

Matti Miestamo

Kielten vertailu kielitypologisessa tutkimuksessa

1 Johdanto

Kielitypologiaa voidaan luonnehtia maailmanlaajuiseksi vertailevaksi kielitieteeksi tai teoreettiseksi kielitieteeksi, jonka empiirinen perusta on maailmanlaajuisessa kielten vertailussa. Typologia pyrkii siis selvittämään ihmiskielen ominaisuuksia ja olemusta kieltenvälisen vertailun pohjalta. Typologista tutkimusta tehdään pääsääntöisesti funktionalistisessa viitekehyksessä, eli kielten rakenteita tarkastellaan suhteessa kielen käyttöön ja merkitykseen, ja näiden nähdään myös pitkälti selittävän kielten rakenteellisia ominaisuuksia. Kielten vertailtavuus ja vertailu ovat keskeisiä teoreettis-metodologisia kysymyksiä kielitypologille, ja tämän luvun tarkoituksena on pohtia näitä kysymyksiä typologisen tutkimuksen näkökulmasta. Valotan aiheesta viime aikoina kirjallisuudessa käytyä keskustelua ja havainnollistan esillä olevia kysymyksiä muun muassa oman negatiota koskevan tutkimukseni kautta.

Typologisen tutkimuksen tyypilliset vaiheet voidaan tiivistää seuraavasti (ks. Stassen 1985: 1–23; Miestamo 2005: 26–50).

I. Aineiston koonti

- Tutkimuskohteen määrittely
- Otoksen valinta
- Datan kerääminen

II. Aineiston analyysi ja yleistysten tekeminen

- Typologinen luokittelu
- Tyyppien yleisyys ja maantieteellinen jakautuminen
- Kielenpiirteiden väliset korrelaatiot

III. Selitysten/motivaatioiden etsiminen

Tällainen esitystapa on tietysti abstraktio ja yksinkertaistus. Tutkimuksen eri vaiheet riippuvat toisistaan ja ovat todellisuudessa osittain päällekkäisiä. Esimerkiksi luokitteluvaiheesta voidaan joutua monta kertaa palaamaan datan keräämiseen, jos analyysin edetessä esiin nousevat lisäkysymykset vaativat tarkempia tietoja kielen ominaisuuksista. Tämän luvun dispositio seurailee typologisen tutkimuksen kulkua painottaen kuitenkin alun metodologisia vaiheita, erityisesti tutkimuskohteen määrittelyä, jossa kysymykset vertailtavuudesta ja vertailusta nousevat selvimmin esiin. Jaksossa kaksi paneudun siis tutkimuskohteen kieltenväliseen määrittelyyn. Kysymys kielen vertailtavuudesta on suorassa yhteydessä teoreettiseen näkemykseen kielen kategorioiden universaaliudesta vs. kielikohtaisuudesta; tätä kysymystä pohdin jaksossa 2.1. ja varsinaista metodologista kysymystä kieltenvälisessä vertailussa käytettävistä määritelmistä jaksossa 2.2. Jaksossa kolme käsittelen vertailtavien kielen valintaa eli typologisia otantamenetelmiä sekä typologin vertailussaan käyttämiä lähteitä ja aineistoon liittyviä erityiskysymyksiä. Jakso neljä kokoaa yhteen tutkimuksen loput vaiheet luokittelusta selittämiseen ja pohtii näihin liittyviä kysymyksiä vertailun kannalta. Jakso viisi on artikkelin yhteenveto.

2 Kielten vertailtavuus ja tutkimuskohteen kieltenvälinen määrittely

Kielten maailmanlaajuisessa vertailussa tutkimuskohde on määriteltävä siten, että määritelmä on sovellettavissa kaikkiin kieliin. Kieltenvälisen identifikaation ongelma – se miten relevantissa mielessä sama ilmiö voidaan tunnistaa erilaisissa kielissä – on keskeinen metodologinen kysymys kielitypologisessa tutkimuksessa. Teoreettinen näkemys kielipiillisten kategorioiden (kuten esim. subjekti, objekti, substantiivi, verbi, nominatiivi, dativi, vertailukonstruktio tai eksistentiaalilause) kieltenvälisestä statuksesta vaikuttaa ratkaisevasti siihen, miten kieliä voidaan verrata keskenään – ja voidaanko niitä ylipäätään verrata. Universalistisen kategoriakäsityksen (*universal categorialism*) mukaan on olemassa kaikille kielille yhteisiä – universaaleja – kieltenvälisiä kategorioita ja kielikohtaiset kategoriat ovat näiden universaalien kategorioiden erilaisia ilmenemisiä. Vastakkaisen, partikularistisen kategorianäkemyksen (*universal particularism*) mukaan universaaleja kielellisiä kategorioita ei ole olemassa, ja kunkin kielen kategoriat ovat puhtaasti kielikohtaisia. Kieltenvälisen vertailun metodologiset valinnat on rakennettava teoreettiselle käsitykselle kielellisten kategorioiden kieltenvälisestä luonteesta. Siksi käsittelen ensin kategorioiden luonteeseen liittyviä kysymyksiä jaksossa 2.1. ja tämän jälkeen itse

tutkimuskohteen määrittelyä jaksossa 2.2. (Samankaltaisia kysymyksiä kontrastiivisen kielitieteen kannalta pohtii Järventausta tässä niteessä, erityisesti jaksossa 2.)

2.1. Kielellisten kategorioiden kieltenvälisestä luonteesta

Universalistinen kategorianäkemyks periytyy länsimaisessa kielitieteessä jo antiikin ajoilta. Myöhäiskeskiajalla toiminut modistien nimellä tunnettu filosofinen suuntaus loi universaalikielioppimallin, joka kuitenkin käytännössä perustui latinan kielen kategorioihin, ja Ranskassa 1600-luvulla vaikuttanut Port Royal -kielioppikoulukunta taas perusti oman universaalin kielioppinsa ranskan kieleen (ks. Itkonen 1991). Eri kielten kielioppeja kirjoitettiin latinan mallin mukaan vielä 1800-luvulle asti. 1800-luvun lopulla ja 1900-luvun alussa noussut strukturalistinen suuntaus otti päinvastaisen, partikularistisen kannan. Etenkin amerikkalaisessa strukturalismissa, jossa teorianmuodostusta ruokki Amerikan alkuperäiskielten tutkimus, partikularistinen näkemys oli vahvoilla. 1900-luvun jälkipuoliskolla kielitieteilijöiden valtavirta kallistui taas universalistiseen kategoriakäsitykseen – erityisesti chomskylaisessa generatiivisessa kielitieteessä. Viime vuosina partikularistinen näkemys on vuorostaan vallannut alaa, varsinkin kielitypologien ja muiden funktionalistisesti suuntautuneiden kielentutkijoiden parissa.

Universalistisen kategorianäkemyksen mukaan on siis olemassa kaikille kielille yhteisiä kieltenvälisiä kategorioita, jotka ovat ontologisesti tai psykologisesti tosia, ja kielikohtaiset kategoriat ovat näiden universaalien kategorioiden ilmentymiä. Kieltenväliset kategoriat, kuten substantiivi, verbi, adjektiivi, subjekti ja objekti, saavat erilaisia ilmiäsuja eri kielissä, mutta kyseessä on aina saman universaalien kategorian kielikohtainen edustaja.

Universalistinen kategoriakäsitys on vallalla oleva näkemys monessa nykykielitieteen haarassa. Esimerkiksi chomskylaisessa generativismissa kielikyvyn ajatellaan olevan synnynnäinen, ja universaalikieliopin kategoriat ovat siten biologisesti määräytyneitä. *Principles and parameters* -lähestymistavassa (ks. Chomsky & Lasnik 1993) universaalikielioppi koostuu universaaleista periaatteista ja kielikohtaisista parametreista. Universaalit kategoriat ovat siis näiden universaalien periaatteiden määrittämiä, ja niiden kielikohtaiset ilmenemät riippuvat kunkin kielen parametrisetuksista. Teorian mukaan synnynnäinen universaalikielioppi selittää eri kielissä havaitut rakenteet, joten universaalit kategoriat ovat oleellinen osa generativisista kielikäsitystä. Chomskylaisen generativismin kaltainen formaali kielioppiteoria ei tule toimeen ilman universaaleja kategorioita, mutta universalistinen kategoriakäsitys ei kuitenkaan rajoitu vain formaaliin, synnynnäistä

kielikykyä olettaisiin teorioihin. Myös monet funktionalistisemmin suuntautuneet tutkijat ja tyypologit ovat universalistisen kategorianäkemyksen kannalla, esim. Lehmann (1989), joka korostaa, että yksittäiskielten kuvauksessa käytettävien kategorioiden on oltava kieltenvälisesti valideja.

