

Lekseemin esiintymistaajuuden vaikutus kielenmuutoksen leksikaaliseen diffuusioon

KARI NAHKOLA
MARJA SAANILAHTI

Jo parikymmentä vuotta on leksikaalisen diffuusion käsitettä hyödynnetty seurattaessa äänteenmuutoksen leviämistä kielimuodon sisällä. Tämä ajatus, joka sinänsä on vanhaa perua (esim. Sturtevant 1917), lähtee siitä, että äänteenmuutos ei alkuvaiheessaan reaalistu kaikissa foneemiset ehdot täyttävissä asemissa, vaan leviää vähitellen sanasta sanaan, morfeemista morfeemiin (erityisesti Wang 1969: 12–13). Kiistattomia esimerkkejä näin etenevästä muutoksesta on eri kielistä runsaasti. Koko ajatus on tosin jyrkästi ristiriidassa sen nuogrammaatikoilta periytyvän käsityksen kanssa, että äänteenmuutokset olisivat säännöllisiä, ts. toteutuisivat periaatteessa samanaikaisesti kaikissa ehdot täyttävissä asemissa. Nykyinen variaatioteoreettinen näkemys on tavallaan synteettinen: jotkin muutokset leviävät todistettavasti morfeemeittain mutta eivät kaikki (Labov 1981: 297–303; Dressler 1982: 94–96).

Monet kielelliset ja kielenulkoiset syyt voivat aiheuttaa sen, että muutos toteutuu yhdessä asemassa nopeammin kuin toisessa. Sanojen frekvenssi on ilmeisesti tässä suhteessa merkittävä tekijä (ks. jo esim. Schuchardt 1885: 58). Jonkinlainen korrelaatio frekvenssin ja muutosalttiuden välillä tuntuu olevan, ja tältä pohjalta pyrkii ns. taajuusherkkyyden hypoteesi (frequency actuation hypothesis, ks. Phillips 1984; suom. termi Nahkola 1987) osaltaan selittämään sanoittaista leviämistä. Monissa muutoksissa on nimittäin havaittu suurtaajuisten sanojen reagoivan herkimmin; toisissa muutoksissa, varsinkin analogisissa, pientaajuiset sanat ovat usein muita edellä. Kumpaa tietä muutos etenee, määräytyy hypoteesin mukaan foneettisten seikkojen, puheen fysiologian perusteella. Sellaiset muutokset, jotka synnyttää jokin luonnollinen foneettinen tekijä, toteutuvat ensin suurtaajuisimmista sanoista, mutta sellaiset muutokset, joiden syyt eivät ole luontevan foneettista laatua, alkavat tiensä leksikon pientaajuisesta päästä. Hypoteesin muotoilu ei ole aivan yksiselitteinen, mutta useimmat muutokset ovat suhteellisen hel-

Lekseemin esiintymistaajuuden vaikutus

posti luokiteltavissa jompaankumpaan ryhmään. Esimerkiksi seuraavassa tarkasteltava yleisgeminaatio sijoittuu kiistatta ensimmäiseen kategoriaan, foneettislähtöisiin, »fysiologisesti» motivoituihin muutoksiin (ks. Nahkola 1987: 11–32). »Ei-fysiologisesti» motivoituun tyyppiin on luettu sellaisia muutoksia, jotka foneettiselta kannalta ovat epätodennäköisiä tai jopa luonnottomia (Phillips 1984: 324–326). Esimerkiksi lainasanojen soinnillisten klusiilien muuttuminen soinnittomiksi jopa soinnillisessa ympäristössä (*pubi* > *pupi*) ei ole fysiologisesti motivoitua.

»Ei-fysiologisesti» motivoituissa muutoksissa on usein kyse pyrkimyksestä isomorfisuuden palauttamiseen, mikä yhdistää ne paitsi analogisiin muutoksiin (ks. Anttila 1972: 101) myös muutosten peruutuksiin eli antimuutoksiin. Tällainen peruutus on esimerkiksi antigeminaatio, joka pyrkii hävittämään geminaatiota. Tässä tutkimuksessa testataan taajuushypoteesin luotettavuutta sekä geminaation että antigeminaation valossa. Koska edellinen osuu lujuimmin suurtaajuiseen osaan sanastoa ja jälkimmäinen pientaajuiseen, olisi eritaajuisten lekseemien geminaatioedustuksessa ilmevä eroja, jos hypoteesi pitää paikkansa.

Systemaattisia huomioita lekseemin taajuuden ja muutosalttiuden mahdollisesta korrelaatiosta ei vielä ole paljon tehty. Eri kielistä tähän mennessä tehdyt havainnot ovat antaneet hypoteesille tukea molemmiin päin: tietoja on sekä »fysiologisesti» motivoitujen muutosten tarttumisesta erityisesti suurtaajuisiin sanoihin että »ei-fysiologisesti» motivoitujen vahvuudesta pientaajuisissa lekseemeissä (ks. esim. Leslau 1969; Fidelholz 1975; Hooper 1976; Phillips 1980, 1983 ja 1984).

Suomessa taajuushypoteesia on testannut Kari Nahkola (1987: 190–237), jonka havainnot tukevat hypoteesia. Muutamissa muissa suomalaisissa tutkimuksissa on esitetty lähinnä hajahuomioiden luonteisia kommentteja eritaajuisten lekseemien muutosalttiudesta. Merkittävää on, että nämä huomiot ovat kaikki vähintäänkin lievästi taajuushypoteesia tukevia. Yleisgeminaatiosta tällaisia havaintoja ovat tehneet Leena Silvennoinen (1980: 94) ja Kaija Jonninen-Niilekselä (1982: 157–159). Tietoja on myös švaavokaalin edustuksesta eritaajuisissa sanoissa (Nuolijärvi 1986: 64–65, 95; Kääriäinen 1988: 142–143). Muita vokaalinmuutoksia koskevia hajamainintoja on ainakin *eA*-yhtymien assimiloitumisesta pitkiksi vokaaleiksi (Silvennoinen 1981: 3–4; Nuolijärvi 1986: 148) sekä *A*:n loppuheitosta (Lahtinen 1981: 35–50; Nuolijärvi 1986: 161–162) samoin kuin eräistä labiaalistumismuutoksista (Nuolijärvi 1986: 98). Kaikissa näissä on kysymyksessä muutos, joka on luokiteltavissa »fysiologisesti» motivoituksi ja jonka havaitaan esiintyvän erityisen vahvana aineiston suurtaajuisimmissa lekseemeissä. Näin taajuushypoteesi ennustaakin. Erityisen selvästi näihin »fysiologisesti» motivoi-

tuihin muutoksiin kuuluvat pikapuheisuudet ja kulumat, jotka myös ovat paljolti keskittyneet erittäin suurtaajuisiin sanoihin (esim. E. Itkonen 1966: 205–206; Rapola 1966: 264–266; myös Mielikäinen 1980). Tukea hypoteesille antavat myös Ulla Nissin (1981: 59) havainnot 3. infinitiivin tunnuksen kadosta illatiivimuodoissa. Tämä muutos kuuluu luontevimmin tyyppiin »eifysiologisesti» motivoitu (*tekeen pro tekemään*), ja Nissin tutkimus osoittaa, että hypoteesin ennustamalla tavalla suurtaajuiset verbit ovat muutoksen toteuttamisessa pääosin jäljessä pienempitaajuisia verbejä.

Ainoa kotimainen tutkimus, jossa taajuusherkkyyshypoteesia on suoraan pyritty testaamaan, on mainittu Nahkolan tutkimus yleisgemaation lingvistikisistä variaatioista (1987). Seuraavilla sivuilla käsitellään toista yleisgemaatiosta tehtyä tutkimusta, jossa niin ikään on haluttu selvittää taajuushypoteesin pitävyyttä. Muuttujaksi on valittu sama äännepiirre kuin Nahkolan tutkimuksessa, jotta havaintoja olisi helpompi verrata. Nyt esiteltävän tutkimuksen perustana on Marja Saanilahden tutkimus yleisgemaation sosiaalisesta vaihtelusta (1988). Tuon tutkimuksen aineistoa tarkastellaan seuraavassa nimenomaan taajuushypoteesin kannalta. Pyrkimys on selvittää, tukeeko gemaation edustus Saanilahden aineistossa hypoteesia vai ei.

