

# Kielen universaalit ja kielen monimutkaisuus

Kolme tapaustutkimusta lauseen  
ydinargumenttien merkinnästä

KAIUS SINNEMÄKI

**Väitöksenalkajaisesitelmä Helsingin yliopistossa 19. lokakuuta 2011**

Väitöskirjani käsittelee kielen monimutkaisuutta kielitypologian näkökulmasta. Kielitypologia on vertailevaa kielitiedettä, joka tutkii maailman kielten monimuotoisuutta eli niiden välisiä yhtäläisyyksiä ja eroavaisuuksia. Mielenkiinnon kohteena ovat erityisesti kielen rakenteiden keskinäiset suhteet sekä kielen rakenteen suhde ihmisen sosio-kulttuurisiin, geneettisiin ja kognitiivisiin tekijöihin. Yleistykset tehdään usein laajojen, jopa kymmeniä tai satoja kieliä kattavien otosten avulla.

Aloitin työni valmistelut kesällä 2004. Aluksi valitsemani aihe tuntui lähes mahdottomalta – olkoonkin, että kielen monimutkaisuudesta oli hiljattain alettu käydä vireää keskustelua. Erityisesti John McWhorterin (2001) artikkeli kreolikielten rakenteellisesta yksinkertaisuudesta muihin kieliin nähden oli herättänyt paljon kärkevääkin keskustelua. Tuo tutkimus, muutaman muun ohella, loi kuitenkin pohjaa ajatukselle kielen monimutkaisuuden vertailusta yleisemmällä tasolla. Lisää uskoa loi Suomen Akatemian halu rahoittaa ohjaajieni monivuotista tutkimusprojektia aihepiiristä, ja tuossa projektissa aloitinkin väitöskirjani työstämisen.

Kielen monimutkaisuus on kiinnostanut tutkijoita, voisi sanoa, jo satojen vuosien ajan. 1800-luvulle asti oli tapana arvottaa monimutkaisuus siten, että se liitettiin korkeaan kehitysasteeseen, kun taas yksinkertaisuus liitettiin primitiivisyyteen ja jopa heikompiin henkisiin kykyihin. 1900-luvun taitteessa ajat muuttuivat, ja moderni käsitys kaikkien ihmisten yhdenvertaisuudesta ja yhtäläisistä henkisistä ja kulttuurisista kyvyistä tuli ansaitusti vallalle. Tutkimus muun muassa Amerikan alkuperäiskielistä lisääntyi ja havaittiin, että kaikissa kielissä on omanlaisiaan monimutkaisia rakenteita ja eri kielten puhujilla on yhtäläiset henkiset kyvyt. Noihin aikoihin strukturalistien

parissa syntyi kaksi ajatusta kielen monimutkaisuudesta, jotka siitä lähtien ovatkin olleet vallalla.

Yhtäältä ajateltiin, että kaikki kielet ovat suurin piirtein yhtä monimutkaisia, sillä yhden osa-alueen yksinkertaisuus näytti aina kompensoituvan monimutkaisuudella jollakin toisella osa-alueella. Esimerkiksi latinan (1) kielessä on erittäin monimutkainen taivutusjärjestelmä, mutta samalla vapaa sanajärjestys; englannin kielessä (2) sitä vastoin on yksinkertainen taivutusjärjestelmä, mutta vastaavasti tiukka sanajärjestys (ks. myös taulukko 1). Tällaisen vastavuoroisen vaihtelun ajateltiin olevan läpituokeva piirre ihmiskielessä.

- (1) Octavianus Cleopatram ama-ba-t.  
 Octavianus.NOM Kleopatra.ACC rakastaa-IPFV-3SG  
 Cleopatram Octavianus amabat.  
 amabat Octavianus Cleopatram.  
 amabat Cleopatram Octavianus.  
 Cleopatram amabat Octavianus.  
 Octavianus amabat Cleopatram.  
 'Octavianus rakasti Kleopatraa.' (Gil 2008: 110–111.)

- (2) Octavian loved Cleopatra.

**Taulukko 1.**

Latinan ja englannin sijasteemit vertailussa (Blake 2001: 4).

	Latina 'herra'	Englanti 'herra'
Nominatiivi	dominus	master
Vokatiivi	domine	
Akkusatiivi	dominum	
Genetiivi	dominī	master's
Datiivi	dominō	
Ablatiivi	domino	

Toinen ajatus oli, että puheyhteisön kulttuuri ei vaikuta kielen rakenteelliseen monimutkaisuuteen mitenkään. Kuuluisin sitaatti tästä lienee Edward Sapirilta (1921: 219), joka totesi: ”Ääretön määrä sekä yksinkertaisia että monimutkaisia kielityyppejä voi esiintyä millä tahansa kulttuurin kehitystasolla.”

