



# SUOMEN NOMINIEN TAIVUTUKSESTA: RYTMI-, SIVUPAINO- JA AGGLUTINAATIOHYPOTEESIEN TESTAUSTA



uomen kieli sallii tietyissä nominaalisissa sanatyypeissä (erityisesti kolmitavuisissa) sijapäätteiden vaihtelua. Parhaiten tämä näkyy monikon genetiivin ja monikon partitiivin muodoissa, mistä ovat osoituksena sanakirjojen antamat paradigmuodot. Näissä listauksissa esitetään nomineista tyypillisesti seuraavat luokat: yksikön nominatiivi, genetiivi, partitiivi ja illatiivi; monikon genetiivi, partitiivi ja illatiivi. Suomen kielen perussanakirja antaa taivutustyyppien mallitaulukossa esimerkiksi seuraavat monikon genetiivin ja monikon partitiivin vaihtoehdot (kukin vaihtoehtojen ryväs toistettu sanakirjan mukaisessa järjestyksessä ja muodossa):

	2. taivutustyyppi	4. taivutustyyppi	13. taivutustyyppi	14. taivutustyyppi
monikon genetiivi	palvelujen palveluiden palveluitten	laatikkojen laatikoiden laatikoitten	katiskoiden katiskoitten katiskojen	solakoiden solakoitten solakkojen
monikon partitiivi	palveluja palveluita	laatikkoja laatikoita	katiskoita katiskoja	solakoita solakkoja

	6. taivutustyyppi		11. taivutustyyppi	
monikon genetiivi	papereiden papereitten paperien	banaanien banaaneiden banaaneitten	omenien omenoiden omenoitten	(omenojen) (omenain)
monikon partitiivi	papereita papereja	banaaneja banaaneita	omenia omenoita	(omenoja)

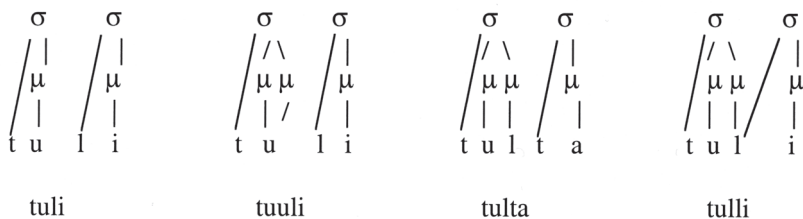
Voimme kysyä muun muassa, minkälaisesta hierarkiasta — jos hierarkiasta ollenkaan — on kyse kunkin vaihtoehtorypään sisällä. Onko ylempänä oleva variantti todella suositumpi kuin sen alla oleva, ja jos on, niin miksi? Vastaavatko tärkeysjärjestykset vaihtoehtojen yleisyyttä, kun tarkastellaan esimerkiksi laajoja korpuksia? Tämän tutkimuksen tarkoituksena onkin tarkastella edellä mainittuja morfologisesti rikkaita (siis suurta vaihtelua sisältäviä) paradigmoja sekä korpusanalyttisesti että optimaalisuusteoriaa (*optimality theory*) apuna käyttäen. Optimaalisuusteoria on yksi formaalin analyysin keinoista, ja mielestämme se sopii varsin hyvin vaihtelun kuvaajaksi. Samalla testamme optimaalisuusteorian rajoja ja pätevyyttä aineistomme vaihtelun kuvaajana.<sup>1</sup>

Taivutus on suomen kielessä epäilemättä keskeistä. Herääkin kysymys, miksi se ei ole esimerkiksi tarkastelumme kohteena olevissa paradigmoissa vakiintunut. Koska kielessä yhtenä vahvana funktionaalis-semioottisena periaatteena on merkityksen ja muodon isomorfia, voimme kuvitella, että myös näissä paradigmoissa kuitenkin tapahtuu vakiintumista. Yhtenä tämän tutkimuksen tarkoituksena on siten etsiä suuntaa, johon taivutus todennäköisimmin vakiintuu, ja antaa näennäiselle satunnaisuudelle nykytietämyksen mukaiset, mahdollisimman objektiiviset rajoitteet.

Analyysissa nojaudumme kolmeen suomen ja muiden kielten affiksaatiosta esitettyyn hypoteesiin: *rytmihypoteesiin*, *sivupainohypoteesiin* ja *agglutinaatiohypoteesiin* (ks. tarkemmin lukua Kohdemuotojen produktiivisuudesta). Tärkeä on myös *morán* käsite: suomessa morán muodostaa tavun ytimenä (*nucleus/peak*) oleva lyhyt vokaali. Sitä edeltävällä konsonantilla (*onset*) ei yleensä ole moraa, sillä se ei vaikuta mitenkään tavun rakuteen (ks. erilaiset painotusjärjestelmät). Ytimen jälkeisellä konsonantilla (*coda*) on kielestä riippuen joko yksi mora tai ei moraa laisinkaan (ks. *weight-by-position*-periaate, Hayes 1989). Mora on tässä tapauksessa vain ja ainoastaan tavun ytimen ominaisuus. Suomen morájärjestelmää voidaan kuvata seuraavalla puurakenteella ( $\mu$  = mora,  $\sigma$  = tavu)<sup>2</sup>:

<sup>1</sup> Tämä tutkimus on osa vuonna 1997 alkanutta kansainvälistä monikielistä mentaalisen leksikon yhteishanketta nimeltä Words in the minds, words in the brain: An integrated multidisciplinary study of the mental lexicon across languages and populations (vastuullinen tutkija professori Gary Libben, yleinen kielitiede, Albertan yliopisto, ks. esim. <http://www.criugm.qc.ca/mentalexiconf2000/group.html>). Hankkeen kielivalikoimaan kuuluvat muiden muassa arabia, bulgaria, englantia, heprea, italia, farsi, flaami, puola, ranska, ruotsi, saksa ja suomi (suomen osalta vastuulliset tutkijat ovat Matti Laine Åbo Akademiasta ja Jussi Niemi Joensuun yliopistosta). Kiitämme myös artikkelin referointia rakentavista parannusehdotuksista. Kieliasua koskevista huomautuksista kiitämme Stefan Werneria.

<sup>2</sup> Puurakennetta ehdotti artikkelin toinen referee, mistä kiitämme häntä. Hyvä lähde morarakenteista on esimerkiksi Kenstowicz 1994.



Moran avulla tavut jaetaan joko heikkoihin tai raskaisiin (vahvoihin) (ks. esim. Tranel 1991; Blevins 1995; Broselow 1995). Heikkoksi suomessa on perinteisesti katsottu yksimorainen tavu eli tavu, joka on siis V tai CV -rakenteinen; useampimoraiset on katsottu vahvoiksi (Sadeniemi 1949; Karlsson 1983: 134).

### VARHAISEMPIA HAVAINTOJA VAIHTELUSTA

Kenttämorfologiassaan Paunonen (1976: 99) tarkastelee vaihtelun syitä taivutustyyppien välisen (syntagmaattisten) vaikutusyhteyksien kautta. Hän tarkastelee eri sijoissa esiintyvää päätevarianttien vaihtelua suhteessa sellaisiin vartalo-tyyppien ominaispiirteisiin kuin vokaalistruktuuriin ja painosuhteisiin. Alaviitteessä hän kuitenkin mainitsee, että kirjoituksen tarkoitus ei ole selvittää suomen taivutustyyppien ominaispiirteiden järjestelmää vaan esittää, miten eri päätevarianttien distribuutiosuhteet valaisevat taivutustyyppien välisiä suhteita (Paunonen 1976: 98). Kiinnostuksemme kohteena olevien kolmitavuisien nominien Paunonen näkee paradigmaattisilta ominaisuuksiltaan rinnastuvan lähinnä kaksitavuisiin nomineihin (esimerkiksi *talojen – harakkojen*). Ainoa kaksi- ja kolmitavuja erottava paradigmaattinen ominaispiirre on siten painosuhteiden erilaisuus: kaksitavuisissa monikkovartalon lopputavu on painoton (esimerkiksi *talojen*), kolmitavuisissa painollinen (esimerkiksi *harakkojen*). Monikon partitiivi- ja genetiivityypit yhdistävät kuitenkin kolmitavuja yksitavuihin ja supistumanomineihin (esim. *maita, harmaiden*): *harakoita, laatikoiden*. Kolmitavuiset muodot, joiden monikkotaivutuksessa esiintyy painollinen diftongi (*harakoi-, laatikoi-*), voivatkin kielitajussa assosioitua yksitavuisille ja supistumanomineille ominaiseen taivutukseen, joten vokaalistruktuuriin verrattuna painosuhteiden osuus on toissijainen. (Paunonen 1976: 99.)

Paradigmaattista vaihtelua tarkastelee myös Itkonen monikon illatiivin astevaihtelun yhteydessä (1979). Itkonen suoritti niinkin varhain kuin vuonna 1956 kokeen, jossa koehenkilöinä oli 144 Helsingin yliopiston opiskelijaa. Koehenkilöiden piti valita kahdesta muodosta mielestään »parempi»; kaikkiaan tehtäväsarja käsitti 142 rinnakkaismuotoista sanaparia (kolmitavuiset sanat monikon partitiivin, genetiivin ja illatiivin muodossa). Kokeen tuloksissa Itkonen (1979: 47) huomasi kaksi ilmeistä säännönmukaisuutta: 1) *a, ä, o, (ö)* -vartaloisissa lyhytpenultimaisissa sanoissa, joissa siis toinen tavu on heikko, enemmistönä ovat äännelailliset muodot (*itA-* ja *iden-*loppuiset, esimerkiksi *myymälöitä, myymälöiden*); 2) pitkäpenultimaisissa sanoissa, joissa toinen tavu on siis vahva, analogiamuodot (*jA-* ja *jen-*loppuiset) ovat vartalon loppuvokaalista riippumatta suositumpia kuin lyhytpenultimaisissa — ja äännelaillisiin muotoihin verraten melkein aina enemmistönä (esimerkiksi *alustoja, alustojen*).

