

# **Suomen kvantiteetti ja kirjoitusprosessi**

## **Mallinnusta ja empiriaa**

Lisensiaatintutkimus

Lokakuu 2007

Ville Pekonen

Kielten laitos (suomen kieli)

Jyväskylän yliopisto

## JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanistinen tiedekunta	Laitos – Department Kielten laitos / suomen kieli
Tekijä – Author Ville Pekonen	
Työn nimi – Title Suomen kvantiteetti ja kirjoitusprosessi – Mallinnusta ja empiriaa	
Oppiaine – Subject Suomen kieli	Työn laji – Level Lisensiaatintutkielma
Aika – Month and year 5.10.2007	Sivumäärä – Number of pages 154
Tiivistelmä – Abstract <p>Lisensiaatintutkielmani käsittelee kirjoituksen oppimista erityisesti pituuden merkitsemisen näkökulmasta. Työn empiirinen aineisto koostuu 230 lapsen kirjoitus-sanelutehtävästä, joista tutkittiin kvantiteetin merkitsemisen hallintaa. Tarkoituksena oli, että ensimmäisen kerran saataisiin tarkka kuva suomen kielen pituusdistinktioiden merkinnästä eri sana-asemissa. Varsinainen päämäärä oli selvittää, missä määrin kirjoitetun kielen oppimisessa vaikuttaa / ennustaa fonologinen kompleksisuus ja eri sana-tyyppien frekvenssi juoksevassa tekstissä (käyttöfrekvenssi) ja sanakirjafrekvenssi (tyyppifrekvenssi).</p> <p>Työ rakentuu kahdeksasta luvusta. Johdannon jälkeiset luvut käsittelevät generatiivista fonologiaa ja deklarativista fonologiaa. Niissä pyrin osoittamaan, että kirjoittamisen perusprosessin tutkimus nojaa valtaosin kielitieteen näkökulmasta liian yksinkertaiseen käsitykseen fonologiasta ja morfologiasta. Kolmannessa luvussa käsittelen morfologian ja pituuden suhdetta. Neljännessä luvussa esittelen autosegmentaalista fonologiaa ja tarkastelen, miten sitä voitaisiin soveltaa kirjoituksen perusprosessin kuvaamiseen. Viidennessä luvussa käsittelen suomen kielen astevaihtelua osana pituuteen liittyviä ongelmia ja pituutta morfologian ja fonologian ongelmana. Kuudennessa luvussa käsittelen konsonanttien ja vokaalien kvantiteetin yleisyyttä sekä juoksevassa tekstissä että <i>Nykysuomen sanakirjan</i> yhdistämättömissä sanoissa. Seitsemäs luku yhdistää kehrittelemäni kielitieteellisen formalismin ja frekvenssin kuvauksen työni empiiriseen aineistoon.</p> <p>Työni tuloksena on, että käyttöfrekvenssiltään pienet ja/tai tunnusmerkkiset sanat ovat kaikkein vaikeimpia oppia kirjoittamaan oikein. Juuri geminaatalliset sanat ovat vaikeimpia kirjoittaa ja myös harvinaisimpia juoksevassa tekstissä.</p>	
Asiasanat – Keywords kirjoittamaan oppiminen, fonologia, autosegmentaalinen fonologia, frekvenssi	
Säilytyspaikka – Depository Fennicum, FE Pekonen	
Muita tietoja – Additional information	

