

# Substantiivien taivutusjärjestelmän produktiivisuudesta

ALEXANDRE NIKOLAEV

## Väitöksenalkajaisitelmä Itä-Suomen yliopistossa 1. huhtikuuta 2011

Produktiivisuus on yksi kielen tärkeimpiä ominaisuuksia. Kielen produktiivisuus erottaa meitä muista eläinlajeista: vain ihminen pystyy tuottamaan ja ymmärtämään periaatteessa rajattoman määrän uusia viestejä. Esimerkiksi ameebojen kommunikointi on kemiallista, mehiläiset käyttävät niin sanottuja tanssikuvioita omassa kommunikoinnissaan, apinat pystyvät oppimaan elekieltä, mutta kaikki tämäntyyppinen viestintä perustuu poikkeuksetta samaan periaatteeseen: kaikki muut lajit käyttävät rajallisia, toisin sanoen epäproduktiivisia merkkijärjestelmiä.

Toinen tärkeä ero on kielen hierarkkinen järjestäytyminen. Kieliopin eri tasot (kuten fonologia, morfologia ja syntaksi) ovat kukin produktiivisia: eri äänteiden (foneemien), morfeemien, sanojen tai lausekkeiden mahdollisten yhdistelmien määrä on tähtitieteellinen. Mitä ylemmän tason osajärjestelmästä on kyse, sitä avoimempi se on.

Tutkimukseni käsittelee kielen morfologiaa ja erityisesti morfologista produktiivisuutta. Yhtäältä morfologinen produktiivisuus on monimutkainen, kompleksinen käsite, toisaalta tapaamme sen ilmentymiä erilaisissa arkitilanteissa. Lasten tekemät taivutusvirheet (esim. *satutin \*käsin*, kun pitäisi sanoa *satutin käden*) osoittavat tyypillisesti lasten pyrkimystä rakenteiden yksinkertaistamiseen ja yhdenmukaistamiseen. Epäproduktiivisen taivutuksen (*käsi* : *käden*) sijaan lapsi käyttää produktiivista mallia (*lasi* : *lasin*). Tämä yksinkertainen esimerkki kielenomaksumisesta heijastaa yleistä tendenssiä, jonka mukaan mitä vähemmän ensikielen morfologia on kehittynyt lapsella, sitä enemmän lapsi yrittää seurata morfologisia universaaleja. Yksi sellaisista universaaleista onkin tunnusmerkittömien kategorioiden suosiminen. *Lasi* on tässä tapauksessa tunnusmerkitön, sillä kaikki sen taivutuspäätteet yksikössä liittyvät muuttumattomaan vartaloon *lasi*. Vartalo ei siis muutu, ja näin sillä ei ole sananmukaisesti omaa tunnusmerkkiä. Tunnusmerkittömyys ymmärretään siis tietynlaisena semioottisena ja kognitiivisena helppoutena, joka johtuu ihmisen ominaisuuksista ja tavastamme hahmottaa maailmaa ja siten myös kieltä.

Kuitenkin siinä, missä lapsi yksinkertaistaa kieltä, myös kieli itse pyrkii yksinkertaistumaan. Osoituksena tästä ovat näiden kahden taivutustyyppin laajuudet: *lasi* : *lasin* -tyyppisiä nomineja on *Suomen kielen perussanakirjan* mukaan yli neljätuhatta, kun taas *käsi* : *käden* -tyyppisiä nomineja on ainoastaan hieman alle sata.