Karlsson (1983) toteaa noin 1,3 miljoonan saneen tekstikorpuksen (Suomen Kuvalehti) perusteella, että CV-loppuisilla monitavuilla on yleensä partitiivin *JA*-päätteen rinnalla vaihtoehtoinen *itA*-muoto. Jälkimmäinen on suosituampi, kun vartalon toiseksi viimeinen tavu on lyhyt, ja edellinen silloin, kun toiseksi viimeinen tavu on pitkä. Korpuksessa on toki 150 (13 %) *JA*-tapausta, joiden vartalopenultima on lyhyt. Itkosen tavoin Karlsson pitää näiden yhteisenä nimittäjänä perusmuodon loppuvokaalin suppeutta (ks. myös Anttila 1997 ja tekstiämme alla). Monitavuisissa monikon genetiivin *-iden* ja *-jen* jakaantuvat Suomen Kuvalehden aineistossa suurin piirtein samalla tavalla kuin monikon partitiivissakin.

Väitöskirjassaan (1997) Arto Anttila tarkastelee suomen monikon genetiivin variaatiota optimaalisuusteorian keinoin. Tämä lähestymistapa on hyvin suosittu myös tutkimusaiheemme kannalta olennaisen suprasegmentaalisen rakenteen (eli rytmin ja painon) tutkimuksessa (esim. Prince ja Smolensky 1993; McCarthy ja Prince 1993; Kager 2000). Optimaalisuusteoria on sellaista näkemystä vastaan, jonka mukaan jokainen pintamuoto generoidaan — sanan kielipölyssä mielessä — tarkasti säädetyn sääntöketjun kautta. Sen sijaan optimaalisuusteoria sallii kaikkien mahdollisten muotojen edustumat, joita se sitten analysoi tietynlaisen suodattimen avulla. Suodatin koostuu hierarkkisesti rakennetuista rajoitteista ja valitsee reaalisesti pintamuodoksi sen, joka rikkoo ylemmän tason rajoitteita vähemmän kuin muut ehdokkaat.<sup>3</sup>

Soveltaessaan väitöskirjassaan optimaalisuusteoriaa myös tässä käsittelemiimme rakenteisiin Anttila näkee suomen nominien taivutuksen variaation syytä sanojen rytmisessä rakenteessa. Analyysimme onkin tavallaan Anttilan tutkimuksen (1997) jatke, sillä myös me tarkastelemme suomen nominien taivutuksen variaatiota käyttöyleisyyden ja optimaalisuusteorian avulla. Anttilan käytössä oli suppea Suomen Kuvalehden vuoden 1987 korpus, joka sisältää noin 1,3 miljoonaa sanetta, kun taas meidän korpusanalyysimme perustuu yli sataan miljoonaan saneeseen (ks. lähemmin lukua Menetelmä). Lisäksi tarkastelemme myös monikon partitiivin esiintymiä. Vielä yksi olennainen ero tutkimustemme välillä on siinä, että tarkastelemme nomineja jaettuina eri homogeenisiin ryhmiin (taivutustyyppiin). Nämä ryhmät vuorostaan jakautuvat joko kahteen tai neljään alaryhmään sanahahmon (tavujen raskauden eli mora-arvojen) perusteella.

Anttila (1997) luokittelee vahvoiksi taivutusmuodoiksi esimerkiksi aitokonsonanttiset *iden*, *itten* ja *ten* -loppuiset monikon genetiivin muodot ja heikoiksi vastaavasti vokaalipohjaiset *en* ja *jen* -loppuiset. Mikäli ei-kaksitavuisen vartalon viimeinen tavu on vahva (*puu*, *potilas*), niin monikon genetiivin päätte on aina vahva (*-iden*, *-itten*). Kaksitavuiset vartalot (*kala*) sekä suuri osa vartaloista, joiden tavujen määrä on parillinen ja viimeinen tavu heikko (*margariini*), saavat monikon genetiivissä heikon päätteen. Näin ollen suurehkoa vaihtelua esiintyy paritontavuisissa, vähintään kolmitavuisissa, CV-tavuun päättyvissä vartaloissa (*naapureiden* ~ *naapurien*).

<sup>3</sup> Olisi täysin intuition vastaista, että kyseinen periaate kuvaisi nopeaa ajallis-paikallista kielentuottoa. Psykolingvistinen kirjallisuus ei tunne kielentuottamisen mallia, jossa puhuja generoisi periaatteessa kaikki mahdolliset muodot, joita sitten enemmän informaatiota sisältävä suodatin («yliminä») siistisi.

Anttila (1997: 12) huomauttaa, ettei vapaaksi nimitetty vaihtelu ole edes naiivien puhujien intuitioiden perusteella täysin vapaata, sillä nämä asettavat usein jonkin variantin tois(t)en edelle mutta ei-lingvisteinä perustelevat oman valintansa impressionistisilla tekijöillä, kuten sillä, että toinen variantti »kuulostaa paremmalta» kuin toinen (ks. esim. Itkonen 1979: 40). Puhujien tiedostamaton yksimielisyys joidenkin muotojen »paremmuudesta» ajan oloon näkyy siinä, että suosituimmalla variantilla on lingvistin käsiin tulevasa korpuksessa suurempi taajuus kuin »huonommalta kuulostavalla» variantilla.

Kolmitavuisten CV-loppuisten vartaloiden käsittelyssä Anttilan (1997) käyttämä suodatin koostuu seuraavista rajoitteista:

- a) Vältä painon ja moran ristikkäisyyttä (mts. 14).
- b) Suppeaan vokaaliin (*i, u, y*) päättyvät vartalot suosivat heikkoa varianttia; väljään vokaaliin (*ä, a*) päättyvät suosivat vahvaa varianttia, ja puolisuppeaan vokaaliin (*o, ö*) päättyvien vartaloiden valinta ei ole yksiselitteinen (mts. 15, 22).

Anttila korostaa sonorisuuden ja tavun mora-arvon korreloivan keskenään ja perustelee tätä näiden luonnollisella läheisyydellä, sillä mitä sonorisempi (avoimempi) vokaali on, sitä pidempi se universaalisti on. Esimerkiksi suomessa lyhyiden monoftongien keskot ovat painollisessa tavussa seuraavat: suppeat 9,5 cs, väliset 10,0 cs ja väljät 10,3 cs (Wiik 1965, taulukko 35, s. 115).

- c) Kolmanneksi viimeisen ja toiseksi viimeisen tavun on oltava erilaisia mora-arvoiltaan (rytmi). (Anttila 1997: 22.)

Anttilan myöhemmässä artikkelissa (2002) tarkasteltavana kieliaineiksena on sellaisen suomen kielen kolmitavuisten sanojen monikon (inessiivin) taivutus, joiden vartalot päättyvät väljään vokaaliin *a*. Nykysuomen sanakirjan aineistoon perustuvassa analyysissään Anttila päätyy mielenkiintoisiin tuloksiin, joissa muun muassa tendenssit kohti muutosta *a > o* (tyypissä *karsina*), kohti *a:n* ja *o:n* synkronista vaihtelua (esimerkiksi tyypissä *ikkuna*) ja vokaalin katoa (esimerkiksi tyypissä *glaukooma*) voidaan esittää kohdevokaalia edeltävän konsonantin ja penultimatavun vokalismien fonologisten piirteiden [korkea] ja [pyöreä] dissimilaation avulla eli pakollisen eron periaatteen avulla (*obligatory contour principle*) (esim. Leben 1973, 1980; McCarthy 1986; ks. tarkemmin s. 538). Mikäli fonologiset syyt ovat heikkoja, ratkaisuun vaikuttavat lisäksi ekstrapfonologiset eli morfologiset ja leksikaaliset eli esimerkiksi sanaluokkaeroihin liittyvät tekijät.<sup>4</sup>

Tutkimuksemme teemaa sivuaa myös Anttilan (1999) esittämä optimaalisuusteorian sovellus. Siinä tekijä käsittelee sananloppuisten *ea:n* ja *ia:n* monoftongiutumista *ee:ksi*

---

<sup>4</sup> Anttilan morfologian emergenssiin liittyvä yleistyys kuuluu seuraavasti (mts. 14): »The Emergence of Morphology: Extraphonological (morphological, lexical) conditions emerge in environments where the phonological conditions are at their weakest.» Ekstrapfonologisista ehdoista on jo G. Karlsson (1978) todennut, että kohdetyypeistä adjektiivit pyrkivät suomessa katoedustumaan (esimerkiksi *kihara : kiharissa*), kun taas perusmuodossa homonymiset substantiivit pyrkivät *a ~ o* -vaihteluun (*kihara : kiharoissa*).

ja vastaavasti *ii*:ksi. Anttila kehittää kolme optimaalisuusteorian rajoitteiksi muunnettua hypoteesia:

1. *ea* monoftongiutuu todennäköisemmin kuin *ia*, esimerkiksi *hopea* ~ *hopee* ja *rasia* ~ *\*rasii*.
2. Monoftongiutuminen on todennäköisempää ei-perusmuodoissa kuin perusmuodoissa, esimerkiksi *mini-ä* ~ *minii* (esimerkiksi autonmerkki *Mini*) ja *miniä* ~ *\*minii*.
3. Monoftongiutuminen on todennäköisempää adjektiiveissa kuin substantiiveissa, esimerkiksi *korea* ~ *koree* ja *Korea* ~ *\*Koree*.

Käyttämällä Paunosen Helsingin puhekielen aineistoa Anttila osoittaa, että Paunosen aineiston määrällinen jakauma tukee näitä rajoituksia.