## Sisällys

<b>1. JOHDANTO .....</b>	<b>6</b>
1.1 TYÖN RAKENNE .....	6
1.2 FONEEMI- JA GRAFEEMITASOJEN SUHTEESTA.....	7
<b>2. GENERATIIVISTA FONOLOGIAA .....</b>	<b>10</b>
2.1 IA- JA IP-FONOLOGIA .....	10
2.2 FONOLOGISESTA JA MORFOLOGISESTA KOMPONENTISTA.....	14
2.3 KUINKA MORFOLOGIA JA FONOLOGIA VAIKUTTAVAT TOISIINSA?.....	16
2.4 GENERATIIVINEN FONOTAKTIikka .....	17
2.5 AFFIKSAALINEN MORFOLOGIA .....	17
2.6 LEKSIKAALINEN FONOLOGIA JA SYKLI.....	19
2.7 LEKSIKAALISTEN PIIRTEIDEN REDUNDANSITTOMUUS JA TUNNUSMERKKISYYS .....	23
2.8 MIKÄ ON FONOLOGINEN LÄPINÄKYVYYS?.....	26
<b>3. DEKLARATIIVINEN FONOLOGIA.....</b>	<b>28</b>
3.1 DEKLARATIIVISEN FONOLOGIAN PERUSOLETUKSET.....	28
3.2 LEKSIKKO DEKLARATIIVISESSA FONOLOGIASSA .....	29
3.3 RAJOITUKSET DEKLARATIIVISESSA FONOLOGIASSA.....	30
<b>4. AUTOSEGMENTAALISTA FONOLOGIAA .....</b>	<b>36</b>
4.1 TAVUNMUODOSTUS FORMAALISTA NÄKÖKULMASTA .....	36
4.2 ERILAISIA TAVUN HIERARKKISEN KUVAUksen MALLEJA.....	39
4.3 MORATEORIA JA SEN SOVELTAMINEN GEMINAATTOIHIN .....	43
4.4. PITUUDEN KOODAAMISONGELMAT KIRJOITUKSESSA FONOLOGISESSA MALLISSA .....	54
4.5 ÄÄNTEIDEN KESTOISTA .....	58
<b>5. ASTEVAIHTELUN TARKASTELUA ERI KUVAUSMENETELMIN.....</b>	<b>71</b>
5.1. ASTEVAIHTELU MORFOFONOLOGIAN ONGELMANA.....	71
5.2 LEKSIKAALINEN FONOLOGIA JA ASTEVAIHTELU .....	77
5.3 ASTEVAIHTELUN KUVAUS DF:ssa.....	82
<b>6. GEMINAATTOJEN FUNKTIONAALISESTA KUORMITUKSESTA .....</b>	<b>85</b>
6.1 ENSIMMÄISEN JALAN RAKENTEIDEN KÄYTTÖFREKVENSsit .....	85
6.2 PITKIEN ENSITAVUN JA TOISEN TAVUN VOKAALIEN KÄYTTÖFREKVENSsi.....	91
6.3 GEMINAATTOJEN TYYPIFREKVENSsi .....	95
6.4 KOLMITAVUISTEN SANATYYPIEN TYYPIFREKVENSsi.....	100
6.5 KAKSITAVUISTEN NOMINIEN KVANTITEETIN FUNKTIONAALINEN KUORMITUS.....	101
6.6 MINIMIPARIT JA NIIDEN FUNKTIONAALINEN KUORMITUS.....	106
6.7 MINIMIPARIEN FREKVENSsi .....	110
6.8 KOONTA SANATYYPIEN FREKVENSsiEN TARKASTELUSTA.....	112
6.9 FREKVENSsi JA TUNNUSMERKKISYYDEN TEORIA .....	116
<b>7. FONOLOGISEN TEORIAN KOKEELLINEN SOVELTAMINEN.....</b>	<b>121</b>
7.1 KOEASETELMA JA MENETELMÄT .....	121
7.2 PITUUDEN GRAFEEMINEN MERKITSEMINEN SANA-ASEMITTAIN .....	123
7.3 TIIVISTYS OIKEINKIRJOITUSTEKSTISTÄ .....	137

<b>8. LOPUKSI</b> .....	<b>145</b>
<b>LÄHTEET</b> .....	<b>155</b>

## Taulukot

Taulukko 1. Tavun pituuden kuvaus Body-mallissa.....	41
Taulukko 2. Tavun pituuden kuvaus riimimallissa.....	42
Taulukko 3. Tavun pituuden kuvaus kolmihaaraisessa mallissa.....	43
Taulukko 4. CVVC <sup>2</sup> .C <sup>2</sup> V-tyyppin nominit Käänteissanakirjan (Tuomi 1980 mukaan)...	49
Taulukko 5. Ylipitkät tavut (malli a).....	53
Taulukko 6. Ylipitkät tavut (malli c).....	53
Taulukko 7. Sanansisäisten konsonanttien ominaiskestit (ms.).....	60
Taulukko 8. Yksinäiskonsonantin ja geminaatan suhteelliset pituudet konsonanteittain.....	61
Taulukko 9. Konsonanttien ominaiskestit ääntöpaikan mukaan.....	62
Taulukko 10. Segmenttien kestot CVCV- ja CVCCV-sanatyypeissä.....	62
Taulukko 11. Umpitavuisuuden vaikutus toisen tavun alkukonsonanttiin.....	63
Taulukko 12. Geminaatta avo- ja umpitavun edellä.....	63
Taulukko 13. Toisen tavun alkukonsonantin kesto.....	64
Taulukko 14. Geminaatan pituus ylipitkissä tavuissa.....	64
Taulukko 15. Lyhyen klusiilin kesto toisen tavun alussa.....	65
Taulukko 16. Pitkän klusiilin kesto toisen tavun alussa.....	66
Taulukko 17. Lyhyen nasaalin kesto toisen tavun alussa.....	66
Taulukko 18. Pitkän nasaalin kesto toisen tavun alussa.....	66
Taulukko 19. Lyhyen klusiilin kesto toisen tavun alussa pitkän vokaalin jälkeen.....	67
Taulukko 20. Klusiiligeminaatan keston pitkän vokaalin jälkeen.....	67
Taulukko 21. Lyhyen nasaalin kesto pitkän vokaalin jäljessä.....	68
Taulukko 22. Nasaaligeminaatan kesto pitkän vokaalin jälkeen.....	68
Taulukko 23. Lyhyt kolmannen tavun tavunalkuinen klusiili.....	69
Taulukko 24. Klusiiligeminaatta toisen ja kolmannen tavun rajalla.....	69
Taulukko 25. Kolmannen tavun tavunalkuinen nasaali.....	70
Taulukko 26. Toisen ja kolmannen tavun tavunrajainen nasaaligeminaatta.....	70
Taulukko 27. Sanatason rajoitukset.....	79
Taulukko 28. Postleksikaalisen tason rajoitukset.....	79
Taulukko 29. Sanataso.....	81
Taulukko 30. Postleksikaalinen taso.....	81
Taulukko 31. Ensimmäisen tavun ja toisen tavun rakenne toisen tavun ensimmäiseen vokaaliin asti (SK).....	87
Taulukko 32. Ensimmäisen tavun ja toisen tavun rakenne toisen tavun ensimmäiseen vokaaliin asti (Holappa).....	88
Taulukko 33. Geminaattojen määrät ja suhteelliset osuudet eri asemissa 1. ja 2. tavun rajalla (SK).....	89
Taulukko 34. Geminaattojen ja yksinäiskonsonanttien määrät ja frekvenssit ensimmäisen ja toisen tavun rajalla (SK).....	90
Taulukko 35. Ensimmäisen ja toisen tavun rajalla olevat yhtymien määrät järjestettynä niiden käyttöfrekvenssin mukaan (SK).....	91
Taulukko 36. Pääpainollisen tavun pitkät vokaalit ja niiden käyttöfrekvenssi (SK).....	92