Partikularistisen kategorianäkemyksen mukaan kieltenvälisiä kategorioita ei ole olemassa, vaan kategoriat ovat aina kielikohtaisia. 1900-luvun alussa noussut strukturalistinen suuntaus, jonka tunnetuin edustaja oli Ferdinand de Saussure, korosti kunkin kielen kategorioiden luonnetta osana juuri sen kielen systeemiä. Kategorioiden rajat määrittyvät suhteessa muihin kielisysteemin osasiin, ja nämä suhteet ovat kielikohtaisia. Partikularistinen kategorianäkemys oli erityisen vahva amerikkalaisessa strukturalismissa, jossa Franz Boasin alulle panemassa antropologisessa lingvistiikassa harjoitettu Amerikan alkuperäiskielten tutkimus oli keskeisessä osassa. Näiden kielten moninaisuus ja erilaisuus suhteessa tuttuihin eurooppalaisiin kieliin sai monet strukturalistit näkemään universaalien kategorioiden olemassaolon mahdottomana. Joos (1957) kiteytti amerikkalaisen strukturalismin näkemyksen sanoihin ”Kielet voivat erota toisistaan rajattomasti ja ennustamattomasti”.¹ Näin radikaali näkemys kielten yhteismitattomuudesta tekee tietysti kielten vertailun mahdottomaksi ja kielitypologisen tai yleislingvistisen tutkimuksen merkityksettömäksi. Jos oletetaan lähtökohtaisesti, ettei ole olemassa mitään kielten erilaisuutta ja samuutta sääteleviä tekijöitä, ei kieltenvälisiin yleistyksiin pyrkivää kielitiedettäkään ole mieltä harjoittaa. Kielitypologia ei kuulunutkaan pääsääntöisesti 1900-luvun alun strukturalistien kiinnostuksen kohteisiin (poikkeuksena kuitenkin esim. Edward Sapir). Vuosisadan jälkipuoliskolla näkemykset muuttuivat, ja kieltenvälisen yhteyksien etsimisessä alettiin taas nähdä järkeä. Monet funktionaalisesti suuntautuneet nykytyypologit ovat kuitenkin partikularistisen näkemyksen kannalla (esim. Lazard 1992, 2005; Dryer 1997; Croft 2001; Gil 2008; Haspelmath 2009, 2010). Kuinka kielten vertailu sitten on mahdollista tai mielekästä tällaisesta teoreettisesta positiosta käsin? Niihin metodologisiin työkaluihin, jotka mahdollistavat kieltenvälisen vertailun myös tätä teoreettista taustaa vasten, paneudun jaksossa 2.2.

Partikularistisen kategoriakäsityksen mukaan jokaisen kielen kategoriat ovat siis kielikohtaisia, eikä niiden taustalla ole ontologisesti tai psykologisesti tosia kieltenvälisiä kategorioita. Dryeriä (1997: 116) mukailten voisimme olettaa seuraavankaltaisten ilmiöiden olevan olemassa kielessä:

- a. Kielellisiä kategorioita yksittäisissä kielissä
- b. Samankaltaisuuksia näiden yksittäiskielten kategorioiden välillä

- c. Funktionaalisia selityksiä näille samankaltaisuuksille
- d. Kielellisiä kategorioita kieltenvälisessä mielessä.

Partikularistinen kategoriakäsitys olettaa kolmen ensimmäisen (a–c) olemassaolon, mutta ei neljättä (d). Jokaisen kielen kategoriat eroavat toisistaan enemmän tai vähemmän. Minkään kielen kategoriat eivät ole identtisiä ominaisuuksiltaan ja rajoiltaan. Esimerkiksi suomen ja ruotsin subjekteilla on kullakin omat syntaktiset ja morfologiset ominaisuutensa, jotka eivät yksi yhteen vastaa toisiaan. Kieltenvälinen vertailu paljastaa samankaltaisuuksia eri kielten kategorioiden välillä, mutta nämä samankaltaisuudet eivät anna perusteita kieltenvälisen kategorioiden olettamiseen. Samankaltaisuudet selittyvät funktionaalisen (semanttisen, pragmaattisen, kognitiivisen) tason yhteyksistä kielten välillä: eri kielillä on samanlaisia merkityksiä välitettävänä, niitä käytetään samankaltaisissa kommunikaatiotilanteissa ja kielenkäyttäjät on varustettu samanlaisilla kognitiivisilla kyvyillä.

Dryerin (1997: 123) mukaan ainoa syy kieltenvälisen kategorioiden olettamiseen on itse asiassa se, että joissakin teorioissa, esim. chomskylaisessa generativismissa, tarvitaan universaaleja kategorioita selittämään yksittäiskielten ominaisuuksia. Funktionalistisesti suuntautunut kielitiede ei tarvitse universaaleja kategorioita, koska kielten ilmiöitä voidaan selittää funktionaalisten ja kognitiivisten periaatteiden avulla. Nähdäkseni psykologisesti tosien kieltenvälisen kategorioiden olettaminen edellyttää näkemystä, jonka mukaan nämä kategoriat ovat synnynnäisiä, siis esimerkiksi jonkinlaista chomskylaista universaalikielioppia, joka on biologisesti määrätty. Kuinka muuten eri kieliyhteisöissä kasvavat lapset voisivat päätyä samoihin universaaleihin kategorioihin? Funktionalistisesti suuntautunut kielitieteilijä, jonka mukaan kielen rakenne on muotoutunut ja muotoutuu kielen käytössä, ei tarvitse eikä myöskään näe mahdollisena eri kielten puhujien jakamia universaaleja kategorioita. Entä jos kieltenväliset kategoriat nähtäisiin ei-diskreetteinä prototyyppien ympärille rakentuvina kategorioina, kuten esimerkiksi merkityskategoriat kognitiivisessa ja funktionalistisessa kielitieteessä usein nähdään? Prototyyppiajattelu ei tekisi kieltenvälisistä kategorioista yhtään sen sopivampia funktionalistiseen ajattelutapaan. Myös niiden täytyisi olla synnynnäisiä, jotta ne voisivat olla kielten välillä jaettuina. Kielten taustalla on monimutkainen funktionaalisten ja kognitiivisten tekijöiden verkosto. Eri tekijät motivoivat erityyppisiä kielenmuutoksia, ja erilaisissa kulttuuris-sosiaalisissa konteksteissa elävät kielet kehittyvät eri tavoin näiden kilpailevien motivaatioiden vaikutuksessa. Tuloksena on paitsi (hyvinkin) erilaisia kielikohtaisia kategorioita eri kielissä, myös samanlaisista

kognitiivis-funktionaalista lähtökohdista johtuvia samankaltaisuuksia eri kielten kielikohtaisten kategorioiden välillä. (Ks. Dryer 1997: 134–135)

Jos kategoriat ovat kielikohtaisia, nousee tietysti kysymys, missä kulkee kielen raja – koska ollaan tekemisissä yhden ja saman kielen eri varianttien kanssa ja koska on kyse kahdesta eri kielestä? Äärimmilleen vietyinä tämä rajanteko voidaan ulottaa kahden eri henkilön puhuman kielimuodon, idiolektin, välille ja kysyä, ovatko kielelliset kategoriat idiolektikohtaisia (vrt. Newmeyer 2007: 146).ⁱⁱ Jos verrataan kahta idiolektiä ja todetaan, että niiden välillä on ero jonkin kieliopillisen kategorian ominaisuuksissa, niin näiltä osin on todettava, että kategoriat eroavat idiolektistä toiseen. Toisilleen sukua olevien kielimuotojen välillä – idiolekteistä eri sosiolekteihin ja murteisiin – kategorioita voi nähdäkseni pitää samoina sikäli kun ne ovat ominaisuuksiltaan samanlaisia. Yhdistävänä tekijänä ovat samasta historiallisesta taustasta nousevat konventiot. Mitä kauempana toisistaan tällaiset kielimuodot ovat, sitä vähemmän samuutta niiden kategorioiden välillä voidaan havaita. Kielen ja murteen eron jatkumoluonne tulee siis tässäkin esiin.

Edellä kirjoitin oletuksen universaaleista kieltenvälisistä kategorioista olevan olennainen tietyissä teoreettisissa lähestymistavoissa kuten chomskylaisessa generativismissa. Teorianmuodostus pohjautuu suurten indoeurooppalaisten kielten, etupäässä englannin, ominaisuuksille, joten teorioiden universaalit kategoriat heijastelevat pääasiassa englannin kielen ominaisuuksia. Teorian pohjana olevien indoeurooppalaisten kielten tutkijoiden onkin helpointa hyväksyä tällaisten teorioiden universaalit kategoriat, jotka eivät ole suuressa ristiriidassa tutkittavista kielistä löytyvien ilmiöiden kanssa. On ymmärrettävää, että partikularistinen kategorianäkemyks on yleisempi monien erilaisten kielten kanssa tekemisissä olevien lingvistien keskuudessa – 1900-luvun alkupuolen amerikkalaisten strukturalistien ja sittemmin kielitypologien. Kuten Haspelmath (2007: 122) toteaa, lähes jokainen uusi kieli, joka saadaan dokumentoiduksi, tarjoaa jotakin uutta ja yllättävää: kategorioita, joiden olemassaolon mahdollisuudesta kielitieteilijöillä ei ennen kielen kuvausta ollut käsitystä. Otan seuraavassa esiin joitakin esimerkkejä siitä, kuinka erilaisia kieliopilliset kategoriat voivat kielten välillä olla.