Fred Karlsson (1993: 48) kuvailee kielen rakennetta ja käyttöä seuraavasti: ”Kielen kaksi perusolottuvuutta ovat sen *rakenne* [– –] ja *käyttö* [– –]. Rakenne ja käyttö eivät ole toistensa vastakohtia tai vaihtoehtoja, vaan rakenne on olemassa, jotta kieltä voitaisiin käyttää tehokkaasti: rakenne palvelee käyttöä”. Taivutustyyppien kannalta tämä tarkoittaa, että jos meillä on tietty taivutustyyppi, joka on hankala (jollaisia kaikki epäproduktiiviset tyypit ovat), niin voisimme odottaa, että tämän tyyppin käyttö myös hankaloituu. Tässä tapauksessa rakenne on tavallaan esteenä kielen tehokkaalle käytölle. Meidän pitää siis olettaa tämän postulaatin olevan kaksisuuntainen: jos tietty rakenne on tiheässä käytössä, niin silloin itse käyttö palvelee rakennetta. Toisin sanoen tiheässä käytössä oleva rakenne pysyy kielessä riippumatta siitä, kuinka hankala se on. Esimerkkinä voi olla edelleen paradigma *käsi*, joka on kielessä erittäin tiheässä käytössä ja on siksi säilynyt meidän päiviimme saakka.

Suomen kieli on kuuluisa morfofonologisista vaihteluistaan. Karlsson (1983) luettelee niiden määräksi puolisen sataa. Nämä vaihtelut synnyttävät paljon muotovariantteja eli allomorfeja sekä sanan vartalossa että taivutuspäätteissä. Eri sanoissa toistuvat morfofonologiset vaihtelut mahdollistavat eri taivutustyyppien eriyttämisen. Käytämme ihmiselle ominaista analogian kykyä ja yhdistämme taivutukseltaan samantyyppiset sanat samoihin taivutusluokkiin.

Morfofonologisten vaihteluiden seurauksena on allomorfia, jota tavataan myös sanojen vartaloissa. Vartaloallomorfia onkin yksi taivutustyyppien järjestäytymisen kriteereistä. Sen sijaan taivutuspäätteissä esiintyvää allomorfiaa on tapana luokitella niin sanotuksi vapaaksi vaihteluksi. Sanakirjat sivuuttavat tällaisen ”vapaan vaihtelun” melkein kokonaan taivutustyyppien luokittelussa. Ne ikään kuin jättävät kielenpuhujan päätettäväksi, kumpi esimerkiksi taivutusmuodoista *palvelujen* tai *palveluiden* on parempi. Väitöskirjani empiiriset osiot näyttävät kuitenkin, että tämä vaihtelun väitetty vapaus on harhaa, sillä se on useimmiten strukturoitua. Tietyn taivutusmuodon paremmuuden takana on rakenne eikä ainoastaan impressionistinen vaikutelma. Kysymys on vain työkaluista, joiden avulla voimme tämän rakenteen nähdä ja kuvata. Siksi yritin käyttääkin tutkimuksissa mahdollisimman laajan spektrin työkaluja ja lähestymistapoja. *Palvelujen* vs. *palveluiden* paremmuuden selvittelyssä auttoi *optimaalisuusteoria*; sen sijaan esimerkiksi eri allomorfien semioottisen helppouden selvittelyssä auttoi *luonnollinen morfologia*. Edellinen teoria on tavallaan chomskylaisen generativismin ”tytär”, kun taas jälkimmäinen on syntynyt eurooppalaisessa kielentutkimustraditiossa. En siis näe ongelmaa siinä, että käytän usein vastakkaisiakin tutkimusparadigmoja, sillä kullakin niistä on omat vahvat puolensa ja useimmiten omat kohteensa. Analogiana voisi olla kahden rakennustyömiehen kiista siitä, kumpi työkalu on parempi – saha vai vasara. En ole tavannut yhtään rakentajaa, joka olisi valinnut ainoaksi työkaluksi vasaran ja onnistunut sen avulla rakentamaan talon.

Lukijalle voisi syntyä kysymys, miksi suomessa on epäproduktiivisia tyypejä. Tämä kysymys sisältää itse asiassa kaksi kysymystä: miten ne ovat syntyneet kieleen

ja miksi ne ovat edelleen olemassa kielessä? Epäproduktiiviset taivutustyyppit ovat syntyneet kieleen suhteellisen pitkän ajan kuluessa kielen käydessä läpi erilaisia muotoa koskeneita muutoksia. On tyyppillistä, että muutokset olivat aluksi erittäin yleisiä eli produktiivisia fonologisia muutoksia, jotka sitten muiden yleisten muutosten kanssa – kuten kielitieteellinen termi kuuluu – salaliitossa synnyttivät epäproduktiivisia muotoja. Onkin vaikea kuvitella sellaista kieltä, joka ikään kuin tarkoituksella synnyttäisi sellaisia elementtejä, joiden pääfunktio olisi vain ja ainoastaan hankaloittaa sen käyttöä. Kielen läpikäydessä erityyppisiä muotoa koskevia muutoksia jotkin produktiiviset tyyppit ovat siis äänneiden muutosten seurauksena muuttuneet epäproduktiivisiksi ja jääneet sen jälkeen kieleen tietynlaisina relikteinä.