Nämä käsitykset kielen monimutkaisuudesta olivat selviä vastavetoja aiemmille rasistisille uskomuksille, ja ajan saatossa niistä tuli yleisesti hyväksytyjä totuuksia. Niitä ei kuitenkaan koskaan tutkittu systemaattisesti. Syy tähän lienee se, että aiheen tutkimus olisi voitu nähdä ihmisten yhdenvertaisuuden potentiaalisena kyseenalaistamisena. Tämä pohjavire on nähtävissä myös siinä kritiikissä, jota kielen monimutkaisuuden nykytutkimus on syyttä saanut osakseen: rasistin viittaa on toisinaan soviteltu, kun

tutkijoiden on oletettu toimivan 1800-luvun hengessä. Tämä oletus on perusteeton, sillä nykytutkimuksessa kielen monimutkaisuutta ei arvoteta.

Maallikolle kysymys kielten monimutkaisuudesta lienee selvempi, ainakin aluksi: totta kai kielten välillä on eroja. Englannin kieli tuntuu saksan rinnalla säännöllisemmältä ja siten myös helpommalta oppia. Niin, ja totta kai suomi on yksi maailman vaikeimmista kielistä, sanotaan, koska ulkomaalaisten tuntuu olevan niin vaikea oppia sitä. Tällaisissa arvioissa meiltä jää herkästi huomaamatta kolme tärkeää seikkaa.

Ensinnäkin oppimisen vaikeus on subjektiivinen kokemus ja siihen vaikuttavat eritoten kielen oppijan kielitausta sekä kielten samankaltaisuus. Englanninpuhujan on helpompi oppia saksaa kuin suomea siitä yksinkertaisesta syystä, että englanti ja saksa ovat lähisukukieliä ja näin ollen niissä on paljon samankaltaisuuksia.

Toiseksi saatamme salaa hykerrellä suomen kielen oletetulla vaikeudella, koska sillä voi tarvittaessa pönkittää omanarvontuntoamme: meissä ja kielessämme täytyy olla jotain aivan poikkeuksellista kielemme vaikeuden tähden. Itseämme pönkittäessä tosin arvotamme vaikeuden käsitteen ja toistamme jo aiemmin tehtyjä virheitä.

Kolmas seikka liittyy rakenteen monimutkaisuuden ja kielenkäytön vaikeuden suhteeseen. Oletettavasti mitä monimutkaisempi esimerkiksi kielen taivutusjärjestelmä on, sitä vaikeampi se todennäköisesti on myös omaksua. Kysymys on tosin tätä moniulotteisempi, sillä se, mikä on äidinkielen oppijalle helppoa, voi olla toisen kielen oppijalle vaikeaa. Otetaanpa vaikka suomen kielen kongruenssi. Suomessa sanotaan *pienessä punaisessa talossa, ei pieni punainen talossa*: inessiivin tunnus *-ssa* ilmaistaan sekä pääsanassa *talo* että sen määreissä. Psykolingvistisen tutkimuksen mukaan tällainen redundanti kongruenssi on hankalaa toisen kielen oppijalle, mutta sitä vastoin hyödyllistä äidinkielen oppijalle (ks. Kusters 2003). Saman rakenteen omaksuminen voi siis vaihdella kielenoppijan tilanteesta riippuen, minkä vuoksi on syytä pitää kielen monimutkaisuus erillään kielenkäytön vaikeudesta, kuten olen tässä työssä tehnyt.

Kun siis puhutaan kielen monimutkaisuudesta, ollaan sellaisen aiheen äärellä, joka tulee lähelle jokaista kielenkäyttäjää, joka voi herättää suuria tunteita, jolla on eettisesti arveluttava historia ja joka on herkästi leimahtava aihe kielentutkijoiden parissa. Aiheen tieteellinen tutkimus on lisäksi leimattu pahamaineisen vaikeaksi, ja siihen ovat ylipäätään kajonneet lähinnä urallaan pitkälle edenneet tutkijat. Kaikkine ristiriitaisuuksineen tutkimuskohde on siis suorastaan fantastinen!