Aineisto on kerätty Virtain Vaskiveden kylästä. Kielenoppaita on 46; heiltä on ollut käytettävissä keskimäärin vajaan tunnin mittaiset nauhoitetut haastattelut. Puhujista vanhimmat ovat syntyneet 1800-luvun loppupuolella, nuorimmat 1970-luvulla. Yleisgemaation sosiolingvistinen asema on Vaskivedellä hyvin samanlainen kuin muuallakin hämäläis-satakuntalaisella gemaatioalueella: vanhoilla gemaatio on vahvaa mutta nuorilla hyvin heikkoa. Keski-ikäisillä idiolekteittaiset erot ovat suuria, sillä tässä ikäryhmässä tapaa kaikenvahvuista gemaatiota: osa puhujista (etenkin miehistä) on säilyttänyt gemaation hyvin, mutta monet ovat vaihtelevassa määrin antaneet periksi yleiskielen tukemalle peruutusmuutokselle, antigemaatiolle, ja luopuneet ainakin osittain gemaation käytöstä. Kaikkiaan variaation asteikko on niin laaja, että edellytykset taajuushypoteesin ennustamalle leksemeittäiselle hierarkialle gemaation vahvuudessa ovat olemassa.

Tutkimusaineistosta on poimittu kaikki yleisgemaation ehdot täyttävät sananmuodot, joita oli 7271. Sananmuodot on tämän jälkeen ryhmitelty leksemeittäin: kunkin lekseemin kaikki gemaation ehdot täyttävät sananmuodot on laskettu yhteen riippumatta siitä, sisältääkö taivutusmuoto pitkän vokaalin vai diftongin (esim. *taloon* ~ *taloissa*). Tällä tavoin on määritetty gemaation keskimääräinen vahvuus kussakin lekseemissä. Vahvuuden mittaamiseksi on olemassa useitakin keinoja. Tässä tutkimuksessa on varianstit, siis gemaation ehdot täyttävässä asemassa sijaitsevat konsonantit, ryhmitelty kahtia siten, että toisessa ryhmässä ovat pidentymättömät kon-

Lekseemin esiintymistäajuuden vaikutus

sonantit, toisessa pidentyneet. Pidennyksen foneettista kvantiteettia ei siis ole otettu huomioon, vaan pidentyneiden ryhmään on luettu paitsi täydet geminaatat myös lyhytalkuiset geminaatat ja puolipitkät yksinäiskonsonantit (esim. *talloon*, *talloon* ja *taloon* vs. *taloon*). Geminaation lekseemeittäistä vahvuutta kuvataan luvulla, joka osoittaa pidentyneen variantin sisältävien muotojen prosenttiosuuden lekseemin kaikista esiintymistä.

TAULUKKO 1. Yleisgeminaation vahvuus suurtaajuisissa lekseemeissä.

	gem.- pros.	N	
1. mitään	51	487	
2. tulla	52	450	
3. enää	57	372	
4. sanoa	37	258	
5. jotain	26	226	
			47 %
6. yleensä	30	224	
7. mennä	51	222	
8. osata	35	164	
9. pitää	42	122	
10. puhua	12	115	
			36 %
11. tehdä	51	111	
12. hyvä	57	97	
13. kevät	61	87	
14. kone	52	81	
15. talo	49	75	
16. usein	44	73	
17. sata	64	66	
18. kukaan	58	59	
19. ruis	87	53	
20. liha	50	52	
			56 %
21. asua	27	51	
22. eläin	14	51	
23. väliin	49	49	
24. mukaan	49	49	
25. talous	4	49	
26. kala	52	48	
27. vene	83	46	
28. oma	54	46	
29. vapaa	34	46	
30. nähdä	64	44	
			45 %
			pientaaj. sanat 48 %

Taulukkoon 1 on koottu aineiston 30 suurtaajuisinta lekseemiä frekvenssien mukaiseen järjestykseen; nämä lekseemit kattavat yli puolet aineistoon sisällyneistä geminaation ehdot täyttäneistä sananmuodoista. Kunkin lekseemin jäljessä on ilmoitettu pidentyneiden varianttien prosenttiosuus tuon lekseemin sananmuodoissa (eli geminoitumisprosentti) sekä niiden sananmuotojen kokonaismäärä (N) aineistossa. Taulukkoon on lisäksi merkitty keskimääräinen geminoitumisprosentti viidelle suurtaajuisimmalle lekseemille (sijat 1–5; nämä kattavat jo neljänneksen kaikista ehdot täyttävistä sananmuodoista). Samoin on keskimääräinen geminoitumisprosentti laskettu viidelle seuraavalle lekseemille (sijat 6–10) sekä kymmenelle näitä seuraavalle (sijat 11–20) ja vielä kymmenelle taulukon viimeiselle lekseemille (sijat 21–30). Myös taulukon ulkopuolelle jäävän, pientaajuisemman sanaston keskimääräinen geminoitumisprosentti on annettu.

Suurtaajuisimpia geminaation ehdot täyttäviä lekseemeitä tutkimusaineistossa ovat pronominit *mitään* ja *jotain*, adverbis *enää* ja *yleensä* sekä joukko-puheessa useasti toistuvia verbejä. Suurtaajuisimmat nominit ovat vasta-sijoilla 12 (adjektiivi *hyvä*) ja 13 (substantiivi *kevät*). Taulukkoon on kaikesta päättäen valikoitunut yleisemminkin katsottuna suurtaajuisin geminaation ehdot täyttävä sanasto. Taulukon kärkipäässä nimittäin ovat ne 7 lekseemiä, jotka Nahkolan tutkimuksessa (1987) täyttivät »erittäin suurtaajuisen» sanan kriteerit, ja muut taulukon lekseemit sisältyvät Nahkolan aineiston »suurtaajuisiin» sanoihin. Koko kielen taajuussanastossa (Saukkonen ym. 1979) sijoittuu kolmannes taulukon lekseemeistä 100 suurtaajuisimman ja lähes kaikki muut 1000 suurtaajuisimman joukkoon. Suomen kielen 100:n taajimpaan käytetyn lekseemin joukossa on vain 4 sellaista, joissa geminaation ehdot täyttyvät joissain sananmuodoissa mutta jotka eivät tutkimusaineistossa yltäneet 30 suurtaajuisimman luetteloon. Nämä lekseemit ovat kielemme suurtaajuisin sana *olla* sekä nominit *sama*, *moni* ja *tapa*. Nimenomaan *olla*-verbistä huomattakoon, että siinä geminaatio on mahdollinen ainoastaan 3. infinitiivin tunnuksettomassa illatiivissa (*oleen*), mikä tekee lekseemistä geminaation kannalta pientaajuisen.

Taulukon 1 perusteella ei korrelaatio lekseemin taajuuden ja geminaation vahvuuden välillä vaikuta lainkaan ilmeiseltä. Kaikkein suurtaajuisimmat viisi sanaa ovat kyllä geminaatiossa edellä viittä seuraavaa, mutta vahvimmillaan muutos on kymmenen seuraavan sanan ryhmässä (sijat 11–20). Kaiken lisäksi geminoituminen on keskimäärin hiukan vahvempaa taulukon ulkopuolelle jäävässä, verrattain pientaajuisessa sanastossa kuin viiden suurtaajuisimman sanan ryhmässä. Aineiston kolme suurtaajuisinta lekseemiä (*mitään*, *tulla* ja *enää*) ovat kyllä geminaatiossa edellä muuta sanastoa: pidentyneiden varianttien keskimääräinen osuus on kolmella taajimpaan käy-

Lekseemin esiintymistajuuden vaikutus

tetyllä lekseemillä 53 % ja muussa sanastossa 47 %. Taulukon kolmen kärkilekseemin jälkeen taajuuskorrelaatio kuitenkin muuttuu käänteiseksi.