## KOHDEMUOTOJEN PRODUKTIIVISUUDESTA

### MENETELMÄ

Produktiivisuus on kielessä melko sumea käsite (ks. esim. Dressler 1998). Yksi klassinen tapa on tarkastella kilpailevien muotojen produktiivisuutta niiden pintamuotojen esiintymisyleisyyden kautta. Muita näkökulmia produktiivisuuteen ovat muun muassa kategorian umpinaisuus eli potentiaalinen soveltuminen uusiin leksemeihin (esim. Nikolaev 2002), sanan tai leksemin vartalon tai kannan käyttö leksikossa yleensä (ns. perheen koko, engl. *family size*, Schreuder ja Baayen 1997). Perheen koolla mitataan niitä tapauksia, joissa sana esiintyy johdetuissa muodoissa ja yhdyssanoissa (esimerkiksi *kana*, *kanala*, *kanavaras*). Tässä analyysissä tukeudumme perinteiseen frekvenssinäkemykseen ja jätämme yhdyssanat laskujemme ulkopuolelle. Yksi syy valintaamme on suomen yhdyssanojen suuri produktiivisuus; suomessahan voimme muodostaa *ad hoc* -yhdyssanoja käyttötärpeen mukaan lähes rajattomasti (vaikkapa *leghornkanavaras* tai *kanavaraspidätys*). Täten perheeseen soveltaminen olisi tuonut analyysiin erittäin avoimen luokan eli yhdyssanat. Halusimme pitäytyä analyysissamme suomen leksikon ytimessä eli morfologisesti yksinkertaisissa perussanoissa ja leksikaalistuneissa johdoksissa.

Tutkimuksen kohteena olevien taivutustyyppien nominien otantamenetelmässä sovelsimme kahta kriteeriä: (1) sanan pitää olla kolmitavuinen, ja (2) mikäli taivutustyyppiin leksikaalinen taajuus ylittää 1 000 sanan määrän, sen tulee esiintyä myös Suomen kielen taajuussanastossa (Saukkonen 1979). Näin esimerkiksi toiseen taivutustyyppiin (*palvelu*) kuuluvien yhdistämättömien sanojen määrästä 1 172 taajuussanastossa esiintyy 178, joista 125 on kolmitavuisia. Tutkimme kunkin sanan taajuutta Kielipankin aineistossa (<http://www.csc.fi>) laskemalla, kuinka monta esiintymää sanalla on monikon genetiivissä *jen-*, *iden-* ja *itten-*loppuisena ja monikon partitiivissa *jA-* sekä *itA-*loppuisena. Emme ottaneet mukaan niitä esiintymiä, joissa oli liitepartikkeli tai possessiivisuffiksi (esimerkiksi *palvelujakin*), sillä lisätavu saattaa sekoittaa rytmihypoteesia. Lisäksi näitä esiintymiä on vähän (keskimäärin alle 5 %), eikä niitä ole pienen taajuuden takia mahdollista jakaa omiin homogeneisiin alaryhmiinsä saati sitten selvittää niiden vaikutusta aineiston kolmitavuisien nominien taivutusmuotoihin.

▷

Korpuksena on siis käytetty Kielipankkiin koottuja ja koodattuja suomenkielisiä tekstejä, pääasiassa lehtitekstejä<sup>5</sup>, joiden yhteinen saneiden määrä on 109 341 835 (elokuu 2003). Analyysissa kaikki sanat jaetaan tavuittain siten, että yksimoraiselle tavulle annetaan arvo *heikko* (H) ja useampimoraiselle tavulle arvo *vahva* (V). Näin esimerkiksi 2. taivutustyyppiin kuuluvat kolmitavuiset sanat jakautuvat neljään ryhmään tavun arvon mukaan: HHH (esim. *pu.he.lu*), VHH (esim. *koh.ta.lo*), VVH (esim. *kil.pai.lu*), HVH (esim. *ti.las.to*). Kaikissa tutkituissa ryhmissä perusmuodon eli yksikön nominatiivin viimeinen (eli kolmas) tavu on siis aina yksimorainen (H). Monikon genetiivin ja monikon partitiivin päätteet voivat joko jättää tämän tavun muuttamatta (*pu.he.lu.jen*, *pu.he.lu.ja*, *koh.ta.lo.jen*, *koh.ta.lo.ja* jne.) tai muuttaa sen laadun siten, että siitä tulee useampimorainen (V) (*pu.he.lui.den*, *pu.he.lui.ta*, *koh.ta.loi.den*, *koh.ta.loi.ta* jne.). Näin kukin yllä mainittu ryhmä (HHH, VHH, VVH ja HVH) jakautuu monikon genetiivin ja monikon partitiivin taivutuksessa kahtia kolmannen tavun moratyyppin perusteella:

	<i>pu.he.lu</i>		<i>kil.pai.lu</i>	
	HHX		VVX	
	/ \		/ \	
mon. gen.	HHHV	HHVV	VVHV	VVVV
	<i>pu.he.lu.jen</i>	<i>pu.he.lui.den</i>	<i>kil.pai.lu.jen</i>	<i>kil.pai.lui.den</i>
mon. part.	HHHH	HHVH	VVHH	VVVH
	<i>pu.he.lu.ja</i>	<i>pu.he.lui.ta</i>	<i>kil.pai.lu.ja</i>	<i>kil.pai.lui.ta</i>
	<i>koh.ta.lo</i>		<i>ti.las.to</i>	
	VHX		HVX	
	/ \		/ \	
mon. gen.	VHHV	VHVV	HVHV	HVVV
	<i>koh.ta.lo.jen</i>	<i>koh.ta.loi.den</i>	<i>ti.las.to.jen</i>	<i>ti.las.toi.den</i>
mon. part.	VHHH	VHVH	HVHH	HVVH
	<i>koh.ta.lo.ja</i>	<i>koh.ta.loi.ta</i>	<i>ti.las.to.ja</i>	<i>ti.las.toi.ta</i>

#### RYTMIHYPOTEESI

Suomen kielen metrisen järjestelmän on perinteisesti ajateltu olevan syllabinen trokee (esim. Hayes 1995). Kanoninen syllabinen trokee on kvantitatiivisesti tasapainoinen ja koostuu kahdesta tavusta, joista ensimmäinen on painollinen (vrt. *jalan* ja Sadeniemen (1949) *puhetahdin* käsitteet).

Esittämämme rytmihypoteesi poikkeaa kuitenkin ajattelusta, jonka mukaan *jalan* (tässä tapauksessa trokeen) sisäinen tavurakenne ja tasapaino ovat ainoat ratkaisutekijät. Oletamme nimittäin rytmihahmotuksen ulottuvan trokeerajan — eli Sadeniemen *puhetahdin* rajan — yli (aineistomme sanoissa trokeeraja kulkee 2. ja 3. tavun välissä). Näin sana-

<sup>5</sup> Aamulehti 1995, 1999; Demari 1995, 1997–2000; Hämeen Sanomat 1999–2000; Hyvinkään Sanomat 1994, 1997; Iltalehti 1996; Kaleva 1998–1999; Kangasalan Sanomat; Karjalainen 1991–1995, 1997–1999; Keski-suomalainen 1999; Kustannusosakeyhtiö Otava 1993; Tekniikan Maailma 1995–1997; Turun Sanomat 1998–1999.

ryhmät VVH ja HVH voidaan kaavamaisesti esitellä samalla tavalla, sillä ryhmässä VVH pitkä toinen tavu pystyy pelkästään suhteellisen painokkuutensa voimalla heikentämään ensi tavun (V) asemaa (Itkonen 1979: 51). Tätä väitettä havainnollistaa karkeasti seuraava kaavio (jossa »X» voi olla joko heikko tai vahva tavu):

$V \nearrow V \searrow H \quad X$	$H \nearrow V \searrow H \quad X$
(kir    jas    to    ja)	(ti    las    to    ja)
(kir    jas    to    jen)	(ti    las    to    jen)

Vahva toinen tavu siis ikään kuin määrää kolmannen, joka vuorostaan määrää päätteen valinnan: *-ja* ja *-jen* (sitien ensimmäisen tavun mora-arvolla ei ole merkitystä ja neljäs tavu on täysin altis kolmannen tavun mora-arvolle). Näin vartalon penultimatavu (tapauksessamme toinen tavu) ohjaa päätteen valintaa.

Samoin on laita ryhmissä HHH ja VHH, joissa vartalon penultimatavu on rytmisesti heikko, mikä puolestaan vahvistaa seuraavaa tavua:

$H \searrow H \nearrow V \quad X$	$V \searrow H \nearrow V \quad X$
(nu    me    roi    ta)	(koh   ta    loi    ta)
(nu    me    roi    den)	(koh   ta    loi    den)

Vaikka *numero*-tyypissä kaksi ensimmäistä tavua ovat heikkoja, ensimmäinen tavu on kuitenkin pääpainollisena kakkostavua »vahvempi». Rytmihypoteesimme kannalta kriittisiä ovat siis ei-perifeeriset tavut. Kaavamaisesti tämä näyttää seuraavalta:

$\diagup \quad X.X.X.X \quad \diagdown$	
produktiivinen sanahahmo	epäproduktiivinen sanahahmo
$X.V.H.X$ <i>ti.las.to.ja</i> <i>ti.las.to.jen</i>	$X.V.V.X$ <i>ti.las.toi.ta</i> <i>ti.las.toi.den</i>
$X.H.V.X$ <i>koh.ta.loi.ta</i> <i>koh.ta.loi.den</i>	$X.H.H.X$ <i>koh.ta.lo.ja</i> <i>koh.ta.lo.jen</i>

Tämä rajoite esiintyy alla olevissa taulukoissa (ks. lukua Tulokset) muodossa

\* $\sigma_{\mu n} \sigma_{\mu n}$  »Vältä kahden tasavahvuisen tavun vierekkäisyyttä.»