On selvää, että tällaista taustaa vasten erityisesti nuoren tutkijan on syytä luovia tiensä tarkasti. Minä valitsin tieni seuraavasti. Väitöskirjallani on kaksi päätavoitetta, teoreettis-metodologinen ja empiirinen. Teoreettis-metodologinen tavoitteeni koskee kielen monimutkaisuuden käsitettä ja sen valjastamista maailman kielten vertailuun; tätä tavoitetta käsittelen erityisesti artikkeliväitöskirjani yhteenveto-osassa. Empiirinen tavoitteeni on tutkia kielen monimutkaisuutta kolmen tapaustutkimuksen avulla; tätä tavoitetta käsittelen väitöskirjani artikkeleissa. Empiiriset kysymykseni tarkastelevat sitä, missä määrin kielen eri osa-alueiden monimutkaisuus vaihtelee systemaattisesti ja missä määrin yhteisön sosiaaliset tekijät voivat vaikuttaa kielen monimutkaisuuteen.

Koska tavoitteeni on lähestyä kielen monimutkaisuutta typologisesti, lähestymistapani tulisi soveltua monimutkaisuuden vertailuun maailman kielissä. Tämä näkökulma tuo omat rajoitteensa. Yksi tärkeä rajoite on, että tutkin kielen monimutkai-

suutta paikallisesti, esimerkiksi sijasysteemin monimutkaisuutta, sen sijaan että tutkisin kielen kokonaiskompleksisuutta eli kielen kaikkien osa-alueiden – muun muassa fonologian, morfologian ja syntaksin – ja niiden välisten suhteiden monimutkaisuutta yhdessä. Väitän, että kielen kokonaiskompleksisuuden tutkimus on mahdotonta, mutta paikallisen monimutkaisuuden tutkimus on sitä vastoin kiintoisaa ja myös tuloksetlista.

Määritän monimutkaisuuden samoin kuin kompleksisuustieteissä yleensä: se tarkoittaa **osien ja niiden välisten suhteiden määrää ja moninaisuutta**. Puran tämän yleisen käsitteen eri alalajeihin filosofi Nicholas Rescherin (1998) luokittelun mukaan ja nivon kielitieteessä käytetyt monimutkaisuuden määritelmät tähän yleisempään viitekehukseen. Tämä strategia selkiyttää kielen monimutkaisuuden käsitettä.

Monimutkaisuuden määritelmä voidaan tehdä yleisemmin sovellettavaksi informaatioteorian avulla. Informaatioteoriassa tärkeä monimutkaisuuden käsite on kuvauksen pituus. Perusajatus kuvauksen pituudessa on, että mitä pidemmän kuvauksen tarkasteltava kohde vaatii, sitä monimutkaisempi se on. Kuvauksen pituutta mitatessani keskityn kohteen rakenteen kuvaukseen eli ns. teholliseen monimutkaisuuteen (Gell-Mann 1995). Näin määriteltynä monimutkaisuus sijoittuu järjestyksen ja epäjärjestyksen välimaastoon ja vastaa parhaiten sitä, mitä monimutkaisuudella yleensä intuitiivisesti ymmärretään.

Kuvauksen pituus on tosin sikäli ongelmallinen, että siihen vaikuttaa käytetty kuvauskieli, jolloin kohteen monimutkaisuus riippuu kuvauskielestä. Ongelma on kuitenkin ohitettavissa, kunhan tarkasteltavien kohteiden monimutkaisuuserot ovat suuria. Kielentutkimuksessa ilmimerkinnän esiintymisen tarkastelu sopii tähän tarkoitukseen hyvin. Esimerkiksi substantiiviobjektin sijanmerkintä on monimutkaisempi latinassa (1) kuin englannissa (2), koska latina käyttää erillistä akkusatiivisijaa, englanti ei. Tällaisen monimutkaisuuseron kuvauksessa kuvauskielellä ei ole juuri merkitystä. Tällaisella kuvaustasolla kielen monimutkaisuutta tarkastellaan lisäksi melko yleisellä tasolla, mutta tällöin voidaan ja on myös hyödyllisempää käyttää kielentutkijoiden kehittämää kuvaustyökäluja kuin esimerkiksi matemaatikoiden työkaluja.

Väitän, että tällaisen teoreettis-metodologisen lähestymistavan avulla kielen monimutkaisuutta voidaan tarkastella kieltenvälisesti, koska fokus on tietystä rakenteellisen monimutkaisuuden alalajissa tietystä paikallisesta kontekstissa. Paikallisen kontekstin voi sitten määrittää eri tavoin, kunhan se on kieltenvälisesti vertailtavissa. Itse määritän kontekstin funktionaalisesti, mikä on tyypillistä käyttämälläni funktionaalis-typologiselle viitekehykselle.

