

ALEXANDRE NIKOLAEV

JUSSI NIEMI

NOMINIEN PARADIGMAATTISTUMINEN SUOMESSA

MILLÄ RAKENTEELLISILLA EHDOKSILLA
KIELENKÄYTTÄJÄT SITOVAT POTENTIAALISET
NOMINIT TAIVUTUSLUOKKIIN?



aivutusmorfologian produktiivisuuden tutkimuksessa on Jean Berkon klassisen artikkelin (Berko 1958) jälkeen käytetty niin sanottua *wug*-testiä, jonka tyypillisessä muodossa naiiveille kielenpuhujille annetaan kohdekielen rakenteeseen soveltuvan potentiaalisen epäsanon ostensiivinen perusmuoto, esimerkiksi substantiiveista yksikön nominatiivi ja verbeistä yksikön 3. persoonan preesens. Pienille lapsille saatetaan koetilanteessa näyttää kuvallisia herätteitä, vanhemmille puhujille riittää yleensä pelkkä kielellinen tehtävänanto. Testattaessa esimerkiksi suomen *o*-loppuisten kaksitavuisten nominien monikkotaivutusta koehenkilö saa seuraavanlaisen ärsykkeen: »*Tässä on yksi peho*». Hänen tulee täydentää — joko suullisesti tai kirjallisesti — annettu kakkoslause »*Tuolla on paljon _____ » peho*-sanan »sopivalla muodolla». Imperfektitaivutuksen testaamisessa koeasetelmassa voi esiintyä seuraavanlaisia testiosioita: »*Maija jantaa joka päivä.*» »*Eilenkin hän _____.*» Epäsanojen käytön motivaationa on aitojen sanojen käyttämiseen liittyvien leksikaalis-semanttisten muuttujien välttäminen. Samalla minimoidaan sanojen käyttöyleisyyden vaikutusta leksikaaliseen prosessointiin (käyttöyleisyyden vahvasta vaikutuksesta sanojen prosessointiin ks. esim. Taft 1979).

Suomen ja ruotsin kieleen on aiemmin sovellettu *wug*-testiä muun muassa Joensuu yliopistossa (ks. J. Niemi 1999; J. Niemi 2006; J. Niemi ja Heikkinen 2000; J. Niemi, Heikkinen ja Järvikivi 2001; S. Niemi ja J. Niemi 2002; S. Niemi 2002; Nikolaev 2002). Mainituissa suomen kielen tutkimuksissa epäsanat laadittiin siten, että (useimmiten) kaksitavuisten aitojen substantiivien tai verbien mahdollinen alkukonsonantti vaihdettiin toi-

seksi. Toinen yleinen keino on konsonantin poisto tai vokaalialkuisen sanojen sananalkuisen konsonantin lisäys. Se, että muuntaminen kohdistui sananalkuiseen konsonanttiin, perustui muun muassa ateoreettiseen havaintoon siitä, että loppusoinnutuksessa alkukonsonantismiltaan eroavat kaksitavuiset sanat sointuvat maksimaalisen hyvin yhteen (esim. *akka* ja *sakka*, *muistaa* ja *luistaa*; vrt. heikommin sointuvat *arka* ja *akka*, *kasakka* ja *sakka*). Lisäksi psykolingvistiikassa on esitetty sananalkuisen konsonantin tai konsonanttiyhdistelmän ja sitä seuraavan sanan rungon olevan puheentuottamisessa erillisiä yksiköitä (Shattuck-Hufnagel 1987). Shattuck-Hufnagel perustelee äänteen asema ja täydennin -malliaan (*slot and filler*) sillä, että MIT:n englanninkielisten puhelipsahdusten aineistossa 91 prosenttia konsonanttien vaihdoista kohdistuu sananalkuiseen konsonanttiin (N = 165/181, mt. taulukko 2.7). Tämän lisäksi kokeellisesti tuotetut lipsahdukset tukevat samaa päätelmää, sillä niihin liittyy suhteellisen paljon sananalkuista konsonanttia koskevia virheitä (mts. 45).¹

Paradigman sisällä sanoille voi kehittyä vahvoja, usean segmentin pituisia rakenneanalogisia siteitä, joita on englannin kielessä tutkinut Joan Bybee kollegoineen. Tätä ilmiötä he kutsuvat skeema-analogiaksi (*schema analogy*). Tämän ajattelutavan mukaan paradigmat voivat järjestyä prototyypin hahmotyypinsä ympärille. (Ks. esim. Paunonen 1976; Bybee 1985; Bybee ja Moder 1983.) Skeema-analogiaa käyttäen on ymmärrettävää esimerkiksi se, että englannissa on jatkuvaa painetta ilmaista verbin *bring* imperfekti muodossa *brung* (pro *brought*), ja joissakin englannin varieteeteissa muutos on jo tapahtunut. Muutospainetta aiheuttaa melko yleisten säännöttömien yksitavuisen verbien suuri ryhmä, jossa infinitiivin tai preesensin nukleusta /l/ vastaa /ʌ/ tai /æ/ ja jota seuraa nasaali. Lisäksi alkukonsonantismien aineksella on omaa vaikutusta.

Tämän suomea koskevan tutkimuksen pääasiallisena tarkoituksena on kokeellisesti testata eri segmenttien tärkeyttä silloin, kun kielenkäyttäjän tehtävänä on sijoittaa hänelle annettu *patsas-* ja *keskus-*tyyppinen *s-*loppuinen nomini hänen mielestään sille kuuluvaan

¹ Toisaalta alkukonsonantismien virheherkkyys puheessa voi johtua myös puheen tuottamisen dynamiikasta eikä konsonanttien fonologisesta statuksesta (tässä tapauksessa siis asemasta joko sananalkuisena tai muissa sana-asemissa). Eikmeyerin ja Schaden (1993) interaktiivisen aktivaation verkkomallin mukaan puheen tuottamisessa sananalkuisen segmentin valinnan hetkellä tämän paikan äänteen saama leksikaalinen aktivaatio nimittäin vasta lisääntyy, eikä siihen vaikuta juuri lainkaan leksikon tasolta tuleva vahvistava aktivaatio. Sanan myöhemmissä äännesegmenteissä väärän äänteen valinnan mahdollisuus on vähäisempi, koska (tyypillisesti) oikea lekseemi estää (inhiboi) kilpailevien lekseemien aktivaatiota ja siten myös niiden aktivaation vaikutusta segmentitasolle. Suomen kielestä Eikmeyerin ja Schaden mukaisia kielen tuottamisen simulaatioita ovat esittäneet Tikkala, Eikmeyer, Niemi ja Laine (1997) ja Niemi, Tikkala, Eikmeyer ja Laine (1999). Voi olla kuitenkin niin, että kielen fonologialla on vaikutusta siihen, minkälaiset sananalkuiset segmentit ovat esimerkiksi häiriöille. Niemi ja Laine (1997) nimittäin havaitsivat, että sananalkuinen ja subsyllabainen CV-jono — niin sanottujen sananmuunnoksien tapaan — on suomen puhelipsahduksissa toimiva yksikkö. Toisena todisteen lähteenä ovat kirjalliset kokeet, joissa suomenkielisiä koehenkilöitä pyydyttiin »vaihtamaan sananalkuja» keskenään. Tuloksena oli, että he erittäin usein vaihtoivat CV-jonoja. Tällöin esimerkiksi jonosta <kärsivä Tyne> tuli helposti <tyrsivä kääne>. Näin silloinkin, kun pelkän alkukonsonantismien tai koko tavun vaihdolla epäsanojen jonosta olisi saanut aidon suomen kielen ilmauksen (esim. *masta nopo* tai *nasta mopo*). (J. Niemi 2004.) Myös Berg ja Niemi (2000) havaitsivat suomen ja saksan sanan ensimmäiseen tavuun kohdistuneessa kokeellisessa tutkimuksessa, että kyseisten kielten puhujat käyttävät hieman erilaisia strategioita: saksan puhujat pyrkivät muun muassa maksimoimaan sananalkuisen konsonantismien (*onset maximization strategy*), kun taas suomen puhujat tyytyvät vain täyttämään kyseisen aseman konsonantilla (*onset filling strategy*).

paradigmaan. *Patsas*- ja *keskus*-tyyppisten nominien valinta ei ole satunnainen. Yksikön genetiivin taivutuksessahan näillä nomineilla on kolme luonnollista vaihtoehtoa, joiden produktiivisuuden aste (joka vuorostaan korreloi agglutinaation asteen kanssa) kasvaa seuraavasti: *patsaan*, *keskuksen*, *Carlosin* (viimeinen siis produktiivisin ja samalla eniten agglutinoiva). Mikäli kielenkäyttäjät taivuttavat *s*-loppuisia epäsanoina viimeksi mainitun tyyppin mukaisesti, voimme olettaa epäsanoinen jakavan samaa kenttää erisnimien kanssa. Mikäli koehenkilöt taas valitsevat jommankumman vaihtoehdoista *patsaan* tai *keskuksen*, on mielenkiintoista nähdä, kuinka potentiaaliset nominit korreloivat taivutustyyppien produktiivisuuden kanssa, siis kuinka tietyn taivutustyyppin produktiivisuus vaikuttaa epäsanoinen paradigmaattistumiseen.

Testissä manipuloimme muutakin kuin alkukonsonantismia. Tutkimme siis sitä, onko suomessa alkukonsonantismin jälkeisiä paradigmaattisia segmenttejä ja assosiaatioita englannissa havaitun skeema-analogian tapaan. Paradigman sisäisenä vaihtelusuhteena tarkastelemme suomen kielen fonotaktisten rajoitteiden mukaan luotujen substantiivien yksikön nominatiivin ja yksikön akkusatiivi-genetiivin suhdetta. Koetta pohjustamme kahdella erillisellä analyysillä: toisaalta tarkastelemme, miten aidot suomen *s*-nominat käyttäytyvät lemmafrekvenssiensä² suhteen, ja toisaalta, miten laaja taivutustyyppi on kyseessä ja miten se vaikuttaa siihen kuuluvien sanojen prosessointiin.³ Nämä molemmat yleisyyttä koskevat muuttujat ovat nimittäin morfo-leksikaalisten prosessien tärkeimpiä kielellisiä muuttujia (esim. Taft 1979). Taivutustyyppin leksikaalista taajuutta tarkastelemme CD-Perussanakirjan aineiston avulla, ja lemmafrekvenssien analyyseissa tukeudumme Kielipankin laajoihin tekstikorpuksiin (yli 130 milj. sanetta).