Taulukko 37. Vokaalien kaksoisgrafeemien määrät ja niiden käyttöfrekvenssi toisessa tavussa juoksevassa tekstissä (SK).....	93
Taulukko 38. Ensimmäisen tavun pitkän vokaalin ja tavorajan geminaatan yhdistelmien määrät juoksevassa tekstissä (SK).....	94
Taulukko 39. Pitkien ja lyhyiden kvantiteettien yhdistelmät (SK).....	94
Taulukko 40. Sanatyyppien tyypifrekvenssit (NS).....	95
Taulukko 41. Yksittäiskonsonanttien ja niiden geminaattojen määrät ja tyypifrekvenssit NS:n sanoissa.....	96
Taulukko 42. Sanatyyppien tyypifrekvenssit <i>Nykysuomen sanakirjan</i> mukaan (sanansisäiset klusiilit). .....	97
Taulukko 43. Sanatyyppien, joissa on sanansisäiset soinnilliset konsonantit, tyypifrekvenssit <i>Nykysuomen sanakirjan</i> mukaan.....	97
Taulukko 44. Pääpainolliset tavujen määrät NS:n mukaan. ....	98
Taulukko 45. Yleisimmät sanarakenteet vähintään kolmitavuisissa sanoissa järjestettynä pääpainollisen tavun yleisyyden mukaan sekä toisen tavun rakenteen yleisyyden mukaan (NS).....	99
Taulukko 46. Pääpainottoman tavun rakenteiden määrät ja tyypifrekvenssit kolmitavuisissa sanoissa (NS). ....	100
Taulukko 47. Konsonanttien kvantiteettioppositio 2. ja 3. tavun rajalla (ei e'-nomineja) (NS).....	101
Taulukko 48. Yksinäisklusiilien ja geminaattojen määrät (NS).....	102
Taulukko 49. Sanansisäisen klusiilin pituuteen perustuvat minimiparit (NS).....	102
Taulukko 50. Sanansisäinen /s/ yksinäiskonsonanttina ja geminaattana (NS).....	103
Taulukko 51. /s/ vs. /ss/ perustuvat minimiparit (NS).....	103
Taulukko 52. Soinnilliset yksinäiskonsonantit ja geminaatat (NS).....	104
Taulukko 53. Sanansisäisen soinnillisen konsonantin pituuteen perustuvat minimiparit (NS). ....	104
Taulukko 54. Minimipareja sanatyypeittäin (NS). ....	105
Taulukko 55. Minimiparit suhteutettuina kaikkiin kaksitavuisiin sanoihin (S = n, m, l, r) (NS).....	106
Taulukko 56. Niiden minimiparisanatyyppien esiintymät, jotka esiintyivät Kielipankin suomenkielisessä aineistossa enemmän kuin kymmenen kertaa (10 >). ....	111
Taulukko 57. Niiden minimiparisanatyyppien esiintymät, jotka esiintyivät Kielipankin suomenkielisessä aineistossa enemmän kuin sata kertaa (100 >). ....	112
Taulukko 58. Pääpainollisen tavun lyhyen vokaalin merkitseminen 1. ....	124
Taulukko 59. Pääpainollisen tavun lyhyen vokaalin merkitseminen 2. ....	124
Taulukko 60. Pääpainollisen tavun pitkän vokaalin merkitseminen 1. ....	125
Taulukko 61. Pääpainollisen tavun pitkän vokaalin merkitseminen 2. ....	125
Taulukko 62. Lyhyen klusiilin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä 1. ....	126
Taulukko 63. Lyhyen klusiilin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä 2. ....	126
Taulukko 64. Lyhyen klusiilin merkitseminen toisen tavun alussa pitkän vokaalin jäljessä 1. ....	127
Taulukko 65. Lyhyen klusiilin merkitseminen toisen tavun alussa pitkän vokaalin jäljessä 2. ....	127
Taulukko 66. Nasaalin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä. ....	127