Kielet voivat pilkkoa jonkin funktionaalisen alueen pieniin osiin, kuten vaikkapa monet afrikkalaiset kielet metrisissä tempusjärjestelmissään, joissa erotetaan useita menneen ajan aikamuotokategorioita sen mukaan, kuinka läheiseen tai kaukaiseen menneeseen tapahtumaan viitataan, tai sitten käyttää laajempimerkityksistä kategoriaa kattamaan koko tämä alue, kuten vaikka suomen preteriti,ⁱⁱⁱ joka voi viitata menneeseen aikaan riippumatta sen läheisyydestä tai kaukaisuudesta. Tai sitten kieli voi jättää kieliopillistamatta koko funktionaalisen alueen,

esim. tempuksen, ja viitata samoihin asiantiloihin esimerkiksi moduksen kautta, jolloin menneeseen aikaa viitataan tyypillisesti reaalilla moduksilla ja tulevaisuuteen irreaalisilla moduksilla. Peruslauseen osallistujien merkintään kielissä voi olla hyvin erilaisia kategorioita sisältäviä järjestelmiä: esimerkiksi nominatiivi-akkusatiivisysteemeissä transitiivilauseen agenttia (1. tekijää) ja intransitiivilauseen ainoaa argumenttia merkitään samalla tavalla (esim. saksan nominatiivilla) ja transitiivilauseen patienttia (1. teon kohdetta) eri tavalla (esim. saksan akkusatiivilla), absoluutiivi-ergatiivisysteemeissä transitiivilauseen patienttia ja intransitiivilauseen ainoaa argumenttia merkitään samalla tavalla (esim. baskin absoluutiivisijalla) ja transitiivilauseen agenttia eri tavalla (esim. baskin ergatiivisijalla), ja aktiivi-statiivisysteemeissä argumenttien merkintä heijastelee suoraan niiden semanttista roolia, niin että intransitiivilauseen ainoaa jäsentä merkitään joko transitiivilauseen agentin tai patientin tavoin sen mukaan, onko se lauseen semantiikassa enemmän agentin vai patientin roolissa (esimerkiksi Kaliforniassa puhutussa keskipomon kielessä on agentti- ja patienttiprotonimit, joita käytetään sekä transitiivisissa että intransitiivisissa lauseissa argumenttien semanttisten roolien mukaisesti).

Äärimmäisenä esimerkkinä siitä, miten kielen kategoriat voivat poiketa totutuista tavoista jäsentää lauseita, voidaan mainita Gilin (2008) analyysi riianindonesiasta: Kielessä ei ole tämän mukaan kuin yksi yleinen sanaluokka ja yksi yleinen syntaktinen konstruktio, jossa sanojen asettaminen vierekkäin ilmaisee, että niiden välillä on jokin semanttinen yhteys. Mitään substantiivin tai verbin kaltaisia sanaluokkia tai subjektin tai objektin kaltaisia lauseenjäsenkategorioita ei ole. Niinpä kanaa ja syömistä tarkoittavien sanojen yhdistelmä *ayam makan* voi tarkoittaa kontekstista riippuen esim. 'syön kanaa', 'kana, joka syö', 'mennään syömään kanaa', 'kananruoka' tai 'kanaravintola'. Yhden sanaluokan ja yhden yleisen konstruktion lisäksi kielessä on vain joitakin kieliopillisia elementtejä, joilla merkityksiä voidaan tarkentaa.

Kun oletetaan kategorioiden olevan kielikohtaisia, niiden nimeäminen on myös kielikohtainen kysymys. Kysymys siitä, nimitetäänkö jonkin kielen ominaisuussanoja sisältävää sanaluokkakategoriaa adjektiiviksi, on viime kädessä puhtaasti terminologinen. Jotkut tutkijat, varsinkin amerikkalaisesta strukturalismista nousseen tagmemiikan edustajat (ks. esim. Pike 1982), korostivat kategorioiden kielikohtaisuutta antamalla niille niiden ominaisuuksia kuvaavien termien sijasta neutraaleja numeroista tai kirjaimista koostuvia nimiä. Näin esimerkiksi suomen kielen substantiivi, verbi ja adjektiivisanaluokkia voitaisiin perinteisten termien sijasta nimittää sanaluokiksi S1, S2 ja S3. Tällainen terminologinen valinta tekee kuitenkin kielioppikuvausten lukemisen ja kielten vertailun vaikeammaksi.

Tagmeemisen teorian mukaan kirjoitettuja kielioppeja, esim. Bolivian alkuperäiskielistä (Matteson toim. 1965), on erittäin vaikea käyttää, jos ei ole syvällisesti perehtynyt tuonaikaiseen tagmemiikan teoriaan. Kuten Haspelmath (2010: 674) toteaa, on järkevää käyttää terminologiaa, jossa kielikohtaisille kategorioille annetaan samoja nimiä kuin funktionaalisesti samankaltaisille kielikohtaisille kategorioille muissa kielissä. Englanninkielisessä typologisessa kirjallisuudessa on yleistynyt konventio, jonka mukaan kielikohtaisiin kategorioihin viitattaessa kategorian nimi kirjoitetaan isolla alkukirjaimella, mutta kieltenvälisistä vertailukäsitteistä tai muuten kieltenvälisistä yleistyksistä puhuttaessa käytetään pientä alkukirjainta.

2.2. Tutkittavan ilmiön määrittely eli miten kieliä verrataan

Edellisessä jaksossa totesin, että monet typologis-funktionaalisessa viitekehyksessä toimivat kiellentutkijat jakavat partikularistisen kategoriakäsityksen, jonka mukaan kielelliset kategoriat ovat kielikohtaisia eikä kieltenvälisiä tai universaaleja kategorioita ole olemassa. Ei ole itsestään selvää, miten kieliä voidaan verrata tällaisen kategorianäkemyksen pohjalta. Tässä jaksossa pyrin valottamaan niitä metodologisia työkaluja, joita typologit ovat tämän ongelman ratkaisemiseksi kehittäneet.

Kysymykseen, kuinka tunnistaa tutkimuksen kohteena oleva ilmiö erilaisissa kielissä, on typologisessa kirjallisuudessa viitattu kieltenvälisen identifikaation ongelmana. Kuten edellisestä jaksosta kävi ilmi, typologista tutkimusta tehdään partikularistisesta kategoriakäsityksestä käsin, jolloin ei ole olemassa kieltenvälisiä universaaleja kategorioita, joihin vertailu voidaan perustaa. Kun tehdään maailmanlaajuista vertailevaa kielitiedettä, kohdataan valtava määrä kielellistä variaatiota, mikä tekee kieltenvälisen identifikaation ongelman erittäin näkyväksi.

Vertailua ei voi tietenkään perustaa yksi yhteen jonkin yksittäisen kielen kategorioihin, koska silloin jäisivät tunnistamatta muiden kielten kategoriat, jotka eroavat vertailun pohjana olevan kielen kategorioista. Emme voi tutkia maailman kielten subjekteja siitä lähtökohdasta, että otamme tutkimukseemme mukaan kaikki ne eri kielten kategoriat, jotka ovat syntaktisilta ja morfologisilta ominaisuuksiltaan identtisiä esimerkiksi englannin kielen subjektikategorian kanssa. Mistään toisesta kielestä ei löydy täsmälleen identtistä subjektikategoriaa, eikä vertailua voida suorittaa. Jos löysennämme kriteerejä siihen suuntaan, että määrittelemme joitakin tyypillisiä subjektin ominaisuuksia, voimme ehkä laajentaa vertailua joihinkin kieliin, joissa on tunnistettavissa ominaisuuksiltaan samankaltainen subjektikategoria. Maailman

kielet kuitenkin eroavat toisistaan suuresti lauseenjäsenkategorioidensa suhteen, eikä maailmanlaajuista vertailua voida tällaiseen yhden kielen kategoriaan pohjautuvaan subjektikäsitteeseen perustaa.

Vertailu ei voi perustua kielikohtaisiin muodollisiin kriteereihin. Emme voi kattavasti tutkia kiellon ilmaisu maailman kielissä määrittelemällä tutkimuksen kohteen muodollisesti niin, että etsisimme kielistä kieltopartikkelirakenteita. Tällöin vertailun ulkopuolelle jäävät kaikki ne eri kielten rakenteet, joissa kiellonmerkitys on jokin muu kuin partikkeli (kuten ruotsin *inte*), siis affiksi (kuten turkin *-mE-*) tai kieltoverbi (kuten suomen persoonamuodoissa taipuva *ei*). Typologisessa kielten vertailussa tarvitaan kaikkiin kieliin soveltuvia määritelmiä vertailuperustaksi (*tertium comparationis*).^{iv} Kielten kategorioiden samankaltaisuus perustuu niiden yhteisiin funktionaalisiin ominaisuuksiin, siis universaalien tason semanttisiin ja pragmaattisiin funktioihin. Niinpä kieltenvälisesti sovellettavilla määritelmillä on yleensä funktionaalinen pohja, jota erilaiset muodolliset rajaukset voivat tarkentaa. Käsittelen nyt vertailuperustana käytettäviä kieltenvälisiä määritelmiä ja niihin liittyviä teoreettismetodologisia kysymyksiä ja havainnollistan niitä muutamien aiemmassa kirjallisuudessa esitettyjen tutkimusten avulla.

