

Edistyneen oppijansuomen konstruktiopirteitä korpusvetoisesti

Avainrakenneanalyysi

ILMARI IVASKA

Väitöksenalkajaisesitelmä Turun yliopistossa 4. syyskuuta 2015

Maahanmuutto on ollut Suomessa tiukassa keskustelussa jo hyvän tovin. Väestön vanhetessa onnistunut työperäinen maahanmuutto on koettu maan hyvinvoinnin edellytykseksi (Maahanmuuton tulevaisuus 2013). Onnistuneen maahanmuuton edellytyksenä voitaneen puolestaan pitää riittävää kielitaitoa, ja nimenomaan sel-laista kielitaitoa, jonka avulla voi toimia korkea koulutusta ja erityisosaamista vaa-tivissa tehtävissä (ks. esim. Forsander & Alitolppa-Niitamo 2000; Schultz-Nielsen 2000). Edelleen, jotta kielitaitoa voidaan kehittää vaadittavalle tasolle, on tiedettävä, mitkä seikat vaativat huomiota vielä edistyneidenkin suomea toisena kielenä käyttä-vien kielessä.

Tämä tutkimus keskittyy edistyneen oppijansuomen rakenteellisiin erityis-piirteisiin. Siinä yritetään tunnistaa ja selittää, miten ja miksi edistyneiden suomen-oppijoiden kieli muuttuu ja poikkeaa ensikielisestä suomesta. Tutkimusaineisto on suomea toisena kielenä oppineiden kirjoittamia akateemisia tekstejä. Tutkimus on osin menetelmällinen, ja siinä käsitellään sitä, miten näihin kysymyksiin voi vastata. Tutkimus koostuu johdanto-osasta sekä viidestä aiemmin julkaistusta artikkelista (Ivaska 2014a, 2014b, 2014c, 2015a, 2015b). Selvitän seuraavaksi ensin, miten kehittä-mäni menetelmän – avainrakenneanalyysin – on tarkoitus toimia. Tämän jälkeen annan kuvan siitä, mitä menetelmän avulla todella saatiin selville.

Aineistoon perustuvassa kielentutkimuksessa – niin kuin empiirisissä tieteissä ylipäätäänkin – tyypillisiä tutkimusasetelmia kuvaavat usein kysymykset ”Miten?” ja ”Miksi?”. Siis vaikkapa seuraavasti: miten jokin kielenilmiö A havaintojen perus-teella toimii, ja miksi se toimii havaitulla tavalla? Suuri osa systemaattisesti kerätyillä

ja dokumentoiduilla sähköisillä kieliaineistoilla eli korpuksilla tehtävästä kielen-
tutkimuksesta on luonteeltaan tällaista.

Tarkemmassa erottelussa voidaan aineistoon perustuvista tutkimusasetelmista
erottaa kaksi toisistaan osin poikkeavaa lähestymistä: aineistopohjainen ja aineisto-
vetoinen tutkimusote – korpuksia käytettäessä on luontevaa puhua korpuspohjaisesta
ja korpusvetoisesta tutkimusotteesta (Tognini-Bonelli 2001; Hardie & McEnery 2010).
Korpuspohjainen tutkimusote vastaa edellä kuvattua ajatusta: tutkijalla on mielessään
omiin havaintoihinsa ja aiempaan tutkimukseen perustuva ilmiö, jonka mekanismeja
hän selvittää niin, että pohjaa selityksensä käsissään olevasta korpuksesta tekemiinsä
havaintoihin. Hän siis esittää kysymykset ”Miten?” ja ”Miksi?”.

Korpusvetoinen tutkimusote pyrkii niin ikään vastaamaan näihin kysymyksiin.
Korpuspohjaisesta tutkimusotteesta poiketen tutkimuskohdetta ei kuitenkaan valita
aiempien havaintojen ja niihin perustuvien hypoteesien perusteella. Sen sijaan annea-
taan aineiston ohjata tutkija jonkin siitä itsestään nousevan ilmiön ääreen. Toisin sa-
noen ensimmäinen esitettävä kysymys onkin ”Mikä?”.