Taulukko 67. Nasaalin merkitseminen toisen tavun alussa pitkän vokaalin jälkeen 1.	128
Taulukko 68. Nasaalin merkitseminen toisen tavun alussa pitkän vokaalin jälkeen 2.	128
Taulukko 69. Klusiiligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jäljessä 1.	129
Taulukko 70. Klusiiligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jäljessä 2.	129
Taulukko 71. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jälkeen 1.	130
Taulukko 72. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jälkeen 2.	130
Taulukko 73. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla pitkän vokaalin jälkeen 1.	131
Taulukko 74. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla pitkän vokaalin jälkeen 2.	131
Taulukko 75. Klusiiligeminaattojen merkitseminen pitkän vokaalin jälkeen 1.	132
Taulukko 76. Klusiiligeminaattojen merkitseminen pitkän vokaalin jälkeen 2.	132
Taulukko 77. Toisen tavun yksinäisvokaalin merkitseminen 1.	132
Taulukko 78. Toisen tavun yksinäisvokaalin merkitseminen 2.	133
Taulukko 79. Toisen tavun pitkän vokaalin merkitseminen 1.	133
Taulukko 80. Toisen tavun pitkän vokaalin merkitseminen 2.	134
Taulukko 81. Lyhyen klusiilin merkitseminen toisen ja kolmannen tavun rajalla 1.	134
Taulukko 82. Lyhyen klusiilin merkitseminen toisen ja kolmannen tavun rajalla 2.	135
Taulukko 83. Lyhyen nasaalin merkitseminen toisen ja kolmannen tavun rajalla 1.	135
Taulukko 84. Lyhyen nasaalin merkitseminen toisen ja kolmannen tavun rajalla 2.	135
Taulukko 85. Klusiiligeminaatan merkitseminen 2. ja 3. tavun rajalla 1.	136
Taulukko 86. Klusiiligeminaatan merkitseminen 2. ja 3. tavun rajalla 2.	136
Taulukko 87. Nasaaligeminaattojen merkitseminen 2. ja 3. tavun rajalla 1.	137
Taulukko 88. Nasaaligeminaattojen merkitseminen 2. ja 3. tavun rajalla 2.	137
Taulukko 89. Tiivistys oikeinkirjoitustestin tuloksista.	139
Taulukko 90. Virheelliset segmentit ja niiden sana-asemat järjestettynä virheiden määrän perusteella.	140
Taulukko 91. Pituusopposition hallinta sanan eri asemissa.	142
Taulukko 92. Vastaukset segmentin tyyppin mukaan laskettuna.	143

## Kuviot

Kuvio 1. IA-fonologian mukainen tasokuvaus.	15
Kuvio 2. Klassisen generatiivisen fonologian mukainen tasokuvaus.	15
Kuvio 3. Kerroksellisen fonologian mukainen tasokuvaus.	16
Kuvio 4. Generatiivisen fonotaktiikan mukainen tasokuvaus.	17
Kuvio 5. Deklaratiivinen fonologia.	30
Kuvio 6. Ydintavutus oikealta vasemmalle.	36
Kuvio 7. Ydintavutus vasemmalta oikealle.	37
Kuvio 8. Yksisuuntainen malli.	38
Kuvio 9. Tavun laajuuden rajoitus.	39
Kuvio 10. "Flat structure".	40