Joseph Greenberg, jota voidaan pitää modernin kielitypologian perustajana, toteaa klassisessa universaaleja käsittelevässä artikkelissaan (Greenberg 1963), että rakenteiden ja kategorioiden vertailu kielten välillä on käytännössä semanttisperusteista. Niinpä esimerkiksi hänen sanajärjestystutkimuksissaan S(ubjekti) tarkoittaa lauseenjäsentä, joka tyypillisesti ilmaisee lauseen agenttia, ja O(bjekti) lauseenjäsentä, joka tyypillisesti ilmaisee patienttia. Tältä pohjalta eri kielistä tunnistetut lauseenjäsenet voivat morfosyntaktisilta ominaisuuksiltaan erota hyvinkin radikaalisti toisistaan. Greenberg vihjaa, että myös tiukempia, muodollisperusteisia määritelmiä voitaisiin ehkä käyttää, mutta tämä edellyttäisi niin yksityiskohtaista analyysiä kunkin kielen lauseenjäsenysoitteesta, ettei se olisi tutkimukseen käytettävissä olevan ajan puitteissa mahdollista. Myöhempi tutkimus on osoittanut Greenbergin semanttisperusteisen metodologisen käytännön teoreettisesti perustelluksi.

Stassenin (1985) mukaan kieltenvälisten määritelmien on oltava semanttisperusteisia, mutta niitä voidaan kuitenkin rajata ja täsmentää myös muodollisin kriteerein. Samoilla linjoilla on Haspelmath (1997) puhuessaan funktionaalisen-formaalisten määritelmistä ja Croft (2003: 17–18) tuodessaan puhtaan ulkoisten (semanttis-pragmaattisten) määritelmien lisäksi esiin myös johdettujen rakenteellisten (*derived structural*) määritelmien tarpeellisuuden.^v

Komparatiiveja käsittelevässä typologisessa tutkimuksessa Stassen (1985: 15) määrittelee komparatiivikonstruktion seuraavasti:

Konstruktio lasketaan komparatiivikonstruktioksi (ja otetaan siksi huomioon luokittelussa), jos sen semanttinen funktio on määrittää asteittainen (l. epäidenttinen) sijainti predikatiivisella asteikolla kahdelle (mahdollisesti kompleksille) objektille.^{vi}

Määritelmän muoto on siis: K on konstruktio, jonka (semanttinen/pragmaattinen) funktio on F. Tämän määritelmän lisäksi Stassen rajaa tutkimuksensa kohteen muodollisin kriteerein niihin komparatiivikonstruktiioihin, joissa vertailtavat objektit on ilmaistu nominaalilausekkeilla. Määritelmän mukaisen komparatiivikonstruktion sisältävät esim. lauseet *puu on korkeampi kuin talo* ja *pidän sinusta enemmän kuin hänestä*, mutta eivät sellaiset lauseet kuin *kenraali oli enemmän ovela kuin rohkea* tai *presidentti on paljon fiksumpi kuin luulet*, joissa molemmissa verrataan saman objektin eri ominaisuuksia, ja jälkimmäisessä vertailtavat eivät ole edes nominaalilausekkeella ilmaistuja.

Havainnollistan muutamalla esimerkillä, kuinka erilaisia konstruktioita tällainen semanttisperusteinen määritelmä voi eri kielistä löytää. Erikielisten esimerkkien ymmärtämisen ja vertailun mahdollistava kolmirivinen esitystapa on ns. Leipzigin glossaussääntöjen mukainen.^{vii}

(1) suomi (uralilainen, Suomi; esimerkit keksittyjä)

- a. kissa on ketterämpi kuin koira
- b. kissa on koiraa ketterämpi

(2) hausa (afroaasialainen, Nigeria; Newman 2000: 93)

jirgī	yā	fi	mōtā	girmā
lentokone	3SG	ylittää	auto	suuruus

'Lentokone on suurempi kuin auto'

(3) hixkaryana (karibilainen, Brasilia; Derbyshire 1979: 67)

kaw-ohra naha waraka kaw naha kaywerye

pitkä-NEG on Waraka pitkä on Kaywerye

'Kaywerye on pidempi kuin Waraka.'

Suomessa (1) on kaksi vaihtoehtoista komparatiivikonstruktiota, jotka eroavat vertailukohdan merkinnän suhteen; esimerkissä (a) sitä merkitään partikkelilla *kuin* ja esimerkissä (b) partitiivisijalla. Hausassa (2) komparatiivi ilmaistaan verbillä, jonka merkitys on 'ylittää', ja vertailukohta on tämän verbin suora objekti. Hixkaryanassa (3) komparatiivikonstruktio koostuu kahdesta rinnasteisesta lauseesta, jotka ilmaisevat vertailtavien vastakkaisia ominaisuuksia. Nämä ovat siis kunkin kielen kieliopillistuneita tapoja ilmaista vertailua. Toki esimerkiksi suomeksi voitaisiin sanoa *kissa ylittää koiran ketteryydessä* tai *kissa on ketterä, koina ei ole ketterä*, mutta näitä tapoja ilmaista vertailua ei voida pitää suomen kielessä (kieliopillistuneina) konstruktioina.

Haspelmath (2009, 2010) nimittää typologiien kieltenvälisessä vertailussa käyttämiä määritelmiä vertailukäsitteiksi (*comparative concepts*) ja korostaa niiden eroja kielikohtaisiin kategorioihin, joita hän kutsuu kuvauskategorioiksi (*descriptive categories*). Haspelmath ottaa näin eksplisiittisesti kantaa kysymykseen, miten typologista vertailua voidaan tehdä partikularistisen kategorianäkemyksen pohjalta. Kuvauskategoriat ovat jaksossa 2.1. puheena olleita yksittäiskielten kielikohtaisia kategorioita, siis kategorioita, jotka ovat osa kunkin yksittäisen kielen kuvausta, ja ne ovat psykologisesti tosia (sikäli kuin kielen kuvaus onnistuu heijastelemaan puhujien kielitajua). Vertailukäsitteet (Haspelmath 2010: 665) ovat vertailevan kielitieteilijän metodologisia työkaluja. Ne eivät ole osa yksittäisten kielten systeemejä eikä niillä siten ole suoraa merkitystä deskriptiota tekeväälle kielitieteilijälle. Ne eivät myöskään ole osa puhujan kielitajua eivätkä siis psykologisesti tosia. Vertailukäsitteet eivät myöskään Haspelmathin mukaan voi olla oikeita tai vääriä, vaan vain paremmin tai huonommin soveltuvia tehtäväänsä vertailun välineinä. Ne ovat kaikkiin kieliin universaalisti sovellettavissa ja muiden universaalisti sovellettavissa olevien käsitteiden kautta määriteltyjä; näitä universaalisti sovellettavissa olevia käsitteitä voivat olla toiset vertailukäsitteet, konseptuaalis-semanttiset käsitteet ja yleistason muodolliset käsitteet (esimerkit toisen vertailukäsitteen ja konseptuaalis-semanttisen käsitteen käytöstä vertailukäsitteen määritelmässä tarjoavat vertailukäsite ”lekseemi” ja konseptuaalis-semanttinen käsite ”deskriptiivinen ominaisuus” vertailukäsitteen ”adjektiivi” määritelmässä jäljempänä;

yleistason muodollinen käsite voi olla esimerkiksi positio l. elementin sijainti suhteessa muihin elementteihin). Vertailukäsitteitä nimitetään usein käytännön syistä samalla tavalla kuin kuvauskategorioita, mutta näiden välillä ei vallitse yksi yhteen -suhde. Huomattakoon, että Lazardin (2005) ajatus ns. arbitraareista käsitteellisistä viitekehysistä (*Arbitrary Conceptual Framework*, ACF), joihin tyypologinen vertailu perustuu, on hyvin samankaltainen Haspelmathin vertailukäsiteajatuksen kanssa.