Korpusanalyysien perusteella tekemissämme epäsanatesteissä manipuloimme siis *s*-paradigmojen aitojen sanojen äänneasua systemaattisesti vaihtamalla eri segmenttejä ja testaamme, kuinka tiivis yhteys vallitsee tietyn muunnoksen läpikäyneen epäsanoinen ja sen »alla olevan» aidon sanoinen välillä. Muutamme esimerkiksi alkukonsonantia sanassa *patsas*, joka CD-Perussanakirjan mukaan kuuluu 41. taivutustyyppiin (esim. *patsas* : *patsaan*), siten että saadaan epäsanoinen *katsas*. Epäproduktiivisen *patsas*-paradigman rinnalle asetamme produktiivisen *keskus*-paradigman (paradigmojen produktiivisuudesta ks. esim. Karlsson 1983: 40; Nikolaev 2002) ja niiden perusteella muodostetut epäsanoinen. Peruskysymyksenä on, voimmeko mitenkään ennustaa, sijoittuuko esimerkiksi *katsas* taivutuslentällä produktiiviseen tyyppiin *keskus*, vai tunnustetaanko jollakin tavalla *katsas*-epäsanoinen alla oikea *patsas*-sana. Entä silloin, kun ärsyksenä on esimerkiksi *atsas* tai *patsus*?

² Lemmafrekvenssillä (käyttöyleisyydellä) tarkoitamme sanoinen esiintymisyleisyyttä (taajuutta) kaikissa taivutusmuodoissa (sanoinen taivutusmuotojen yhteenlaskettua frekvenssiä).

³ Taivutustyyppin laajuudella tarkoitamme sen leksikaalista taajuutta eli sitä, kuinka paljon lekseemejä siihen kuuluu.

KÄYTTÖYLEISYYTTÄ KOSKEVIA HAVAINTOJA

TAIVUTUSTYYPIN LEKSIKAALINEN TAAJUUS

Ensimmäisenä vaiheena kohdeparadigmojen (siis CD-Perussanakirjan paradigmojen 41 (*patsas*) ja 39 (*keskus*)) sisältämien lekseemien muodon analyysissa tarkastelemme sanojen leksikaalista taajuutta tavumäärän mukaisesti (taulukko 1). Vartalon tavumäärällähän on suomen taivutusmorfologiassa selvää ohjausvaikutusta (ks. esim. yksitavuiset *maa* : *maita*, kaksitavuiset *laiva* : *laivoja*, monitavuiset *sikala* : *sikaloita* : *sikaloja*).⁴ Lisäksi tavu prosodisena jaksottajana tuntuu luontevalta ensiaskeleelta suomen sanansisäisen, äännettä laajemman rakenteen analyysissa.

Lemmat	<i>patsas</i>		<i>keskus</i>	
	N	%	N	%
1-tavuiset	2	0,5	1	0,04
2-tavuiset	138	31,4	318	12,1
3-tavuiset	235	53,4	2213	84,2
4-tavuiset	55	12,5	88	3,3
5-tavuiset	9	2,0	6	0,2
6-tavuiset	1	0,2	2	0,1
Yhteensä	440	100	2 628	100

Taulukko 1. Taivutustyyppien 41 (*patsas*) ja 39 (*keskus*) leksikaalinen taajuus tavumäärän mukaan.⁵

Havaitsemme taulukosta 1, että yli puolet (eli 53,4 %) *patsas*-taivutustyyppiin kuuluvista nomineista on kolmitavuisia ja että *keskus*-taivutustyyppissä kolmitavuiisten nominien osuus on vieläkin suurempi (84,2 %). Kaksitavuiset nominit ovat toiseksi yleisimpiä: *patsas*-tyypissä 31,4 % ja *keskus*-tyypissä 12,1 %. Oletamme, että mitä enemmän sanassa on tavuja, sitä helpommin siihen assosioidaan sen nojaan tehty epäšana. Tästä syystä valitsimme kokeita varten kaksitavuisia sanoja ja kontrollisanaksi yhden kolmitavuisen, jonka avulla halusimme testata intuitiivista oletustamme, että kolmitavuinen epäšana on läpinäkyvämpi kaksitavuiseen verrattuna.

⁴ Vaihtelevien muotojen optimaalisteoreettisista analyyseista ks. esim. Prince ja Smolensky 1993; McCarthy ja Prince 1993; teorian soveltamisesta suomen kieleen ks. Anttila 1997, 2003; Nikolaev ja Niemi 2005.

⁵ Taajuus on laskettu CD-Perussanakirjan perusteella yhdyssanat poissulkien.

Kummassakin tyypissä tavujen muodon fonologinen skaala on laajin ensimmäisen tavun kohdalla, ja se suppenee sanan loppua kohti mentäessä (taulukot 2 ja 3). Samalla huomaamme, että tavuaseman suositumpien muotojen osuus kasvaa melkoisen jyrkästi. Nämä tendenssit esiintyvät kaikissa tavumäärällä luokitelluissa ryhmissä — eli kaksitavuista viisitavuisiin — molemmissa taivutustyypeissä. Havainnollistamme nyt tekemiämme yleistyksiä kolmitavuisten lemموjen esimerkillä. Säästääksemme tilaa esitämme vain 15 yleisintä eli ensimmäistä tavua sekä kolme viimeistä eli harvinaisinta tavua. (Täydelliset analyysit ovat saatavissa tämän artikkelin ensimmäiseltä kirjoittajalta.)

1. tavu		f	2. tavu		f	3. tavu		f
1	<i>a</i>	4	1	<i>ti</i>	11	1	<i>kas</i>	164
2	<i>toi</i>	4	2	<i>di</i>	11	2	<i>käs</i>	50
3	<i>la</i>	3	3	<i>mi</i>	9	3	<i>las</i>	10
4	<i>ha</i>	3	4	<i>ha</i>	8	4	<i>as</i>	8
5	<i>huo</i>	3	5	<i>lo</i>	8	5	<i>gas</i>	1
6	<i>juo</i>	3	6	<i>ni</i>	8	6	<i>läs</i>	1
7	<i>ko</i>	3	7	<i>pi</i>	7	7	<i>nas</i>	1
8	<i>ku</i>	3	8	<i>do</i>	6			
9	<i>mal</i>	3	9	<i>ja</i>	6			
10	<i>o</i>	3	10	<i>no</i>	6			
11	<i>rai</i>	3	11	<i>si</i>	6			
12	<i>tai</i>	3	12	<i>tei</i>	6			
13	<i>va</i>	3	13	<i>vo</i>	6			
14	<i>ar</i>	2	14	<i>hi</i>	5			
15	<i>e</i>	2	15	<i>ho</i>	5			
.			
.			
.			
173	<i>ä</i>	1	73	<i>tö</i>	1			
174	<i>är</i>	1	74	<i>vei</i>	1			
175	<i>ää</i>	1	75	<i>vu</i>	1			
Yht.		235	235			235		

Taulukko 2. Taivutustyyppi 41 (*patsas*). Kolmitavuisten lemموjen tavujen distributio.

1. tavu		f	2. tavu		f	3. tavu		f
1	<i>ko</i>	28	1	<i>ka</i>	110	1	<i>tus</i>	695
2	<i>va</i>	25	2	<i>ta</i>	65	2	<i>us</i>	464
3	<i>si</i>	24	3	<i>la</i>	55	3	<i>tys</i>	338
4	<i>ku</i>	23	4	<i>ku</i>	49	4	<i>dus</i>	151
5	<i>a</i>	22	5	<i>sa</i>	48	5	<i>ys</i>	99
6	<i>ki</i>	19	6	<i>pa</i>	46	6	<i>dys</i>	82
7	<i>e</i>	18	7	<i>kah</i>	31	7	<i>nus</i>	73
8	<i>pa</i>	17	8	<i>ra</i>	31	8	<i>mus</i>	65
9	<i>sa</i>	17	9	<i>sah</i>	31	9	<i>nys</i>	42
10	<i>vi</i>	16	10	<i>ki</i>	30	10	<i>mys</i>	38
11	<i>ka</i>	15	11	<i>ri</i>	29	11	<i>lus</i>	32
12	<i>pu</i>	15	12	<i>kä</i>	28	12	<i>rus</i>	26
13	<i>rä</i>	14	13	<i>pu</i>	27	13	<i>lys</i>	23
14	<i>ta</i>	14	14	<i>ja</i>	26	14	<i>nos</i>	21
15	<i>tu</i>	14	15	<i>ky</i>	26	15	<i>rys</i>	19
.
.
.
862	<i>zoo</i>	1	354	<i>vys</i>	1	30	<i>tes</i>	1
863	<i>än</i>	1	355	<i>vyy</i>	1	31	<i>tis</i>	1
864	<i>ö</i>	1	356	<i>äl</i>	1	32	<i>vis</i>	1
Yht.		2 213			2 213			2 213

Taulukko 3. Taivutustyyppi 39 (*keskus*). Kolmitavuisten lemموjen tavujen distributio.

Molemmassa taivutustyyppissä on siis yksi vakioaines, joka on leimallinen juuri kyseiselle taivutustyyppille: 41. taivutustyyppissä (*patsas*) tällaisena aineksena on loppusegmentti *-As* ja taivutustyyppissä 39 (*keskus*) *-Us*.

<i>patsas</i>	esim.	N	%	<i>keskus</i>	esim.	N	%
-as	<i>rikas</i>	326	74,1	-us	<i>korpus</i>	1659	63,1
-äs	<i>eräs</i>	90	20,5	-ys	<i>kehys</i>	661	25,2
-is	<i>ruumis</i>	13	3,0	-os	<i>johdos</i>	120	4,6
-at	<i>tikkaat</i>	5	1,1	-is	<i>roskis</i>	73	2,8
-es	<i>kirves</i>	3	0,7	-es	<i>kyljes</i>	31	1,2
-os	<i>uros</i>	1	0,2	-as	<i>lihas</i>	29	1,1
-us	<i>vantus</i>	1	0,2	-ös	<i>päätos</i>	27	1,0
-yt	<i>rynttyyt</i>	1	0,2	-et ⁶	<i>siskokset</i>	25	1,0
				-äs	<i>teräs</i>	3	0,1
Yht.		440	100			2628	100
(-As)		416	94,6)	(-Us)		2320	88,3)

Taulukko 4. Taivutustyyppien 41 ja 39 leksikaaliset taajuudet loppusegmentin mukaan jaoteltuina.