Tähän asti olen puhunut taivutustyypeistä ikään kuin ne olisivat jokin kiistämättömän fakta tai itsestäänselvyys. Näin ei kuitenkaan ole. Aloitan väitöskirjani historiallisella katsauksella siihen, miten eri suomen kieliopeissa on kuvattu taivutustyyppiä. Ensimmäiset suomen kielioipit etsivät suomesta eri deklinaatioita ainoastaan sillä perusteella, että tällaisia oli latinan kielessä. Latinan kielen tapaan jokaisessa deklinaatiossa oli omat päätteensä. Latinan kielessä sanat *käsi* ja *lasi* olisivat kuuluneet samaan deklinaatioon, sillä kummankin sanan yksikön nominatiivi on *i*-loppuinen. Osittain latinan kielioppimallin dominoinnin seurauksena Wikströmistä (1832) ja Lönnrotista (1841–42) alkaen monet kielentutkijat ovat luopuneet deklinaatiojaosta kokonaan. Morfofonologia vaihteluita selitetään säännöillä, joita leimaa kuitenkin synkronisen ja diakronisen tason osittainen sekoitus. 1900-luvulla taas tapahtuu tietynlainen paluu taivutustyyppiin.

Kuinka monta eri taivutustyyppiä suomen nomineilla loppujen lopuksi on? On vieläkin tutkijoita, jotka pitävät taivutustyyppiä ylipäänsä tarpeettomana käsitteenä. Toinen äärimmäisyys on esimerkiksi *Nyky-suomen sanakirja*, joka erottaa peräti 82 eri taivutustyyppiä. Eriyttämällä taas pintamuotoja päästään kymmenkuntaan taivutustyyppiin. Voimme siis todeta, että taivutustyyppi on erittäin näkökulmariippuvainen käsite.

Haluaisin korostaa, että tutkimukseni tarkoituksena ei missään tapauksessa ollut esittää uutta taivutustyyppiklassifikaatiota tai kritisoida olemassa olevia. Käytän tutkimuksen empiirisissä osioissa *Suomen kielen perussanakirjan* tarjoamaa klassifikaatiota, joka on laajempi kuin minkä tahansa suomen kielioipin mutta suppeampi ja nykyaikaisempi kuin *Nyky-suomen sanakirjan* tarjoama.

Sanakirjojen paradigmatklassifikaatioiden vertailu valaisee tutkijan intuition ja korpuksen suhdetta. Näin laajasta korpuksesta tarkistettuna jotkin *Suomen kielen perussanakirjan* sanoista kuuluvat eri tyyppiin, kuin miten *Perussanakirja* niitä luokittelee.

Korpuksella on tutkimuksessani suuri merkitys. Käytin korpusta melkein kaikissa empiirisissä osioissa: sekä kielen rakenteen kuvauksen yhteydessä että esimerkiksi testien ärsykesanoja valitessani. Mitä suurempia sanojen massoja jouduin tarkistamaan korpuksesta, sitä automaattisempi laskentatapani oli. Silloin kun laskin esimerkiksi lemmataajuuksia tuhansille sanoille, käytin Kielipankin Lemmie2-ohjelmaa, joka mahdollisti omien Perl-skriptien käytön. Tästä automaattisuudesta johtuvat tietyt virheet ja epätarkkuudet, sillä mikään korpus ei ole ideaalinen – ei edes käsin jäsennetty ja koodattu. Näin korpuksessa saattaa esiintyä yksittäisiä virheitä. Kuitenkin artikkeleissa yritin tehdä kompromissin ja suhteuttaa määrän ja laadun sillä periaatteella, että mitä

tärkeämpi merkitys oli korpuksen antamalla tiedoilla, sitä enemmän puutuin aineiston käsin evaluointiin. Tärkeimmissä tapauksissa, kuten esimerkiksi testisanojen valinnassa kuudetta artikkelia varten, turvauduin arviointipaneeliin, joka koostui Itä-Suomen yliopiston yleisen kielitieteen opettajista ja opiskelijoista.