Esimerkiksi vertailukäsitteen adjektiivinen Haspelmath (2010: 670) määrittelee seuraavasti: ”Adjektiivinen on lekseemi, joka ilmaisee kuvailevaa ominaisuutta ja jota voidaan käyttää substantiivin tarkoitteen rajaamiseen”.^{viii} Tämä vertailukäsitteen määritelmä sisältää vertailukäsitteen lekseemi (joka on tietysti vielä erikseen määriteltävä), ja konseptuaalis-semanttiset käsitteet ”deskriptiivinen ominaisuus” ja ”rajata viittausta”. Tämän vertailukäsitteen avulla voidaan poimia kielistä vertailuun adjektiiveja riippumatta siitä, onko näissä kielissä erityistä adjektiivisanaluokkaa kielikohtaisena kategoriana. Määritelmä siis sisällyttää tutkimukseen ominaisuutta ilmaisevia lekseemejä paitsi kielistä, joissa on erillinen adjektiivisanaluokka (kielikohtaisena kategoriana), myös kielistä, joissa nämä lekseemit kuuluvat (kielikohtaisiin) verbi- tai substantiivisanaluokkakategorioihin.

Edellä on todettu, että kieltenvälisten määritelmien on oltava semanttisperusteisia. Jos haluamme laajentaa tätä ajatusta fonologian suuntaan, voimme sanoa, että vertailun on perustuttava substanssiin (vrt. Haspelmath 2007: 126). Fonologian tapauksessa tämä substanssi on fonetiikkaa. Foneemien rajat akustisessa ja artikulatorisessa avaruudessa ovat kielikohtaisia, mutta eri kielten foneemeja voidaan verrata niiden foneettisten ominaisuuksien pohjalta. Substanssivaatimus ei kuitenkaan ole ehdoton. Kuten on jo käynyt ilmi, monia semanttisperusteisia määritelmiä tarkennetaan muodollisilla kriteereillä, ja on myös vertailukäsitteitä, joiden määritelmässä semantiikan osuus jää pieneksi. Kuten Haspelmath (2007: 126) toteaa, tällaisten käsitteiden, joita ovat esimerkiksi sija tai adpositio, kieltenvälinen määrittely on vaikeaa, ja niiden käytössä esiintyykin tavallista enemmän vaihtelua eri tutkijoiden välillä.

Omassa negatiivista käsittelevässä tutkimuksessani olen käsitellyt standardikieltoa. Sitä voidaan lyhyesti luonnehtia kielten perustavoiksi kieltää deklarativisia päälauseita, joiden predikaattina on verbi. Tarkemmin olen määritellyt kyseisen vertailukäsitteen seuraavasti (Miestamo 2005: 42):

Standardikieltokonstruktio on konstruktio, jonka funktio on modifioida verbaalista, deklarativista päälauseita, joka ilmaisee proposition p, siten, että näin modifioitu lause

ilmaisee proposition, jolla on päinvastainen totuusarvo kuin propositiolla p , siis $\sim p$, tai proposition, jota käytetään lähinnä vastaamaan propositiota p , jos proposition $\sim p$ ilmaisevaa lausetta ei voi kielessä muodostaa, ja joka on produktiivinen ja yleinen keino ilmaista tätä funktiota kielessä.^{ix}

Vertailukäsitteen semanttisena pohjana on negaatio määriteltynä totuusarvon muutoksen kautta. Sitä rajataan edelleen vertailukäsitteellä deklaratiiivinen verbaalinen päälause (joka itsessään koostuu useammasta vertailukäsitteestä). Lisäksi otetaan huomioon mm. *se*, *ettei* myönnön ja kiellon välillä ole kaikissa kielissä yksi yhteen -suhdetta.

Standardikiellon määritelmällä voimme tunnistaa standardikieltokonstruktioita tutkimuksen otokseen valituista kielistä. Suomen standardikieltokonstruktiossa on kiellonmerkitsimenä kieltoapuverbi *e-* ja leksikaalinen verbi *on ei-*finiittisessä muodossa: preesensissä konnegatiivi^x ja preteritissä menneen ajan partisiippi (4).

(4) suomi (uralilainen, Suomi; esimerkit keksittyjä)

- a. King Arthur haukkuu parvekkeella.
- b. King Arthur ei hauku parvekkeella.
- c. King Arthur haukkui parvekkeella.
- d. King Arthur ei haukkunut parvekkeella.

Vertailukäsitteen standardinegaatio alle mahtuu siis suomessa yksi kieltokonstruktio (ellei leksikaalisen verbin eron perusteella pidetä preesensin ja preteritin kieltokonstruktioita erillisinä).

Komisyrjäänissä (5) eri aikamuodoissa on täysin erilaiset kieltorakenteet. Preesensissä ja preteritissä on kieltoapuverbirakenne, ja kieltoapuverbin muoto vielä eroaa näiden kahden aikamuodon välillä (preesensissä *o-* ja preteritissä *e-*). Perfektissä ja pluskvamperfektissä taas on rakenne, jossa kieltopartikkeli *abu* toimii kiellonmerkitsimenä.

(5) komisyrjääni (uralilainen, Venäjä; Rédei 1978: 105–109)

- | | |
|-----------------|-------------------|
| a. <i>šet-ę</i> | b. <i>o-z šet</i> |
| antaa-3SG.PRES | NEG-3 antaa |
| 'hän antaa' | 'hän ei anna' |

c. <i>šet-i-s</i>	d. <i>e-z</i> <i>šet</i>
antaa-PST-3SG	NEG.PST-3 antaa
'hän antoi'	'hän ei antanut'
e. <i>šet-em-a</i>	f. <i>abu</i> <i>šet-em-a</i>
antaa-PERF-3SG	NEG antaa-PERF-3SG
'hän on antanut'	'hän ei ole antanut'
g. <i>šet-em-a</i> <i>veli</i>	h. <i>abu</i> <i>šet-em-a</i> <i>veli</i>
antaa-PERF-3SG olla.PST.3SG	NEG antaa-PERF-3SG olla.PST.3SG
'hän oli antanut.'	'hän ei ollut antanut'

Komisyrjäänin ja suomen vertailu havainnollistaa sitä, ettei kielikohtaisten kategorioiden ja vertailukäsitteiden välillä vallitse yksi yhteen -vastaavuus.

Maungissa standardikieltorakenne muodostetaan kieltopartikkelilla *marig* (6c). Myöntölauseissa voidaan tehdä ero realis- ja irrealismoduksen välillä (6a vs. b). Kiellossa voi kuitenkin käyttää vain irrealismuotoa – realismuoto ja kieltopartikkeli eivät esiinny yhdessä – joten modusero katoaa kiellossa.

(6) maung (australialainen, Australia; Capell & Hinch 1970: 67)

a. <i>ŋi-udba</i>	b. <i>ni-udba-ji</i>
1SG > 3-panna	1SG > 3-panna-IRR.NPST
'Panen (sen).'	'Voin panna (sen).'
c. <i>marig ni-udba-ji</i>	
NEG 1SG > 3-panna-IRR.NPST	
'En pane / voi panna (sitä).'	

Tässä nähdään uusi esimerkki siitä, kuinka kategorioiden rajat eroavat eri kielten välillä, mutta esimerkki havainnollistaa myös sitä standardikiellon määritelmässä huomioon otettua seikkaa, ettei myöntölauseesta aina voida muodostaa suoraan vastaavaa kieltolauseita.

Määritelmän mukaisesti standardikieltorakenteita käytetään väitelauseiden kiellossa. On toki mahdollista, että samaa kieltorakennetta käytetään myös esimerkiksi käskylauseiden kiellossa kuten bagirmissa (7), jossa kieltopartikkeli *eli* on sekä väite- että käskylauseen

kiellonmerkitsin. Käskylauseiden kiellot kuitenkin usein eroavat väitelauseiden kiellosta, esimerkkinä purépecha (8), jossa standardikieltokonstruktiossa esiintyy kieltopartikkeli 'no ja käskylauseen kieltokonstruktiossa kieltopartikkeli 'ašř .

(7) bagirmi (nilosaharalainen, Tšad; Stevenson 1969: 91, 93, 95)

a. je j-ab eli
me 1PL-mennä NEG
'Emme menneet.'

b. ab eli
mennä NEG
'Älä mene!'

(8) purépecha (taraskalainen, Meksiko, Chamoreau 2000: 112, 242)

a. 'no pi'ri-šiN-ti
NEG laulaa-HAB-ASS.3
'Hän ei laula.'

b. 'aši 'xua-ø-rini 'sani
NEG tuoda-IMP-2 > 1 vähän
'Älä tuo minulle vähän!'

Jälleen näemme, kuinka kuvauskategorioiden rajat ovat kielikohtaisia, ja vertailukäsitteen (standardikielto) ja kuvauskategorioiden (bagirmin lausekielto vs. purépechan erilliset väitelauseen ja käskylauseen kieltokonstruktiot) vastaavuus ei ole yksi yhteen.