Taulukosta 4 näkyy, että prototyypinen *patsas*-paradigmaan kuuluva sana on *As*-loppuinen (94,6 % tapauksista) ja kilpailijaparadigman *keskus* prototyypinen edustaja on *Us*-loppuinen (88,3 % tapauksista). Sen sijaan ei-terminaalisen (ei-viimeisen) tavun hajonta on paljon laajempi (ks. taulukkoja 2 ja 3). Voimme kyllä todeta, että *patsas*-tyypissä esimerkiksi kolmitavuisen nominin yleisin ensimmäinen tavu on *a-* tai *toi-*, yleisin toinen on *ti-* tai *di-*; samoin *keskus*-tyypissä yleisin ensimmäinen tavu on *ko-* ja yleisin toinen tavu on *ka-*. Emme kuitenkaan voi antaa näille yleisimmille ensimmäiselle ja toiselle tavulle prototyypin segmentin asemaa, sillä kyseisten tavujen »yleisyys» on verrattain alhainen: näissä paikoissa olevia *a-* ja *toi-*-tavuja on 4/235 (1,7 %), *ti-* ja *di-*-tavuja 11/234 (4,7 %); *ko-*-tavuja on 28/2213 (1,2 %) ja *ka-*-tavuja 110/2213 (4,9 %) (vrt. loppusegmentin yleisyyteen: *As*-loppuja on 95,7 %, *Us*-loppuja 88,3 %). Näin ollen ainoana kyseisten taivutustyyppien prototyypisinä segmentteinä voidaan pitää *As*-loppusegmenttiä (taivutustyyppissä *patsas*) ja *Us*-loppusegmenttiä (taivutustyyppissä *keskus*). Se, että jälkimmäiseen tyyppiin voi kuulua *As*-loppuisia sanoja (esim. *lihas*), ei ratkaisevasti vaikuta prototyypisyyden jakaumaan, sillä näitä on *keskus*-tyypissä ainoastaan 1,2 %.

Mora⁷ on tavun raskautta kuvaava muuttuja. Jos jaetaan edelleen taulukossa 1 esitetyt ryhmät tavujen mora-arvon perusteella siten, että yksimoraiselle tavulle annetaan klassi-

⁶ Yksikössä 11 *us*-loppuisia sanaa (esim. *kaverukset*), 4 *is*-loppuisia (esim. *anopikset*), 3 *es*-loppuisia (esim. *lehdekset*), 3 *ys*-loppuisia (esim. *kälykset*), 2 *as*-loppuisia (esim. *kaimakset*), 2 *os*-loppuisia (esim. *langokset*).

⁷ Moran käsitteestä ks. esimerkiksi Tranel 1991; Blevins 1995; Broselow 1995; tavujen mora-arvojen vaikutuksesta suomen nominien taivutukseen ks. Nikolaev ja Niemi 2005.

seen tapaan heikko arvo (H) ja useampimoraiselle tavulle vahva arvo (V), saamme seuraavan taulukon:

	<i>patsas</i>	%	<i>keskus</i>	%
2-tavuiset	V-V	83-100	V-V	74-100
3-tavuiset	V-H-V	70-94-100	V-H-V	66-60-100
4-tavuiset	V-H-H-V	53-69-90-100	H-V-H-V	58-63-64-100

Taulukko 5. Taivutustyyppien 41 ja 39 yleisimmät sanahahmot mora-arvojen perusteella luokiteltuina.⁸

Kaksitavuisien ja kolmitavuisien nominien kohdalla (siis 84,8 %:ssa *patsas*-tyyppiin kuuluvista nomineista ja 96,3 %:ssa *keskus*-tyyppiin kuuluvista nomineista) kilpailijaparadigmat eivät eroa ratkaisevasti toisistaan. Molempien paradigmojen kolmitavuiset nominit pyrkivät olemaan tyyppiä V-H-V, mikä vastaa universaalia tasavahvuisten vie-
reisten elementtien välttämisen periaatetta (*obligatory contour principle*; esim. Leben 1973, 1980; McCarthy 1986).

LEMMOJEN KÄYTTÖYLEISYYS

Edellisessä luvussa tarkastelimme kilpailijaparadigmoja niiden leksikaalisten taajuuksien kautta. Tältä kannalta tarkasteltuna produktiivisen tyyppin leksikaalinen taajuus on odotetustikin suurempi kuin epäproduktiivisen: *keskus*-tyyppiin kuuluu 2 628 lemmaa ja *patsas*-tyyppiin vain 440. Tässä luvussa tarkastelemme, miten *keskus*-taivutustyyppin produktiivisuus ja vastaavasti *patsas*-taivutustyyppin epäproduktiivisuus ilmenee silloin, kun tyyppien leksikaaliset taajuudet suhteutetaan lemموjen käyttöyleisyyteen. Lemmafrequenssien laskemiseen käytämme Suomen Kielipankin korpusta.⁹

Taulukko 6 selvittää nominien käyttöyleisyyttä (f). Taulukon tulkintaa helpottanee seuraava esimerkki: *keskus*-tyypin käyttöyleisyyden maksimiarvo lankeaa kohtaan 143 495, mikä tarkoittaa, että yleisin *keskus*-tyyppiin kuuluva nomini (tässä tapauksessa

⁸ Taulukkoa luetaan siten, että esimerkiksi *patsas*-tyypin 3-tavuisissa sanoissa 1. tavu on vahva 70 %:ssa sanoista, 2. tavu heikko 94 %:ssa tapauksista ja 3. tavu on aina vahva, onhan se *Vs*-loppuinen. Esitämme taulukossa ainoastaan kaksi-, kolmi- ja nelitavuisien nominien yleisimmät sanahahmot mora-arvojen perusteella, sillä yksi-, viisi- ja kuusitavuisien nominien leksikaalinen taajuus on edellisiin verrattuna vähäinen (ks. taulukko 1 s. 49).

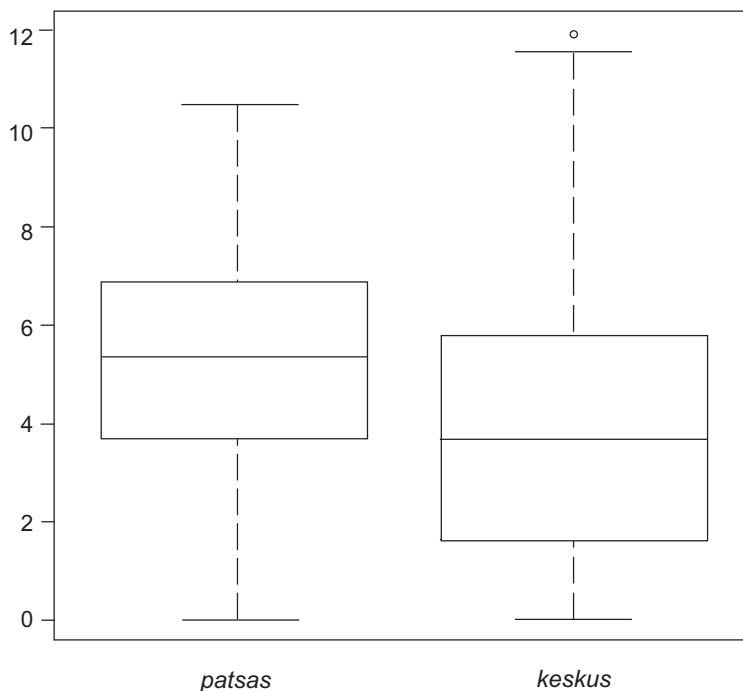
⁹ Korpuksena on käytetty Kielipankkiin koottuja suomenkielisiä lehtitekstejä, joiden saneiden yhteenlaskettu määrä on 131 406 087 (lokakuu 2004).

substantiivi *hallitus*) esiintyy käytetyssä korpuksessa 143 495 kertaa eri muodoissa: *hallitus*, *hallituksemme*, *hallitusta* ja niin edelleen. Samalla tavalla olemme laskeneet lemmafrekvenssit jokaiselle *patsas*-tyyppiin kuuluvalla 440 sanalle ja jokaiselle *keskus*-tyyppiin kuuluvalla 2 628 sanalle. Pitää kuitenkin mainita, että kirjoitettu kieli on sekundaarinen puhuttuun kieleen verrattuna, ja näin esimerkiksi sanalla *hallitus* tuskin olisi niin korkea frekvenssi puhekielessä. Toki jos tarkoitus on verrata näitä kahta taivutustyyppiä keskenään, tekstikorpus on relevantti aineisto.

	<i>patsas</i>	suhde	<i>keskus</i>
leksikaalinen taajuus	440	1 : 5,97	2 628 (taulukosta 1)
näistä korpuksessa	391	1 : 5,74	2 248
lemmafrekvenssit			
minimiarvo	0		0
maksimiarvo	40 727	1 : 3,52	143 495
summa	723 476	1 : 3,70	2 680 144
mediaani	231	5,77 : 1	40
keskiarvo	1 850	1,55 : 1	1 192
keskihajonta	5 369	1 : 1,12	6 001

Taulukko 6. Taivutustyyppien leksikaalinen taajuus ja lemmafrekvenssit.

Suurin osa CD-Perussanakirjassa kahteen tarkasteltavaan tyyppiin kuuluvista lemmoisista esiintyy myös Kielipankin korpuksessa: *patsas* 391/440 (88,8 %) ja *keskus* 2 248/2 628 (85,5 %). *Patsas*-taivutustyyppin yleisin sana on *valmis* (40 727 esiintymää) ja *keskus*-taivutustyyppin yleisin sana jo mainittu *hallitus* (143 495 esiintymää). Seuraavalla kuviolla havainnollistamme taulukon 6 lukuja. Reaalilukujen asemesta käytämme kuviossa logaritmisia arvoja. Jos nimittäin käyttäisimme reaalilukuja, kuviosta tulisi vaikeaselkoinen, sillä jotkin luvut ovat erittäin suuria (mikä taas johtuu korpuksen laajuudesta), ja näin poikkeavat arvot vaikeuttaisivat kuvion lukemista. Logaritmin avulla supistamme jakautuman: esimerkiksi *hallitus*-sanan logaritminen frekvenssi on 11,87 — vertaa absoluuttiseen frekvenssiin 143 495.



Kuvio 1. Taivutustyyppien logaritmiset lemmafrekvenssit.¹⁰

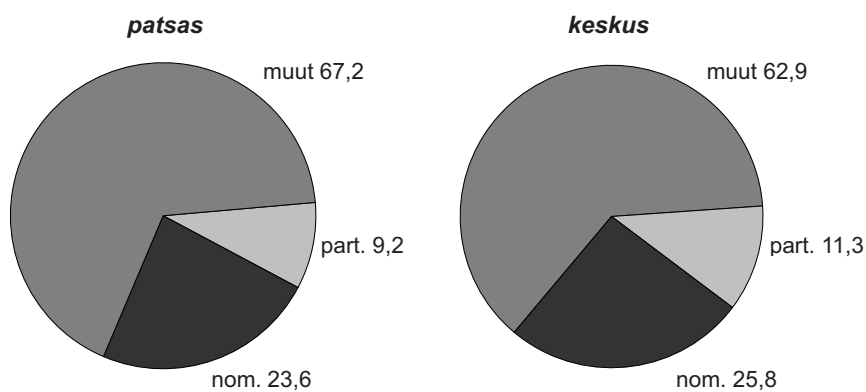
Kuviosta 1 ja taulukosta 6 huomataan kaksi säännönmukaisuutta: epäproduktiivisen taivutustyyppin *patsas* leksikaalinen taajuus on huomattavasti pienempi kuin produktiivisen tyyppin *keskus*, mutta sen lemmafrekvenssi on suurempi (mediaani *patsas*-ryhmässä on 231 ja *keskus*-ryhmässä 40). Tämä on periaatteessa tautologia, sillä epäproduktiivisen taivutustyyppin sanojen määrä ei tyypillisesti enää kasva ja sanat ovat »elossa» vain niiden suhteellisen korkean käyttöyleisyyden ansiosta.