Yksi tapa mikä-kysymyksen esittämiseen on etsiä ilmiöitä, jotka ovat aineistolle jol-
lain lailla ominaisia tai erityisiä. Tällöin lähtökohtana pidetään usein määrällisiä eroja
tutkittavan aineiston ja jonkin vertailuaineiston välillä. Vaihtoehtoisesti voidaan etsiä
eroja saman aineiston eri osien välillä. Ajatusta voidaan havainnollistaa lasten tiede-
keskuksista tutun fakiirinpedin avulla. Kun fakiirinpeti asetetaan henkilön kasvoille,
pedin neulat muotoutuvat kasvojen kolmiulotteiseksi luonnokseksi. Kun fakiirin-
pedille asetetaan käsi, hahmottuu siitä vastaavasti käsi. Kun tällaista kuvaa katsoo ensi-
kerran, on selvää, että kyseessä on käsi. Mutta vasta, kun verrataan kahta fakiirinpetiä
toisiinsa, voidaan todella tarkastella sitä, miten kädet eroavat toisistaan. Edelleen, kun
suurempaa otantaa kahden eri ryhmän käsistä hahmottuvia profileja verrataan toi-
siinsa, voidaan paikantaa ne kädenpiirteet, jotka ovat ominaisia tai erityisiä jommalle-
kummalle näistä joukoista.

Tässä tutkimuksessa korpusvetoinen tutkimusote on tällaista fakiirinpedin sovit-
tamista erilaisiin osa-aineistoihin ja niiden piirteiden tunnistamista, jotka tyypillisesti
erottavat näitä osa-aineistoja toisistaan. Käytännössä aloin tarkastelun laskemalla sa-
nojen morfologisten muotojen ja niiden yhdistelmien frekvenssejä eri teksteissä. Tein
tämänkaltaista vertailua kolmesta eri näkökulmasta:

Ensinnäkin vertasin ensikielenään suomea käyttävien ja jotain muuta kieltä ensim-
mäisenä kielenä oppineiden suomenkielisiä kirjoitelmia keskenään. Tällä tavoin py-
rin paikantamaan sellaisia kielenpiirteitä, joiden frekvenssi tyypillisesti erottaa vielä
edistyneitäkin suomenoppijoita ensikielisistä kielenkäyttäjistä. Toiseksi vertasin ensi-
kielenään japania, unkaria, tsekkia, liettuaa ja venäjää käyttäviä suomenoppijoita toi-
siinsa. Pyrin tällä tavoin paikantamaan sellaisia kielenpiirteitä, joita ensikieletään eri-
laiset suomenoppijat käyttävät keskenään eri määriä. Kolmanneksi vertasin useita sa-
moilta kielenoppijoilta eri aikoina kerättyjä tekstejä toisiinsa. Tässä kohdin pyrin pai-
kantamaan sellaisia kielenpiirteitä, joiden käyttöfrekvenssi useimmiten muuttuu ajan
kuluessa vielä edistyneidenkin suomenoppijoiden teksteissä.

Näin käsiin saatavat listat eivät kuitenkaan vielä itsessään kerro, mikä on se kielen
konstruktio – avainrakenne – joka on eron taustalla. Tämän selvittämiseksi on paneu-

duttava tarkemmin listalle päässeiden avainmuotojen käyttöön – katsomalla sitä, missä niitä esiintyy, mitä niitä ennen ja niiden jälkeen usein on ja mitä niillä usein ilmaistaan.

- (1) Opettaja voi *puhua* helpotettua kieltä
- (2) siksi on helpompi *rekonstruoida*
- (3) koska opettajalla ei ole mahdollisuutta *selittää* oppilaille kasvokkain

Yksi esiinnoisseista avainmuodoista on esimerkeissä 1–3 näkyvä A-infinitiivi eli vanhastaan ensimmäinen infinitiivi, siis verbin perusmuoto. Muotoa esiintyy ensikielenään suomea käyttävien teksteissä kaksi kertaa niin paljon kuin suomea toisena kielenä kirjoittavien teksteissä.