▷



Rajoite muistuttaa tasavahvuisten viereisten elementtien välttämisen periaatetta (*obligatory contour principle*, esim. Leben 1973, 1980; McCarthy 1986). Emme kuitenkaan pidä käyttämämme rajoitetta tämän periaatteen ilmentymänä, sillä jälkimmäinen ei koske ainoastaan toista ja kolmatta tavua, vaan mitä tahansa kahden tavun jonoa. Esittämämme rytmihypoteesin mukaan tarkasteltavien sanojen ensimmäinen ja toinen tavu voivat olla tasavahvuisia, samoin kolmas ja neljäs tavu. Rajoitetta pitää siis tarkentaa: »Vältä kahden tasavahvuisen tavun vierekkäisyyttä trokeejalan rajalla», mikä on suomen tapaisille trokeekielille harmonisuusperiaate.<sup>6</sup>

#### SIVUPAINOHYPOTEESI

Toisena tekijänä, jonka voidaan ihmiskielen yleisiin periaatteisiin nojautuen olettaa rytmien rinnalla vaikuttavan tuloksiin, on *sivupaino*. Suomen sivupainon määräytymisessä on tärkeätä havaita, että pitkissä sanamuodoissa (joissa tavuluku on viisi tai enemmän), sivupaino lankeaa kolmannelle (eli viisitavuisten muotojen antepenultima-asemassa olevalle) tavulle, jos kolmas ja neljäs tavu ovat heikkoja mora-arvoltaan (esimerkiksi *sá.ta.mà.na.kin*). Jos neljäs tavu on kuitenkin vahva ja edeltävä tavu heikko, sivupaino siirtyy neljännelle tavulle (eli viisitavuisten penultimatavulle, esimerkiksi *sá.ta.ma.kin.ko*). (Suomen sivupaino ei koskaan lankea sananloppuiselle tavulle.)

Sadeniemen (1949: 69) mukaan suomen sivupaino on hahmotuskysymys. Samoin kuin neljän ryhmässä annetut objektiivisesti tasaiset ärsykkeet tavallisesti jakautuvat ryhmiin 2 + 2, suomessa niin sanottu *puhetahti* on yleensä kahden tavun mittainen: se alkaa (sivupainollisesta tavusta ja käsittää seuraavat tavut painolliseen tavuun saakka. Mikäli rytmihypoteesin perusteella vältettäviä (epäproduktiivisia) sanahahmoja ei ole (taivutus-tyypit 4, *laatikko*, ja 14, *kännykkä*), niin xVHx ja xHVx -hahmojen välillä oletamme jälkimmäisen olevan produktiivisempi: useampimorainen tavu on luonnostaan painokkaampi kuin yksimorainen, ja koska kaikilla aineistomme taivutusmuodoilla sivupainollinen tavu on kolmas, voidaan olettaa sen pyrkivän olemaan pikemminkin vahva (siis useampimorainen) kuin heikko: xHVx, mikä on myös universaali periaate (*attraction of stress by a heavy syllable*, Kager 1995). Taulukoissa tämä tunnusmerkkisyysrajoite esiintyy muodossa

\* $\sigma_{\mu}$     »Vältä heikkoja sivupainollisia tavuja.»

#### AGGLUTIONAATIOHYPOTEESI

Erityisesti kielenomaksumisen tutkimuksessa on havaittu, että omaksumisen varhaisvaiheessa lapset hallitsevat varhemmin tai paremmin sellaiset muodot, joissa morfologinen yhdistäminen perustuu agglutinaatioon eli muodosteisiin, joissa ei esiinny vartalonvaihtelua tai morfofonologisia muutoksia. Lisäksi agglutinaatio yleistyy helposti muihin affiksaatiotyyppeihin (spontaanista tuotoksesta ja agglutinaatiohypoteeseista ks. Niemi ja Niemi 1987, suomen kokeista Niemi 1999).

<sup>6</sup> Suomen kielen painotuksesta ks. esim. McCartney 1999.

Agglutinaatiohypoteesia voidaan soveltaa aineistossamme ainoastaan 4. ja 6. taivutustyyppin kohdalla, sillä vain näissä aineistomme tyypeistä taivutusmuodot voivat olla sekä agglutinatiivisia (*laatikko* : *laatikkojen*, *paperi* : *paperien*) että ei-agglutinatiivisia (*laatikko* : *laatikoiden*, *paperi* : *paperiden*). Tyyppin *paperien* sanojen luokittelu agglutinoinviksi saattaa ensi näkemältä vaikuttaa kiistanalaiselta. Tiukoin kriteerein muotohan ei ole puhtaasti agglutinoinva. Kuitenkin puhuja (ei lingvisti), joka operoi pintamuodoilla, voi mieltää sen seuraavalla tavalla koostuvaksi: *paper-ilen*. Näin siis salkkumorfiksi tässä analyysissä katsottu *-en* edustaisi sekä monikollisuutta että genetiivin sijaa. Itkonen (1979) perusteleeikin tyyppin *paperien* suosiota tietynlaisella hahmotustendenssillä, nimittäin pyrkimyksellä valita vaihtoehtoja, jotka jäsenyvät mahdollisimman jyrkkäräjäisesti vaihtelemattomaan sanavartaloon ja vaihtelemattomaan pääteainekseen. Lisätukea ratkaisumme saa Niemen, Järvikiven ja Häggin (1997) suorittamista kokeista, joissa koehenkilöt (lingvistisesti naiivit puhujat) katkoivat morfologisesti erilaisia sanoja vetäen kynällä pystyviivan yksittäisen esiintyvän sanan sellaiseen kohtaan, jossa he katsoivat »suurimman sanan sisäisen rajan sijaitsevan». <sup>7</sup>Yksi tulos tällaisista kokeista on (sekä englannista että suomesta), että tavalliset kielenkäyttäjät katkovat sanoja usein ei-lingvistisesti, muun muassa hahmottaen vartalon tai kannan päättyvän *paperilen* -tapaisen jaon mukaisesti.

Agglutinaatiohypoteesia ilmaiseva rajoitus on muotoa

\* ei-aggl. »Vältä vartalonmuutoksia suffiksaatiossa.»

## TULOKSET

Lopullisista tuloksistamme puuttuu 11. taivutustyyppi eli *omena*-paradigma, koska CD-Perussanakirjan mukaan tämä luokka sisältää ainoastaan 46 sanaa (yhdyssanoja lukuun ottamatta). Näistäkin kymmenen esiintymän (monikon genetiivissä ja monikon partitiivissa) rajan ylittää vain 11 sanaa, jotka vuorostaan jakautuvat sanahahmoltaan neljään ryhmään. Näin alhaisten lukujen perusteella on mahdotonta laskea luotettavaa frekvenssijakaumaa.

Tulokset on alla esitetty optimaalisuusteoriaa ja sen notaatioita käyttäen. Rajoitteet ovat siis:

\* $\sigma_{\mu n} \sigma_{\mu n}$  »Vältä kahden tasavahvuisen tavun vierekkäisyyttä.»

\* $\sigma_{\mu}$  »Vältä heikkoja sivupainollisia tavuja.»

\* ei-aggl. »Vältä vartalonmuutoksia suffiksaatiossa.»

Näiden keskinäinen järjestys on \* $\sigma_{\mu n} \sigma_{\mu n}$  > \* $\sigma_{\mu}$  > \* ei-aggl (vasemmalla tärkeämpi, oikealla vähemmän tärkeä). Mikäli ehdokas rikkoo rajoitteen, se merkitään asteriskilla. Asteriskia seuraava huutomerkki taulukoissa tarkoittaa, että ehdokas rikkoi tärkeimmän rajoitteen ja että »voittaja» on valittu jo siinä vaiheessa. Näin suluisissa olevat asteriskit eivät enää vaikuta voittajan valintaan.

Optimaalisuusteorian keinoin pyrimme siis ennustamaan, kumpi vaihtelun kahdesta osapuolesta on suositumpi esitettyjen hypoteesien kannalta (hypoteesit on muunnettu rajoitteiksi). Kunkin muototyyppin kohdalla esitämme myös leksikaalisen frekvenssin (eli

<sup>7</sup> Tekijät soveltavat suomenkieliseen aineistoon Bruce Derwingin englannin kielessä tekemiä »sanojen pätkimiskokeita» (ks. esim. Derwing 1992).

montako sanaa kuuluu CD-Perussanakirjan mukaan kuhunkin muototyyppiin). Kielipankin avulla lasketut pintafrekvenssit<sup>8</sup> (joiden prosentit ovat tärkeämpiä ja informatiivisempia kuin pelkät lukumäärät, ks. selitys Menetelmä-alaluvussa) tukevat optimaalisuusteoria-analyysiamme.

Prosenttisuhteet on alla esitettävissä kuvioissa ja taulukoissa laskettu siten, että monikon genetiivin kaikille esiintymille jokaisen sanan (X-muuttuja) kohdalla on annettu arvo 100 %, samoin monikon partitiivin esiintymille (tällä tavalla kaikki X<sub>i</sub> -muuttujat on standardoitu), ja sitten näistä prosenteista on laskettu ryhmän yhteiset prosentit. Prosenttisarakeessa ilmoitetut luvut (ks. taulukot alla) lasketaan siis prosenteista (tarkemmin sanottuna jokaisen muuttujan kahden variantin — esimerkiksi *paperien* vs. *papereiden* — prosenttisuhteesta) eikä suoraan pintafrekvensseistä (siksi N = tietyn ryhmän pintafrekvenssien summa %). Analysoimme siis standardoituja lukuja (sarake %) emmekä suoraan pintafrekvenssejä (sarake N). Jos kohdesanan kaikkien monikon genetiivin ja monikon partitiivin esiintymien yhteinen pintafrekvenssi on alle kymmenen, sitä ei lasketa tuloksissa; muutoin yllä mainitulla tavalla standardoitu muuttuja (jonka N on alle 10) rinnastettaisiin muihin muuttujiin (sanoihin), joiden frekvenssi on suurempi ja näin luotettavampi.