Selitystä taajuushypoteesin heikolta vaikuttavalle ennustusvoimalle voi yrittää etsiä puhtaan lingvistikista, taajuuden ulkopuolisista tekijöistä. Eri konsonanttien geminoitumisherkyys ei ensinnäkään ole yhtä suuri (ks. Nahkola 1987: 172–181). Taajuuskorrelaation huonoutta saattaisi näin ollen selittää heikosti geminoituvan konsonantin sisältävien lekseemien mahdollinen kasautuminen sanaston suurtaajuiseen osaan. Taulukossa 2 on

TAULUKKO 2. Yleisgemiaation vahvuus suurtaajuisissa lekseemeissä verrattuna geminaatioasemaisen konsonantin pidentymiseen pientaajuisessa sanastossa.

	gem.- pros.	N	vert.- pros.	
1. mitään	51	487	48	
2. tulla	52	450	52	
3. enää	57	372	53	
4. sanoa	37	258	53	
5. jotain	26	226	48	
				47 %
6. yleensä	30	224	52	
7. mennä	51	222	53	
8. osata	35	164	24	
9. pitää	42	122	48	
10. puhua	12	115	38	
				36 %
11. tehdä	51	111	50	
12. hyvä	57	97	46	
13. kevät	61	87	46	
14. kone	52	81	53	
15. talo	49	75	52	
16. usein	44	73	24	
17. sata	64	66	48	
18. kukaan	58	59	50	
19. ruis	87	53	50	
20. liha	50	52	38	
				56 %
21. asua	27	51	24	
22. eläin	14	51	52	
23. väliin	49	49	52	
24. mukaan	49	49	50	
25. talous	4	49	52	
26. kala	52	48	52	
27. vene	83	46	53	
28. oma	54	46	50	
29. vapaa	34	46	43	
30. nähdä	64	44	50	
				45 %
				pientaaj. sanat 48 %

suurtaajuisten sanaston geminoitumisprosenttien ja esiintymämäärien (N) jäljessä ilmoitettu vertailuluku, joka osoittaa geminaatioasemaisen konsonantin keskimääräisen geminoitumisprosentin taulukon ulkopuolisessa, pienempi-taajuudessa sanastossa (taulukossa 'vert.pros.'). Esimerkiksi lekseemien *mitään* ja *jotain* geminoitumista verrataan *t:n* keskimääräiseen geminoitumiseen pientaajuudessa sanastossa. Tällä tavoin voidaan poistaa ne häiriöt, joita konsonanteittaiset erot geminoitumisalttiudessa mahdollisesti aiheuttavat taajuushypoteesin toteutumiseksi.

Taulukosta näkyy, että 16 lekseemissä konsonantti on geminoitunut paremmin kuin pientaajuudessa sanastossa. Kahdessa tapauksessa geminoitumisprosentti on sama, ja 12 lekseemissä jää konsonantin geminoituminen sanaston suurtaajuudessa osassa heikommaksi kuin pientaajuudessa. Aineisto antaa tällä tavoin tarkasteltuna lievää tukea taajuushypoteesille. Silti on huomattava, että hypoteesia tukevat lekseemit ovat pääosin taulukon sijoilla 11–30. Kaikkein suurtaajuusimmassa sanakymmenikössä vain kolme lekseemiä nousee vertailulukunsa yläpuolelle geminaation vahvuudessa, vaikka hypoteesi ennustaa juuri tämän sanaryhmän olevan muutoksessa pisimmällä.

Edes konsonanteittain tarkasteltuna ei taajuuskorrelaatio siis käänny hypoteesin kannalta kovin edulliseksi. Voidaan kuitenkin osoittaa, että enemmän kuin geminoituvan konsonantin foneettinen laatu on geminaation toteutumista ohjaamassa se, onko geminaation äännekontekstina pitkä vokaali vai diftongi (Nahkola 1987: 226–228). Pitkä vokaali on nimittäin geminaatiolle diftongia otollisempi äänneympäristö. Tutkimusaineistossa geminaatio on erityisen heikkoa silloin, kun muutoksen laukaisijaksi sattuu alkuperäinen *i*-loppuinen diftongi (kuten sananmuodossa *taloissa*) tai *U*-loppuinen supistumadiftongi (esim. *tapaus*); itse asiassa geminaatio likimain puuttuu näistä äänneympäristöistä.

Erilaisten äännekontekstien suosiollisuuserot käyvät selittämään joidenkin suurtaajuisten sanojen heikohkoa geminoitumista. Aineistossa viidenneksi taajimpaan esiintynyt lekseemi, pronomini *jotain*, geminoituu verrattain heikosti, mitä osaltaan selittää se, että siinä geminaation kontekstina on aina diftongi. Toinen selittävä tekijä on lekseemin kaksimerkityksisyys. Primaarissa (yleiskielen kannalta ainoassa) merkityksessään *jotain* geminoituu kohtalaisen hyvin. Lekseemillä on kuitenkin myös uudempi, sekundaari merkitys 'noin, suunnilleen', jota osa puhujista käyttää taajaan määränilmauksissa (*jotain viis kilometriä*), ja tässä merkityksessä esiintyessään lekseemi jää lähes aina geminoitumatta. (Vastaavia havaintoja Nahkola 1987: 233.)

Taulukossa sijoilla 14 ja 15 olevien nominien *kone* ja *talo* geminoitumisprosenttia laskee se, että ne esiintyvät tässä aineistossa poikkeuksellisen

Lekseemin esiintymistaajuuden vaikutus

usein monikkomuodossa, ts. diftongillisessa asussa (pitkävokaalisten ja diftongillisten muotojen suhteellisesta osuudesta ks. esim. Nahkola 1987: 226–228; ks. myös esim. Räsänen 1979: 32–33). *Kone*-lekseemillä keskimääräinen geminoitumisprosentti on pitkävokaalisissa sananmuodoissa 56 mutta supistumadiftongin sisältävissä monikkomuodoissa vain 44. Vielä suurempi on hajonta *talo*-lekseemillä: pitkävokaalisista muodoista pidentyneitä on 75 %, kun taas alkuperäisdiftongin (eli erityisen epäsuotuisan geminaatiokontekstin) sisältävistä monikkomuodoista vain 4 % on pidentyneitä. Kun näillä lekseemeillä monikkomuotojen osuus on peräti runsas kolmannes kaikista esiintymistä, saa taajuushypoteesin näennäisen huono ennustusvoima lingvistisen selityksen. Niin ikään sijoilla 22 ja 25 olevat nominat *eläin* ja *talous* kärsivät geminaatiota hylkivästä äännerakenteestaan: niissä ei pitkävokaalisia sananmuotoja voi esiintyä.

Edellä käsitellyt lingvistiset näkökohdat selittävät jossain määrin taajuushypoteesin puutteellista toteutumista tutkimusaineistossa. Tällaisenaan korrelaatiota ei kuitenkaan rohkene varauksetta tulkita hypoteesia tukevaksi. Jokin yleinen selitys korrelaation vaisuudelle olisi löydettävä. Tutkimusaineisto saattaa kyllä sisältää merkittävääkin tukea taajuushypoteesille, mutta taajuuskorrelaation etsiminen koko aineiston perusteella laaditun taajuushierarkian mukaan ei ehkä johda selkeään tulokseen tutkimusaineiston sosiolingvistisen heterogeenisuuden vuoksi. Koko aineistoa kuvaava taulukko voi kätkeä sisäänsä sellaisia taajuuskorrelaatioita ohjaavia tekijöitä, jotka ovat seurausta geminaation vahvuuden suurista eroista puhujaryhmien välillä. Vanhojen geminaatio on vahvaa, keski-ikäisten horjuvaa ja nuorten heikkoa. Täten lekseemin keskimääräisen geminoitumisprosentin kannalta on ratkaisevaa se, mistä puhujaryhmästä lekseemin esiintymät enimmältään tulevat. Näin ollen on välttämätöntä selvittää, esiintyvätkö taulukoiden 1 ja 2 lekseemit suunnilleen yhtä taajaan kaikissa puhujaryhmissä.

Jo ylimalkainen katsaus suurtaajuisten lekseemien esiintymiseen eri puhujaryhmissä osoittaa, että monissa tapauksissa jakautuminen on hyvin epätaisaista. Esimerkiksi *ruis* ei esiinny nuorten puheessa kertaakaan, kun taas *puhua*-lekseemin esiintymistä noin 80 % on peräisin nuorilta. Koska koko aineiston kannalta suurtaajuisten sanojen käyttötaajuus vaihtelee geminaatioltaan eritasoisten puhujaryhmien kesken melkoisesti, on taajuuskorrelaatiota etsittävä kustakin puhujaryhmästä erikseen, jotta kokonaistaulukoihin 1 ja 2 kasaantuneet distribuution vinoumat eivät peittäisi alleen mahdollista taajuuskorrelaatiota.