3 Otanta ja lähteet

Tässä jaksossa käsittelen typologisen vertailun metodologiaa otannan eli vertailtavien kielten valinnan ja vertailussa käytettävien lähteiden osalta. Kielitypologiset tutkimukset perustuvat laajoihin, useita kymmeniä tai satoja kieliä käsittäviin otoksiin. Otoksen koko ja otantamenetelmät riippuvat tutkimuskysymyksestä – esimerkiksi siitä, kuinka paljon työtä

tutkittavan ilmiön selvittäminen kussakin kielessä vaatii. Mitä työläämpi yksittäisen kielen analysointi on, sitä vähemmän kieliä otos voi sisältää. Suurilla otoksilla saadaan kattavampi kuva maailman kielistä kuin pienillä, mutta hävitään vastaavasti siinä, kuinka syvälle kunkin kielen ominaisuuksiin voidaan pureutua; molempia näkökulmia – sekä laajaa että syvää – tarvitaan ihmiskielen olemuksen ymmärtämiseksi. Yleistypologinen tutkimus, joka ei keskity johonkin tiettyyn kielikuntaan tai maantieteelliseen alueeseen eikä muutenkaan rajaa osaa kielistä fokuksensa ulkopuolelle, tarvitsee otoksen, joka mahdollisimman hyvin edustaa maailman eri kielikuntia ja maantieteellisiä alueita.

Otantamenetelmiä voidaan luokitella eri perustein. Voidaan esimerkiksi tehdä ero todennäköisyys- ja variaatio-otannon välillä (Rijkhoff ym. 1993). Todennäköisyysotannon tavoitteena on luoda otoksia, jotka olisivat päteviä tilastollisten laskelmien pohjaksi. Variaatio-otokset taas pyrkivät tuomaan mahdollisimman hyvin esiin tutkittavan ilmiön koko variaation maailman kielissä. Otannassa on tärkeää edustavuus ja riippumattomuus. On tärkeää, että otoksen kielet edustavat olemassa olevia tai olleita maailman kieliä tai kaikkia mahdollisia ihmiskieliä mahdollisimman hyvin. Otoksen edustavuutta saadaan tietysti kasvatettua kasvattamalla otoskokoa, siis lisäämällä otoksen kielten määrää. Liian suuret otokset eivät kuitenkaan ole tutkimuksen käytännön kannalta toivottavia, joten otoksen kokoa pyritään rajoittamaan. Tällöin edustavuutta saadaan kasvatettua huolehtimalla otoskielten riippumattomuudesta. Ositetussa otannassa otoskielet valitaan tiettyjen periaatteiden perusteella relevanteista kieliryhmistä. Näitä kieliryhmiä ovat etupäässä eri kielikunnat ja niiden haarat sekä eri maantieteelliset alueet. Kun huolehditaan eri kielikuntien ja maantieteellisten alueiden tasaisesta edustuksesta, siis otoskielten genealogisesta ja arealisesta riippumattomuudesta, saadaan myös otoksen yleistä edustavuutta kasvatettua suhteessa otoksen kokoon. Todennäköisyysotannassa on variaatio-otantaa tärkeämpää, että kielet ovat toisistaan mahdollisimman riippumattomia esim. genealogisesti ja arealisesti, koska monet tilastolliset testit edellyttävät otosyksikköjen olevan toisistaan riippumattomia.

Omassa negatiota koskevassa tutkimuksessani olen käyttänyt variaatio-otosta, jossa otos on ositettu genusten ja makroalueiden tasolla. Genus (Dryer 1989) on genealoginen ryhmä, jonka ajallinen syvyys ei ylitä 3500–4000 vuotta. Esimerkiksi indoeurooppalaisen kielikunnan haarat – romaaniset, germaanisets, slaavilaiset jne. kielet – muodostavat kukin oman genuksensa. Dryerin (2011) luokittelussa erotetaan maailmanlaajuisesti 511 genusta. Otosmenetelmäni perusideana on ottaa jokaisesta genuksesta yksi kieli, sikäli kun otoksen kielistä on lähteitä saatavilla. Makroalueet (Dryer 1989) taas ovat maanosankokoisia kieliä, joiden kielten välillä on muinaisista kontakteista tai kielisukulaisuudesta johtuvia

samankaltaisuuksia. Otosmenetelmässäni kunkin makroalueen edustusta voidaan rajata suhteessa sen genealogiseen diversiteettiin. Siis alueilta, joiden genusten lukumäärä on suuri, valitaan suhteessa enemmän kieliä, ja alueilta, joilla on vähemmän genuksia, kieliä valitaan suhteessa vähemmän. Otantamenetelmistä on kielitypologisessa tutkimuksessa kirjoitettu suhteellisen paljon, ks. tarkemmin Bakker (2010) ja siinä mainitut lähteet.

Typologin datan keruussa käyttämät lähteet riippuvat tutkimuksen kohteesta ja otoksen koosta. Tavallisimmin lähteinä käytetään kielten kielioppeja tai muita saatavilla olevia tutkimuksia. Jonkin verran käytetään myös kyselyomakkeita (esim. Dahl 1985). Sähköinen viestintä helpottaa myös tutkimuskohdetta koskevien kysymysten lähettämistä eri kielten asiantuntijoille, jos käsillä olevat lähteet eivät anna riittävää kuvaa aiheesta. Monet typologit tekevät myös kenttätyötä ja tutkivat yksittäisiä kieliä, mutta laajaan kielten vertailuun oma yksittäiskielen tutkimus ei tietenkään anna mahdollisuuksia (yksittäiskielen tutkimuksen suhteesta typologiaan ja kielten vertailuun ks. Sandman tässä niteessä). Viime aikoina on myös ruvettu kehittämään korpusten käyttöä laajemmassa kielten vertailussa (esim. Cysouw & Wälchli 2007),^{xi} mutta tällaisella metodologialla on omat rajoituksensa siinä, millaisiin tutkimuskysymyksiin sitä voidaan käyttää.

On tärkeää huomata, että toisten kielitieteilijöiden tekemän tutkimuksen käyttö lähteenä ei tarkoita, että typologi kopioisi analyysin sellaisenaan, vaan lähteen pohjalta muodostetaan aina oma analyysi, mahdollisesti useampia saman kielen lähteitä verraten ja lähdekritiikkiä harjoittaen. Typologi on tietysti pitkälti riippuvainen yksittäiskielen tutkijoiden tekemistä analyyseistä, ja kriittinen suhtautuminen lähteisiin on oleellista kieliä vertailtaessa. Kielioppien ja muiden lähteiden laatu ja luonne ovatkin keskeisiä aiheita typologien metodologisissa pohdinnoissa. Lähteiden luotettavuutta on harjaannuttava arvioimaan. Lähteiden vertailtavuutta arvioitaessa on otettava huomioon myös, että kielistä on kirjoitettu erilaisista näkökulmista: erilaisissa teoriakehyksissä, erilaisten aineistojen pohjalta, eri tarkoituksia palvelemaan. Esimerkiksi kirjoitetun ja puhutun kielen suhde referenssikieliopin tausta-aineistossa on herättänyt viime aikoina keskustelua. Kieliopit pohjautuvat pääasiassa puhuttuihin kielimuotoihin – maailman kielistä vain pientä osaa käytetään kirjoitetussa muodossa – mutta Euroopan kielten kielioppien taustalla vaikuttaa vahvasti kirjoitettu standardikieli. On todettu, että tällaiset erot aineiston luonteessa voivat vaikuttaa yleistyksiin, joita esimerkiksi Euroopan kielistä tehdään verrattuna muiden maanosien kieliin (ks. Himmelmann 2000: 10–11).

Eräs ongelma on myös lähteiden saatavuus. Kielikunnat ja maantieteelliset alueet eroavat suuresti siinä, kuinka laajalti ja syvällisesti niiden kieliä on tutkittu ja kuinka hyvin lähteitä

alueen tai kielikunnan kielistä on saatavilla. Lähteiden saatavuus vaikuttaa otoskielten valintaan, ja datan keräämisvaiheesta saatetaan joutua palaamaan täydentämään otosta, jos jonkin kielen lähde osoittautuu riittämättömäksi ja kielen tilalle otokseen täytyy valita uusi, korvaava kieli. Lähteiden huono saatavuus voi johtaa joidenkin alueiden ja kielikuntien aliedustukseen otoksessa. Eri kielten lähteitä voi etsiä esim. teoksista Dryer & Haspelmath (2011), Lewis (2009) ja Hammarström (2007).

Käsitettä dokulekti on käytetty viittaamaan kielimuotoon, joka on kuvattu tai dokumentoitu esimerkiksi kieliopissa tai korpuksessa.^{xii} Esimerkiksi *Iso suomen kielioppi* (Hakulinen ym. 2004) edustaa yhtä dokulektiä ja jokin toinen suomen kielen kielioppi toista dokulektiä. Vertaillen kieliä kielioppien tms. lähteiden perusteella, typologi siis oikeastaan vertailee dokulektejä. Typologin kieliotos siis koostuu dokulekteistä, ja kieltenväisiä yleistyksiä tehdään dokulektien pohjalta.