### Toinen taivutustyyppi

Aloitamme aineiston läpikäynnin toisesta taivutustyyppistä (eli *palvelu*-paradigmasta), josta astevaihtelu ja loppuvokaalin vaihtelu puuttuvat ja joka on muita laajempi. Tulos on seuraava:

-jen	67,5 %	vs.	-idenlitten	32,5 %	(27 897 vs. 6 009)
-jA	67,6 %	vs.	-itA	32,4 %	(24 022 vs. 7 695)

Tulos ei kerro sinänsä muuta kuin sen, että *-jen* ja *-jA* -loppuiset muodot ovat suositumpia kuin *-idenlitten* ja *-itA* -loppuiset ja että kumpikin sija käyttäytyy samalla tavalla. Paljon kuvaavampaa tietoa saadaan hyödyntämällä moran käsitettä.

**Taulukko 1.** Toisen taivutustyyppin (*palvelu*) aineisto optimaalisuusteorian mukaisesti analysoituna.

1.1) HHx (*pu.he.lu*). Leksikaalinen taajuus: 13

ehdokkaat	*σ <sub>μn</sub> σ <sub>μn</sub>	*σ <sub>μ</sub>	*ei-aggl.	%	N
1.1.1) Monikon genetiivi					
HHV ( <i>puheluiden</i> )				71,7	(690)
HHH ( <i>puhelujen</i> )	*!	(*)		28,3	(521)
1.1.2) Monikon partitiivi					
HHV ( <i>puheluita</i> )				76,1	(1 075)
HHH ( <i>puheluja</i> )	*!	(*)		23,9	(895)

<sup>8</sup> Tiettyjen taivutusmuotojen käyttöyleisyys.

1.2) VHx (*koh.ta.lo*). Leksikaalinen taajuus: 34

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
1.2.1) Monikon genetiivi VHV ( <i>kohtaloiden</i> )				74,9	(5 175)
VHH ( <i>kohtalojen</i> )	* !	(*)		25,1	(8 784)
1.2.2) Monikon partitiivi VHV ( <i>kohtaloita</i> )				75,7	(6 529)
VHH ( <i>kohtaloja</i> )	* !	(*)		24,3	(8 192)

1.3) VVx (*kil.pai.lu*). Leksikaalinen taajuus: 56

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
1.3.1) Monikon genetiivi VVV ( <i>kilpailuiden</i> )	* !			4,1	(136)
VVH ( <i>kilpailujen</i> )		(*)		95,9	(13 291)
1.3.2) Monikon partitiivi VVV ( <i>kilpailuita</i> )	* !			2,1	(81)
VVH ( <i>kilpailuja</i> )		(*)		97,9	(11 645)

1.4) HVx (*ti.las.to*). Leksikaalinen taajuus: 22

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
1.4.1) Monikon genetiivi HVV ( <i>tilastoiden</i> )	* !			0,2	(8)
HVH ( <i>tilastojen</i> )		(*)		99,8	(5 301)
1.4.2) Monikon partitiivi HVV ( <i>tilastoita</i> )	* !			0,3	(10)
HVH ( <i>tilastoja</i> )		(*)		99,7	(3 290)

Kuten taulukosta 1 näkyy, frekvenssit puoltavat optimaalisuusteorian ennustamaa tulosta, tarkemmin sanoen rytmihypoteesin mukaisen taivutuksen produktiivisuutta: esimerkiksi HHV (*puheluiden, puheluita*), VVH (*kilpailujen, kilpailuja*).

Monikon genetiivin ja monikon partitiivin välillä frekvensseissä ei ole merkitsevää eroa, mikä tarkoittaa, että neljännen tavun mora-arvolla (V/H) ei myöskään ole eroa frek-

venssien jakautumisen kannalta. Siksi kuvion mora-arvot kattavat vain kolme ensimmäistä tavua: HHH, HHV ja niin edelleen (pro HHHH / HHHV, HHVH / HHVV jne.). Neljän-  
nen eli viimeisen tavun näkymättömyys viittaa lopputavun ekstrametrisyys -ilmiöön (*final  
consonant extrametricality*; ks. esim. Kenstowicz 1994).

Rytmiyhypoteesin kannalta »huonoista» vaihtoehdoista tyytit *puhelujen/puheluja*  
(HHH) ja *kohtalojen/kohtaloja* (VHH) vetävät puoleensa noin neljäsosan esiintymistä, kun  
taas tyytit *kilpailuiden/kilpailuita* (VVV) ja *tilastoiden/tilastoita* (HVV) ovat selvästi epä-  
suosittuja. Tämä ero selittyy mielestämme sonorisuuden periaatteella: kustakin ryhmästä  
omana alaryhmänään nimittäin erottuvat *U*-loppuiset nominit (*puhelu, kysely* jne.). Jos  
erotetaan *U*-loppuiset nominit kustakin ryhmästä, niin taulukon 1 frekvenssit muuttuvat  
seuraavasti (monikon genetiivin ja monikon partitiivin esiintymät liitetty yhteen; plus tar-  
koittaa, että rytmiyhypoteesi puoltaa ehdokasta, ja miinus, että ehdokas rikkoo rytmiyhypo-  
teesin asettaman rajoitteen) (taulukko 2).

**Taulukko 2.** Toisen taivutustyyppin aineisto eriteltynä *U*-loppuisiin ja ei-*U*-loppuisiin sa-  
noihin.

	HHx			VVx	
	HHH	HHV		VVH	VVV
- <i>U</i> ( <i>puhelu</i> )	75 %	25 %	( <i>kilpailu</i> )	97,6 %	2,4 %
muut ( <i>numero</i> )	5,1 %	94,9 %	( <i>kirjasto</i> )	95,8 %	4,2 %
	VHx			HVx	
	VHH	VHV		HVH	HVV
- <i>U</i> ( <i>palvelu</i> )	67,5 %	32,5 %	( <i>kokeilu</i> )	99 %	1 %
muut ( <i>kohtalo</i> )	5,2 %	94,8 %	( <i>varasto</i> )	99,9 %	0,1 %
Rytmi:	-	+		+	-

Tulokset tukevat Itkosen ja Karlssonin tekemiä havaintoja siitä, että suppeaan *U*-  
vokaaliin päättyvät sanat käyttäytyvät taivutuksessa eri tavalla kuin esimerkiksi *O*-vokaali-  
liin päättyvät sanat, joiden penultimatu on lyhyt. Tässä pätee myös Anttilan käyttämä  
rajoite, jonka mukaan suppeaan vokaaliin (*i, u, y*) päättyvät vartalot suosivat heikkoa va-  
rianttia (*jen/ja*) (ks. lukua Varhaisempia havaintoja vaihtelusta). Muiden kuin *U*-loppuis-  
ten sanojen kohdalla pelkästään rytmiyhypoteesi näyttää tukevan frekvenssien jakautumis-  
ta. *U*-loppuisten sanojen kohdalla suosituin on analoginen muoto, joka rytmiyhypoteesin  
kannalta on kuitenkin epäproduktiivinen vaihtoehto. Suppea loppuvokaali *U* siis on tässä  
tapauksessa vaikutukseltaan tärkeämpi kuin rytmiyhypoteesin antama ideaalihakmo xHVx.

#### *Kolmastoista taivutustyyppi*

Kolmannentoista taivutustyyppin nominit ovat *a*-loppuisia. Monikon genetiivissä ja mo-  
nikon partitiivissa *a* nousee *o*:ksi (esimerkiksi *vaahtera* : *vaahteroita*). Koska nominien  
määrä on tässä taivutustyyppissä huomattavasti pienempi kuin toisessa (CD-Perussana-  
kirjassa 153 nominia ilman yhdyssanoja), ainoana otannan kriteerinä oli sanan kolmi-

tavuisuus (142 sanaa). Erona taivutustyyppien välillä on ainoastaan se, että 13. taivutustyyppissä sanojen loppuvokaali *a* muuttuu *o*:ksi; muuten 13. taivutustyyppi vastaa 2. taivutustyyppiä sanahahmojakaumaltaan: kaikki sanat jakautuvat neljään ryhmään. Jakauma on siis samanlainen kuin edellisen alaluvun taulukossa (vrt. taulukkoja 1 ja 3).

**Taulukko 3.** Kolmannentoista taivutustyyppin (*vaahtera*) aineisto optimaalisuusteorian mukaisesti analysoituna.

3.1) HHx (*si.ka.la*). Leksikaalinen taajuus: 31

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
3.1.1) Monikon genetiivi HHV ( <i>sikaloiden</i> )			(*)	99,6	(315)
HHH ( <i>sikalojen</i> )	* !	(*)	(*)	0,4	(1)
3.1.2) Monikon partitiivi HHV ( <i>sikaloit</i> a)			(*)	85,3	(213)
HHH ( <i>sikaloj</i> a)	* !	(*)	(*)	14,7	(15)

3.2) VHx (*vaah.te.ra*). Leksikaalinen taajuus: 22

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
3.2.1) Monikon genetiivi VHV ( <i>vaahteroiden</i> )			(*)	86,7	(127)
VHH ( <i>vaahterojen</i> )	* !	(*)	(*)	13,3	(113)
3.2.2) Monikon partitiivi VHV ( <i>vaahteroit</i> a)			(*)	88,8	(108)
VHH ( <i>vaahteroj</i> a)	* !	(*)	(*)	11,2	(51)

3.3) VVx (*sai.raa.la*). Leksikaalinen taajuus: 39

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
3.3.1) Monikon genetiivi VVV ( <i>sairaaloiden</i> )	* !		(*)	35,3	(972)
VVH ( <i>sairaalojen</i> )		(*)	(*)	64,7	(295)
3.3.2) Monikon partitiivi VVV ( <i>sairaaloit</i> a)	* !		(*)	44,9	(455)
VVH ( <i>sairaaloj</i> a)		(*)	(*)	55,1	(245)

▷

3.4) HVx (*ka.tis.ka*). Leksikaalinen taajuus: 22

ehdokkaat	* $\sigma_{\mu n} \sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
3.4.1) Monikon genetiivi					
HVV ( <i>katiskoiden</i> )	* !		(*)	34,1	(59)
HVH ( <i>katiskojen</i> )		(*)	(*)	65,9	(108)
3.4.2) Monikon partitiivi					
HVV ( <i>katiskoita</i> )	* !		(*)	37,8	(84)
HVH ( <i>katiskoja</i> )		(*)	(*)	62,2	(145)

Kuten toisessakin taivutustyyppissä, tässä paradigmassa rytmihypoteesi näyttää ensisijaiselta frekvenssien jakauman selittäjältä. Morfologisesti produktiivinen on siis se vaihtoehto, jota puoltaa rytmihypoteesi. Pintafrekvenssien prosentit tukevat jälleen optimaalisuusteorian ennustamaa tulosta.

