jakautumista. Kaiken kaikkiaan tässä artikkelissa päästään muita syvem-
mälle löydösten vertaamisessa aikaisemman tutkimuksen tuloksiin.

Laaja-alainen katsaus toisen kielen oppimisen tutkimukseen

Väitöstutkimuksen johdanto-osa koostuu teoreettisten näkemysten (käyttöpohjaisuus, konstruktiot analyysiyksiköinä ja toisen kielen oppimisen yleiset taustatekijät, tekstilaji) sekä metodisten ratkaisujen (korpusvetoisuus, avainrakennanalyysi) esittelystä ja perustelusta. Tulokset on koottu hyvin tiiviisti noin 13 sivulle, mikä puoltaa paikkaansa, koska tulosten yksityiskohdat ovat luettavissa artikkeleista. Johdanto-osa osoittaa hyvää perehtymistä toisen kielen oppimisen keskeisiin suuntauksiin ja tutkimuskysymyksiin. Valitut käyttöpohjaisuuden ja toisen kielen oppimisen tutkimuksen suuntauksat on esitelty asiantuntevasti ja olennaiseen keskittyen. Laajalta alueelta on valikoitu tämän työn kannalta keskeisimmät käsitteet ja ilmiöt. Suuri määrä tietoa ja laaja lukeneisuus on tiivistetty hyvin. Vain muutama asia ei suoraan liity artikkeleihin ja työn aihepiireihin. Tekstilajien käsittely, vaikka sinänsä toisen kielen oppimisen kannalta tärkeä seikka, on jossain määrin tarpeeton, kun koko ai-

neisto edustaa yhtä tekstilajia, tenttivas-
tausta. Kyseisen tekstilajin kannalta taas tehtävänantoa olisi voinut käsitellä hie-
man syvemmin, vaikka sen huomioon ot-
taminen tutkimuksessa ei ollutkaan kor-
puksen ominaisuuksien vuoksi mahdol-
lista. Tehtävänannollahan on monenlaisia
sekä kielellisiä että tekstin rakenteeseen
liittyviä vaikutuksia tuotettuun tentti-
vastaukseen.

Toinen aavistuksen tarpeettomalta vai-
kuttava kohta on fraseologian tutkimuk-
sen esittely pääasiassa morfologiaa käsit-
televässä tutkimuksessa, vaikka valintaa
voidaan perustella korpustutkimuksen
tradition näkökulmasta. Fraseologian tut-
kimus on toki myös suhteessa valittuun
konstruktionäkökulmaan, vaikka se ei
työstä suoraan käykään ilmi. Voi näet aja-
tella, että fraseologian tutkimus, joka ta-
vallisesti sijoitetaan sanaston tutkimuk-
sen yhteyteen, ja kielen rakenteen alalle
kuuluva konstruktioikielioppi ja muut
konstruktio pohjaiset näkemykset kielestä
ovat itse asiassa samalla jatkumolla.

Psykolinguistien paljon tutkima kysy-
mys eksplisiittisen ja implisiittisen tiedon
roolista kielenoppimisessa ei myöskään
suoraan liity korpustutkimukseen, mutta
on hyvä esimerkki Ivaskan taidosta valita
laajasta, yli kolmen vuosikymmenen ai-
kana kertyneestä tutkimustiedosta olen-
naisin ja tiivistää se puoleen sivuun. Kaik-
kien kolmen tässä mainitun aihepiirin
mukanaolo väitöskirjassa on ymmärret-
tävä myös siltä kannalta, että opinnäyte-
työssä on tapana esitellä taustoja hieman
laajemmin kuin muussa tutkimuksessa.