Voimme esittää nyt kaksi perusteisiin menevää kysymystä: (1) Minkä periaatteen voimme ajatella toimivan suomen *keskus*-taivutustyyppin produktiivisuuden alla? (2) Jos periaate löytyy, onko meillä mitään perustetta ajatukselle, että se olisi universaalinen? Morfologiassa isomorfian tendenssiin perustuvan vastauksen näihin kysymyksiin antaa Dressler (1985). Hän on taivutusta koskevilla analyyseillaan esittänyt universaalinen hypoteesin, jota voi kutsua periaatteeksi »jos lisää morfin, lisää myös tavu». Kielessähän on yhtenä vahvana funktionaalisen-semioottisena periaatteena merkityksen ja muodon isomorfia. Taivutukseen sovellettuna voimme ajatella, että kun vartaloon lisätään morfologista ainesta (esim. akkusatiivi-genetiivin päätte *n*), on kognitiivisesti optimaalista — ainakin tavutukseen paljolti nojautuvissa fusionaalis-agglutinatiivisissa kielissä, kuten suomessa

¹⁰ Kuvion katkoviivat rajaavat vaihteluvälin (pienin arvo, suurin arvo). Laatikkojen ala- ja ylätasot ovat alaja yläkvartiileja (alakvartiili on se arvo, jota pienempiä on 25 % havainnoista ja yläkvartiili on se arvo, jota pienempiä on 75 % havainnoista). Laatikkojen sisällä olevat vaakajanaat ovat mediaaneja (niitä muuttujan arvoja, joita pienempiä ja suurempia on puolet havainnoista). Y-akseli kuvaa pintamuotojen käyttöyleisyyttä logaritmisilla arvoilla laskettuna. Yksittäinen piste on poikkeava arvo.

— myös lisätä tavu (vrt. *kes.kus* : *kes.kuk.sen* vs. *pat.sas* : *pat.saan*).¹¹ Kyseinen produktiivisuuden periaate on osoittautunut myös kokeellisesti toimivaksi jopa morfologialtaan erilaisessa kielessä (tai oikeastaan erilaista kieltä puhuvien miellissä), sillä suomea tyystin hallitsemattomat ummikkovenäläiset pyrkivät taivuttamaan suomen sanojen tapaisia epä-sanoja tällä isomorfian periaatteella (ks. Nikolaev 2002; suomenkielisten lasten varhais-morfologiasta ja isomorfiatendenssistä ks. esim. J. Niemi ja S. Niemi 1987: 71).

s-nominien paradigmojen fonologinen samanlaisuus yksikön nominatiivissa ja partitiivissa (esim. *patsasta* vs. *keskusta*) aiheuttaa sen, että näissä taivutusmuodoissa paradigmojen välinen kilpailu tavallaan eliminoituu. Onkin mielenkiintoista kysyä, mikä on tällaisten isomorfiavertailussa neutraalien muotojen suhde sellaisiin muotoihin, jotka mahdollistavat paradigmojen kilpailun (esim. *patsaan* vs. *keskuksen*; vrt. Paunonen 1976, joka kenttämorfologiassaan tarkastelee muun muassa koko paradigman hahmottamista lapsilla yksittäisten sananmuotoesiintymien kautta). Toisin sanoen voimme perustellusti kysyä, onko jälkimmäisiä enemmän kuin yksikön nominatiivin ja partitiivin muotoja ja onko kummassakin taivutustyyppissä samanlainen suhde näiden taivutusmuotojen välillä. Vastataksemme näihin kysymyksiin laskimme Kielipankin korpuksessa yllä lueteltujen pintamuotojen käyttöyleisyydet molemmissa taivutustyypeissä jokaisen lemman kohdalla. Karsimme pois sellaiset yksikön nominatiivin muodot, joissa on possessiivisuffiksi (esim. *keskukseni*, *patsaani*), sillä näissä muodoissahan paradigmatilpailu ei eliminoidu. Kuvio 2 sisältää analyysimme päätulokset. Kuviossa suhteet on annettu prosenteissa, jotka vuorostaan on laskettu kunkin muodon keskiarvosta.¹²



Kuvio 2. Pintamuotojen käyttöyleisyydet (%) sijan mukaan (nom. = yksikön nominatiivi, part. = yksikön partitiivi).

¹¹ Diakronisesti myös *patsas*-tyypin muoto oli kolmitavuinen: *patsaan* < **patsahan*. Eteläpohjalaisessa murteessa on tähän saakka säilynyt sanansisäinen *h* esimerkiksi *kirves*-sanan taivutusmuodoissa. Samaten arkaisitiset rekisterit ovat saattaneet säilyttää *h*:llisiä muotoja laajemminkin nykypuhujien tietoisuudessa (esim. Mustassa Saarassa lauletaan vieläkin, että *Ei taivahassa kuolon vaaraa...* ja Porilaisten marssin 1. säkeistössä *viel' on Suomi voimissaan, voi vainolaisen hurmehella peittää maan*).

¹² Pintamuotojen käyttöyleisyyksien keskiarvot:

	<i>patsas</i>	<i>keskus</i>
yks. nom.	436,86	308,26
yks. part	169,49	134,52
muut	1249,95	749,98

Paradigmojen välinen kilpailu eliminoituu (eli yksikön nominatiivi ja yksikön partitiivi esiintyvät) *patsas*-taivutustyyppissä 32,8 prosentissa (N = 237 089) kaikista tapauksista ja vastaavasti *keskus*-taivutustyyppissä 37,1 prosentissa (N = 994 940) tapauksista. Näin neutraalien muotojen suhde paradigmakilpailun mahdollistaviin muotoihin ei eroa merkittävästi kyseisten taivutustyyppien välillä eikä todennäköisesti vaikuta vinouttavasti kummankaan paradigman hyödyksi tai haitaksi.

TAIVUTUSKOKEET

Edellisen luvun yleisyyslaskelmat osoittavat, että tarkasteltavien taivutustyyppien välillä on kaksi selvää eroa. Ensimmäisenä yleistyksenä on jo aiemmin todettu kieliin yleisesti sopiva havainto, jonka mukaan produktiivinen paradigma (*keskus*) sisältää enemmän lemmoja kuin sen epäproduktiivinen kilpailija (*patsas*), kun taas lemmafrekvenssillä mitattuna tulos on päinvastainen. Tätä asiantilaa voidaan havainnollistaa seuraavalla suhdetaulukolla:

	<i>patsas</i>	:	<i>keskus</i>
leksikaalinen taajuus	1	:	5,97
lemmafrekvenssi (mediaani)	5,77	:	1

Toinen havainto koskee sitä merkittävää eroa, joka vallitsee sanan nominatiivimuodon VC-loppuosassa. Kummassakin paradigmassa on nimittäin selvä prototyyppinen edustumansa: prototyyppinen *patsas*-taivutustyyppiin kuuluva sana on *As*-loppuinen (94,6 %) ja vastaavasti *keskus*-taivutustyyppiin kuuluva *Us*-loppuinen (88,3 %). Kyseisiä loppusegmenttejä voisikin kutsua temaattisiksi segmenteiksi, jotka osoittavat sanan taivutusluokkaa.

Tarkoituksenamme on tämän tutkimuksemme lopuksi selvittää, miten loppuosan prototyyppisyys (vs. marginaalisuus) vaikuttaa näiden taivutustyyppien pohjalta muodostettujen epäsanojen prosessointiin (taivutukseen). Hypoteesimme on, että mikäli epäsanon loppusegmentti ei ole kategoriansa prototyyppinen edustaja (eli tässä tapauksessa epä-sana ei ole *As*- tai *Us*-loppuinen), epä-sana on neutraali. Taivutuskentällä se oletusarvoisesti mielletään produktiiviseen taivutustyyppiin (*keskus*) kuuluvaksi, sillä kielenkäyttäjät tietänevät, että ilman vahvaa fonologista motivaatiota *patsas*-paradigmaan ei sijoiteta uusia lemmoja, tässä tapauksessa luomiamme epäsanvoja. Ilmiötä voidaan kuvata magneetikenttää käyttävän analogian avulla: kaksi kilpailevaa paradigmaa ovat magneetin kaksi napaa. Kun kenttään pääsee neutraali epä-sana (sillä ei ole omaa fonologista latausta), niin produktiivinen napa vetää sen puoleensa. Mikäli taas epä-sana on *As*- tai *Us*-loppuinen, niin sillä on jo oma latauksensa, joka vaikuttaa paradigmojen (napojen) kilpailuun. Hypoteesimme mukaan siis epäproduktiiviseksi väitetty *patsas*-paradigma ei olisikaan täysin sulkeinen, vaan kielenkäyttäjät voisivat periaatteessa taivuttaa antamiamme epäsanvoja (uudissanvoja?) myös *patsas*-paradigman mukaisesti.

Seuraavilla sivuilla yritämmekin mitata kontrolloidusti laadittujen epäsanojen »latausta», toisin sanoen selvittää, kuinka hyvin niiden taivutuksessa tunnistetaan alla olevia

▷

oikeita sanoja. Tätä tarkoitusta varten muodostimme kahdesta kilpailevasta paradigmat kaksikielisen sanaryhmää, joihin sovelsimme tiettyjä muunnos algoritmeja. Tulokseksi saimme kuusi epäsanaryhmää, jotka annoimme koehenkilöille taivutettaviksi.

Testisanat valittiin siten, että kummassakin ryhmässä oli viisi konsonanttialkuista ja viisi vokaalialkuista sanaa. Toisen ryhmän sanojen loppusegmentti ei saanut olla *Us* (nämä sanat edustavat siis *patsas*-taivutustyyppiä), kun taas toisen ei saanut olla *As* (*keskus*-taivutustyyppi)¹³.

<i>patsas</i>	<i>keskus</i>
<i>patsas</i>	<i>keskus</i>
<i>raskas</i>	<i>jänis</i>
<i>potilas</i>	<i>kiitos</i>
<i>taivas</i>	<i>panos</i>
<i>kiivas</i>	<i>päärros</i>
<i>armas</i>	<i>aines</i>
<i>oras</i>	<i>ilves</i>
<i>äyräs</i>	<i>ostos</i>
<i>eväs</i>	<i>intos</i>
<i>aulis</i>	<i>aunus</i>

Taulukko 7. Taivutuskokeiden epäsanojen laadinnassa käytetyt aidot sanat.