Kun puhutaan korpuksesta, haluaisin mainita myös leksikaaliset tietokannat, jotka *olisivat* erittäin tärkeitä psykolingvistiksessä ja ylipäänsä lingvistiksessä tutkimuksessa. Käytin tarkoituksella konditionaalia sanassa *olisivat*, koska tällä hetkellä tämäntyyppisiä tietokantoja löytyy vain ja ainoastaan muutamasta kielestä. Suomen kieli ei valitettavasti kuulu näiden joukkoon. Hiljattain on ilmestynyt suomen kielen WordNet (<http://www.ling.helsinki.fi/en/lt/research/finnwordnet/>), joka on erittäin tärkeä askel kohti suurta leksikaalista tietokantaa. Suomessa on myös käytettävissä ohjelma, joka laskee Turun Sanomien korpuksen perusteella sanojen bigrammi- ja trigrammitaajuu-  
det. Mutta jos lingvisti tarvitsee sellaisia muuttujia kuin esimerkiksi fonologisten naapureiden määrä, morfologinen perhekoko, taivutuksellinen entropia, hän joutuu laske-  
kemaan ne kaikki itse. Jos vähän ennakkoin tulevaa, niin väittäisin, että joku muukin huomaa tällaisen suomea koskevan tietokannan tekemisen ja julkaisun erittäin mielekkääksi tehtäväksi.

Miksi ylipäänsä pitäisi laskea sanoille lukuisia muuttujia? Vastaus on: Kielen prosessoinnin kuvaus edellyttää useamman muuttujan tuntemista, ennen kuin voimme selittää tai ennustaa, miten ja millä kriteereillä esimerkiksi sanat jakautuvat eri taivutustyyppeihin. Olisi nimittäin naiivia olettaa, että mentaalissa leksikossamme sanat ja niiden taivutustyytit noudattaisivat täsmälleen samoja periaatteita, joilla kuvaamme kielen rakennetta. Kielen kuvauksen kannalta on helpompi, jos merkitys ja muoto on eristetty toisistaan. Toisaalta, silloin kun haluamme ymmärtää, miten ihminen prosessoii kieltä ja miten sanat ovat jakautuneet meidän mentaalissa leksikossamme, huomaamme, että nämä muuttujat, joista osa on semanttisia, kuten esim. morfologinen perhekoko, ja osa taas leksikaalisia, kuten esimerkiksi bigrammitaajuus, ovat kaikki vuorovaikutussuhteissa keskenään. Näin merkityksen ja muodon jyrkkä jako on ainoastaan ikään kuin mukavuuskysymys kielen grammatikaalisessa kuvauksessa.

Deskriptiivisillä kielioppeilla on vahvat perinteet ja omat tehtävänsä. Ne eivät kuvaa mentaalista leksikkoa ja sen järjestäytymistä. Eikä niiden tarvitsekaan kuvata sitä, sillä eri tutkimuksissa saadun evidenssin perusteella voimme todeta, että mentaalinen leksikko sisältää paljon redundanssia ja memoroituja muotoja. Lisäksi kielen reaaliaikaisessa prosessoinnissa rakenteelliset rajat ovat sumeita. Niinpä emme voi siis odottaakaan, että ajallis-paikallinen mieli toimisi kuin kielioppikirjan lukujen ja kappaleiden jako. Täytyy myöntää, että termi *sumea raja* on jo pitkään ollut muodissa. Minusta tämä osoittaa enemmän meidän suhteellisen heikkoja tietojamme mentaalista leksikosta kuin sitä, että rajat ovat tosiaan sumeita eikä niitä voisi koskaan selvittää. Olemme vasta raapaisseet mentaalisen leksikon pintaa, ja näin kaikki johtopäätökset sen rakenteesta ja järjestäytymisestä ovat toistaiseksi pikemmin suuntaa antavia eivätkä missään tapauksessa aksiomia.