Seuraavassa puhujat on jaettu neljään ryhmään geminaation vahvuuden perusteella. Ryhmäjaon pohjana on puhujittain laskettu geminaatioindeksi, jonka arvo asteikossa 0,0–3,0 on sitä korkeampi, mitä vahvempaa puhujan

gemmaatio on (indeksien laskemisesta ks. Saanilahti 1988: 21 sekä siinä mainittua kirjallisuutta). Indeksiluokkien rajat ovat seuraavat:

puhujaryhmä I:	indeksi 0,0—0,5
puhujaryhmä II:	indeksi 0,6—1,5
puhujaryhmä III:	indeksi 1,6—2,5
puhujaryhmä IV:	indeksi 2,6—3,0

Gemmaation väistyminen tutkimusalueelta jaksottuu ikäryhmittäin niin selvästi, että gemmaation vahvuuden perusteella tehty ryhmäjako seuraa varsin tarkoin puhujien ikäryhmittäistä jakaumaa. Ryhmässä I on 14 nuorta, heikkogemmaatioista puhujaa. Ryhmään II kuuluvat 10 puhujaa ovat enimmäkseen keski-ikäisiä naisia, joiden puheessa antigemmaation vaikutus näkyy niin ikään sangen vahvana. Ryhmässä III on 11 puhujaa, joista useimmat ovat syntyneet vuosisadan alkupuolella mutta mukana on myös kolme keski-ikäistä miestä. Tällä ryhmällä gemmaatio on vielä vahvaa, vaikka peruu- tusmuutoksenkin vaikutus on jo havaittavissa. Ryhmään IV lukeutuvista 11 puhujasta useimmat ovat syntyneet 1800-luvun loppupuolella, mutta ryhmään kuuluu myös kaksi keski-ikäistä miestä. Ryhmän gemmaatio on vahvaa, ja vain muutamalla puhujalla on merkkejä edes lievistä muutoksen häiriintymisestä.

Kun gemmaation ehdot täyttävien sananmuotojen jakautumista tarkastellaan eri ryhmien kesken, havaitaan, että edellä mainittujen *ruis-* ja *puhua-*lekseemien lisäksi on muitakin tapauksia, joissa lekseemin suurtaajuisuus perustuu lekseemin erittäin taajaan käyttöön yhdessä puhujaryhmässä. Koska tällaiset taajuushuiput ovat useimmiten peräisin ryhmästä I, saa monen lekseemin yllättävän vaisu geminoituminen (koko aineiston kannalta taajuuteen suhteutettuna) luonnollisen selityksen. Seuraavassa lekseemien frekvenssiä tarkastellaan erikseen puhujaryhmittäin. Kullekin ryhmälle on laadittu erillinen taajuustaulukko, jossa on mukana ryhmän 20 suurtaajuusinta lekseemiä. Lekseemin jäljessä on taulukossa annettu kolme lukua. Ensimmäinen osoittaa lekseemin geminoitumisprosentin, ts. pidentyneiden varianttien osuuden lekseemin niissä sananmuodoissa, jotka täyttävät gemmaation ehdot. Seuraavana on annettu gemmaation ehdot täyttävien sananmuotojen absoluuttinen määrä kyseisellä lekseemillä tuossa puhujaryhmässä. Kolmantena on vertailuluku, joka ilmoittaa gemmaatioasemaisen konsonantin geminoitumisprosentin pientaajuudessa sanastossa kyseisellä puhujaryhmällä. Lisäksi taulukoissa on ilmoitettu keskimääräinen geminoitumisprosentti ryhmän a) viidelle suurtaajuusimmalle lekseemille, b) viidelle seuraavalle lekseemille (sijat 6—10), c) kymmenelle taulukon viimeiselle lekseemille

Lekseemin esiintymistäajuuden vaikutus

(sijat 11–20) sekä d) pientaajuiselle eli taulukon ulkopuolelle jäävälle sanastolle.

TAULUKKO 3. Yleisgemmaation vahvuus puhujaryhmällä I suurtaajuisissa lekseemisissä verrattuna geminaatioasemaisen konsonantin pidentymiseen pientaajuisessa sanastossa.

	gem.- pros.	N	vert.- pros.	
1. tulla	16	207	8	
2. mitään	13	192	7	
3. sanoa	2	149	0	
4. jotain	3	138	7	
5. yleensä	9	125	8	9 %
6. enää	7	113	0	
7. mennä	7	101	0	
8. puhua	0	91	0	
9. osata	5	81	0	
10. pitää	4	45	8	5 %
11. asua	5	41	0	
12. tehdä	11	36	7	
13. talous	0	27	8	
14. vapaa	33	24	15	
15. musiikki	0	23	0	
16. kukaan	4	22	7	
17. hyvä	0	22	0	
18. usein	9	21	0	
19. kone	5	21	0	
20. eläin	0	17	8	7 %
				pientaaj. sanat 6 %

Ryhmän kaksi suurtaajuisinta lekseemiä *tulla* ja *mitään* ylittävät geminaation vahvuudessa vertailulukunsa suhteellisen selvästi. Niin ikään viiden suurtaajuisimman lekseemin keskimääräinen geminoitumisprosentti on korkeampi kuin muiden taajuusluokkien. Aivan tasaisesti ei geminaation vahvuus tunnu lekseemien taajuutta seuraavan, sillä geminoitumisprosentti on matalin toisessa taajuusluokassa (sijat 6–10). Taajuuskorrelaatio näkyy taulukossa 3 kuitenkin selvempänä kuin koko aineistoa kuvaavissa taulukoissa 1 ja 2: 12 lekseemiä ylittää vertailulukunsa geminaation vahvuudessa, ja vain 5 tapauksessa jää suurtaajuisen lekseemin geminaatio heikommaksi kuin pientaajuisesta sanastosta määritetty vertailulukku. Näistäkin tapauksista kolmelle on löydettävissä lingvistinen selitys: geminaatiolle verrattain epäsuotuisa

äännekonteksti vaikuttaa ainakin osin siihen, että lekseemit *jotain* (sijalla 4), *talous* (13) ja *eläin* (20) jäävät geminaation vahvuudessa jälkeen pientaajuisten sanaston keskimääräisestä geminoitumisprosentista. *Jotain*-lekseemin geminaatiota heikentää ratkaisevasti se, että tässä puhujaryhmässä lekseemillä on hyvin usein merkitys 'noin, suunnilleen'; tässä merkityksessä esiintyessään *jotain* ei geminoidu juuri koskaan.

Koska geminaatio on ryhmässä I hyvin heikkoa, eivät lekseemeittaiset vahvuuserot voi olla suuria. Erot ovat kuitenkin järjestelmällisiä juuri siihen tapaan kuin taajuushypoteesi ennustaa: suurtaajuiset sanat sisältävät geminaatiota enemmän kuin pienempitaajuiset. Puhujaryhmittäisen tarkastelun merkitys tulee hyvin ilmi, kun luodaan silmäys muutamiin niistä lekseemeistä, jotka koko aineistoa kuvaavissa taulukoissa 1 ja 2 ilmensivät taajuuteensa nähden yllättävän huonoa geminoitumista. Taulukossa 4 on mainittu kahdeksan tällaista lekseemiä. Niille on yhteistä, että 45–79 % niiden kaikista esiintymistä on peräisin ryhmästä I, heikon geminaation puhujilta. Niinpä ei ole lopultakaan yllättävää, että noiden lekseemien geminaatio on jäänyt koko aineistonkin kannalta heikohkoksi. Taulukossa 4 on lekseemin jäljessä mainittu ensin sen sijoitus koko aineiston taajuuslistassa (taulukot 1 ja 2) sekä toiseksi prosenttiluku, joka osoittaa, miten suuri osa lekseemin kaikista esiintymistä on peräisin ryhmästä I.

TAULUKKO 4. Lekseemeitä, joiden esiintymistä keskimääräistä suurempi osuus on ryhmästä I.

tulla	2.	46 %
sanoa	4.	58 %
jotain	5.	61 %
yleensä	6.	56 %
mennä	7.	45 %
puhua	10.	79 %
talous	25.	55 %
vapaa	29.	52 %

Nuorten heikko geminaatio ilmensi siis melko hyvin taajuuskorrelaatiota. Taulukko 5 kuvaa ryhmän II edustusta; laadintaperiaate on sama kuin taulukossa 3.