- (4) Suomessa verbit *voidaan jakaa* 0-, 1-, 2- tai 3-paikkaisiin verbeihin sen mukaan, moneenko tahoon ne vaikuttavat.
- (5) Mordvalainen eeos *voidaan luokitella* seuraaviin ryhmiin: 1) mytologinen ja 2) historiallinen.

Tarkemmassa tyyppillisen käytön analyysissä kuitenkin selvisi, että kyseinen muoto itsessään ei ole millään lailla näitä kielimuotoja erottava tekijä eikä minkään yksittäisen sanan infinitiivi myöskään selittänyt frekvenssiero. Sen sijaan erottavaksi tekijäksi ilmenivät moniosaiset verbiketjut, erityisesti sellaiset verbiketjut, jotka ilmaisevat jonkinlaista mahdollisuutta – kuten esimerkin 4 *voidaan jakaa* ja esimerkin 5 *voidaan luokitella*.

Tällaisia ilmauksia on siis suomea toisena kielenä kirjoittavien akateemisissa teksteissä vähemmän kuin ensikielenään suomea käyttävien vastaavissa teksteissä. Lähemmässä tarkastelussa ja peilattaessa tuloksia aiempaan tutkimukseen selvisi niin ikään, että tällainen mahdollisuuden ilmaiseminen – ja siis samalla muiden vaihtoehtoisten tulkintojen olemassaolon epäsuora toteaminen – on tyyppillistä akateemiselle kielenkäytölle (ks. Hyland 1999). Kyseessä on keino, jolla ilmaistaan kirjoittajan omaa suhdetta esitettyyn tietoon. Edellä esitetyn kaltaisissa mahdollisuuden ilmauksissa tavoitteena lienee usein se, että kirjoittaja etäännyttää oman mielipiteensä ilmauksen merkityssisällöstä ja sen oikeellisuudesta: kun siis katsomme esimerkkejä 4 ja 5, mahdollisuuden ilmaus osoittaa sen, että kirjoittajan mukaan esitellyt jaot tai luokittelut ovat yksi mahdollisuus mutta että muitakin mahdollisuuksia saattaa olla ja että ne eivät välttämättä ole oikeita tai ehdottomia.

Tilanne on samankaltainen useissa tässä tutkimuksessa avainrakenteiksi havaituissa konstruktioissa: monet niistä ovat jollain tavalla joko tyyppillisiä tai epätyypillisiä akateemiselle kirjoittamiselle. Näin on laita niissä konstruktioissa, jotka erottavat ensikielisiä ja ei-ensikielisiä kirjoittajia toisistaan, niissä konstruktioissa, jotka erottavat ensikieleltään erilaisia oppijoita toisistaan, ja niissä konstruktioissa, joiden käyttö muuttuu pitkäikäistarkastelun aikana. Esittelen seuraavaksi esimerkin kustakin ryhmästä.

(6) Suomen kielen taito on osoitettu *heikkentyvän*

Esimerkin 6 kaltaisen VA-partisiipin eli niin sanotun ensimmäisen partisiipin gettiiviä käytetään usein esimerkin kaltaisessa referatiivikonstruktiossa, perinteisesti sanottuna lauseenvastikkeena. Konstruktiio on verraten harvinainen ylipäätään. Ensikielenään suomea käyttävien teksteissä sitä kuitenkin esiintyy harvakseltaan tasaisesti, kun taas suomea toisena kielenä kirjoittavien 275:ssä tarkastellussa tekstissä niitä esiintyy yhteensä vain viisitoista kappaletta. Konstruktiio on verraten monimutkainen niin morfologiselta rakenteeltaan kuin morfosyntaktiselta funktioltaankin. Sen merkityksen täsmällisyys ja ilmauksen lyhyys tekevät konstruktion kuitenkin hyvin käyttökelpoiseksi akateemisen kirjoituksen kaltaisessa, yksiselitteisyyttä ja ilmauksen ekonomisuutta suosivassa käyttöympäristössä. Näin ollen onkin luontevaa ajatella käyttöeron johtuvan nimenomaan tekstilajille tyypillisen, rakenteellisesti kompleksisen konstruktion tyypillisen käytön osittaisesta oppimisesta.