Kuvio 11. Binaarinen jako 1.....	40
Kuvio 12. Binaarinen jako 2.....	41
Kuvio 13. Pitkän ja ylipitkän tavun merkinnän ekonomiaa.....	42
Kuvio 14. Kolmihaarainen malli.....	43
Kuvio 15. Prosodinen hierarkia.....	44
Kuvio 16. Ajoitusyksiköt autosegmentaaliossa teoriassa.....	44
Kuvio 17. Moraan perustuva tavumalli.....	45
Kuvio 18. Kolmimorainen malli.....	46
Kuvio 19. Erilaisia tapoja kuvata ylipitkä tavu.....	47
Kuvio 20. Ylipitkän tavun kuvaus haarautuvana morana.....	50
Kuvio 21. Nelimorainen tavu.....	51
Kuvio 22. Ylipitkät tavut.....	52
Kuvio 23. Tavunmuodostus morateoriassa.....	54
Kuvio 24. Kirjoitus ja autosegmentaalinen malli A.....	55
Kuvio 25. Kirjoitus ja autosegmentaalinen malli B.....	56
Kuvio 26. Kirjoitus ja autosegmentaalinen malli B.....	56
Kuvio 27. Mora ajoitusyksikkönä 1.....	57
Kuvio 28. Mora ajoitusyksikkönä 2.....	58
Kuvio 29. Painottoman ja sivupainollisen tavun pituusvaihtelut.....	72
Kuvio 30. Leksikaalinen optimaaliteoria.....	78
Kuvio 31. Sanatason ja postleksikaalisen tason painohierarkia.....	80
Kuvio 32. Kaksitavuisten sanojen rakenteen kuvaus mora-asemittain.....	107
Kuvio 33. Grafeemien suhteellinen yleisyys morittain merkittynä.....	108
Kuvio 34. Grafeemien suhteellinen yleisyys morittain merkittynä (NS:n kaikki sanat). .....	109
Kuvio 35. Pääpainolliset tavut NS:n yhdistämättömissä sanoissa (prosentteina)....	109
Kuvio 36. Sanatyyppien yleisyys tasoittain juoksevassa tekstissä (SK).....	113
Kuvio 37. Vokaalien ja konsonanttien kvantiteettityyppien jakauma niiden yleisyyden perusteella juoksevassa tekstissä (SK).....	114
Kuvio 38. Fonologinen kompleksisuus 1.....	117
Kuvio 39. Fonologinen kompleksisuus 2.....	118
Kuvio 40. Tunnuksmerkkisyysjatkumo juoksevasta tekstistä.....	119
Kuvio 41. Tunnuksmerkkisyysjatkumo <i>Nykysuomen sanakirjassa</i> .....	120
Kuvio 42. Pitkä ja lyhyt konsonantti kuvattuna morateoriaa käyttäen.....	146
Kuvio 43. Pitkä ja lyhyt konsonantti kuvattuna ilman morateoriaa.....	146
Kuvio 44. Grafeemin ja segmentin vastaavuus lyhyissä konsonanteissa.....	147
Kuvio 45. Grafeemin ja segmentin vastaavuus pitkissä konsonanteissa.....	147
Kuvio 46. Tavurakenne ja frekvenssi.....	148

# 1. Johdanto

## 1.1 Työn rakenne

Lisensiaatintutkielmani käsittelee kirjoituksen oppimista erityisesti pituuden merkitsemisen näkökulmasta. Työssäni kuvaan kirjoitusprosessia empiiriseen aineistoon soveltaen kielitieteellisen formalismin näkökulmasta. Samalla selvitän, mikä merkitys kielitieteellisellä näkökulmalla on kirjoitusprosessin monitieteisessä tutkimuksessa. Työn empiirinen aineisto koostuu 230 tois- ja kolmasluokkalaisen lapsen kirjoitusanelutehtävästä, joista tutkittiin kvantiteetin merkitsemisen hallintaa. Sanelutehtävä koostui 28 epäsanasta, jotka oli tehty systemaattisesti siten, että voitiin tutkia kvantiteetin merkintä eri sana-asemissa ja erilaisissa sanatyypeissä. Tarkoituksena oli, että ensimmäisen kerran saataisiin tarkka kuva suomen kielen pituusdistinktioiden merkinnästä eri asemissa. Varsinainen päämäärä oli selvittää, missä määrin kirjoitetun kielen oppimisessa vaikuttaa eri sanatyypin frekvenssi juoksevassa tekstissä (käyttöfrekvenssi) ja sanakirjafrekvenssi (tyyppifrekvenssi). Hypoteesina on, että pitkän ja lyhyen kvantiteetin merkinnän oppimisessa ainakin sanatyypin käyttöfrekvenssillä on merkitystä. Tämän hypoteesin mukaan kirjoittamaan oppija oppii nopeammin ja helpommin kirjoittamaan sellaiset sanatyypit, joita tulee useammin vastaan lukiessa ja kirjoittaessa. Toinen työni hypoteesi on, että kvantiteetin merkinnän vaikeuksissa on kyse siitä, että kirjoittamaan oppivilla lapsilla on tendenssi jättää kvantiteetti huomiotta. Tunnetusti sekä pitkää että lyhyttä kvantiteettia merkitään yleisesti yhdellä grafeemilla.