Taulukossa 7 esitetyistä sanoista muodostimme epäsanat kolme seuraavaa algoritmia käyttäen:

1. muutettiin alkukonsonantti/alkuvokaali toiseksi (taulukon 8 sarake 1)
2. poistettiin alkukonsonantti/lisättiin alkuvokaalin eteen konsonantti (taulukon 8 sarake 2)
3. vaihdettiin loppusegmentti oppositiossa olevan ryhmän prototyypiseen loppusegmenttiin (taulukon 8 sarake 3).

Näillä muunnoksilla saimme aikaan kolme epäsanaryhmää kustakin taivutustyyppistä (yhteensä siis 60 epäsanana; ks. taulukkoa 8 viereisellä sivulla). Kolmannella tyypillä, jossa siis kilpailevan paradigman prototyypinen VC-loppu liitetään aitoon sanan alkuosaan, testaamme mielestämme maksimaalisen tiukasti leksikaaliseen taajuuteen perustuvan analyysimme tuottamaa hypoteesia näiden VC-jaksojen paradigmaattisesta merkityksestä.

¹³ Mahdollista kritiikkiä valittujen sanojen suhteen ennustaan mainitsemme jo tässä, että sanan alhainen pinta-muotoyleisyys, erisnimiluonne tai kahdenlainen taivutus (esim. *armas* : *armaan* ja *Armas* : *Armaksen*) eivät vinouta testituloksia (ks. standardoitujen residuaalien testejä jäljempänä). Väitettä tukevat myös korrelaatiotestit, jotka eivät osoita merkittävää korrelaatiota vastausten ja pinta-frekvenssien välillä (ks. korrelaatiotestejä jäljempänä).

	1	2	3
<i>patsas</i>	<i>katsas</i>	<i>atsas</i>	<i>patsus</i>
<i>patsas</i>	<i>katsas</i>	<i>atsas</i>	<i>patsus</i>
<i>raskas</i>	<i>laskas</i>	<i>askas</i>	<i>raskus</i>
<i>potilas</i>	<i>kotilas</i>	<i>otilas</i>	<i>potilus</i>
<i>taivas</i>	<i>kaivas</i>	<i>aivas</i>	<i>taivus</i>
<i>kiivas</i>	<i>tiivas</i>	<i>iivas</i>	<i>kiivus</i>
<i>armas</i>	<i>ormas</i>	<i>sarmas</i>	<i>armus</i>
<i>oras</i>	<i>eras</i>	<i>koras</i>	<i>orus</i>
<i>äyräs</i>	<i>öyräs</i>	<i>päyräs</i>	<i>äyrys</i>
<i>eväs</i>	<i>öväs</i>	<i>söväs</i>	<i>evys</i>
<i>aulis</i>	<i>oulis</i>	<i>taulis</i>	<i>aulus</i>
<i>keskus</i>	<i>teskus</i>	<i>eskus</i>	<i>keskas</i>
<i>keskus</i>	<i>teskus</i>	<i>eskus</i>	<i>keskas</i>
<i>jänis</i>	<i>vänis</i>	<i>änis</i>	<i>jänäs</i>
<i>kiitos</i>	<i>piitos</i>	<i>iitos</i>	<i>kiitas</i>
<i>panos</i>	<i>tanos</i>	<i>anos</i>	<i>panas</i>
<i>piirros</i>	<i>kiirros</i>	<i>iirros</i>	<i>piirras</i>
<i>aines</i>	<i>äines</i>	<i>taines</i>	<i>ainas</i>
<i>ilves</i>	<i>ulves</i>	<i>tilves</i>	<i>ilväs</i>
<i>ostos</i>	<i>estos</i>	<i>postos</i>	<i>ostas</i>
<i>intos</i>	<i>untos</i>	<i>tintos</i>	<i>intas</i>
<i>aunus</i>	<i>ounus</i>	<i>saunus</i>	<i>aunas</i>

Taulukko 8. Aidot pohjasanat ja niistä muodostetut epäsanat.

Kokeeseen osallistui 46 Joensuun Normaalikoulun 14-vuotiasta oppilasta. Kaikki koehenkilöt olivat yksikielisiä suomen puhujia ja kotoisin samalta alueelta (Joensuusta ja lähikunnista), joten ryhmä oli kieli-, koulutus- ja ikärakenteeltaan sekä murretaustaltaan homogeeninen. Koe oli suullinen yksilöttesti, jonka vastaukset nauhoitettiin ja purettiin jälkepäin. Jokaisen koehenkilön piti taivuttaa taulukossa 8 esitetyt 60 epäsanaa yksikön genetiivin akkusatiivissa (*Minä otan...*) hänelle lausutusta listasta, johon epäsanat oli kirjoitettu yksikön nominatiivissa. Lausekontekstin tarkoituksena oli mahdollisimman minimaalisin keinoin mutta kuitenkin yksiselitteisesti kehottaa koehenkilöä taivuttamaan epäsanat oikeiden sanojen tapaan. Näin koehenkilöä ei ole pakotettu valitsemaan kahdesta vaihtoehdosta (esim. *katsaan* tai *katsaksen*) parempaa, vaan vastausten skaala voisi olla vaikka kuinka laaja¹⁴. Sanat oli järjestetty satunnaisjärjestykseen siten, ettei mikään sana-

¹⁴Toki käytännössä kävi niin, että jos jokin muu kuin *patsas*- tai *keskus*-paradigmojen mukainen taivutus esiintyi (näin oli vain 0,62 %:ssa vastauksista; N = 17), se oli konsonanttiloppuisten vierassanojen, erisnimien tai kirjainlyhennesanojen mukainen, esimerkiksi *katsas* : *katsasin*, vrt. *Carlos* : *Carlosin*).

listan järjestys enää toistunut jossakin muussa sanalistassa. Näin vältettiin epäsanojen järjestyksen mahdollinen vaikutus tuloksiin.

Alla esitetyt tulokset olemme saaneet laskemalla yhteen kunkin ryhmän vastaukset siten, että jokaisen epäsanana kohdalla esimerkiksi *katsaan* ja *katsaksen* -varianttien summan pitäisi olla 46 (koehenkilöiden määrä; yksi koehenkilö – yksi vastaus). Mikäli luku on pienempi kuin 46, tämä tarkoittaa sitä, että vähintään yksi koehenkilö taivutti epäsanana jollakin toisella tavalla (esim. *Minä otan katsasin*), ja tällaiset vastaukset vähennetään luvusta 46. Summia verrattiin tilastollisesti keskenään khiin neliö -merkitsevyydestin avulla (χ^2). Ero on merkitsevä, jos p-arvo on pienempi kuin 0,05. Tulokset esitetään taulukossa 9.

<i>patsas</i>	<i>katsas</i>	katsaan	katsaksen	<i>atsas</i>	atsaan	atsaksen	<i>patsus</i>	patsuun	patsuksen
<i>patsas</i>	<i>katsas</i>	8	38	<i>atsas</i>	9	36	<i>patsus</i>	2	44
<i>raskas</i>	<i>laskas</i>	11	35	<i>askas</i>	14	32	<i>raskus</i>	1	45
<i>potilas</i>	<i>kotilas</i>	19	26	<i>otilas</i>	18	28	<i>potilus</i>	2	44
<i>taivas</i>	<i>kaivas</i>	6	40	<i>aivas</i>	3	43	<i>taivus</i>	1	45
<i>kiivas</i>	<i>tiivas</i>	7	39	<i>iivas</i>	8	37	<i>kiivus</i>	3	43
<i>armas</i>	<i>ormas</i>	8	38	<i>sarmas</i>	7	39	<i>armus</i>	0	46
<i>oras</i>	<i>eras</i>	9	37	<i>koras</i>	10	36	<i>orus</i>	1	45
<i>äyräs</i>	<i>öyräs</i>	9	37	<i>päyräs</i>	9	37	<i>äyrys</i>	4	42
<i>eväs</i>	<i>öväs</i>	8	37	<i>söväs</i>	5	41	<i>evys</i>	3	43
<i>aulis</i>	<i>oulis</i>	2	42	<i>taulis</i>	4	40	<i>aulus</i>	2	42
yht.		87	369		87	369		19	439
%		19%	81%		19%	81%		4%	96%
<i>keskus</i>	<i>teskus</i>	teskuun	teskuksen	<i>eskus</i>	eskuun	eskuksen	<i>keskas</i>	keskaan	keskaksen
<i>keskus</i>	<i>teskus</i>	1	45	<i>eskus</i>	2	44	<i>keskas</i>	7	38
<i>jänis</i>	<i>vänis</i>	2	44	<i>änis</i>	2	43	<i>jänäs</i>	8	38
<i>kiitos</i>	<i>piitos</i>	2	44	<i>iitos</i>	2	44	<i>kiitas</i>	6	40
<i>panos</i>	<i>tanos</i>	3	43	<i>anos</i>	2	44	<i>panas</i>	6	40
<i>piirros</i>	<i>kiirros</i>	2	43	<i>iirros</i>	1	45	<i>piirras</i>	9	37
<i>aines</i>	<i>äines</i>	3	43	<i>taines</i>	5	38	<i>ainas</i>	7	39
<i>ilves</i>	<i>ulves</i>	3	43	<i>tilves</i>	5	41	<i>ilväs</i>	4	42
<i>ostos</i>	<i>estos</i>	4	42	<i>postos</i>	4	42	<i>ostas</i>	7	39
<i>intos</i>	<i>untos</i>	2	44	<i>tintos</i>	3	42	<i>intas</i>	7	39
<i>aunus</i>	<i>ounus</i>	2	44	<i>saunus</i>	3	43	<i>aunas</i>	9	37
yht.		24	435		29	426		70	389
%		5%	95%		6%	94%		15%	85%

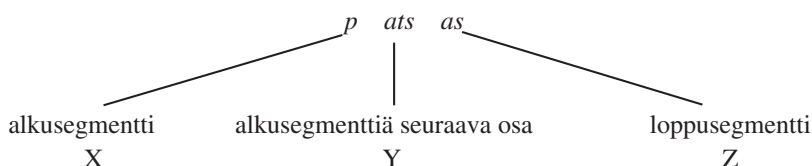
Taulukko 9. Epäsanojen taivutuskokeen tulokset.