(7) *Sanat järvi lampi meri ja joki* kuvaavat kaikki erilaisia vesistöjä

Esimerkissä 7 näkyvän listauskonstruktion käytön voidaan niin ikään ajatella olevan tavallista akateemisille teksteille. Niissä ilmaistaan usein erilaisia luokitteluja ja kategorioita. Konstruktiio on kuitenkin siinä mielessä epätyypillinen suomelle, että substantiivin määritteet tulevat vasta pääsanansa jälkeen – esimerkiksi *sanat*-sanaa seuraavat sen määritteet *järvi*, *lampi* ja *joki*. Konstruktion käytön yleisyys erottaa eri ensikieliryhmiä toisistaan. Tuntuu luontevalta ajatella, että tällaisen kohdekielen yleisestä mekanismista poikkeavan konstruktion soveltaminen on todennäköisempää, mikäli jokin toinen hallittu kieli tukee vastaavaa mallia.

(8) Suurin osa tehtävistä *oli* lapsille vaikea, mutta vaikeimmat *olivat* sanojen tavun poistuminen ja lisääminen. Kun lapset *lisäsivät* tavut sanoihin, jotkin sanat *olivat* vaihtunut toisiksi sanoiksi. Sanojen tavun poistumisessa tehtävässä *oli* myös nähtävissä samoin.

Esimerkissä kursivoituina näkyvät preteritit eli imperfektit ovat selvästi yleisempiä suomea toisena kielenä kirjoittavien teksteissä kuin ensikielisten teksteissä. Preteritien määrä kuitenkin vähenee kuudentoista kuukauden tarkastelujakson aikana. Samanaikaisesti preesens-muodot yleistyvät yhtä paljon. Jakson lopussa frekvenssit vastaavat toisella kielellään kirjoittavien ja ensikielisten kirjoittajien teksteissä toisiaan, eli ryhmät siis käyttävät aikamuotoja samalla tavalla.

Suomen preteriti on luonteeltaan narratiivinen, ja se korostaa tapahtumien ajallista järjestystä ja peräkkäisyyttä (ISK 2004: § 1531). Akateemisessa suomessa tiedosta ja tutkimuksen tuloksista kirjoitetaan tyypillisesti preesensissä, mistä syystä menneen ajan aikamuodon käyttö korostaa esimerkin lailla tehtyä tutkimusprosessia siitä saadun tiedon sijaan. Käytön väheneminen puolestaan kuvaa luultavasti tyypillisen käytön vähittäistä oppimista.

Kun kielitaito kasvaa, myös käyttötilanteet muuttuvat ja monimutkaistuvat sekä sisäiseltä rakenteeltaan että erilaisten hallittavien käyttötilanteiden määrän osalta. Tämän tutkimuksen perusteella edistyneen oppijansuomen avainilmiöt liittyvät usein tekstilajiin eli tyypilliseen tapaan käyttää kieltä tietynkaltaisessa käyttötilanteessa. Useat havaituista eroista liittyvät samanaikaisesti myös kielen kompleksisuuteen. Kuten edellä olevat esimerkit osoittavat, akateeminen kirjoittaminen sisältää runsaasti hyvin hienovaraisia merkitysvivahteita ja käytänteitä.

Valtaosa konstruktioista on sellaisia, että edistyneet oppijat kyllä sinänsä ne ymmärtävät ja osaavat ne muodostaa. Niiden käyttötilannekohtainen eli kontekstuaalinen tyypillinen käyttö kuitenkin opitaan vain kokemuksen kautta eli tutustumalla erilaisiin kielenkäytön tilanteisiin ja käyttämällä kieltä näin saatavien mallien mukaisesti. Mekanismi on siis samankaltainen kuin kielenoppimisessa ylipäätäänkin: oppiminen vaatii syötöstä eli kielellisiä malleja, joita oppija voi alkaa käyttämään ja soveltamaan. (Ks. esim. Bates & MacWhinney 1987; Tomasello 1992; Lieven, Pine & Baldwin 1997.) Kohdekielimäistä kielenkäyttöä – niin laadullisesti kuin määrällisestikin – edeltää aina kokeileminen, yritys ja erehdys. Kieltä opitaan siis käyttöpohjaisesti havainnoimalla, kokeilemalla ja mukautumalla – konteksti kerrallaan.