Aluksi käyn läpi työni keskeiset käsitteet, teoriataustaa ja aikaisempaa työhöni kiinteästi liittyvää tutkimusta. Työni rakentuu kahdeksasta luvusta. Johdannon jälkeinen luku käsittelee generatiivista fonologiaa ja kolmas luku deklarativista fonologiaa. Näissä pyrin osoittamaan, että kirjoittamisen perusprosessin tutkimus nojaa valtaosin kielitieteen näkökulmasta liian yksinkertaiseen käsitykseen fonologiasta ja morfologiasta. Kolmannessa luvussa käsittelen morfologian ja pituuden suhdetta. Neljännessä luvussa esittelen autosegmentaalista fonologiaa ja tarkastelen, miten sitä voitaisiin käyttää kirjoituksen perusprosessin kuvaamiseen. Viidennessä luvussa



käsittelen suomen kielen astevaihtelua osana pituuteen liittyviä ongelmia ja astevaihtelua morfologian ja fonologian ongelmana. Kuudennessa luvussa käsittelen konsonanttien ja vokaalien kvantiteetin yleisyyttä sekä juoksevassa tekstissä että *Nykysuomen sanakirjan* yhdistämättömissä sanoissa. Seitsemäs luku yhdistää kehrittelemäni kielitieteellisen formalismin kuvauksen työni empiiriseen aineistoon. Luvussa käsittelen kirjoitustestin tulokset ja tulkiten ne.

Empiirinen frekvenssiaineistoni koostuu kirjoitetusta tekstistä ja sanakirja-aineistosta. Kirjoitustestin aineisto on luettujen sanojen kirjoitusta, se koostuu siis grafeemeista. Fonologia, foneemi ja foneemin ja grafeemin välinen suhde ovat kuitenkin kompleksisia asioita, joten pyrin ensin määrittelemään fonologiaa ja tarkastelemaan sen autonomisuutta, riippumista tai riippumattomuutta kielen muista tasoista, ja itse fonologian monitasoisuutta. Pyrin myös pitämään grafeemin ja foneemin käsitteen erossa toisistaan. Tutkimukseni painopiste on pituuden fonologiassa, sen yhteyksissä morfologiaan ja pituuden fonologian monitasoisessa kuvauksessa.

## 1.2 Foneemi- ja grafeemitasojen suhteesta

Fonologia, fonologinen tieto ja foneemi ovat ongelmallisia käsitteitä niin kielitieteessä kuin käsitteet omaksuneessa lukemisen tutkimuksessa. Uppstad & Tonnessen (2005, 60) ovat koonneet erilaiset käsitykset fonologian ontologiasta ja kustakin käsityksestä seuraavista implikaatioista lukemisen ja kirjoittamisen tutkimukseen seuraavalla tavalla. Ensimmäisen käsityksen mukaan fonologia on lingvistinen ilmiö, kirjoitettu kieli ei ole. Tämä käsitys implikoi, että leksikkoon on pääsy ainoastaan puheen kautta ja että fonologinen rakenne on suoraan relevantti kirjoitettuun kieleen nähden, mutta ei toisin päin. Toisen käsityksen mukaan fonologia on luonnollinen, kirjoitettu kieli on artefakti. Tämä käsitys implikoi sitä, että puhe on sisäsyntyistä, kirjoitus on artefakti ja puhe kuuluu ihmisenä olemiseen, mutta kirjoitettu kieli on ulkoista. Kolmannen käsityksen mukaan fonologia tarkoittaa segmentaalista, lineaarista rakennetta. Käsitys implikoi että prosodia on perifeerista, eikä se ole fonologisen rakenteen osa ja äänteelliset yksiköt ovat lineaarisesti järjestyneitä. Neljännen mallin mukaan fonologia on platonista. Tämä käsitys implikoi että fonologia on abstraktia, se on kaukainen ilmiö verrattuna foneettiseen, akustiseen

ääntämykseen. Fonologiselle tasolle täytyy olla pääsy lukemaan opittaessa ja fonologiaan pääsee ainoastaan fonologisten yksiköiden tunnistamisen tietoisuuden kautta. Viidennen käsityksen mukaan fonologia on hyvin representationaalista. Käsitys implikoi että on olemassa spesifinen mentaalinen taso, mihin fonologinen rakenne voidaan totuudenmukaisesti sijoittaa psykologisen reaalisuuden ja kommunikatiivisen relevanssin vuoksi ja representaatio on kaukainen ja kvalitatiivisesti erilainen kuin puhe.

Oman tutkimukseni kannalta erityisesti Uppstadin & Tonnessenin (2005, 60) viiden kohdan listan kolmas käsitys ansaitsee lisämaininnan jo tässä vaiheessa. Kohdan mukaan fonologia tarkoittaa lineaarista, segmentaalista rakennetta, johon nähden prosodia on toisarvoista. Tällainen näkemys on vastakkainen tämän työn näkökulman kanssa, jonka mukaan prosodia on keskeinen ilmiö fonologiassa ja fonologia on monitasoinen ilmiö. Myös toinen kohta on liian rajaava. On olemassa yksi kieli, joka ilmenee sekä puhuttuna että kirjoitettuna kielenä. Ensimmäinen kohta on tutkimukseni kannalta melko relevantti: fonologia voidaan käsittää abstraktina rakenteena, joka on kirjoituksen kaltainen, mutta jolla ei ole vastinetta puheessa. Neljäs kohta tulee lähelle ensimmäistä ja korostaa ensimmäisen lailla fonologian abstraktisuutta. – Palaan näihin käsityksiin vielä työni lopussa.