Rytmihypoteesiin huonoimmin sopivien alaryhmien frekvenssit jakautuvat hieman eri tavalla kuin toisessa taivutustyyppissä. VVV- (*sairaaloiden/sairaaloita*) ja HVV- (*katiskoiden/katiskoita*) alaryhmät ovat suosittumia vaihtoehtoja HHH- (*sikalojen/sikaloja*) ja VHH- (*vaahterojen/vaahteroja*) alaryhmiin verrattuna. Näin oletus, että kolmas tavu pyrkii sivupainollisena olemaan mora-arvoltaan vahva, näyttää pitävän paikkaansa. Emme kuitenkaan voi esittää selitystä sille, miksi vasta tästä taivutustyyppistä alkaen kolmannen tavun painollisuus nousee esiin tarkasteltaessa prosenttilukuja, kun toisessa taivutustyyppissä (ks. taulukko 1) sillä ei ole merkitystä.

Sijalla ei näissäkään tuloksissa ole merkitsevää eroa. Poikkeuksena on ryhmä HHx (*sikala*), jonka alaryhmässä HHH monikon partitiivin esiintymiä on enemmän kuin monikon genetiivin esiintymiä (14,7 % vs. 0,4 %). Tähän vaikuttaa kuitenkin yhden lekseemin eli *peseta*-sanan pintafrekvenssi (*pesetoja* [15] vs. *pesetoita* [2]). Sana on ryhmässä ainoa *tA*-loppuinen, ja sen taivutusta voidaan selittää analogialla: muiden ryhmien kaikki *tA*-loppuiset nominit suosivat nimittäin myös *jen-* ja *JA*-päätteitä (esimerkiksi ryhmässä VVx *keskusta* : *keskustojen* [145] vs. *keskustoiden* [0], *keskustoja* [28] vs. *keskustoita* [0] ).

*Kuudes taivutustyyppi*

Otannan kriteerit olivat 6. taivutustyyppissä (*paperi*) samat kuin 2. taivutustyyppissä käytetyt. Näin CD-Perussanakirjassa esiintyvien (6. tyyppiin kuuluvien) sanojen määrä 1 142 (ilman yhdyssanoja) supistui 165:een (Taajuussanastossa esiintyvään) sanaan, joista 143 on kolmitavuisia.

Kuudes taivutusparadigma eroaa yllä tarkastelluista tyypeistä siten, että monikon genetiivissä ja monikon partitiivissa on seuraava oppositio:

- en vs. -eiden/eitten (*paperien* vs. *papereiden/papereitten*)
- ejA vs. -eitA (*papereja* vs. *papereita*).

Sanahahmoltaan tyyppin taivutusmuodot jakautuvat samaan tapaan kuin 2. ja 13. taivutustyyppin muodot. Kuudes taivutustyyppi on *i*-loppuinen, ja sen tuloksia tarkasteltaes-

sa voidaan soveltaa kolmatta hypoteesia, nimittäin agglutinaatiohypoteesia, sillä monikon genetiivin *en*-loppuiset muodot ovat — omaksumamme tulkinnan mukaan (ks. yllä) — agglutinoivia (*paperien*) ja muut ei-agglutinoivia (*papereiden*, *papereja*, *papereita*) (taulukko 4).

**Taulukko 4.** Kuudennen taivutustyyppin (*paperi*) aineisto optimaalisuusteorian mukaisesti analysoituna.

4.1) HHx (*pa.pe.ri*). Leksikaalinen taajuus: 22

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
4.1.1) Monikon genetiivi HHV ( <i>papereiden</i> )			(*)	36,2	(2 777)
HHH ( <i>paperien</i> )	* !	(*)		63,8	(2 684)
4.1.2) Monikon partitiivi HHV ( <i>papereita</i> )			(*)	86,6	(3 934)
HHH ( <i>papereja</i> )	* !	(*)	(*)	13,4	(554)

4.2) VHx (*lää.kä.ri*). Leksikaalinen taajuus: 80

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
4.2.1) Monikon genetiivi VHV ( <i>lääkäreiden</i> )			(*)	52,3	(6 568)
VHH ( <i>lääkärien</i> )	* !	(*)		47,7	(4 821)
4.2.2) Monikon partitiivi VHV ( <i>lääkäreitä</i> )			(*)	98,5	(6 954)
VHH ( <i>lääkärejä</i> )	* !	(*)	(*)	1,5	(181)

4.3) VVx (*ken.raa.li*). Leksikaalinen taajuus: 26

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
4.3.1) Monikon genetiivi VVV ( <i>kenraaleiden</i> )	* !		(*)	5	(225)
VVH ( <i>kenraalien</i> )		(*)		95	(3 199)
4.3.2) Monikon partitiivi VVV ( <i>kenraaleita</i> )	* !		(*)	24,6	(856)
VVH ( <i>kenraaleja</i> )		(*)	(*)	75,3	(2 115)

▷



4.4) HVx (*ki.vää.ri*). Leksikaalinen taajuus: 15

ehdokkaat	* $\sigma_{\mu n} \sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
4.4.1) Monikon genetiivi HVV ( <i>kivääreiden</i> ) HVH ( <i>kiväärien</i> )	* !	(*)	(*)	2,2 97,8	(32) (2 852)
4.4.2) Monikon partitiivi HVV ( <i>kivääreitä</i> ) HVH ( <i>kiväärejä</i> )	* !	(*)	(*) (*)	15,5 84,5	(66) (1 918)

Rytmiyhypoteesi on jälleen primaarinen tämänkin paradigman sisäisen vaihtelun ja frekvenssijakauman selityksessä (taulukko 4).

*Paperien*-ehdokas rikkoo ensisijaisen rajoitteen \* $\sigma_{\mu n} \sigma_{\mu n}$ , mutta sitä tukee agglutinaatorajoite. Optimaalisuusteoria ei kuitenkaan salli hierarkiassa kolmannella sijalla olevan rajoitteen nostamista etusijalle *paperi*-tyypissä. Ongelman voisi ratkaista muokkaamalla koko analyysia siten, että agglutinaatorajoite tulisi ensisijaiseksi, mutta siinä tapauksessa analyysi ei pystyisi enää selittämään yhtäkään muuta varianttia, joten pitäydymme omaksumassamme rajoitehierarkiassa. Vaihtoehtoinen ratkaisu voisi olla *paperi*-ehdokkaan jättäminen analyysin ulkopuolelle, mihin emme kuitenkaan löydä perusteltua syytä. Siksi voimme todeta vielä yhden poikkeuksen teoriasta: ensisijaisen rajoitteen hylkäämä ehdokas *paperien* esiintyy taajimpaan.

Muissa ryhmissä tuloksillemme tyypillinen hierarkia rytmi > paino > agglutinaatio ei horju. Korkeimmat taajuudet on niillä alaryhmillä, joille sekä rytmiyhypoteesi että agglutinaatiohypoteesi antavat positiivisen arvon (VVH, esim. *kenraalien*; HVH, esim. *ki-väärien*). Vaihtoehdoissa, joissa agglutinaatio ei vaikuta (*papereja, papereita, kenraaleja, kenraaleita, lääkärejä, lääkäreitä, kiväärejä, kivääreitä*), sivupainohypoteesi näyttää olevan toissijainen tekijä. Toisin sanoen pareissa, joissa molemmat ehdokkaat rikkovat rytmirajoitteen (*papereja* vs. *kenraaleita* ja *lääkärejä* vs. *kivääreitä*), hivenen suositumpi on variantti, jonka kolmas tavu on mora-arvoltaan vahva (*kenraaleita, kivääreitä*).

#### *Neljäs taivutustyyppi*

Neljännän taivutustyyppin (*yksikkö*) ero yllä tarkasteltuihin verrattuna on se, että se sisältää kvantitatiivista astevaihtelua. Otannan kriteerit olivat samanlaiset kuin 13. taivutustyyppissä (*vaahtera*) käytetyt, ja näin CD-Perussanakirjan mukaan neljänteen taivutustyyppiin kuuluvista sanoista 263 (luku ei sisällä yhdyssanoja) on jäänyt 190 kolmitavuista tarkasteltavaksi.

Hahmoltaan sanat jakautuvat tässä paradigmassa kahteen pääryhmään: VVx (130 sanaa, esim. *yksikkö*) ja HVx (60 sanaa, esim. *kolikko*), jotka vuorostaan jakautuvat alaryhmiin:

VV<sub>x</sub> → VVH (*yksikköjen*) vs. VHV (*yksiköiden*); HV<sub>x</sub> → HVH (*kolikkojen*) vs. HHV (*kolikoiden*) (taulukko 5).