Siksi minusta on tärkeä jatkaa tutkimusta nimenomaan kokeellisempaan suuntaan eli tehdä enemmän behavioraalisia ja neurofysiologisia kokeita suomen morfologiasta. Tässä yhteydessä täytyy vähän kertoa koehenkilöistä. Useammassa väitöskirjan artik-

kelissa koehenkilöinä olivat neurologisesti terveet kielenpuhujat. Ainoastaan yhdessä artikkelissa (Niemi, Nikolaev & Hugdahl 2009) mukana oli kielihäiriöisiä koehenkilöitä (nimittäin SLI/FLI-puhujia). Heidän testituloksensa osoittavat, että heidän mentaalinen leksikkonsa noudattaa aivan toisenlaisia järjestäytymisperiaatteita kuin tyypillisesti kehittyneillä koehenkilöillä, sillä SLI/FLI-koehenkilöiden tuloksia emme voi selittää yhdelläkään käyttämällämme psykolingvistisellä muuttujalla. He kumminkin suoriutuvat arkitilanteissa kielellisesti normaalilla tavalla.

Täytyy huomauttaa, että emme (valitettavasti) voi mennä suoraan ihmisen päähän katsomaan, miten kieli on siellä järjestäytynyt. Normaaleilla kielenkäyttäjillä tehdyt kokeet voivatkin vain epäsuorasti valaista joitakin kohteita. Toisesta näkökulmasta katsottuna voimme huomata kielihäiriöissä sellaista, mitä emme välttämättä olisi huumanneet normaalissa populaatiossa. Siksi on tärkeää suunnitella ja toteuttaa lingvistiä kokeita myös lingvistiksi mielenkiintoisissa kielihäiriöpopulaatioissa. Jos jokin kielen komponentti synnyttää vaikeuksia tietyn neurologisen sairauden potilailla, saamme voida linkittää kyseisen kielen komponentin tiettyihin aivoalueisiin ja siten ehkä saada neuraalista lisävalaistusta kielen psykologisen prosessoinnin mallintamiselle.

## Lähteet

- KARLSSON, FRED 1983: *Suomen kielen äänne- ja muotorakenne*. Helsinki: WSOY.
- 1993: Lingvistiikka. – Eero Hyvönen, Ilkka Karanta & Markku Syrjänen (toim.), *Tekoälyn ensyklopedia* s. 47–52. Helsinki: Gaudeamus.
- LÖNNROT, ELIAS 1841–42: *Bidrag till Finska språkets Grammatik*. Helsinki: SKS. – SWK 3, 1836–1845.
- NIEMI, JUSSI – NIKOLAEV, ALEXANDRE – HUGDAHL, KENNETH 2009: Impaired attention to phonological input, dysfunctional lexical "frequency counter/s" and abnormal grammatical morphology. A possible causal chain? – Jordan Zlatev, Mats André, Marlene Johansson Falck & Carita Lundmark (toim.), *Studies in language and cognition* s. 257–271. Newcastle: Cambridge Scholars Publishing.
- Nykysuomen sanakirja*. Helsinki: WSOY 1951–1961.
- Suomen kielen perussanakirja*. Helsinki: Edita Oyj 1990–1994.
- SWK = *Suomen Wanhat Kieliopit* 1–10. Toim. Kalevi Wiik. Turku: Turun yliopisto, Fonetiikka 1987.
- WIKSTRÖM, M. W. 1832: *Försök till en Finsk Grammatik, framställande en enda Declination och en enda Conjugation*. Wasa. – SWK 2, 1818–1832.

**Alexandre Nikolaev: Suomen nominien taivutusjärjestelmän produktiivisuudesta: korpuslingvistisiä ja kokeellisia tutkimuksia.** Moniste. Joensuu: Itä-Suomen yliopisto 2011.

Kirjoittajan yhteystiedot: etunimi.sukunimi@uef.fi