Lekseemin esiintymistaajuuden vaikutus

TAULUKKO 5. Yleisgeminaation vahvuus puhujaryhmällä II suurtaajuisissa lekseemeissä verrattuna geminaatioasemaisen konsonantin pidentymiseen pientaajuisessa sanastossa.

	gem.- pros.	N	vert.- pros.	
1. mitään	42	104	37	
2. tulla	50	95	43	
3. enää	39	85	52	
4. jotain	47	57	37	
5. mennä	61	31	52	
				48 %
6. sanoa	51	31	52	
7. yleensä	38	29	43	
8. usein	33	27	13	
9. eläin	0	24	43	
10. kone	30	23	52	
				32 %
11. pitää	17	23	37	
12. liha	20	20	21	
13. kevät	26	19	26	
14. hyvä	37	19	26	
15. koti	38	16	37	
16. tehdä	38	16	30	
17. hakea	60	14	30	
18. oma	29	14	20	
19. ruis	46	13	30	
20. talous	0	13	43	
				30 %
				pientaaj. sanat 30 %

Puhujaryhmällä II erottuu ensimmäinen taajuusluokka (taulukon sijat 1–5) geminaation vahvuudessa selvästi muista. Viiden suurtaajuisimman lekseemin esiintymistä on liki puolessa konsonantti pidentynyt, kun muussa sanastossa pidentymien osuus jää alle kolmannekseen. Taajuusluokkien geminaatioprosentti laskee johdonmukaisesti kohti matalamman frekvenssin taajuusluokkia. Tulos on taajuushypoteesin mukainen. Ero toisen ja kolmannen taajuusluokan välillä (sijat 6–10 ja 11–20) olisi suurempikin, ellei sijalle 9 sattuva lekseemi *eläin* laskisi toisen taajuusluokan geminoitumisprosenttia. Tämä lingvistikista syistä geminaatiolle epäsuosiollinen lekseemi jää ryhmässä II kokonaan geminaation ulkopuolelle, mikä laskee geminaation vahvuutta toisessa taajuusluokassa. Muiden tämän luokan lekseemien keskimääräinen geminoitumisprosentti on noin 38. Myös kolmas taajuusluokka erottuisi pientaajuisesta sanastosta ilman sijalla 20 olevaa lekseemiä *talous*, joka niin ikään jää lingvistikista syistä geminaation ulkopuolelle. Muiden tämän luokan lekseemien geminoitumisprosentti on keskimäärin noin 35.

KARI NAHKOLA — MARJA SAANILAHTI

Puhujaryhmän II geminaatio toteuttaa siis sangen hyvin taajuushypoteesin ennustamaa suoraa korrelaatiota lekseemin taajuuden ja geminaation vahvuuden välillä. Taulukko 6 kertoo puhujaryhmän III tilanteen.

TAULUKKO 6. Yleisgeminaation vahvuus puhujaryhmällä III suurtaajuisissa lekseimeissä verrattuna geminaatioasemaisen konsonantin pidentymiseen pientaajuisessa sanastossa.

	gem.- pros.	N	vert.- pros.	
1. enää	97	107	83	
2. mitään	89	92	72	
3. yleensä	61	59	72	
4. tulla	96	52	72	
5. mennä	94	36	83	
				88 %
6. sanoa	97	32	83	
7. talo	53	32	72	
8. kevät	81	31	75	
9. hyvä	76	29	75	
10. vene	88	26	83	
				79 %
11. sata	80	25	72	
12. kukaan	100	22	80	
13. liha	73	22	61	
14. kala	59	22	72	
15. ruis	100	20	80	
16. mukaan	90	20	80	
17. kone	89	18	83	
18. usein	78	18	67	
19. nähdä	94	17	80	
20. perä	100	16	82	
				85 %
				pientaaj. sanat 71 %

Tässäkin ryhmässä geminoitumisprosentti on korkein ensimmäisessä taajuusluokassa (sijat 1–5). Geminaation vahvuus ei tosin heikkene tasaisesti lekseemien taajuuden pienentyessä, sillä kolmannen taajuusluokan geminoitumisprosentti on korkeampi kuin toisen luokan. Syynä tähän on sijalla 7 (siis toisessa taajuusluokassa) olevan *talo*-lekseemin huono geminoituminen. Tässä puhujaryhmässä *talo* sattuu esiintymään poikkeuksellisen taajaan monikkovartaloisena: noin puolet esiintymistä on diftongillisia muotoja, mikä — lingvistikista syistä — laskee lekseemin keskimääräisen geminoitumisprosentin matalaksi. Lekseemin pitkävokaaliset asut sisältävät tässä ryhmässä aina pidentyneen konsonantin (eli geminoitumisprosentti on 100), mutta monikkomuotojen geminoitumisprosentti on vain 6. Ilman tätä epätavallista

Lekseemin esiintymistajuuden vaikutus

monikkovartaloiden kasaumaa olisi *talo*-lekseemin geminaatio taajuusluokansa muiden lekseemien tasolla. Kaikkiaan kuitenkin puhujaryhmässä III peräti 17 taulukon 20 suurtaajuisesta lekseemistä ylittää geminaation vahvuudessa ne geminoitusmisarvot, jotka vastaavilla kosoanteilla on pientaajuisessa sanastossa.

Puhujaryhmän III geminaatio toteuttaa siis sekin taajuushypoteesin ennusteen hyvin. Taulukossa 7 tarkastellaan ryhmän IV edustusta.

TAULUKKO 7. Yleisgeminaation vahvuus puhujaryhmällä IV suurtaajuisissa lekseemeissä verrattuna geminaatioasemaisen konsonantin pidentymiseen pientaajuisessa sanastossa.

	gem.- pros.	N	vert.- pros.	
1. mitään	97	99	95	
2. tulla	97	96	96	
3. enää	100	67	100	
4. mennä	100	54	100	
5. sanoa	98	46	100	
				98 %
6. pitää	85	40	95	
7. tehdä	100	34	91	
8. osata	96	27	85	
9. hyvä	96	27	92	
10. kevät	96	23	92	
				94 %
11. talo	85	20	96	
12. sata	90	20	95	
13. ruis	100	20	91	
14. väliin	100	20	96	
15. kone	95	19	100	
16. vene	94	16	100	
17. jotain	93	15	95	
18. luja	100	14	98	
19. peläätä	100	13	96	
20. vetää	77	13	95	
				93 %
				pientaaj. sanat 93 %

Myös tässä geminaatioltaan vahvimmassa ryhmässä on ensimmäinen taajuusluokka edellä muita. Keskimääräinen geminoitusprosentti laskee sitä mukaa kuin siirrytään kohti pienemmän taajuuden luokkia. Erot taajuusluokkien välillä eivät ole suuria, koska tässä puhujaryhmässä geminaatio on kauttaaltaan vahvaa. Taulukossa 7 on kaksi lekseemiä, joiden geminoituminen kärsii hiukan siitä, että näillä lekseemeillä on erilaisia merkityssisältöjä, jotka eivät tunnu olevan geminaatioltaan samanarvoisia (ks. ed. lekseemi

jotain). Sijalla 6 oleva *pitää* geminoituu verrattain heikosti silloin, kun sen merkitys on 'täytyä', varsinkin jos samassa lauseessa lekseemi esiintyy myös toisessa merkityksessä, esim. *siitä pitää pittää lujasti kiinni*. Samoin sijalla 20 oleva lekseemi *vetää* geminoituu suhteellisen huonosti merkityksessä 'johattaa', kuten lauseessa *se vetää lausuntapiiriä*.

Kaikkiaan voi todeta, että puhujaryhmittäinen tarkastelu tukee taajuushypoteesia ainakin jossain määrin: kaikissa ryhmissä suurtaajuisten sanojen geminaatioprosentti nousee korkeammaksi kuin pientaajuisten sanaston, ja vahvinta geminaatio on ensimmäisessä taajuusluokassa, kaikkein suurtaajuissimmilla lekseemeillä.

Taajuushypoteesin testaamista vaikeuttaa se, että esiintymistaajuutta voidaan mitata monella eri tavalla. Ensimmäinen valinta on tehtävä jo ratkaistaessa, minkä sanaotoksen pohjalta operoidaan. Tässä tutkimuksessa taajuudet on laskettu tietyn rajallisen tutkimusaineiston sisällä, mutta kuten edellä todettiin, vaikka pohjana olisi ollut koko Suomen kielen taajuus-sanaston (Saukkonen ym. 1979) aineisto, olisi osapuilleen sama lekseemijoukko muodostanut suurtaajuisten sanojen taulukon.