Jatkuvasti karttuva korpus on luon-
nollisesti hankala pitkässä työprosessissa,
koska eri artikkeleita kirjoitettaessa ai-
neistoa on ollut eri määrä. Ivaska rapor-
toi kyllä tarkasti kulloinkin artikkeleissa
käytetyn aineiston määrän ja tekee niistä
koonnin johdannossa (s. 61). Aineisto si-
nänsä on riittävä, vaikka korpustutkimuk-
sessa on aina tietenkin sitä parempi, mitä
enemmän aineistoa on. Vain kolmannen

artikkelin kohdalla aineiston koko rajoittaa hieman tutkimuksen tuloksia ja niiden käytettävyyttä, koska siinä aineisto joudutaan jakamaan osa-aineistoihin kirjoittajan äidinkielen perusteella. Tämäkin ongelma lievittyy korpuksen kasvaessa.

Piiloilmiöt esiin avainrakenne-analyysillä

Väitöskirjan keskeinen menetelmä on avainrakenneanalyysi. Siinä ”yhdistyvät vertailtavien kieliaineistojen tilastollisesti merkitsevien frekvenssierojen paikallistaminen aineistovetoisesti ja havaittujen erojen kvantitatiivis-kvalitatiivinen analyysi niiden tyypillisen käytön kannalta” (artikkeli 2, s. 168). Analyysiprosessissa on useita vaiheita, ja siinä käytetään monia tilastollisia menetelmiä, joiden hallinta ja soveltaminen tähän tarkoitukseen on sinänsä jo huomattava saavutus. Menetelmällisesti avainrakenneanalyysi on keksintö, jonka avulla löydetään tutkittavaksi ilmiöitä, joita ei ole välttämättä ennen huomattu tutkia. Menetelmällä on toki kansainväliset juuret, kuten käy ilmi kollostruktio-käsitteen esittelystä (s. 28–29), mutta kirjoittajan oma rooli avainrakenneanalyysin kehittäjänä jää vaatimattomuuden jalkoihin, sillä oikeastaan vasta päätännössä (ja sielläkin passiivimuodossa, s. 78) hän myöntää käytännössä kehittäneensä tämän menetelmän.

Avainrakenneanalyysin avulla löytyneisiin uusiin oppijankielen ilmiöihin on paneuduttu huolellisesti ja perusteellisesti siitä huolimatta, että artikkeleiden erilaisten tutkimuskysymysten vuoksi joudutaan käsittelemään hyvin monentasoisia seikkoja ja suhteuttamaan niitä toisen kielen oppimisen tutkimukseen laajasti. Avainrakenneanalyysin toimivuuden osoittavat artikkelien tulokset, jotka ovat linjassa aiempien havaintojen kanssa. Menetelmä ei siis tuota tuloksia, jotka olisivat aivan aiemman tutkimuksen vastaisia, vaan aikaisemman tutkimuksen voi kat-

soa validoivan tässä esitellyn menetelmän tuloksia. On toivottavaa, että menetelmä leviäisi myös muiden laajoja aineistoja käyttävien tutkijoiden käyttöön.

Kansainvälisesti keskeisin tulos lienee, että työ antaa uuden näkökulman lingvistisen kompleksisuuden tutkimukseen, erityisesti frekvenssin ja kompleksisuuden välisen suhteen pohdiskeluun. Ivaska määrittelee aiempia tutkimuksia tarkemmin, mitä morfosyntaktinen kompleksisuus tällä kielitaidon tasolla tarkoittaa ja millaisia ovat aikamuotojen ja modaali-ilmausten erot oppijankielen ja syntyperäisen kielen välillä. Työssä on hyödynnetty myös useissa viimeaikaisissa suomi toisena kielenä -alan tutkimuksissa käytettyä DEMfad-mallia (ks. esim. Martin ym. 2010), jonka tavoitteena on mahdollistaa juuri frekvenssin, tarkkuuden ja kompleksisuuden yhteyksien tutkimista. Tämän mallin kehittämiseen Ivaskan tutkimus antaa uutta näkökulmaa.

Edistyneiden suomenoppijoiden kielen tutkiminen on tärkeää myös toisen kielen oppimisen yleisen tutkimuksen kannalta. Akateemisten oppijoiden englantia on tutkittu paljon, koska englantia on monissa osissa maailmaa keskeinen tai ainoa akateemisten opintojen kieli. Työ antaa tähän kansainväliseen tutkimukseen arvokkaan lisän sekä tuomalla mukaan uuden ja rakenteeltaan huomattavasti englannista poikkeavan kohdekielen että kehittämällä menetelmän, joka on sovellettavissa muihinkin kieliin.