**Taulukko 5.** Neljännen taivutustyyppin (*yksikkö*) aineisto optimaalisuusteorian mukaisesti analysoituna.

5.1) VV<sub>x</sub> (*yk.sik.kö*). Leksikaalinen taajuus: 130

ehdokkaat	*σ <sub>μn</sub> σ <sub>μn</sub>	*σ <sub>μ</sub>	*ei-aggl.	%	N
5.1.1) Monikon genetiivi VHV ( <i>yksiköiden</i> ) VVH ( <i>yksikköjen</i> )		* !	(*)	71,7 28,3	(2 400) (856)
5.1.2) Monikon partitiivi VHV ( <i>yksiköitä</i> ) VVH ( <i>yksikköjä</i> )		* !	(*)	76,1 23,9	(2 813) (679)

5.2) HV<sub>x</sub> (*ko.lik.ko*). Leksikaalinen taajuus: 60

ehdokkaat	*σ <sub>μn</sub> σ <sub>μn</sub>	*σ <sub>μ</sub>	*ei-aggl.	%	N
5.2.1) Monikon genetiivi HHV ( <i>kolikoiden</i> ) HVH ( <i>kolikkojen</i> )		* !	(*)	52,1 47,9	(297) (249)
5.2.2) Monikon partitiivi HHV ( <i>kolikoita</i> ) HVH ( <i>kolikkoja</i> )		* !	(*)	55,1 44,9	(456) (205)

Sivupainohypoteesi on ensisijainen, ja agglutinaatiohypoteesi on toiseksi todennäköisin jakauman selittäjä. Tulos eroaa siis muiden taivutustyyppien (2., 13., 6.) tuloksista: rytmihypoteesia on mahdotonta enää soveltaa, sillä yllä olevasta kaaviosta puuttuvat rytmin kannalta vältettävät hahmot HHH, VVV, VHH ja HVV.

Sijoilla ei ole nykyään merkitsevää eroa. Ensimmäisen tavun moratyypillä on kuitenkin merkitystä neljännen taivutustyyppin tuloksissa: ryhmä HV<sub>x</sub> (*kolikko*) ei yksiselitteisesti suosi sivupainohypoteesin mukaista taivutusta (HHV, *kolikoiden* tai *kolikoita*), vaan prosentit jakautuvat melko tasaisesti kahden ryhmän välillä (HVH vs. HHV). HVH-ryhmä (*kolikkojen/kolikkoja*) on näin suositumpi kuin VVH-ryhmä (*yksikköjen/yksikköjä*). Jos verrataan näitä sanahahmoja keskenään, niin HVH-hahmo vastaa periaatetta, jon-

ka mukaan kieli välttää identtisiä tai samankaltaisia vierekkäisiä elementtejä (*obligatory contour principle*). VVH-sanahahmo taas rikkoo sen. Kuitenkaan muissa ryhmissä ei ole merkitsevää eroa ensimmäisen tavun mora-arvoissa (V/H), joten tätäkin tulosta voimme pitää poikkeuksena.

#### Neljästoista taivutustyyppi

Neljännestätoista taivutustyyppissä (*kännykkä*) tapahtuu astevaihtelua (vrt. 4. taivutustyyppi, *laatikko*) ja myös loppuvokaali A:n nousua O:ksi (vrt. 13. taivutustyyppi, *katiska*). Otannan ainoana kriteerinä oli (kuten 13. taivutustyyppissäkin) sanan kolmitavuisuus. Näin taivutustyyppin 231 sanasta (CD-Perussanakirja) jäi tarkasteltavaksi 227 sanaa. Kuten edellisessä, 4. paradigmassa, sivupainohypoteesi on tässäkin ensisijainen ja selittää yksin frekvenssijakauman (taulukko 6). Tulokset jakautuvat siis nyt sivupainohypoteesin mukaisesti. Neljännen tavun mora-arvoissakaan (V/H) ei ole merkitsevää eroa.

**Taulukko 6.** Neljäntoista taivutustyyppin (*kännykkä*) aineisto optimaalisuusteorian mukaisesti analysoituna.

6.1) VVx (*kän.nyk.kä*). Leksikaalinen taajuus: 117

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
6.1.1) Monikon genetiivi VHV ( <i>kännyköiden</i> ) VVH ( <i>kännykköjen</i> )			(* (*	99 1	(1 019) (9)
6.1.2) Monikon partitiivi VHV ( <i>kännyköitä</i> ) VVH ( <i>kännykköjä</i> )		* !	(* (*	99,8 0,2	(1 771) (4)

6.2) HVx (*te.lak.ka*). Leksikaalinen taajuus: 110

ehdokkaat	* $\sigma_{\mu n}$ $\sigma_{\mu n}$	* $\sigma_{\mu}$	*ei-aggl.	%	N
6.2.1) Monikon genetiivi HHV ( <i>telakoiden</i> ) HVH ( <i>telakkojen</i> )			(* (*	95,1 4,9	(1 124) (59)
6.2.2) Monikon partitiivi HHV ( <i>telakoita</i> ) HVH ( <i>telakkoja</i> )		* !	(* (*	95,7 4,3	(1 216) (42)

## PÄÄTELMÄ

Päätulokseksi toteamme, että optimaalisuusteorian keinoin asetetut rajoitteet selittävät vaihtelun näennäistä satunnaisuutta varsin hyvin. Luonnollista on, että jokaisesta teoreettisesta väittämästä on aina olemassa poikkeuksia.<sup>9</sup> Esittämämme analyysi sotiikin yhdessä tapauksessa optimaalisuusteorian periaatteita vastaan: muissa tapauksissa toimiva hierarkia horjuu *paperien*-ryhmässä, jossa ensisijaisen rajoitteen hylkäämä ehdokas saa kuitenkin suuremman frekvenssin. Jos oletetaan tarkoituksenamme olevan optimaalisuusteorian rajojen ja pätevyden testaaminen, täytyy siis myöntää, että yllä mainitussa kohdassa käyttämämme optimaalisuusteoreettinen viitekehys vuotaa. Vaihtoehtona on analyysin muunnos (esimerkiksi käyttämämme rajoitehierarkian muunnos), mutta tällöin syntyisi vielä enemmän tuloksia, joita analyysi ei pystyisi enää selittämään.

Myös frekvenssien jakauma alaryhmittäin puoltaa valittua strategiaa ryhmitellä sanahahmoja tavun raskauden (mora-arvoltaan vahva/heikko) perusteella. Olemme seuranneet tässä edeltäjiemme jälkiä (Paunonen 1976; Itkonen 1979; Karlsson 1983; Anttila 1997, 1999, 2002), mutta tuoneet myös uutta aikaisemmin esitettyyn tarkastelemalla nomineja homogeenisiin ryhmiin (so. taivutustyyppeihin) kuuluvina ja jakamalla kunkin ryhmän kahteen tai neljään alaryhmään sanahahmon perusteella. Laaja korpus (yli sata miljoonaa sanaa) on mahdollistanut luotettavien pintafrekvenssien laskemisen sekä monikon genetiivin että monikon partitiivin edustumien pohjalta. Näiden muotojen välillä esiintyy vahvaa korrelaatiota. Olemme soveltaneet analyysissa myös uutta rajoitehierarkiaa, jossa vahvemmassa heikompaan olemme testanneet rytmin, sivupainon ja agglutinaation vaikutusta paljon vaihtelua sisältäviin monikon genetiivin ja monikon partitiivin taivutusmuotoihin.

Rytmiä voidaan pitää sanan sisäisen hahmotuksen helpottajana ja siten sanan prosessoinnin yksinkertaisuuden motivaationa; optimaalisuusteorian hierarkiassa se osoittautuikin muita tekijöitä korkeammaksi. Prosodisen jäsentelyn (fraseerauksen) ja rytmin on osoitettu olevan myös puheen ymmärrettävyyden kannalta tärkeää (Klatt 1976; Nakatani ja Schaffer 1978). Näin ollen rytmihypoteesiamme voi pitää suomen kielessä produktiivisena periaatteena, joka rajoittaa pintamuotoja (esimerkiksi produktiivisen sanahahmon aineistomme sanojen kohdalla on oltava xVHx tai xHVx), minkä seurauksena vuorostaan on tietyn päätteen suosituimmuus.

Yksi Anttilan (1997) käyttämistä rajoitteista — vältä painon ja moran ristikkäisyyttä — jäi tässä analyysissa toissijaiseksi tekijäksi, joka muuttuu primaariseksi vasta sen jälkeen, kun kumpikaan ehdokas ei riko ensimmäistä rajoitetta (\* $\sigma_{\mu n} \sigma_{\mu n}$ ). Näin on laita esimerkiksi taivutustyypeissä 4 (esimerkiksi *yksiköiden* 71,7 % vs. *yksikköjen* 28,3 %) ja 14 (esimerkiksi *kännyköiden* 99 % vs. *kännykköjen* 1 %).

Agglutinaatio osoittautui useassa tapauksessa merkityksettömäksi tekijäksi siitä yksinkertaisesta syystä, että se joko on mukana kaikissa tietyn taivutustyyppin muodoissa (eli monikon genetiivissä ja monikon partitiivissa, taivutustyyppi 2) tai puuttuu kaikista tietyn taivutustyyppin monikon genetiivin ja partitiivin muodoista (taivutustyyppit 13 ja 14).

<sup>9</sup> Lukuun ottamatta sellaisia mielestämme triviaaleja perusväittämiä kuin »jokaisessa kielessä on vokaaleja ja konsonantteja».

Tapauksissa, joissa sillä näyttää olevan merkitystä tulosten jakauman selittäjänä (taivutustyyppi 4, *laatikko*, ja 6, *paperi*), voittaja on selvinnyt jo ennen kuin ehdokkaat evaluoidaan agglutinaatorajoitteen perusteella; näin rajoite ei varsinaisesti vaikuta voittajan valintaan. Sitä voidaan kuitenkin pitää toissijaisena tekijänä (poikkeuksena tyyppi *papereiden* vs. *paperien*), sillä huonoja ehdokkaita keskenään verrattaessa sillä on selvää vaikutusta taajuusjakaumaan (*kenraalien* [aggl.] 95 % [3 199] vs. *kenraaleja* [ei-aggl.] 75,3 % [2 115]).