4 Kieltenväliset yleistyksiset ja selittäminen

Kun tutkimuskohteen määritelmän mukainen data on kerätty kustakin otoksen kielestä, päästään varsinaiseen analyysivaiheeseen. Toki datankeruu- ja analyysivaiheet lomittuvat toisiinsa siten, että dataa kerätessään tutkija jo hahmottelee erilaisia analyyseja ja mahdollisia luokitteluja, ja varsinaiseen analyysiin päästyään hän joutuu palaamaan lähteisiin, jotta löytäisi lisädataa ja pystyisi näin vastaamaan analyysissa nousseisiin kysymyksiin. Dataa kerätessä ja analysoitaessa rakennetaan tietokantaa, johon kunkin otoskielen relevantit ominaisuudet koodataan. Analyysivaiheessa kiinnitetään huomiota kieltenvälisiin yhtäläisyyksiin ja eroihin, jäsennellään aineistoa ja tehdään typologinen luokittelu erilaisista rakennetyypeistä, joita aineistossa esiintyy, ja etsitään kaikista kielistä päteviä universaaleja yleistyksiä tai tilastollisia tendenssejä. Tunnetuimpia typologisia luokitteluja lienevät kielten luokittelu niiden perussanajärjestyksen pohjalta S(ubjekti)O(bjekti)V(erbi)-, SVO-, VSO-, VOS-, OVS- ja OSV-tyyppeihin sekä klassisen morfologisen typologian mukainen jaottelu isoiviin, agglutinatiivisiin, flekteeraaviin ja polysynteettisiin rakenteisiin/kieliin.^{xiii} Runsaasti esimerkkejä typologisista luokitteluista tarjoaa Dryer & Haspelmath (2011), jonka jokainen kartta sisältää typologisen luokittelun.

Luokittelu ei ole itsetarkoitus, vaan se tehdään aina jotakin tarkoitusta varten. Luokittelun on siis oltava hyödyllinen esim. teorianmuodostuksen ja selitysten etsimisen kannalta. Tutkimuksen tarkoituksena on paitsi kuvata kieltenvälistä variaatiota, myös ymmärtää, miksi

kielissä on havaittuja eroja ja samankaltaisuuksia. Luokittelun tarkoituksena on edesauttaa tätä ymmärrystä. Typologista luokittelua voidaan toki tehdä hyvin monenlaisten kriteerien mukaan. Esimerkiksi maailman kielten kieltorakenteet voitaisiin jakaa eri tyyppeihin sen mukaan, alkaako kiellonmerkitsin vokaalilla vai konsonantilla, ja edelleen sen mukaan, minkä tyyppisellä vokaalilla tai konsonantilla se alkaa. Tällainen luokittelu ei kuitenkaan olisi teoreettisesti perusteltu eikä auttaisi meitä ymmärtämään kieltorakenteissa esiintyvää vaihtelua. Omassa tutkimuksessani (Miestamo 2005) olen luokitellut kieltorakenteet sen mukaan, onko myöntö- ja kieltolauseiden välillä rakenteellisia eroja (asymmetriaa) negaationmerkitsimien lisäksi, ja edelleen, millaista asymmetriaa esiintyy. Näille rakenteellisen asymmetrian eri tyypeille olen sitten etsinyt motivaatioita kiellon ja myönnön funktionaalisen tason semanttisista ja pragmaattisista eroista.

Kieltenvälisiin yleistyksiin kuuluvat varsinaisen luokittelun lisäksi myös havainnot eri tyyppien yleisyydestä ja maantieteellisestä jakautumisesta sekä eri kielenpiirteiden välisten korrelaatioiden etsiminen. Korrelaatioissa on kyse implikationaalisista universaaleista, siis väittämistä, jotka kertovat kahden loogisesti toisistaan riippumattoman kielenpiirteen välisestä riippuvuudesta ja pätevät kaikista tai tilastollisesti merkittävästä osasta kieliä. Tunnetuimpia implikationaalisia universaaleja ovat sanajärjestysuniversaalit, joissa eri lause- ja lauseketyyppien sanajärjestysominaisuudet korreloivat keskenään, esimerkiksi: jos kielen hallitseva sanajärjestys on VSO, sen adpositiot ovat prepositioita, ja jos kielen hallitseva sanajärjestys on SOV, sen adpositiot ovat postpositioita (ks. Greenberg 1963).

Viimeisenä vaiheena tutkimuksen metodologisessa kronologiassa on havaittujen ilmiöiden selittäminen. Typologista tutkimusta tehdään tavallisimmin funktionaalis-kognitiivisessa viitekehyksessä ja myös selityksiä kielen ilmiöille etsitään kielen rakenteen ulkopuolelta, kielen semanttisista ja pragmaattisista funktioista ja kielenkäyttäjien kognitiosta. Edellä viittasin lyhyesti negaatorakenteiden selityksiin omassa tutkimuksessani, mutta en tässä mene syvemmälle selittämisen problematiikkaan.

Olen tässä luvussa esitellyt typologisen tutkimuksen kulkua. Koska kyseessä ei ole typologisen metodologian kokonaisesitys, olen sivuuttanut monet kohdat lyhyellä maininnalla ja paneutunut pidemmälti seikkoihin, joiden kohdalla vertailun problematiikka nousee erityisesti esiin. Kielten vertailunäkökulman kannalta on oleellista muistaa, että itse kieltenvälisen tutkimuskohteen määrittelyn ja vertailtavuuden lisäksi myös kieltenväliset yleistyksen – luokittelut, universaalit, korrelaatiot – sekä selitykset on määriteltävä ja muotoiltava niin, että ne pätevät yli yksittäiskielten rajojen, siis vertailukäsitteitä käyttäen.

5 Lopuksi

Tässä kirjoituksessa olen käsitellyt kielten vertailua kielitypologisen tutkimuksen näkökulmasta. Olen keskittynyt vertailun ydinkysymykseen: millaisten vertailuperustojen avulla kielten ilmiöitä voidaan kieltenvälisesti tunnistaa ja näin asettaa vierekkäin vertailtavaksi, siis millaisia ovat ne kieltenväliset määritelmät, joiden perusteella eri kielistä saadaan poimittua toisiaan vastaavat ilmiöt vertailuun. Tämän käytännön metodologisen kysymyksen taustalla on teoreettisempi kysymys kielellisten kategorioiden luonteesta universaaleina vs. kielikohtaisina. Suuressa osassa nykyistä typologista tutkimusta on taustaoletuksena näkemys kielellisten kategorioiden kielikohtaisuudesta – oletetaan siis, että kategoriat ovat kielikohtaisia ja niiden väliset samankaltaisuudet johtuvat yhteisestä funktionaalisesta (semanttisesta/pragmaattisesta) taustasta. Kielten vertailussa käytetään kieltenvälisiä vertailukäsitteitä, jotka eivät viittaa kielikohtaisiin kategorioihin eivätkä ole yhteneviä näiden kanssa. Ne ovat kielitieteilijän vertailua varten kehittämiä käsitteitä, joiden ei ole tarkoituskaan olla psykologisesti tosia. Varsinaisen vertailtavuuden ja vertailun lisäksi olen käsitellyt myös muita typologisen tutkimuksen metodologian kysymyksiä kielten vertailun näkökulmasta: kieliotoksen valintaa, lähteiden arviointia ja käyttöä sekä kielenpiirteiden typologista luokittelua. Esiin nostamani kysymyksiä olen havainnollistanut myös oman negaatiota koskevan typologisen tutkimukseni avulla.

Esimerkeissä käytetyt lyhenteet

- 1 ensimmäinen persoona
- 2 toinen persoona
- 3 kolmas persoona
- ASS assertiivi
- HAB habituaali
- IMP imperatiivi
- IRR irrealis
- NEG negaatio
- NPST ei-mennyt aika
- PERF perfekti

PL monikko
PRES preesens
PST mennyt aika
SG yksikkö

Lähteet

Bakker, Dik 2010: Language sampling. – Jae Jung Song (toim.), *The Oxford Handbook of Linguistic Typology* s. 100–127. Oxford: Oxford University Press.

Capell, A. – Hinch, H. E. 1970: *Maung grammar*. The Hague: Mouton.

Chamoreau, Claudine 2000: *Grammaire du purépecha*. Munich: Lincom Europa.

Chomsky, Noam – Lasnik, Howard 1993: Principles and Parameters Theory. – J. von Stechow, A. Jacobs, W. Sternefeld & T. Vennemann (toim.), *Syntax: An international handbook of contemporary research* s. 505–569. Berlin: Walter de Gruyter.

Croft, William 2001: *Radical construction grammar*. Oxford: Oxford University Press.

Croft, William 2003: *Typology and universals* (2. laitos). Cambridge: Cambridge University Press.

Cysouw, Michael – Wälchli, Bernhard (toim.) 2007: *Parallel texts. Using translational equivalents in linguistic typology*. Teemanumero, *Sprachtypologie & Universalienforschung* 60 (2).

Dahl, Östen 1985: *Tense and aspect systems*. Oxford: Blackwell.

Derbyshire, Desmond C. 1979: *Hixkaryana*. Amsterdam: North-Holland Publishing Company.

Dryer, Matthew 1989: Large linguistic areas and language sampling. – *Studies in Language* 13 s. 257–292.