Toinen tärkeä valinta on tehtävä päätettäessä, minkä yksiköiden taajuuksia lasketaan. Tässä tutkimuksessa on laskettu lekseemien taajuuksia. Varsinkin suomen kaltaisessa agglutinoivassa kielessä, jossa allomorfinen vaihtelu on runsasta, voisivat laskettavina yksiköinä kuitenkin yhtä hyvin olla allomorfit tai tapauksen mukaan pikemminkin erilaiset vartalot, alivartalot tai muut määrämuotoiset konstruktiot (esim. vartalo + tietynlainen suffiksiaines). Tällöin esimerkiksi lekseemin *talo* asemesta olisivat laskettavina yksiköinä erikseen muodosteet *taloo-* ja *taloi-*. On vaikea arvioida, kumpi laskuperuste olisi kielipsykologisesti perustellumpi, ts. kumpien taajuus on puhujan kannalta olennaista, lekseemien vai niiden erilaisten reaalistumien. Pulma on oikeastaan sama kuin morfologian kuvauksessa: onko kielipsykologisesti oikea vaihtoehto kuvauksen pohjaksi morfeemi vai allomorfi.

Mikäli taajuudet lasketaan erikseen samaan lekseemiin perustuville erilaisille allomorfeille, vartaloille tai muille vastaaville muodosteille, joudutaan uuden valinnan eteen. Tarjolla on kaksi päävaihtoehtoa. Ensinnäkin voimme laskea yhteen kaikki saman konstruktion sisältävät sananmuodot, esim. toisaalta kaikki *taloo*-esiintymät, toisaalta kaikki *taloi*-esiintymät. Toinen vaihtoehto on jakaa perusrakenteeltaan samanlaiset (ts. toisaalta pitkän vokaalin, toisaalta diftongin sisältävät) muodot edelleen sananmuodon kokonaisuuden perusteella esimerkiksi niin, että lasketaankin ainoastaan kokonaisen sananmuotojen taajuuksia. Tällä tavoin eri taivutusmuodot muodostaisivat omat ryhmänsä: yhdessä ryhmässä olisivat lekseemin illatiivimuodot

Lekseemin esiintymistaajuuden vaikutus

(esim. *taloon*), toisessa vain partitiivimuodot (*taloo*) jne. Tällöinen sananmuototasolle ulottuva ryhmittely ei tosin monessa tapauksessa vaikuttaisi lopputuloksiin paljonkaan jo pelkästään sen vuoksi, että eri taivutusmuotojen suhteellinen frekvenssi on hyvin erilainen esimerkiksi suomen kielessä. Niinpä nomineissa nominatiivi- ja partitiivimuodot ohjailisivat suurilla taajuuksillaan tutkimustuloksia samalla tavalla kuin jos jakaisimme sananmuodot vain kahtia *taloo*-ryhmään ja *taloi*-ryhmään. On syytä korostaa myös sitä, että aina tuloksiin ei ratkaisevasti vaikuta sekään, onko laskennan perusteena ollut lekseemi vai jokin ankarammin rajattu yksikkö: monissa tapauksissa jokin lekseemin rakennetyypeistä on muita selvästi suurtaajuisempi (esim. *taloo*- esiintyy paljon taajempaan kuin *taloi*-). Tämä suurtaajuinen muoto joka tapauksessa sanelee tutkimuksen tuloksen, olipa se laskettavana yksikkönä yksinään tai suurena enemmistönä kokonaisen lekseemin esiintymistä.

Mikäli lasketaan lekseemien (eikä erilaisten sublekseemisten muotojen) esiintymistaajuuksia, kolmas etukäteen ratkaistava seikka on se, mitkä lekseemin esiintymistä ylipäätään otetaan huomioon. Tässä tutkimuksessa mukaan otettiin vain ne sananmuodot, joissa tarkasteltavan muutoksen, yleisgeminatian, ehdot täyttyvät. Periaatteessa kuitenkin vaihtoehtona olisi määrittää lekseemien keskinäiset taajuuserot niin, että laskettaisiin kaikki kunkin lekseemin esiintymät riippumatta siitä, täyttävätkö ne seurattavan muutoksen ehdot vai eivät. Laskettavien lekseemien joukko ei suurenisi: huomioon otettaisiin vain lekseemit, joilla muutoksen ehdot täyttyvät edes josakin sananmuodossa. Lekseemien keskinäiset taajuussuhteet kyllä muuttuisivat. Aineiston suurtaajuisimmat lekseemit voisivat tuolloin olla sellaisiakin, joilla muutoksen ehdot täyttyviä sananmuotoja esiintyisi verrattain harvassa.

Lekseemien erilaiset morfofonologiset ominaispiirteet voivat saada aikaan sen, että tarkasteltavan muutoksen ehdot täyttyvät toisilla lekseemeillä huomattavastikin useammin kuin toisilla lekseemeillä. Esimerkiksi yleisgeminatian ehdot täyttävien taivutusmuotojen taajuus on huomattavan suuri vaikkapa *o*-vartaloisilla nomineilla (kuten *talo*), joilla (monissa murteissa) geminatian ehdot täyttyvät mm. frekvenssiltään suuressa yksikön partitiivissa (*taloo*) sekä monikkomuodoissa (*taloissa*, *taloilla* jne.). Geminatian alaisten muotojen suhteellinen osuus jää pieneksi etenkin konsonanttivartalon omaavilla *e*-vartaloilla (kuten *tuli*, *käsi*, *vesi*), joilla muutoksen ehdot täyttyvät lähinnä illatiivimuodoissa (*tuleen*, *tuliin*). Verbeillä eroja syntyy esimerkiksi sen mukaan, täyttyvätkö geminatian ehdot preesensin ohella myös imperfektimuodoissa (*sanoin*) vai eivät (*tulin*). Omana ryhmäänään ovat vielä ne lekseemit, joissa muutoksen ehdot täyttyvät aina, kuten

geminaatiossa eräät partikkelit (*enää, yleensä*) sekä *vapaa*-tyyppiset supistumanominit. Huomattakoon, että jos lekseemin kaikki esiintymät olisi tässä tutkimuksessa otettu laskuihin mukaan, olisi suurtaajuimpana lekseeminä ollut *olla*-verbi, jonka esiintymistä todellisuudessa vain ani harva täyttää geminaation ehdot: tutkimusalueella geminaatio on mahdollinen 3. infinitiivin tunnuksettomassa illatiivimuodossa (*oleen*).

Olipa laskettavana yksikkönä lekseemi tai jokin ankarammin rajattu yksikkö, kuten (ali)vartalo, voidaan taajuuden ja muutosalttiuden korrelaatiota etsiä vasta sen jälkeen, kun muut variaatiota kanavoivat tekijät ovat edes pääosiltaan selvillä. Esimerkiksi seurattaessa yleisgeminaation vahvuuden korrelointia sanan esiintymistaajuuden kanssa on oltava selvillä siitä, että pitkä vokaali on muutokselle suosiollisempi lingvistinen asema kuin diftongi ja että *U*-loppuinen diftongi on muutokselle erittäin epäsuosiollinen. Kun lingvistiset vaihtelurajoitukset ovat tiedossa, kyetään monia taajuuskorrelaation näennäisiä häiriöitä selittämään lingvistisin syin. Edellä tällaisia tapauksia olivat esimerkiksi lekseemin *talous* yllättävän heikko geminoituminen esiintymistaajuuteen nähden sekä *talo*-lekseemin matalahko geminaatio-prosentti puhujaryhmässä III, jossa poikkeuksellisesti liki puolet tuon lekseemin sananmuodoista oli monikkomuotoja (diftongillisia).