Alussa esitettyjen oletusten lisäksi myös analogialla on tietyissä tapauksissa oma merkityksensä. Esimerkiksi 13. taivutustyyppin omassa ryhmässä (HHx) ainoa *ta*-loppuinen sana *peseta* saa 15 äännelaillista (*pesetoja*) esiintymää, mikä on ristiriidassa rytmihypoteesin ja ryhmän muiden sanojen kanssa. Tämän idiosynkrasian voimme kuitenkin selittää analogialla: muiden ryhmien *ta*-loppuiset sanat suosivat samaten *jen-* ja *ja*-päätteitä (vrt. *keskustojen/keskustoja* vs. *?keskustoiden/?keskustoita*).

Hieman yleisemmälle tasolle mennäksemme voimme todeta, että suomen nominien taivutusjärjestelmän muodostavat allomorfit ryhmittyvät eri taivutustyypeihin, joiden lukumäärä vaihtelee lähteestä riippuen aina 5:stä (Cannelin 1932) 85:een (Nykysuomen sanakirja). Ei ole kuitenkaan mitään absoluuttista taivutustyyppien luokittelua: jokainen on konventionaalinen ja riippuu siitä, kuinka tarkasti allomorfeja eriytetään paradigmoihin. Sanakirjojen tarjoama luokittelu on ensinnäkin tarkoitettu palvelemaan käytännöllisiä tarkoituksia (Karlsson 1983: 202). Käyttämämme Suomen kielen perussanakirjan paradigmaluokittelu on auttanut meitä tarkastelemaan sanoja suhteellisen homogeenisina ryhminä. Kukin tarkastelemamme paradigma sisältää sanoja, jotka jakautuvat paradigman sisällä joko neljään (taivutustyyppit 2, 4 ja 13) tai kahteen (taivutustyyppit 14 ja 6) alaryhmään mora-arvonsa perusteella. Näiden uusien ryhmien sisäistä järjestäytymistä (morfologista variaatiota) selittää kolme yllä tarkasteltua rajoitetta, jotka vuorostaan pohjautuvat todennäköisesti universaaliin tai suomen kielessä produktiiviseen ilmiöön ja joiden rikkominen taas johtaa pintamuodon suhteelliseen epäproduktiivistumiseen.

## LÄHTEET

- ANTTILA, ARTO 1997: *Variation in Finnish phonology and morphology*. Julkaisematon väitöskirja. Yleinen kielitiede, Stanfordin yliopisto.
- 1999: *Derived environment effects in colloquial Helsinki Finnish*. Käsikirjoitus, Boston University. (Myös: roa-406, Rutgers Optimality Archive, <http://roa.rutgers.edu/>)
- 2002: Morphologically conditioned phonological alternations. – *Natural language and linguistic theory* 20 s. 1–42.
- BLEVINS, JULIETTE 1995: The syllable in phonological theory. – John Goldsmith (toim.), *The handbook of phonological theory* s. 206–244. Cambridge, MA: Blackwell.
- BROSELOW, ELLEN 1995: Skeletal positions and moras. – John Goldsmith (toim.), *The handbook of phonological theory* s. 175–205. Cambridge, MA: Blackwell.
- CANNELIN, K. 1932: *Finska språket*. Helsingfors: Schildts.
- CD-Perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Helsinki: Kotimaisten kielten tutkimuskeskus 1997.

- DERWING, BRUCE 1992: A 'pause-break' task for eliciting syllable boundary judgements from literate and illiterate speakers: Preliminary results for five diverse languages. – *Language and speech* 35 s. 219–235.
- DRESSLER, WOLFGANG 1998: What is the core of morphology? – Jussi Niemi, Terence Odlin & Janne Heikkinen (toim.), *Language contact, variation, and change: Studies in languages* 32 s. 15–32. Joensuu: Joensuun yliopisto.
- HAYES, B. 1989: Compensatory lengthening in moraic phonology. – *Linguistic inquiry* 20 s. 253–306.
- 1995: *Metrical stress theory: Principles and case studies*. Chicago: The University of Chicago Press.
- ITKONEN, TERHO 1979: *Retkiä nykysuomeen*. Tarkistettu ja laajennettu laitos kirjasta »Nykysuomen tutkimusta ja huoltoa». Helsinki: Suomalaisen Kirjallisuuden Seura.
- KAGER, R. 1995: Book review of Bruce Hayes, *Metrical stress theory: Principles and case studies*. – *Phonology* 12 s. 437–464.
- 2000: *Optimality theory*. Cambridge: Cambridge University Press.
- KARLSSON, FRED 1983: *Suomen kielen äänne- ja muotorakenne*. Helsinki: WSOY.
- KENSTOWICZ, M. 1994: *Phonology in generative grammar*. Cambridge, MA: Blackwell.
- Kielipankki: <http://www.csc.fi>.
- KLATT, DENNIS 1976: Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. – *Journal of the Acoustical Society of America* 59 s. 1208–1221.
- LEBEN, W. R. 1973: *Suprasegmental phonology*. Väitöskirja, MIT.
- 1980: A metrical analysis of length. – *Linguistic inquiry* 10 s. 497–509.
- MCCARTHY, J. 1986: OCP Effects: Gemination and Antigemination. – *Linguistic inquiry* 17 s. 207–263.
- MCCARTHY, J. – PRINCE, A. 1993: *Prosodic morphology I: Constraint interaction and satisfaction*. Käsikirjoitus. University of Massachusetts, Amherst, ja Rutgers University.
- MCCARTNEY, S. 1999: Ternarity in Finnish Stress. – S. Smythe, S. McCartney & G. Coelho (toim.), *Proceedings of SCIL 11*. MIT Working Papers in Linguistics 45. Cambridge: The MIT Press.
- NAKATANI, LLOYD – SCHAFFER, JUDITH 1978: Hearing »words» without words: Prosodic cues for word perception. – *Journal of the Acoustical Society of America* 63 s. 234–245.
- NIEMI, JUSSI – NIEMI, SINIKKA 1987: Acquisition of inflectional marking: A case study of Finnish. – *Nordic journal of linguistics* 10 s. 59–89.
- NIEMI, JUSSI 1999: Production of grammatical number in specific language impairment: An elicitation experiment on Finnish. – *Brain and language* 68 s. 262–267.
- NIEMI, JUSSI – JÄRVIKIVI, JUHANI – HÄGG, MINNA 1997: Morphological and syllabic awareness in native speakers: Transparency of grammatical constructs. Kongressiesitelmä, 16th International Congress of Linguists, Pariisi, 20.–25. heinäkuuta 1997.
- NIKOLAEV, ALEXANDRE 2002: Eräiden suomen taivutustyyppien produktiivisuudesta. – *Puhe ja kieli* 22 s. 113–124.
- Nykysuomen sanakirja*. Helsinki: WSOY 1951–1961.
- PAUNONEN, H. 1976: Allomorfiin dynamiikkaa. – *Virittäjä* 80 s. 82–107.
- PRINCE, A. – SMOLENSKY, P. 1993: *Optimality theory: Constraint interaction in generative grammar*. Rutgers University, New Brunswick, ja University of Colorado, Boulder.

▷

- SADENIEMI, MATTI 1949: *Metriikkamme perusteet*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- SAUKKONEN, PAULI – HAIPUS, MARJATTA – NIEMIKORPI, ANTERO – SULKALA, HELENA 1979: *Suomen kielen taajuussanasto*. Helsinki: WSOY.
- SCHREUDER, R. – BAAYEN, R. H. 1997: How complex simplex words can be. – *Journal of memory and language* 37 s. 118–139.
- Suomen kielen perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 55. Helsinki: Kotimaisten kielten tutkimuskeskus 1990–1994.
- TRANEL, BERNARD 1991: CVC Light syllables, geminates and moraic theory. – *Phonology* 8 s. 291–302.
- WIIK, KALEVI 1965: *Finnish and English vowels*. Annales Universitatis Turkuensis B:94. Turku: University of Turku.

INFLECTION OF FINNISH NOMINALS:  
TESTING OF RHYTHM, SECONDARY STRESS  
AND AGGLUTINATION HYPOTHESES

The article analyses the free variation claimed to be typical in plural genitives and partitives of Finnish trisyllabic nominals. On the basis of a large text corpus (approx. 109 million words), the writers calculated the surface frequencies of lexemes in certain trisyllabic paradigms classified by the Suomen kielen perussanakirja ('A basic dictionary of Finnish') as containing highly variable competing forms. These were paradigm 2 (example: *palvelu* 'service'), paradigm 4 (*laatikko* 'box/drawer'), paradigm 13 (*katiska* 'weir/(fish) trap'), paradigm 14 (*solakka* 'slim'), paradigm 6 (*paperi* 'paper') and paradigm 11 (*omena* 'apple'). Each paradigm was then divided into sub-groups on the basis of the mora values of the syllables, since the writers assumed that what appeared to be free variation was not entirely free but guided by a number of principles that are reflected in, for example, syllable weight.

The variation was analysed with the aid of three hypotheses covering rhythm, secondary stress and agglutination. Each of these factors presents its own constraints on the eventual surface forms of the paradigms, and violations of these constraints are manifested as a lowered degree of productivity. This, in turn, is reflected in the preferential use of different forms by native speakers and through the relative frequency of use of the competing forms.

The rhythm hypothesis may be seen as an inherent principle of spoken Finnish, and it turned out to be the primary explanation in a majority of the instances studied. The secondary stress hypothesis was in some cases the primary and in others the secondary explanation of the variation at hand, while the agglutination hypothesis was applicable only to the inflectional paradigms of *laatikko* and *paperi*, where, in any case, its role remained secondary. By applying these hypotheses to an optimality theory analysis and comparing the outcome of this with the frequencies determined from the corpus analysis, the writers also investigate other structural factors affecting the rank order of the paradigmatic variants. ■

Kirjoittajien yhteystiedot (address):

*Yleinen kielitiede*

*Joensuun yliopisto*

*PL 111*

*80101 Joensuu*

Sähköposti: *alexandre.nikolaev@joensuu.fi*

*jussi.niemi@joensuu.fi*