Lähteet

- BATES, ELIZABETH – MACWHINNEY, BRIAN 1987: Competition, variation and language learning. – Brian MacWhinney (toim.), *Mechanisms of language acquisition* s. 157–193. Hillsdale: Lawrence Erlbaum Associates.
- FORSANDER, ANNIKA – ALITOLPPA-NIITAMO, ANNE 2000: *Maahanmuuttajien työllisyminen ja työhallinto – keitä, miten ja minne?* Työhallinnon julkaisu 242. Helsinki: Työministeriö.
- HARDIE, ANDREW – MCEENERY, TONY 2010: On two traditions in corpus linguistics, and what they have in common. – *International Journal of Corpus Linguistics* 15 s. 384–394.
- HYLAND, KEN 1999: Disciplinary discourses. Writer stance in research articles. – Christopher N. Candlin & Ken Hyland (toim.), *Writing. Texts, processes and practices* s. 99–121. London: Longman.
- ISK = HAKULINEN, AULI – VILKUNA, MARIA – KORHONEN, RIITTA – KOIVISTO, VESA – HEINONEN, TARJA RIITTA – ALHO, IRJA 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura. Verkkoversio. <http://scripta.kotus.fi/visk>.
- IVASKA, ILMARI 2014a: Edistyneen oppijansuomen avainrakenteita. Korpusnäkökulma kahden kielimuodon tyypillisiin rakenteellisiin eroihin. – *Virittäjä* 118 s. 161–193.
- 2014b: Mahdollisuuden ilmaiseminen S1-suomea ja edistynyttä S2-suomea erottavana piirteenä. – *Lähivertailuja. Lähivördlusi* 24 s. 47–80.
- 2014c: The corpus of Advanced Learner Finnish (LAS2). Database and toolkit to study academic learner Finnish. – *Apples – Journal of Applied Language Studies* 8 s. 21–38.
- 2015a: Tracing crosslinguistic influences in structural sequences. What does key structure analysis have to offer? – *Learner Corpus Research: LCR2013 Conference Proceedings* s. 23–44. Bergen Language and Linguistics Studies 6. <https://bells.uib.no/bells/article/view/807> (5.10.2015).

- 2015b: Longitudinal changes in academic learner Finnish. A key structure analysis. – *International Journal of Learner Corpus Research* 1 s. 210–241.
- LIEVEN, ELENA – PINE, JULIAN – BALDWIN, GILLIAN 1997: Lexically-based learning and early grammatical development. – *Journal of Child Language* 24 s. 187–219.
- Maahanmuuton tulevaisuus 2013* = Valtioneuvoston periaatepäätös maahanmuuton tulevaisuus 2020 -strategiasta. 13.6.2013. Sisäministeriö.
http://www.intermin.fi/fi/kehittamishankkeet/maahanmuutto_2020 (14.5.2014).
- SCHULTZ-NIELSEN, MARIE-LOUISE 2000: Hvilke individuelle faktorer har betydning for integrationen på arbejdsmarkedet? – Gunnar Viby Mogensen & Poul C. Matthiessen (toim.), *Integration i Danmark omkring årtusindskiftet. Invandrerne møde med arbejdsmarkedet og velfærdssamfundet* s. 127–159. Aarhus: Aarhus Universitetsforlag.
- TOGNINI-BONELLI, ELENA 2001: *Corpus linguistics at work*. Studies in Corpus Linguistics 6. Amsterdam: John Benjamins.
- TOMASELLO, MICHAEL 1992: *First verbs. A case study of early grammatical development*. Cambridge: Cambridge University Press.

Ilmari Ivaska: *Edistyneen oppijansuomen konstruktiopiirteitä korpusvetoisesti: avainrakenneanalyysi*. Turun yliopiston julkaisuja C 409. Turku: Turun yliopisto 2015. Väitöskirja on luettavissa osoitteessa <http://urn.fi/URN:ISBN:978-951-29-6204-4>.

Kirjoittajan yhteystiedot:
etunimi.sukunimi@utu.fi