Lekseemien morfofonologisten ominaisuuksien erilaisuudesta voi myös seurata, että kaikki lekseemit eivät joudu samassa mitassa kärsimään lingvistisesti epäsuosiollisten vaihtelurajoitusten vaikutuksista. Tämä on otettava huomioon, mikäli taajuuslaskelmien perustana on kokonainen lekseemi. Yleisgeminaation kannalta ovat periaatteessa edullisimmassa asemassa ne lekseemit, joissa geminaatioasemainen konsonantti esiintyy aina pitkän vokaalin edellä, ei koskaan diftongin, esim. *hyvä* : *hyvää, hyvään*, mutta *hyvisä*. Sellaiset lekseemit, joilla monikkomuodoissa syntyy diftongi (*talo* : *taloissa*), ovat sikäli huonommassa asemassa, että niiden geminoitumisprosenttia pääsevät pienentämään nuo lingvistiseltä rakenteeltaan epäsuosiolliset alivartalot. Samalla tavoin verbeillä ovat parhaassa asemassa ne, joilla ei esiinny diftongillisia muotoja edes imperfektissä (*tulla* : *tulin*); asetelma ei ole yhtä otollinen, jos myös verbin imperfektimuodot täyttävät geminaation ehdot syntyvän diftongin takia (*sanoa* : *sanoin*). Tosin yleisgeminaation tapauksessa tämä epäsymmetria, jossa toisilla lekseemeillä on vain lingvistisesti suosiollisia asuja mutta toisilla lisäksi epäsuosiollisia, ei vaikuttanut tuloksiin kovin paljon. Diftongillisten sananmuotojen osuus on nimittäin tavallisesti suhteellisen vähäinen (poikkeuksia on, kuten *talo*-lekseemi puhujaryhmässä III). Nahkolan tutkimuksessa (1987) oli morfofonologisen vaihtelun aiheuttama epäsymmetria eliminoitu suurelta osin: lingvistisen suosiollisuuden kannalta erilaiset sananmuodot pidettiin erillään, nimenomaan pit-

Lekseemin esiintymistäajuuden vaikutus

kävokaaliset oli erotettu diftongillisista. Taajuuskorrelaatiosta saadut havainnot olivat olennaisesti samanlaisia kuin nyt tehdyssä tutkimuksessa, nimittäin että muutoksen vahvuus korreloi suoraan sanan taajuuden kanssa mutta vahvuuserot eivät ole suuria. Koska yleisgemmaatioitapauksista suuri enemmistö on pitkän vokaalin sisältäviä sananmuotoja, ei tätä nimenomaisesti muutosta tarkasteltaessa ole suurta merkitystä edes sillä, lasketaanko esiintymistäajuudet jakamattomista lekseemeistä vai eri tavoin ryhmitellyistä sananmuodoista (esim. pitkän vokaalin sisältävät erotettuina diftongin sisältävistä). Kaikkien kielenmuutosten osalta tilanne ei ole näin yksinkertainen, vaan tulokset voivat hahmottua hyvinkin erilaisiksi sen mukaan, kumpi laskuperuste valitaan. Kunkin muuttujan osalta ratkaisua on harkittava erikseen.

Taajuuden määrittämiseen vaikuttaa myös se, miten suhtaudutaan homonymiaan ja polysemiaan eli miten paljon semanttista tietoa hyödynnämme laskettavia yksiköitä rajattaessa. Edellä on selvät homonyymit laskettu eri lekseemeiksi, mikä todennäköisesti on puhujan kannalta psykologisesti oikea ratkaisu (esim. sananmuoto *sataa* on merkityksensä perusteella luettu joko verbin *sataa* tai nominin *sata* esiintymäksi). Mutkikkaampi ilmiö taajuuslaskelmien kannalta on polysemia (samoin kuin polysemialta vaikuttava »epäselvä» homonymia). Tässä tutkimuksessa on kaikki polyseemisen pesyeen sanat luettu saman lekseemin esiintymiksi, mutta toisinkin rajan voi vetää. Jos esiintymien merkitykset ovat lähellä toisiaan, tuntuu perustellulta lukea tapaukset samaan lekseemiin, esimerkiksi konkreetti ja abstrakti merkitys samalle lekseemille, kuten *vetää kuormaa* ja *vetää lausuntapiiriä*. Vaikka tämmöistenkin tapausten kesken tuntuu eroja muutosaltiudessa toisinaan olevan, on toisaalta syytä välttää hyvin hienojakoiseksi haarovaa semanttista jaottelua, johon helposti joudutaan, mikäli tällaiset tapaukset luetaan kahdeksi eri lekseemiksi. Kun esiintymien syntaktis-semanttinen ero on äskeistä suurempi, joudutaan laskuperustetta harkitsemaan tarkemmin, esimerkiksi *pitää*-verbin esiintymät lauseessa *sitä pitää pittää kiinni*. Tässä tutkimuksessa tuollaisetkin tapaukset on luettu samaan lekseemiin, mutta semanttinen etäisyys on riittävä siihenkin, että esiintymiä tarkasteltaisiin kahden eri lekseemin sananmuotoina.

Muutoksen vahvuusasteen mittaaminen ei sekään ole aina ongelmaton. Jos kyseessä nimittäin on yleisgemmaation kaltainen, foneettisessa jatkumossa tapahtuva vähittäinen muutos, ei ole ilman muuta selvää, miten muutoksen tuottamia erilaisia variantteja tulisi ryhmitellä. Tällainen jatkumomuuttuja (käsitteestä ks. esim. Berdan 1978: 149; Nahkola 1988: 5–6) tuottaa lukemattoman määrän hiukan toisistaan eroavia variantteja, jotka muodostavat vanhan ja uuden äännekannan väliin foneettisen jatkumon. Kun

siis esimerkiksi geminaatioasemassa voi esiintyä paitsi pidentymätön yksinäiskonsonantti tai täysi geminaatta myös kestoltaan jonnekin niiden välille sijoittuva konsonantti, joudutaan päättämään, millä kriteereillä ja moneenko ryhmään nuo lukuisat variantit sijoitetaan. Nahkola (1987) määritteli kullekin yksikölle erityisen geminoitumisindeksin, joka otti varsin tarkasti huomioon sen, minkä kestoisia variantteja lekseemin eri esiintymissä oli. Nyt tehdyssä tutkimuksessa variantit on yksinkertaisuuden vuoksi jaettu kahtia niin, että toisessa ryhmässä ovat pidentymättömät yksinäiskonsonantit ja toisessa kaikki pidentyneet variantit (riippumatta pidennyksen foneettisesta asteesta). Kahtiajako voitaisiin tehdä niinkin, että esimerkiksi täydet geminaatat muodostaisivat oman ryhmänsä ja kaikki näitä lyhyemmät segmentit toisen ryhmän. Mahdollinen olisi myös kolmijako, jossa foneettisen jatkumon ääripäät saisivat omat ryhmänsä ja kaikki kestoltaan näiden väliin sijoittuvat segmentit olisivat kolmannessa ryhmässä.

Varianttien ryhmittelykriteerit on siis erikseen päätettävä, mikäli tarkasteltavana on jatkumomuuttuja. Joissain tapauksissa ratkaisu voi suorastaan vaikuttaa tutkimuksen tuloksiin. Nyt tehdyssä tutkimuksessa eivät erilaiset ryhmittelytavat olisi tuoneet juurikaan muutoksia lekseemeittäisiin geminoitumisarvoihin, koska aineistossa oli perin niukalti jatkumon sisään, foneettisten ääriasteiden välille, sijoittuvia variantteja. Tämän syynä on se, että tutkitut puhujat jakautuvat karkeasti kolmeen tyyppiin geminaatioedustuksensa perusteella ja kaikille näille tyypeille on ominaista foneettisten väliasteiden vähyys. Puhujien geminaatio on nimittäin a) kauttaaltaan vahvaa, b) kauttaaltaan heikkoa tai c) taantumassa. Taantuva yleisgeminaatio on laadullisesti harppauksellinen muutos: puhujalla esiintyy pääasiassa vain foneettisia ääri vaihtoehtoja, ei juuri lainkaan jatkumon sisään sijoittuvia variantteja. Etenemässä oleva geminaatio tuottaa viljalti myös noita foneettisia väliasteita, joten tuollaisessa tilanteessa olisi varianttien ryhmittelyperuste harkittava tarkoin.

Taajuuskorrelaation olemassaolon vahvistaminen ja hypoteesin osoittaminen niin muodoin oikeaksi vaativat vielä lisätutkimuksia. Edellä esitetyt havainnot samoin kuin yleisgeminaatiosta aikaisemmin tehty tutkimus (Nahkola 1987) tuntuvat kyllä tukevan hypoteesia, vaikka taajuusluokkien erot eivät olekaan suuret. Myös muutamat muut meillä ja muualla tehdyt tutkimukset asettuvat hypoteesin puolelle. On tosin huomattava, että useimmiten tutkimuksissa esitetään lekseemien taajuudesta vain ylimalkaisia huomioita: aineistosta on harvoin tehty täsmällisiä frekvenssilaskelmia, ja taajuuteen viitataan lähinnä kommentoimalla erikseen joidenkin hyvin suurtaajuisten lekseemien edustusta. Näissä hajahuomioissakin hypoteesi saa tavallisesti

Lekseemin esiintymistaajuuden vaikutus

tukea, ja ehkä vielä tärkeämpää on se, että hypoteesia vastaan olevia havaintoja ei ole esitetty käytännöllisesti katsoen ollenkaan.