Dryer, Matthew 1997: Are grammatical relations universal? – Joan Bybee, John Haiman & Sandra Thompson (toim.), *Essays on language function and language type, dedicated to T. Givón* s. 115–143. Amsterdam: Benjamins.

Dryer, Matthew 2011: Genealogical language list. – Matthew Dryer & Martin Haspelmath (toim.), *The world atlas of language structures online*. Munich: Max Planck Digital Library. <http://wals.info/languoid/genealogy>.

Dryer, Matthew – Haspelmath, Martin (toim.) 2011: *The world atlas of language structures online*. Munich: Max Planck Digital Library. <http://wals.info/>.

Gil, David 2008: How complex are isolating languages? – Matti Miestamo, Kaius Sinnemäki & Fred Karlsson (toim.), *Language complexity: Typology, contact, change* s. 109–131. Amsterdam: Benjamins.

Greenberg, Joseph H. 1963: Some universals of grammar with particular reference to the order of meaningful elements. – Joseph H. Greenberg (toim.), *Universals of language. Report of a conference held at Dobbs Ferry, New York, April 13–15, 1961*, s. 73–113. Cambridge, MA: M.I.T. Press.

Hakulinen, Auli – Vilkuna, Maria – Korhonen, Riitta – Koivisto, Vesa – Heinonen, Tarja Riitta – Alho, Irja 2004: *Iso suomen kielioppi*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Hammarström, Harald 2007: *Handbook of descriptive language knowledge: A full-scale reference guide for typologists*. Munich: Lincom Europa. [päivitetty versio <http://haraldhammarstrom.ruhosting.nl/lwib.pdf>]

Haspelmath, Martin 1997: *Indefinite pronouns*. Oxford: Oxford University Press.

Haspelmath, Martin 2007: Pre-established categories don't exist – consequences for language description and typology. – *Linguistic Typology* 11 s. 119–132.

Haspelmath, Martin 2009: Pourquoi la typologie des langues est-elle possible? – *Bulletin de la Société de Linguistique de Paris* CIV s. 17–38.

Haspelmath, Martin 2010: Comparative concepts and descriptive categories in cross-linguistic studies. – *Language* 86/3 s. 663–687.

Himmelmann, Nikolaus P. 2000: Towards a typology of typologies. – *Sprachtypologie und Universalienforschung* 53 s. 5–12.

Itkonen, Esa 1991: *Universal history of linguistics: India, China, Arabia, Europe*.
Amsterdam: Benjamins.

Janhunen, Juha 1977: *Samojedischer Wortschatz. Gemeinsamojedische Etymologien*.
Helsinki: Helsingin yliopisto.

Joos, Martin (toim.) 1957: *Readings in linguistics I*. Chicago: University of Chicago Press.

Lazard, Gilbert 1992: Y a-t-il des catégories interlangagières? – Susanne Anschütz (toim.), *Texte, Sätze, Wörter und Moneme: Festschrift für Klaus Heger* s. 427–434. Heidelberg: Heidelberg Orientverlag. [uudelleenjulkaistu teoksessa Lazard, Gilbert 2001: *Études de linguistique générale*. Leuven: Peeters]

Lazard, Gilbert 2005: What are we typologists doing? – Zygmunt Frajzyngier, Adam Hodges & David Rood (toim.), *Linguistic diversity and language theories* s. 1–23. Amsterdam: Benjamins.

Lehmann, Christian 1989: Language description and general comparative grammar. – Gottfried Graustein & Gerhard Leitner (toim.), *Reference grammars and modern linguistic theory* s. 133–162. Tübingen: Niemeyer.

Lewis, M. Paul (toim.) 2009: *Ethnologue: Languages of the world* (16. laitos). Dallas: SIL International. [verkkoversio: <http://www.ethnologue.com/>]

Matteson, Esther (toim.) 1965: *Gramáticas estructurales de lenguas bolivianas 1-3*. Riberalta: Summer Institute of Linguistics.

Miestamo, Matti 2005: *Standard negation: The negation of declarative verbal main clauses in a typological perspective*. Berlin/New York: Mouton de Gruyter.

Newman, Paul 2000: *The Hausa language, an encyclopedic reference grammar*. New Haven: Yale University Press.

Newmeyer, Frederick J. 2007: Linguistic typology requires crosslinguistic formal categories. – *Linguistic Typology* 11 s. 133–157.

Newmeyer, Frederick J. 2010: On comparative concepts and descriptive categories: A reply to Haspelmath. – *Language* 86/3 s. 688–695.

Pike, Kenneth L. 1982: *Linguistic concepts: An introduction to tagmemics*. Lincoln, Nebraska: University of Nebraska Press.

Rédei, Károly 1978: *Syrjänische Chrestomathie mit Grammatik und Glossar*. Wien: Verband der wissenschaftlichen Gesellschaften Österreichs.

Rijkhoff, Jan – Bakker, Dik – Hengeveld, Kees – Kahrel, Peter 1993: A method of language sampling. – *Studies in Language* 17 s. 169–203.

Stassen, Leon 1985: *Comparison and universal grammar*. Oxford: Blackwell.

Stassen, Leon 2010: The problem of cross-linguistic identification. – Jae Jung Song (toim.), *The Oxford Handbook of Linguistic Typology* s. 90–99. Oxford: Oxford University Press.

Stevenson, R. C. 1969: *Bagirmi Grammar*. Khartoum: Sudan Research Unit, University of Khartoum.

Viitteet

- i. ”Languages can differ without limit and in unpredictable ways”, suom. kirjoittajan.
- ii. Croft (2001) vie kategorioiden spesifisyyden detaljitasolle myös kielen sisällä. Hänen mukaansa kategoriat ovat paitsi kielikohtaisia, myös kielen sisällä konstruktiospesifejä. Siis esimerkiksi englannin aktiivikonstruktion subjektilla on eri ominaisuudet kuin englannin passiivikonstruktion subjektilla, ja näissä tapauksessa on kyse kahdesta eri kieliopillisesta relaatiosta: englannin aktiivikonstruktion subjektista ja englannin passiivikonstruktion subjektista.
- iii. Käytän termiä preteriti suomen menneisyyteen viittaavasta aikamuodosta, jota perinteisesti kutsutaan paremmin aspektuaalisia ominaisuuksia kuvaavalla termillä imperfekti.
- iv. Latinankielistä termiä *tertium comparationis* on käytetty erityisesti kontrastiivisessa kielitieteessä, ks. Järventausta tässä niteessä.
- v. Newmeyer (2007, 2010) perustelee formaalien kieltenvälisten kategorioiden välttämättömyyttä sillä, että monissa typologisissa tutkimuksissa kieltenväliset määritelmät ja yleistyksen viittaavat myös muotoon. Kuten jäljempänä käy ilmi, muodosta puhuminen ei kuitenkaan edellytä kieltenvälisiä muodollisia kategorioita.
- vi. ”[A] construction counts as a comparative construction (and will therefore be taken into account in the typology) if that construction has the semantic function of assigning a graded (i.e. non-identical) position on a predicative scale to two (possibly complex) objects.”, suom. kirjoittajan.
- vii. <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>
- viii. ”An adjective is a lexeme that denotes a descriptive property and that can be used to narrow the reference of a noun”, suom. kirjoittajan.
- ix. ”A standard negation construction is a construction whose function is to modify a verbal declarative main clause expressing a proposition p in such a way that the modified clause expresses the proposition with the opposite truth value to p, i.e. $\sim p$, or the proposition used as the closest equivalent to $\sim p$ in case the clause expressing $\sim p$ cannot be formed in the language, and that is (one of) the productive and general means the language has for performing this function.”, suom. kirjoittajan.
- x. Termi konnegatiivi viittaa kiellonmerkitsimen kanssa esiintyvään muotoon. Se on peräisin Janhuselta (1977).

- xi. Näissä tutkimuksissa käytetään ns. paralleelikorpuksia, joissa on kyse saman tekstin käännöksistä eri kielille. Useimmin käytettyjä tekstejä ovat suurelle määrälle kieliä käännetyt raamatun osat.
- xii. Termin alkuperästä ks. <http://www.glottopedia.de/index.php/Doculect>.
- xiii. Isoloivissa rakenteissa jokainen morfeemi on erillinen sana. Agglutinatiivisissa ja flekteeraavissa rakenteissa morfeemit yhdistyvät kaksi- tai useampimorfeemisiksi sanoiksi, mutta edellisissä morfeemit on helppo erottaa toisistaan, niiden muodossa ei ole vaihtelua eikä yhteen morfeemiin ole kumuloitunut useampia merkityksiä, kun taas jälkimmäisissä morfeemirajat ovat hämärtyneet, morfeemeilla on useampia muotovariantteja tai niihin on kumuloitunut useampia merkityksiä; polysynteettisessä rakenteessa sanat koostuvat suuresta määrästä morfeemeja ja vastaavat kokonaisia monimutkaisia lauseita.