Tuloksista näkyy, että kaikki sanaryhmät suosivat noin 80–95 prosentin enemmistöllä produktiivista *keskus* : *keskuksen* -taivutusta. Toisaalta tuloksissa ryhmien välillä on kuitenkin syvemmin analysoituna eroja, joita tarkastelemme alla tilastollisesti. Tuloksia voi havainnollistaa seuraava yhtälö:

$$(katsas \approx atsas \approx keskas) \neq (teskus \approx eskus \approx patsus).$$

Yhtälössä \approx tarkoittaa, että erolla ryhmien välillä ei ole tilastollista merkitystä, ja \neq vastaavasti tarkoittaa, että ryhmien välillä on tilastollisesti merkitsevä ero (siis suluissa olevat vasemmanpuoliset ryhmät eroavat merkitsevästi oikeanpuolisista).¹⁵

Jos edelleen käytetään epäsanojen muodostamisen yhteydessä käytettyä algoritmia, niin esimerkiksi *patsas*-ryhmälle voidaan antaa seuraavat arvot:



Näin *patsas* on yhtä kuin järjestetty kolmikko (X, Y, Z). Muuntamalla alkusegmenttiä saadaan ryhmä *katsas*, joka on yhtä kuin järjestetty kolmikko (X₁, Y, Z). Edelleen ryhmä *atsas* = (X₀, Y, Z) ja ryhmä *patsus* = (X, Y, *Z), jossa *Z tarkoittaa Z:n komplementtia eli *Us*-loppusegmenttiä. Järjestettyinä kolmikkoina ryhmät siis näyttävät seuraavilta:

<i>patsas</i>	<i>katsas</i>	<i>atsas</i>	<i>patsus</i>
(X, Y, Z)	(X ₁ , Y, Z)	(X ₀ , Y, Z)	(X, Y, *Z)

Vastaavasti:

<i>keskus</i>	<i>teskus</i>	<i>eskus</i>	<i>keskas</i>
(X, Y, *Z)	(X ₁ , Y, *Z)	(X ₀ , Y, *Z)	(X, Y, Z)

Merkitsevyydestien avulla rakennettu yhtälö (*katsas* \approx *atsas* \approx *keskas*) \neq (*teskus* \approx *eskus* \approx *patsus*) on siis nyt muotoa

$$((X_1, Y, Z) \approx (X_0, Y, Z) \approx (X, Y, Z)) \neq ((X_1, Y, *Z) \approx (X_0, Y, *Z) \approx (X, Y, *Z)).$$

¹⁵ Pearsonin khiin neliötesti (Yatesin korjauksin) antaa seuraavat tulokset:

aineisto: *teskus* vs. *eskus*

$$X = 0,3587, df = 1, p = 0,5492$$

aineisto: *patsus* vs. *keskas*

$$X = 30,9886, df = 1, p = 2,595e-08$$

aineisto: *katsas/atsas* vs. *patsus*

$$X = 48,2372, df = 1, p = 3,776e-12$$

aineisto: *teskus* vs. *patsus*

$$X = 0,3813, df = 1, p = 0,5369$$

aineisto: *eskus* vs. *patsus*

$$X = 1,8441, df = 1, p = 0,1745$$

aineisto: *katsas/atsas* vs. *keskas*

$$X = 2,097, df = 1, p = 0,1476$$

aineisto: *teskus* vs. *keskas*

$$X = 24,0001, df = 1, p = 9,633e-07$$

aineisto: *eskus* vs. *keskas*

$$X = 17,7346, df = 1, p = 2,540e-05$$

aineisto: *katsas* vs. *teskus*

$$X = 39,8762, df = 1, p = 2,706e-10$$

aineisto: *atsas* vs. *eskus*

$$X = 31,9522, df = 1, p = 1,58e-08$$

▷

Ainoa jäsen, joka erottaa järjestetyt kolmikot toisistaan, on Z (As-loppusegmentti) ja Z:n komplementti *Z (Us-loppusegmentti).

Tämän luvun alussa esittämämme hypoteesi sai siis tukea tuloksista, sillä ainoa epäsanana jakso, joka merkitsevästi vaikuttaa epäsanana taivutukseen (prosessointiin) on juuri yleisyyslaskelmissa prototyypiseksi (siis yleiseksi) osoittautunut sananloppuinen VC-jono eli As- tai Us-loppusegmentti.

Jos edellä esittämäämme yhtälöä lähestytään tiukemmin kriteerein, siitä tulee ainakin astetta monimutkaisempi, sillä kussakin ryhmässä on viisi konsonantialkuista ja viisi vokaalialkuista sanaa (ks. taulukkoa 7 s. 58). Tämän perusteella jokainen ryhmä pitää jaksaa kahtia ja antaa kullekin omat arvonsa X_2 , X_3 ja niin edelleen. Merkitsevyydestien tulosten mukaan yhdessäkin ryhmässä ei kuitenkaan ole merkitsevää eroa konsonantti-alkuisten ja vokaalialkuisten vastausten summien välillä.¹⁶

Mielenkiintoista on, että ainoa tavumäärältään eroava epäsanana sai hieman erilaisia tuloksia kuin muut: epäsanamme olivat muutoin kaksitavuisia, paitsi *potilas* (*kotilas*, *otilas*, *potilus*). Ryhmissä *katsas* ja *atsas* tämä epäsanana käyttäytyy eri tavalla kuin ryhmän muut epäsanat, mikä näkyy standardoitujen residuaalien testissä (taulukko 10).

<i>katsas</i>	<i>katsaan</i>	<i>katsaksen</i>	<i>atsas</i>	<i>atsaan</i>	<i>atsaksen</i>
katsas	-0,26	0,13	atsas	0,14	-0,07
laskas	0,75	-0,36	askas	1,76	-0,86
kotilas	3,55	-1,73	otilas	3,11	-1,51
kaivas	-0,94	0,46	aivas	-1,95	0,95
tiivas	-0,6	0,29	iivas	-0,2	0,1
ormas	-0,26	0,13	sarmas	-0,6	0,29
eras	0,08	-0,04	koras	0,41	-0,2
öyräs	0,08	-0,04	päyräs	0,08	-0,04
öväs	-0,2	0,1	söväs	-1,27	0,62
oulis	-2,21	1,07	taulis	-1,52	0,74

Taulukko 10. Ryhmien *katsas* ja *atsas* standardoidut residuaalit.

¹⁶ Pearsonin khiin neliötesti (Yatesin korjauksin) antaa seuraavat tulokset:

<i>katsas</i>	$X = 2,6344$, $df = 1$, $p = 0,1046$
<i>atsas</i>	$X = 3,6363$, $df = 1$, $p = 0,05653$
<i>teskus</i>	$X = 0,382$, $df = 1$, $p = 0,5365$
<i>eskus</i>	$X = 3,8254$, $df = 1$, $p = 0,05048$
<i>patsus</i>	$X = 4e-04$, $df = 1$, $p = 0,9845$
<i>keskas</i>	$X = 0,0224$, $df = 1$, $p = 0,881$

Mikäli taulukon 10 kentällä ilmenevä luku on suurempi kuin 2, niin tulos on odotettua suurempi. Jos se on pienempi kuin -2 , niin tulos on odotettua pienempi. Kolmitavuiset *kotilaan* (3,55) ja *otilaan* (3,11) -vastaukset saivat odotettua suuremmat frekvenssit, mikä tarkoittaa, että koehenkilöt tunnistivat kolmitavuisissa epäsanoina niiden alla olevat oikeat sanat helpommin kuin kaksitavuisissa epäsanoina. Tämän perusteella voimme olettaa, että mitä enemmän sanassa on tavuja, sitä latautuneempi¹⁷ on sen nojaan tehty epä-sana, jossa on muunnettu pelkästään alkusegmentti (vrt. Bybeen skeema-analogiaan). Sen sijaan loppusegmenttiryhässä kolmitavuisen *potilus* käyttäytyy samalla tavalla kuin ryhmän muut kaksitavuiset sanat, mikä taas kertoo prototyypin VC-loppujonon vahvasta vaikutuksesta epäsanojen prosessointiin.

Ryhmässä *katsas* on myös yksi epä-sana, joka sai odotettua pienemmät frekvenssit: *ouliin* : *ouliin* ($-2,21$), minkä syyksi oletamme sen, että sana on ryhmän ainoa, joka ei ole *As*-loppuinen ja näin neutraalina sijoittuu helposti taivutuslentällä produktiiviseen taivutus-tyyppiin.

Kuten kirjoituksemme alussa totesimme, sanan käyttöyleisyys on prosessoinnin kuvauksessa tärkeimpiä tekijöitä, sillä on saatu vahvaa näyttöä siitä, että sanan suhteellisen korkea käyttöyleisyys helpottaa sen prosessointia (ks. esim. Taft 1979). Voidaan siis olettaa oikeiden sanojen käyttöyleisyyden vaikuttavan myös niiden perustalle rakennettujen epäsanojen prosessointiin. Taulukossa 11 listamme käyttämiemme aitojen sanojen käyttöyleisyydet (Kielipankin korpuksista).

<i>patsas</i>	f	<i>keskus</i>	f
<i>patsas</i>	2 786	<i>keskus</i>	19 642
<i>raskas</i>	13 598	<i>jänis</i>	1 756
<i>sotilas/potilas</i>	12 494/15 574	<i>kiitos</i>	8 172
<i>taivas</i>	7 301	<i>panos</i>	4 351
<i>kiivas</i>	3 243	<i>piirros</i>	1 931
<i>armas</i>	1 754	<i>aines</i>	4 415
<i>oras</i>	328	<i>ilves</i>	8 955
<i>äyräs</i>	298	<i>ostos</i>	2 282
<i>eväs</i>	3 893	<i>intos</i>	0
<i>aulis</i>	2 563	<i>aunus</i>	646

Taulukko 11. Aitojen sanojen lemmafrekvenssit (f).

¹⁷ Latauksella tarkoitamme epäsanatunnettavuutta eli sitä, kuinka hyvin koehenkilöt tunnistavat epäsanatunnettavuuden alla olevan oikean sanan.

Suorittamamme korrelaatiotestit eivät kuitenkaan osoita merkitsevää korrelaatiota vastausten ja lemmafrekvenssien välillä.¹⁸

Ainoat kaksi ryhmää, jotka lähestyvät korrelaatiossa merkitsevyystasoa, ovat *katsaan* ($p = 0,065$) ja *atsaan* ($p = 0,07$). Toki varsinaista korrelaatiota ei näissäkään ryhmissä esiinny. Lisäksi jos poistetaan yksi poikkeava arvo (ainoa kolmitavuinen *sotilas/potilas*), niin positiivinen korrelaatio (0,60; 0,59) muuttuu negatiiviseksi.