LÄHTEET

- ANTTILA, RAIMO 1972: An introduction to historical and comparative linguistics. New York.
- BERDAN, ROBERT 1978: Multidimensional analysis of vowel variation. — Linguistic variation (toim. D. Sankoff) s. 149–160. New York.
- DRESSLER, WOLFGANG 1982: A semiotic model of diachronic process phonology. — Perspectives on historical linguistics (toim. W. Lehmann — Y. Malkiel) s. 93–131. Amsterdam.
- FIDELHOLZ, JAMES 1975: Word frequency and vowel reduction in English. — Papers from the eleventh regional meeting, Chicago Linguistic Society s. 200–213. Chicago.
- HOOPER, JOAN 1976: Word frequency in lexical diffusion and the source of morpho-phonological change. — Current progress in historical linguistics (toim. W. Christie Jr.) s. 95–105. Amsterdam.
- ITKONEN, ERKKI 1966: Kieli ja sen tutkimus. Helsinki.
- JONNINEN-NIILEKSELÄ, KAIJA 1982: Eräitä äänne- ja muoto-opillisia piirteitä. — Tampereen puhekieli tutkimuskohteena (toim. K. Jonninen-Niilekselä) s. 121–160. Folia fennistica & linguistica 6. Tampere.
- KÄÄRIÄINEN, MERJA 1988: »Vokaaleja on ylimääräisiä». Svaavokaali Virtain Vaskiveden murteessa. — Virtolaista venytystä (toim. M. Saanilahti—K. Nahkola) s. 113–189. Tampereen yliopisto, Virtain tutkimuksia 3. Tampere.
- LABOV, WILLIAM 1981: Resolving the Neogrammarian controversy. — Language 57 s. 267–308.
- LAHTINEN, LEENA 1981: *a:n* ja *ä:n* loppuheitto Jyväskylän puhekielessä. — NPMJy 3 (toim. A. Mielikäinen) s. 35–50. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 26. Jyväskylä.
- LESLAU, WOLF 1969: Frequency as determinant of linguistic change in the Ethiopian languages. — Word 25 s. 180–189.
- MIELIKÄINEN, AILA 1980: Pikapuhemuodot Jyväskylän puhekielessä. — NPMJy 1 (toim. A. Mielikäinen) s. 123–148. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 20. Jyväskylä.
- NAHKOLA, KARI 1987: Yleisgeminaatio. SKST 457. Mänttä.
- 1988: Suomen kielen tyyppioppia. Opera fennistica & linguistica 3. Tampere.
- NISSI, ULLA 1981: III infinitiivin illatiivi ja inessiivi jyväskyläläisten ja eteläpohjalaisien puhekielessä. — NPMJy 3 (toim. A. Mielikäinen) s. 51–65. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 26. Jyväskylä.
- NPMJy = Nykysuomalaisen puhekielen murros. Jyväskylän osatutkimus.
- NUOLIJÄRVI, PIRKKO 1986: Kolmannen sukupolven kieli. SKST 436. Mänttä.
- PHILLIPS, BETTY 1980: Old English *an* ~ *on*: A new appraisal. — Journal of English Linguistics 14 s. 20–23.
- 1983: ME diphthongization, phonetic analogy, and lexical diffusion. — Word 34 s. 11–24.
- 1984: Word frequency and the actuation of sound change. — Language 60 s. 320–342.
- RAPOLA, MARTTI 1966: Suomen kielen äännehistorian luennot. SKST 283. Helsinki.

KARI NAHKOLA — MARJA SAANILAHTI

- RÄSÄNEN, SEPPÖ 1979: Huomioita suomen sijojen frekvensseistä. — *Sananjalka* 21 s. 17–34.
- SAANILAHTI, MARJA 1988: »Vieläkö tulee kallaa?» Yleisgemmaation sosiaalinen vaihtelu Virtain Vaskivedellä. — *Virtolaista venytystä* (toim. M. Saanilahti—K. Nahkola) s. 13–111. Tampereen yliopisto, Virtain tutkimuksia 3. Tampere.
- SAUKKONEN ym. = SAUKKONEN, PAULI—HAIPUS, MARJATTA—NIEMIKORPI, ANTERO—SULKALA, HELENA 1979: Suomen kielen taajuussanasto. Porvoo.
- SCHUCHARDT, HUGO 1885: Über die Lautgesetze: Gegen die Junggrammatiker. — Englanninkielinen käännös teoksessa: Schuchardt, the Neogrammarians, and the transformational theory of phonological change (toim. Th. Venemann—T. Wilbur) s. 39–72. Frankfurt.
- SILVENNOINEN, LEENA 1980: Laaja-alaiset itämurteisuudet Jyväskylän puhekielessä. — *NPMJy* 1 (toim. A. Mielikäinen) s. 89–122. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 20. Jyväskylä.
- 1981: *a, ä* -loppuiset vokaaliyhtymät Jyväskylän puhekielessä. — *NPMJy* 3 (toim. A. Mielikäinen) s. 1–34. Jyväskylän yliopiston suomen kielen ja viestinnän laitoksen julkaisuja 26. Jyväskylä.
- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia.
- STURTEVANT, E. H. 1917: *Linguistic change*. Chicago.
- WANG, WILLIAM 1969: Competing changes as a cause of residue. — *Language* 45 s. 9–25.

On the effect of lexeme frequency on the lexical diffusion of a language change

KARI NAHKOLA and MARJA SAANILAHTI

During the past twenty years research on the spread of language change within a given form has drawn attention to the phenomenon known as lexical diffusion. The term refers to the way in which a change is not initially realized in all of its theoretically possible environments but spreads gradually from word to word, morpheme to morpheme. An important factor in this progress is the frequency of the lexical items concerned. There seems to be some kind of correlation between the frequency of a word and its susceptibility to change, and it is on this basis that the so-called Frequency Actuation Hypothesis seeks to account for lexical diffusion.

According to this hypothesis, sound changes caused by a natural, surface-phonetic factor are first realized in the most frequent words, whereas changes which do not have natural phonetic causes start in least frequent words. The

present article discusses a situation where a phonetically triggered sound change, primary gemination, is in competition with the corresponding reverse change, anti-gemination. Since the hypothesis predicts that such reverse changes should start from relatively infrequent words, and that gemination itself should first manifest itself in more frequent items, it ought to follow that gemination would be differently represented in words of different frequencies, if the hypothesis is correct.

The study supports the frequency hypothesis to some extent: gemination is indeed more in evidence the more frequent a word is, although the differences are not very large.

One difficulty in testing the hypothesis is that frequency of occurrence can be measured in several different ways. An initial selection is already made in the choice of sample: are the frequencies to

Lekseemin esiintymistaajuuden vaikutus

be defined on the basis of a given corpus or the whole of the language? A second important choice concerns the units in terms of which the frequencies are calculated. The study reported here is based on lexeme frequencies. Especially in a language like Finnish, however, where there is a good deal of allomorphic variation, one might equally well take allomorphs, or rather different stems, substems or other constructions of a given form (e.g. stem + a given suffix).

If frequencies are calculated separately for the various allomorphs or equivalent units of the same lexeme, a further choice arises, with two main alternatives. Either we can take as belonging to the same unit all word-forms containing a given phonetic structure; or we can divide these forms further so that each complete word-form itself counts as a separate unit. In this latter case all the inflected forms of a lexeme would be taken as separate units in the frequency calculation.

If the calculations are based on lexemes (and not on the forms of different sublexemes), it must be decided which

occurrences of the lexeme are to be included. The present study only included the word-forms which satisfied the conditions for the change in question. The alternative would have been to define the relative frequencies on the basis of all the occurrences of a given lexeme regardless of whether such conditions were satisfied or not. The total set of lexemes included would not increase: only lexemes where the conditions were satisfied in at least one word-form would be included. On the other hand, the relative frequencies of the lexemes might change radically. In such a case, lexemes counted as frequent might also include those where word-forms actually satisfying the conditions for change were themselves rare.

All in all, the testing of the frequency hypothesis is an extremely complex matter, and there is thus not yet sufficient evidence of its validity. Studies which have looked at the correlation between word frequency and susceptibility to change, however, do support the hypothesis, as do the observations on Finnish discussed here.