Kuinka autonominen kielen taso fonologia sitten on? Luonnollisesti myös kielitieteessä on pohdittu fonologian luonnetta ja abstraktisuutta etenkin suhteessa morfofonologiaan, jota yllä kuvatussa luokituksessa ei käsitellä. Paul Kiparsky käsittelee klassisessa artikkelissaan "How Abstract Is Phonology?" (1968) abstraktin ja konkreettisen morfofonemiikan – artikkelissa fonologia samaistetaan morfofonemiikkaan generatiivisen käsityksen mukaan – eroja. Morfofoneemiset teoriat voidaan jakaa karkeasti kolmeen luokkaan sen mukaan, kuinka abstraktia morfofonemiikan katsotaan olevan. Konkreettisinta näkemystä edustaa IA-teoria ja abstraktisinta Lambin (1965) ja Fudgen (1967) abstrakti morfofonemiikka. Seuraavassa luonnehdin lyhyesti näitä ääripäitä ja niiden väliin sijoitettavaa kompromissia.

1) Morfofoneemit ovat kaikki niiden konkreettiset esiintymät (allomorfit) käsittäviä listoja: Tätä käsitystä voidaan nimittää konkreettiseksi morfofonemiikaksi. Tätä

kantaa edustivat esimerkiksi amerikkalainen 1940-luvun fonologia (IA-malli) ja Prahan koulukunnan pioneerit.

2) Edelliselle mallille vastakkaisen näkemyksen mukaan morfofoneeminen representaatio on luonnoltaan puhtaasti klassifikatorista (Lamb 1965; Householder 1965; Fudge 1967). Tämän käsityksen mukaan morfofonemiikka on täysin abstraktia. Siinä ei ole lainkaan foonisia elementtejä ja ne representoituvat täysin neutraaleina entiteetteinä.

3) Kolmas malli sijoittuu kahden edellisen väliin. Tässä mallissa pyritään yhdistämään piilevien muotojen täysin konkreettinen fonemiikka ja täysin abstrakti morfofoneemiikka. Kiparsky (1968, 1) kutsuu tätä prosessimorfonemiikaksi. Tähän malliin kuuluu lukuisia suuntauksia, mutta niille on yhteistä distinktiivisten piirteiden luettelo sekä käsitys, jonka mukaan pintamuodot on johdettu abstrakteista syvämuodoista (esim. Jakobson 1948; Chomsky & Halle 1968).

Suomessa tunnetusti pitkän ja lyhyen foneemin oppositio on paitsi puhtaasti fonologinen ilmiö (*matto - mato*) myös morfofonologinen ilmiö (astevaihtelu) ja jossain määrin myös vapaasti vaihteleva ilmiö etenkin uusissa lainasanoissa (*me(e)dia, in(t)ternetti*), jotka etsivät vielä suomalaista muotoaan ja joissa erilaiset rajoitukset kilpailevat. Edelleen fonologisesti lähinnä vapaaseen vaihteluun kuuluvat arki- ja virallisen kielen erot, kuten yleisgeminatio (*menee - mennee, tulee - tullee*), jotka jopa "peittävät alleen" morfologisia oppositioita. Näin voidaan siis hyvällä syyllä kysyä, kuinka abstraktia fonologia on, kuinka paljon se ottaa huomioon morfologiaa ja kuinka fonologista tietoa esimerkiksi pituudesta hyödynnetään kirjoitusprosessissa pituuden merkitsemisessä. Käsitystä fonologiasta ei voi siis erottaa morfologian ja morfofonemiikan määrittelystä ja kuvauksesta.

## 2. Generatiivista fonologiaa

### 2.1 IA- ja IP-fonologia

Tässä luvussa käsittelen generatiivista fonologiaa Chomskyn & Hallen klassisesta teoksesta *Sound Pattern of English* (1968) lähtien. Tarkoitukseni on pyrkiä osoittamaan, että fonologia ja varsinkin fonologian suhde kielen muihin osa-alueisiin on mutkikas. Kirjoituksen omaksumisen tutkimukseen tämä liittyy siten, että kirjoituksen ja lukemisen tutkimuksessa olisi tärkeää kyetä muodostamaan konkreettisia hypoteeseja siitä, miten kielen eri osajärjestelmät, kuten fonologia ja morfologia, toimivat yhdessä ihmisen prosessoidessa kieltä kirjoitettaessa tai luettaessa. Samalla pyrin osoittamaan, että IA-fonologian mukainen käsitys kielestä ei ole riittävä kirjoittamisen ja lukemisen tutkimuksessa.