Testaan malliani vain yhdellä funktionaalisella alueella rajaten tutkimukseni lauseen pääjäsenten eli tekijän/subjektin ja teon kohteen/objektin merkintään. Tekijän ja teon kohteen erottelussa käytetään kolmea muotokeinoa: sijanmerkintää, kongruenssia ja sanajärjystä. Sijanmerkintää käytetään muun muassa latinassa (1) ja sanajärjystä englannissa (2). Kongruenssi tarkoittaa, että tekijä tai teon kohde on koodattu myös verbiin (esim. tekijä latinassa).

Tutkin lauseen pääjäsenten erottelussa esiintyvää monimutkaisuuden vaihtelua kolmen tapaustutkimuksen avulla. Ensimmäisessä artikkelissa tutkin sijanmerkinnän, kongruenssin ja sanajärjestyksen monimutkaisuutta ja niitä mahdollisesti tasa-

painottavia tendenssejä. Mittaan sijanmerkinnän ja kongruenssin monimutkaisuutta suhteessa niiden ilmimerkintään ja sanajärjestyksen monimutkaisuutta suhteessa sen rajoituksiin. Aineisto tulee 50 kielen otoksesta ja perustuu lähinnä referenssi-kielipeista saatuun materiaaliin. Tuloksen mukaan sijanmerkinnän monimutkaisuus korreloi käänteisesti sanajärjestyksen monimutkaisuuden kanssa. Toisin sanoen: **vähän sijanmerkintää – rajoitettu sanajärjestys, paljon sijanmerkintää – vapaa sanajärjestys.** Tämä tulos todentaa aiempia alustavia käsityksiä ja antaa viitteitä sellaisista prosessointimekanismeista, jotka saattavat rajoittaa kielen monimutkaisuutta tietyllä alueella.

Toisessa artikkelissani tutkin lauseen pääjäsenten merkinnän monimutkaisuutta suhteessa puheyhteisön kokoon. Mittaan kielen monimutkaisuutta yksi merkitys – yksi muoto -periaatteen avulla; pitäytyminen siihen on yksinkertaista, mutta poikkeamat siitä lisäävät kuvauksen pituutta. Aineisto tulee 50 kielen otoksesta ja sen perusteella pienissä kielissä on lähes 13 kertaa suurempi todennäköisyys käyttää monimutkaisempaa tekijän ja teon kohteen erottelutapaa kuin suurissa kielissä. Toisin sanoen: **suurempi puhujayhteisö – yksinkertaisempi merkintä, pienempi puhujayhteisö – monimutkaisempi merkintä.** Tulos antaa viitteitä siitä, että kielen rakenne sopeutuu sosiaaliseen ympäristöön, vastoin yleisiä käsityksiä. Syy tähän sopeutumiseen lienee se, että suuret kielet vetävät puoleensa suhteessa enemmän sitä toisena kielenä oppivia puhujia, joiden on useissa tutkimuksissa osoitettu suosivan läpinäkyviä ja yksinkertaisia rakenteita (esim. Kusters 2003). Puheyhteisön koko toimii tässä heuristina sosiaaliselle tyypille yleisemminkin.

Kolmannessa artikkelissani käsittelem nollamerkinnän ja sanajärjestyksen välistä suhdetta. Nollamerkinnällä tarkoitan sitä, että kieli ei käytä tekijän ja teon kohteen erottelussa sijanmerkintää tai kongruenssia, kuten esim. monet Kaakkois-Aasian kielet. Aiempien alustavien käsitysten mukaan nollamerkintää esiintyy erityisesti kielissä, joilla on subjekti-verbi-objekti (SVO) -sanajärjestys. Esittämäni mallin mukaan nollamerkintä on yksinkertaisempaa kuin ilmimerkintä ja SVO-sanajärjestys on yksinkertaisempi kuin muut sanajärjestykset.

Tästä artikkelista muodostui tapaustutkimuksistani teknisesti vaativin. Nollamerkinnän jakauma oli niin vinoutunut sekä genealogisesti että alueellisesti, että oli vaikea löytää riittävän luotettavaa tapaa todentaa tai kumota sitä hypoteesia, että nollamerkintä korreloisi SVO-sanajärjestyksen kanssa. Tämä riippumatta siitä, että aineistossani on noin 850 kieltä ja se kattaa 95 % sellaisista alemman luokittelutason kieliperheistä, joihin arvovaltainen *World Atlas of Language Structures* -teos (Haspelmath ym. 2005) jakaa maailman kielet. Vinoutunut jakauma saattoi olla merkki siitä, että nollamerkinnän ja SVO-sanajärjestyksen oletettu korrelaatio ei olisikaan ollut riippumaton genealogisesta periytymisestä tai kielikontaktien aiheuttamasta alueellisesta diffuusiosta. Jos näin olisi ollut, oletettu universaali korrelaatio olisikin ollut vain paikallinen tendenssi. Ongelmaan löytyi lopulta ratkaisu Balthasar Bickelin (2008) kehittämästä menetelmästä mallintaa typologisia jakaumia tilastollisten monimuuttujamenetelmien, erityisesti usean muuttujan logistisen regression avulla. Käytin lisäksi hänen tapansa soveltaa Joseph Greenbergin (1978) ideaa kielen universaaleista diakronisina preferensseinä.