Tulos on varsin mielenkiintoinen, sillä se tavallaan sotii sitä periaatetta vastaan, jonka mukaan sanan käyttöyleisyys helpottaa sen prosessointia. Tämä on kuitenkin vain näennäinen ristiriita, sillä epäsanojen ja oikeiden sanojen luonteiden välillä on pikemmin laatuero kuin aste-ero. On luonnollista, että strukturaalisella tasolla sanojen käyttöyleisyys ei ole relevantti tekijä epäsanojen kohdalla. Ratkaisua siis pitää etsiä semantiikasta, jossa oikeilla sanoilla ja epäsanoina on tyystin erilainen status. Vaikka epäsanoina siis taivutetaan useimmiten neologismien tapaan, niitä ei pidä sekoittaa neologismeihin, sillä epäsanat eivät kannata merkitystä samalla tavalla kuin jälkimmäiset.

PÄÄTELMIÄ

Voimme siis todeta, että epäsanat säilyttävät jossain määrin sen pohjana olevan oikean sanan latausta, mutta syy ei ole semanttinen. Esimerkiksi kaksitavuista epäsanana prosessoitessaan koehenkilö ei ainakaan suoranaisesti assosioi sitä sen alla olevaan oikeaan sanaan. Kolmesta luonnollisesta taivutusvaihtoehdosta — *patsaan*, *keskuksen*, *Carlosin* — viimeksi mainittu on eniten agglutinoiva ja näin rakenteellisesti produktiivisin. Silti *Carlosin*-tyyppisiä vastauksia oli vain 0,62 % ($N = 17$), mikä osoittaa kielenkäyttäjien mieltävän epäsanat pikemmin yleisnimiksi kuin vierassanoiksi, erisnimiksi tai kirjainlyhenneiksi.

Aitojen nominien loppusegmenttejä käyttämällä saatu epäsanat on ateoreettisen oletuksen mukaan helpommin yhdistettävissä sen pohjana olevaan nominiin kuin epäsanat, jossa on muunnettu alkusegmentti. Väitettä tukevat muun muassa suomen affektiivisten sanojen (kirosanojen) muunnokset, joiden tarkoitus on lieventää pohjana olevan kirosanan »tehokkuutta», esimerkiksi *vitjat*, *vitsi*, *vitsu*, *vitura*, *vitskale*; *perskule*, *perskatti*, *perskutti*, *perhana*; *saakeli*, *samperi*, *saakutti*, *saaplari*, *saateri* jne. Näiden yhteisenä nimittäjänä on alkusegmentin koskemattomuus, millä pyritään säilyttämään selviä viitteitä muunnoksen alla olevaan affektiiviseen sanaan. Sananalkujen muistamisen suhteellista helpoutta osoittaa myös se, että akronyymit ovat tyypillisesti initialismeja. Siten muun

¹⁸ Korrelaatiotesti:

aineisto: *katsaan* ja *patsas*

$t = 2,1378$, $df = 8$, $p = 0,065$

korrelaatiokerroin 0,6029643

aineisto: *atsaan* ja *patsas*

$t = 2,0812$, $df = 8$, $p = 0,07098$

korrelaatiokerroin 0,5926708

aineisto: *patsaan* ja *patsas*

$t = -0,6899$, $df = 8$, $p = 0,5098$

korrelaatiokerroin -0,2369685

aineisto: *teskuun* ja *keskus*

$t = -1,2635$, $df = 8$, $p = 0,2420$

korrelaatiokerroin -0,4078799

aineisto: *eskuun* ja *keskus*

$t = -0,1384$, $df = 8$, $p = 0,8934$

korrelaatiokerroin -0,04885858

aineisto: *keskaan* ja *keskus*

$t = -1,2765$, $df = 8$, $p = 0,2376$

korrelaatiokerroin -0,4113578

muassa Sosialististen neuvostotasavaltojen liitto on SNTL (vrt. potentiaaliseen finalismiin *NONO). Samaa alkusegmenttien kaavaa noudattavat myös initialismien leksikaalistuneet laajentumat, kuten esimerkiksi Supo (vrt. *Lusi). Sananalkujen mnemonista etua käytetään hyväksi myös psykolingvistiikassa ja psykologiassa esimerkiksi leksikaalisissa kuvannimeämistehtävissä, joissa kohdetarkoitteen nimen löytymistä voidaan helpottaa antamalla joko semanttinen tai fonologinen vihje ja joissa tuo fonologinen vihje on sananalkuinen (esim. 1. tavu; menetelmästä tarkemmin, ks. esim. Laine, Koivuselkä-Sallinen, Hänninen ja Niemi 1990).

Väitettä tukee myös Mortonin logogeenimalli (1969). Se perustuu muun muassa foneemimonitorointikokeisiin, joissa koehenkilöä pyydetään painamaan nappia silloin, kun hän kuulee jonkin tietyn äänteen annettussa, yleensä lyhyen lauseen tapaisessa kontrolloidussa ärsykkeessä. Tämän tyyppisissä kokeissa on osoitettu, että kuulijat tunnistavat sanan yleensä jo silloin, kun siitä on kuultu niin paljon, että kuultu jakso vastaa yksiselitteisesti vain yhtä sanaa. Kuulijan ei tarvitse siis aina kuulla koko sanaa ennen leksikaalista hakua (*retrieval*). Logogeenimallin mukaan ajatellaan, että leksikko on järjestynyt sananalkuisten kohorttien mukaan. Lisäksi edeltävä konteksti vaikuttaa monitorointiin muun muassa siten, että semanttinen sukulaiskonteksti nopeuttaa hakua.

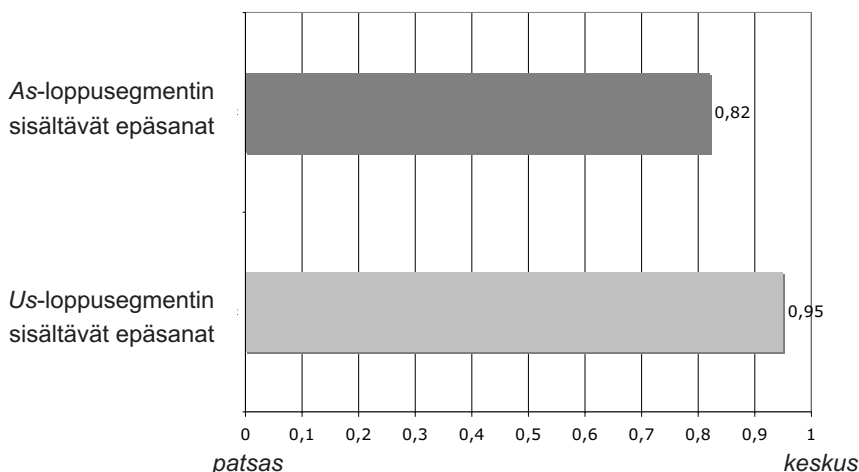
Saamamme tulokset eivät kuitenkaan osoita muunnetun segmentin paikan (joko sanan alussa tai lopussa) vaikuttavan epäsanojen *taivutukseen*. Sen sijaan segmentin prototyypisyys vaikuttaa merkittävästi epäsanojen prosessointiin.

Olemme mielestämme testanneet morfologiaa niin sanotun dissosiativisen morfologian (esim. Spencer 1991) viitekehyksessä ilman sanamuotojen liittämistä mahdollisiin merkitysrakenteisiin. Uskomme siis, että tuloksemme heijastavat sanamuotojen rakennetta, eivät merkityksiä: mikäli epäsanassa on jokin deklinaatiota osoittava prototyyppinen (temaattinen) segmentti (kuten esim. *patsas*-taivutustyyppissä *As*-loppusegmentti), juuri se antaa epäsanalle tietyn latauksen. Eli silloin kun korostamme sanan alkukonsonantismien tärkeyttä, liikumme leksikaalis-semanttisella pohjalla; silloin, kun koehenkilö joutuu sijoittamaan kuulemansa epäsanatiettyyn paradigmaan, hän liikkuu pikemmin dissosiativisen morfologian pohjalla. Näin selitämme näennäisen ristiriidan alkukonsonantismien ja sananloppujen tärkeyden välillä.

Koetta pohjustavassa analyysissämme toteamme, että *patsas*-tyypin epäproduktiivisuus ilmenee kaikkein selvimmin tyyppin leksikaalisessa taajuudessa ja tyyppin lemmafrekvensseissä. Edellinen on pieni (varsinkin *keskus*-tyyppiin verrattuna) samalla kun jälkimmäinen on suurempi kuin *keskus*-tyypillä. Toisin sanoen *patsas*-tyyppisiä sanoja on paljon vähemmän kuin *keskus*-tyyppisiä, ja ne ovat vielä »elossa» suhteellisen korkean käytöleyisyytensä ansiosta.

Johdannossa mainituissa suomen kielen tutkimuksissa tavallisin epäsanojen muodostamiskeino oli (useimmiten) kaksitavuisten aitojen substantiivien tai verbien alkukonsonantin vaihto toiseksi (konsonantin poisto ja vokaalialkuisten sanojen sananalkuisen konsonantin lisäys mukaan lukien). Tällä keinolla epäšana tavallaan vapautetaan sen alla olevan aidon sanan paradigmasta, minkä jälkeen sitä taivutetaan produktiivisen taivutustyyppin mukaisesti kartoittaen samalla, mikä epäsanasta kilpailevista paradigmoista on produktiivisin. Tämä tapahtuu kuitenkin vain sillä ehdolla, ettei epäšana sisällä yhtään prototyyppistä segmenttiä, joka kuuluisi epäproduktiiviseen taivutustyyppiin. Mikäli epäsanassa tällainen segmentti kuitenkin on, epäsanalla voidaan todeta olevan oma latauksensa, jota

periaatteessa voi jopa mitata esimerkiksi käyttämiemme testien avulla. Tämä niin sanottu lataus on kuitenkin sen verran pieni, ettei se vaihda epäsanana »polarisaatiota», vaan hiukan etäännyttää sitä produktiivisesta taivutustyyppistä (ks. kuvio 3, joka havainnollistaa taulukon 9 (s. 60) keskiarvoja).



Kuvio 3. Taulukon 9 keskiarvot. X-akselin 0 edustaa epäproduktiivista taivutustyyppiä *patsas* : *patsaan* ja 1 — vastaavasti — produktiivista *keskus* : *keskuksen*.