Kaiken kaikkiaan väitöskirjan viisi artikkelia ja johdanto-osa kytkeytyvät luontevasti toisiinsa ja muodostavat pakottoman kokonaisuuden, jossa on nähtävissä jatkumo artikkelista toiseen. Työn suurin anti tutkimuskentälle on metodinen, mutta eri artikkeleiden sisällölliset tulokset ovat nekin monessa kohtaa uusia ja hyödyllisiä. Metodi on myös sovellettavissa monenlaisten muiden kieliaineistojen tutkimukseen, kunhan vain aineiston koko on riittävän suuri. Sen avulla nou-

sevat esiin oppijankielen ilmiöt, joita ei havainnoimalla tai intuition varassa havaitisi.

MAISA MARTIN
etunimi.sukunimi@jyu.fi

Lähteet

Eurooppalainen viitekehys. Kielten oppimisen, opettamisen ja arvioinnin yhteinen

eurooppalainen viitekehys 2014. Helsinki: Sanoma Pro.

MARTIN, MAISA – MUSTONEN, SANNA – REIMAN, NINA – SEILONEN, MARJA 2010: On becoming an Independent User. – Inge Bartning, Maisa Martin & Ineke Vedder (toim.), *Communicative proficiency and linguistic development. Intersections between SLA and language testing research*. Eurosla Monograph Series 1. <http://www.eurosla.org/monographs/EM01/EM01home.php> (20.5.2016).

Epäkanoniset subjektit konstruktioissa

Marja-Liisa Helasvuo & Tuomas Huumo (toim.): *Subjects in constructions – canonical and non-canonical*. Amsterdam: John Benjamins 2015. 324 s. ISBN 978-90-272-0438-7.

Marja-Liisa Helasvuon ja Tuomas Huumon toimittaman kokoomateoksen *Subjects in constructions – canonical and non-canonical* aiheina ovat tavalla tai toisella epäkanoniset subjektit. Yhteisenä viitekehystenä kaikille kirjan kymmenelle luvulle on konstruktionäkökulma yhdistettynä kognitiiviseen tai diskurssifunktionaaliseen kielioppiin. Kirja jakautuu kolmeen osaan. Ensimmäisen osan (*Grammatical and discourse perspectives on non-canonical subjects*) kolme lukua käsittelevät suomen ja viron epäkanonisten subjektien kieliopillisia ja diskursiivisia piirteitä. Keskimmäinen osa (*Stretching the limits of subjecthood*) on kahta muuta sekakoosteisempi: sen neljä lukua käsittelevät ranskan ja englannin muodollisia subjekteja, viron passiivilauseiden agent-

teja, säätilaa ilmaisevia predikaatteja ja niiden subjekteja sekä epäkanonisen subjektin sisältävien lauseiden ilmaisemaa tapahtumarakennetta kieltenvälisesti. Kolmannen osan (*Subjects in networks of constructions*) kolme lukua keskittyvät suomen genetiivisubjekteihin ja venäjän datiivisubjekteihin konstruktioiden verkostoina.

Kanoniset ja ei-kanoniset subjektit

Subjektin käsitteen määrittäminen katavalla ja kieltenvälisen vertailun mahdollistavalla tavalla on yksi funktionaalisen kielitieteen kiistellyimmistä kysymyksistä. Epäkanonisen subjektin käsite ei ole ainkaan ongelmattomampi: väljästi määritellen epäkanonisina subjekteina voidaan pitää argumentteja, jotka täyttävät osan subjektisuuden semanttisista, diskursiivista tai kieliopillisista kriteereistä mutta eivät kaikkia. Toisinaan subjektina pitämiseen voi riittää lähinnä kieliopillinen perinne tai ajatus siitä, että jokaisessa lauseessa