Riippumatta siitä, millä tavoin mallinsin kielen universaaleja, tulos oli sama: **morfologinen nollamerkintä on monin verroin todennäköisempää SVO-kielissä kuin kielissä, joilla on jokin muu sanajärjestys.** Tämä tulos todensi sen, mikä oli aiemmin vain alustava käsitys. Selitän tämän tuloksen funktionaalisten motivaatioiden avulla: nollamerkintää käyttävissä kielissä SVO-sanajärjestys on sekä ekonomisin että ikonisin vaihtoehto ja lisäksi helpompi prosessoida kuin muut sanajärjestykset.

Väitöskirjani päätavoite on ollut luoda sellainen lähestymistapa kielen monimutkaisuuteen, jota voitaisiin soveltaa kieltenväliseen tutkimukseen laajemminkin. Testasin malliani kolmen empiirisen tapaustutkimuksen avulla, ja tulokset osoittivat mallini käyttökelpoisuuden. Tämän vuoksi uskon, että ehdottamaani lähestymistapaa voidaan soveltaa myös muiden kieliopillisten ilmiöiden typologiseen tutkimukseen. Kieltenvälinen tutkimus vaatii lisäksi käsitteiden selkeää määrittelyä, ja tästä selkeydestä on hyötystä myös soveltavalle tutkimukselle, erityisesti kielenoppimisen tutkimukselle.

Kielen monimutkaisuuden tutkimusta on kritisoitu siitä, että se ei olisi ylipäättään relevanttia kielitypologialle. Väitän vastaan. Tulokseni osoittavat, että kielen monimutkaisuuden tutkimus liittyy suoraan kielitypologian perimmäisiin tavoitteisiin eli kielen rakenteiden keskinäisten suhteiden sekä kielen rakenteen ja kognitiivisten, geneettisten ja sosiaalisten tekijöiden välisen suhteen tutkimukseen. Voisin jopa väittää, että heti kun typologit ovat ottaneet kielen monimutkaisuuden tutkimuksen tosissaan, on löydetty tuloksia, jotka ovat täysin linjassa funktionaalis-typologisen tutkimuksen perusoletusten ja päätavoitteiden kanssa. Kaiken kaikkiaan kielen monimutkaisuus on kiintoisa, haasteellinen ja varsin produktiivinen tutkimuskohde.

## Lähteet

- BICKEL, BALTHASAR 2008: A general method for the statistical evaluation of typological distributions. <http://www.spw.uzh.ch/bickel-files/papers> (1.12.2011).
- BLAKE, BARRY 2001: *Case*. 2. painos. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.
- GIL, DAVID 2008: How complex are isolating languages. – Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (toim.), *Language complexity. Typology, contact, change* s. 109–131. Studies in Language Companion Series 94. Amsterdam: John Benjamins.
- GELL-MANN, MURRAY 1995: What is complexity? – *Complexity* 1 s. 16–19.
- GREENBERG, JOSEPH 1978: Diachrony, synchrony and language universals. – Joseph Greenberg, Charles Ferguson & Edith Moravcsik (toim.), *Universals of human language 1. Method and theory* s. 61–92. Stanford: Stanford University Press.
- HASPELMATH, MARTIN – DRYER, MATTHEW S. – GIL, DAVID – COMRIE, BERNARD (toim.) 2005: *The world atlas of language structures*. Oxford: Oxford University Press.
- KUSTERS, WOUTER 2003: *Linguistic complexity. The influence of social change on verbal inflection*. LOT Dissertation Series 77. Leiden: University of Leiden.
- MCWHORTER, JOHN 2001: The world's simplest grammars are creole grammars. – *Linguistic Typology* 5 s. 125–156.
- RESCHER, NICHOLAS 1998: *Complexity. A philosophical overview*. New Brunswick: Transaction.

SAPIR, EDWARD 1921: *Language. An introduction to the study of speech*. New York: Harcourt, Brace and Company.

**Kaius Sinnemäki:** *Language universals and linguistic complexity. Three case studies in core argument marking*. Helsinki: Helsingin yliopiston nykykielten laitos 2011. <http://urn.fi/URN:ISBN:978-952-10-7259-8>.

Kirjoittajan yhteystiedot:  
sukunimi@gmail.com