Käyttämämme analyysi pätee ainoastaan kaksitavuisiin epäsanoihin. Tavujen ja äänne-segmenttien lukumäärän kasvaessa epäsanasta tulee tavallaan läpinäkyvämpi, jolloin se asosioidaan helpommin sen alla olevaan oikeaan sanaan (vrt. Bybeen skeema-analogiaan, jossa paradigmata voivat järjestyä prototyypin hahmotyypinsä ympärille). Analyysimme tukee Bybeen hypoteesia. Emme kuitenkaan operoi hahmotyypin käsitteellä, vaan segmentin, jonka prototyypisyys määrää viime kädessä epäsanana käyttäytymisen taivutus-kentässä. Ei ole myöskään sattumaa, että kyseinen segmentti sisältyy useimmiten loppu-tavuun; taulukot 2 ja 3 (s. 50 ja 51) havainnollistavat suomen nomineille tyypillistä tavu-jen hajontaa: ensimmäisessä tavussa hajonta on laajin ja lopputavussa suppein. Silti epä-sanojen prosessoinnin yhteydessä tärkeäksi todettu deklinaatiota osoittava prototyypinen (temaattinen) segmentti ei välttämättä aina ole loppusegmentti. Niemi, Heikkinen ja Järvi-kivi ovat *Miksei kania vaikka munoja?* -nimisessä tutkimuksessaan (2001) saaneet psyko-lingvistisiä (behaviooraalisia) todisteita sille, että kielentutkijoiden lisäksi myös naiivit kiel-tenpuhujat pitävät suomen *a*-loppuisten substantiivien tapaisten epäsananojen ensitavun la-biaalisuutta taivutusparadigmaa vahvasti määrittelevänä morfologisena indeksinä.

Näin voimme todeta, että prototyypinen segmentti on yksi tärkeimmistä paradigmaan sitovista tekijöistä, jotka vaikuttavat potentiaalisten sanojen prosessointiin. Mikäli tämä segmentti edustaa produktiivista paradigmaa, on loogista, että segmentin sisältävä epä-sana mielletään produktiiviseen paradigmaan kuuluvaksi. Sama tapahtuu myös silloin, kun yksikään epäsanana segmenteistä ei ole prototyypinen paradigman edustaja (ja segmentit ovat näin neutraaleja). Sen sijaan tilanteessa, jossa epäsanasta »kilpailee» kaksi paradigmaa ja epäsanassa on epäproduktiivista paradigmaa edustava prototyypinen segmentti, kyseinen segmentti piilomuuttujan tapaan vaikuttaa epäsanana prosessointiin.

LÄHTEET

- ANTTILA, ARTO 1997: *Variation in Finnish phonology and morphology*. Julkaisematon väitöskirja. Yleinen kielitiede, Stanfordin yliopisto.
- 2003: Variation and phonological theory. – J. Chambers, P. Trudgill & N. Schilling-Estes (toim.), *The handbook of language variation and change* s. 208–243. Oxford: Blackwell.
- BERG, THOMAS – NIEMI, JUSSI 2000: Syllabification in Finnish and German. – *Journal of Phonetics* 28 s. 187–216.
- BERKO, JEAN 1958: The child's learning of English morphology. – *Word* 14 s. 150–177.
- BLEVINS, JULIETTE 1995: The syllable in phonological alternations. – 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.
- BYBEE, JOAN 1985: *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
- BYBEE, JOAN – MODER, CAROL LYNN 1983: Morphological classes as natural categories. – *Language* 59 s. 251–270.
- CD-Perussanakirja*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 94. Helsinki: Kotimaisten kielten tutkimuskeskus 1997.
- DRESSLER, WOLFGANG 1985: *Morphology: The dynamics of derivation*. Ann Arbor: Karoma.
- EIKMEYER, HANS-JÜRGEN – SCHADE, ULRICH 1993: The role of computer simulation in neurolinguistics. – *Nordic Journal of Linguistics* 16 s. 153–169.
- KARLSSON, FRED 1983: *Suomen kielen äänne- ja muotorakenne*. Helsinki: WSOY.
- Kielipankki: <http://www.csc.fi/>.
- KIPARSKY, PAUL 2003: *Finnish noun inflection*. – D. Nelson, S. Manninen (toim.), *Generative approaches to Finnic and Saami linguistics*. Stanford: CSLI Publications.
- LAINEN, MATTI – KOIVUSELKÄ-SALLINEN, PÄIVI – HÄNNINEN, RITVA – NIEMI, JUSSI 1990: *Bostonin nimentätesti*. Helsinki: Psykologien kustannus.
- LEBEN, WILLIAM 1973: *Suprasegmental phonology*. Väitöskirja, MIT.
- 1980: A metrical analysis of length. – *Linguistic Inquiry* 10 s. 497–509.
- MCCARTHY, JOHN 1986: OCP Effects: Gemination and antigemination. – *Linguistic Inquiry* 17 s. 207–263.
- MCCARTHY, JOHN – PRINCE, ALAN 1993: *Prosodic morphology I: Constraint interaction and satisfaction*. Käsikirjoitus. University of Massachusetts, Amherst, ja Rutgers University. – <http://roa.rutgers.edu/view.php3?id=590>.
- MORTON, JOHN 1969: Interaction of information in word recognition. – *Psychological Review* 76 s. 165–178.
- NIEMI, JUSSI 1999: Production of grammatical number in specific language impairment: An elicitation experiment on Finnish. – *Brain and Language* 68 s. 262–267.
- 2004: Onset + rhyme or first mora + end: Intrasyllabic structure in Finnish. – *Italian Journal of Linguistics* 16 s. 373–385.
- 2006: Paradigm competition in inflection: An experimental note on Finnish verbs. – Mickael Suominen, Antti Arppe, Anu Airola, Orvokki Heinämäki, Matti Mies-

- tamo, Urho Määttä, Jussi Niemi, Kari Pitkänen & Kaius Sinnemäki (toim.), *A man of measure: A Festschrift in honour of Fred Karlsson on his 60th birthday* s. 227–235. Helsinki: The Linguistic Association of Finland.
- NIEMI, JUSSI – HEIKKINEN, JANNE 2000: Paradigmatic and syntagmatic inflectional morphology in specific language impairment. – *Journal of Language and Verbal Behaviour* 3 s. 80–85.
- NIEMI, JUSSI – HEIKKINEN, JANNE – JÄRVIKIVI, JUHANI 2001: Miksei kania vaikka munoja? – Helena Sulkala & Leena Nissilä (toim.), *XXVII Kielitieteen päivät Oulussa 19.–20.5.2000* s. 191–197. Acta Universitatis Ouluensis, B Humaniora, 41. Oulu: Oulun yliopisto.
- NIEMI, JUSSI – LAINE, MATTI 1997: Slips of the tongue as linguistic evidence: Finnish word initial segments and vowel harmony. – *Folia Linguistica* 31 s. 161–175.
- NIEMI, JUSSI – NIEMI, SINIKKA 1987: Acquisition of inflectional marking: A case study of Finnish. – *Nordic Journal of Linguistics* 10 s. 59–89.
- NIEMI, JUSSI – TIKKALA, ANNELI – EIKMEYER, HANS-JÜRGEN – LAINE, MATTI 1999: Naturalistic data and computer simulations of language production: Are word onsets special? – J. Järvi­kivi & J. Heikkinen (toim.), *Out loud: Proceedings of the 19th Meeting of Finnish Phoneticians* s. 60–64. Studies in Languages 33. Joensuu: University of Joensuu.
- NIEMI, SINIKKA 2002: Schemas and paradigms in Swedish plural formation. – *Brain and Language* 81 s. 464–474.
- NIEMI, SINIKKA – NIEMI, JUSSI 2002: Emergent nature of morphological paradigms: Plural inflection in Swedish and Finnish. – Renate Pajusalu & Tiit Hennoste (toim.), *Tä­hendusepüüdja / Catcher of the meaning. A Festschrift to Haldur Õim on the occasion of his 60th birthday* s. 285–296. Publications from the Department of General Linguistics 3. University of Tartu.
- NIKOLAEV, ALEXANDRE 2002: Eräiden suomen taivutustyyppien produktiivisuudesta. – *Puhe ja kieli* 22 s. 113–124.
- NIKOLAEV, ALEXANDRE – NIEMI, JUSSI 2005: Suomen nominien taivutuksesta: rytmi-, sivu­paino- ja agglutinaatiohypoteesien testausta. – *Virittäjä* 109 s. 482–505.
- PAUNONEN, HEIKKI 1976: Allomorfien dynamiikkaa. – *Virittäjä* 80 s. 82–107.
- PRINCE, ALAN – SMOLENSKY, PAUL 1993: *Optimality theory: Constraint interaction in generative grammar*. Rutgers University, New Brunswick, ja University of Colorado, Boulder.
- SHATTUCK-HUFNAGEL, STEFANIE 1987: The role of word-onset consonants in speech production planning: New evidence from speech error patterns. – E. Keller & M. Gopnik (toim.), *Motor and sensory processes of language*. Hillsdale, NJ: Erlbaum.
- SPENCER ANDREW 1991: *Morphological Theory*. Oxford: Blackwell.
- TAFT, MARCUS 1979: Recognition of affixed words and the word frequency effect. – *Memory & Cognition* 7 s. 263–272.
- TIKALA, ANNELI – EIKMEYER, HANS-JÜRGEN – NIEMI, JUSSI – LAINE, MATTI 1997: On the production of Finnish nouns: A psycholinguistically motivated connectionist model. – *Connection Science* 9 s. 295–314.
- TRANEL, BERNARD 1991: CVC Light syllables, geminates and moraic theory. – *Phonology* 8 s. 291–302.

PARADIGMATISATION OF NOUNS IN FINNISH
 HOW DO SPEAKERS STRUCTURALLY ASSOCIATE POTENTIAL NOUNS WITH
 PARTICULAR INFLECTIONAL CATEGORIES?

The Wug Test, in which linguistically naive speakers are required to inflect non-words, or pseudowords, has been used extensively in studies of inflectional morphology since the classic study by Jean Berko (1958). One motivation for the use of pseudowords is to avoid the potential effects of intervening lexical-semantic variables, which can arise when using real words.

The main aim of the study was to test the significance of a set of sub-lexical segments in the decisions made by speakers who were required to place a potential *s*-final Finnish noun in a specific paradigm. In these pseudoword tasks, which are based on the writers' analysis of corpus data, the phonological structure is manipulated by systematically changing various segments, and the strength of the link between the pseudoword and the 'underlying' real word is tested. This 'wug' task is preceded by two analyses: an analysis of the surface frequency of Finnish *s*-final nouns, and an analysis of the lemma frequency of the target paradigms and its effect on the processing of lexical items in these paradigms.

The results show that pseudowords are affected by the underlying real lexical items. The writers claim that the reason for this is not semantic but instead probably reflects formal associations. More specifically, the results indicate that if the pseudoword carries a prototypical (thematic) segment of an inflectional paradigm (e.g. final *-as/äs* in the *patsas* paradigm), this segment will have a considerable influence on the assignment of the pseudoword to a specific paradigm. ■

Yhteystiedot (addresses):

Yleinen kielitiede

Joensuun yliopisto

PL 111

80101 Joensuu

Sähköpostit: *alexandre.nikolaev@joensuu.fi*

jussi.niemi@joensuu.fi