Generatiivisen fonologian koulukunnan voi ajatella perustuvan Noam Chomskyn ja Morris Hallen ajatuksiin 1950-luvun lopussa. Generatiivisen fonologian lähtökohtana oli premissi, jonka mukaan kielen fonologinen rakenne heijastaa syntyperäisen kielenpuhujan kielellistä kompetenssia. Syntyperäinen kielenpuhuja pystyi Chomskyn & Hallen mukaan tuottamaan rajatusta määrästä kielellistä informaatiota rajattoman määrän lauseita. Tätä kielikykyä kuvasi kieliopin säännöstö, jonka avulla oli mahdollista spesifioida (generoida) rakennekuvaukset rajattomalle määrälle kielen kieliopillisia lauseita. Chomsky ja Halle (1968) pyrkivät tutkimaan suoraan kielellistä kompetenssia, joka mahdollistaa foneettisen signaalin (Saussuren parole) tuottamisen. Näin ollen generatiivista fonologiaa voidaan pitää osana 1900-luvun jälkimmäisellä puoliskolla virinnyttä kiinnostusta kognitiivisiin tieteisiin yleensäkin.

Generatiivisessa standarditeoriassa (Chomsky & Halle 1968) sääntö on yhdensuuntainen: se siis kuvaa fonologista operaatioita morfofonologisesta syvämuodosta kohti pintamuotoa. Säännön ulkopuolisia tapauksia koskee ns. ”Elsewhere Condition”, jonka mukaan sääntö koskee vain sitä ympäristöä, mitä se kuvaa.

$A \rightarrow B / \_C$  Elsewhere  $A = A$

Chomsky ja Halle vastustivat ja hylkäsivät työssään suuren määrän generatiivista fonologiaa edeltävistä fonologisista teorioista. Tulilinjalle joutuivat etenkin amerikkalaisen strukturalismin edustajat Edward Sapir (1884—1942) ja Leonard Bloomfield (1887—1949). Sapir'n "Item and Process" -mallissa (IP-malli) abstrakti fonologinen edustus oli käännettävissä foneettiseksi representaatioksi prosessien, kuten segmenttien poistojen, segmenttien lisäysten ja äänteenmuutosten avulla. Sapir itse oli painottanut kielellisten representaatioiden psykologista reaalisuutta, mutta ei osoittanut kiinnostusta niiden formalisointia kohtaan.

Bloomfieldiläinen koulukunta taas kehitti Sapir'n mallista poiketen "Item and Arrangement" -mallia (IA-malli), joka painotti kielellisten analyysien eksplisiittistä rakennetta. IA-fonologia perustui käsitykseen, jonka mukaan fonologisen tutkimuksen pääkohde on kyetä formalisoimaan kontrastin käsite (kuitenkin usein esiteoreettiselta pohjalta). Aspiraatio on kontrastiivinen esimerkiksi korean kielessä, mutta ei englannissa. IA-mallissa foneemi määriteltiin luokaksi, joka koostuu foneista (foneettinen segmentti). Foneemin allofonit ovat fooneja, jotka ovat komplementaarisisessa distribuutiossa tai vapaassa vaihtelussa toistensa kanssa. Jos äänne on esimerkiksi komplementaarisisessa distribuutiossa toisen kanssa, se ei ole itsenäinen foneemi, vaan jonkun foneemin allofoni. Allofoneja ei johdettu foneemeista fonologisten prosessien avulla, vaan foneemin ja allofonin käsitteet nojasivat pikemminkin "äänteellisen vastaavuuden käsitteeseen". Nykyään vallinneeksi yksimielinen käsitys, että amerikkalaisen strukturalismin mukainen foneemikäsitys ei ole riittävä kielen ilmiöiden kuvauksessa.

Kirjoituksen (perusprosessin) tutkimus propagoi kuitenkin nykyäänkin (ehkä tietämättään) IA-mallin puolesta. Kirjoitusprosessia ei nähdä prosessina, vaan se kuvataan äänneiden automaattisena fonologisena koodaamisena grafeemeiksi. Suomessa kirjoituksen oppiminen nähdään vielä yksiviivaisemmin kuin muissa kielissä, joissa foneemin ja grafeemin suhde ei ole yhtä yksinkertainen kuin suomessa. Koska suomen kielen oikeinkirjoitusjärjestelmässä äänteen ja sitä vastaavan kirjainmerkin suhde on useimmissa tapauksissa yksi yhteen, oletetaan myös kirjoitusprosessin olevan yksinkertaisesti "lineaarisen" foneemisen esityksen koodaamista

vastaaviksi kirjaimiksi. Tällöin fonemisaation kompleksinen luonne jää analyysissä kokonaan pois, ja kielellisen esityksen lineaarinen luonne ylikorostuu: kieli nähdään suoraviivaisena foneettisena esityksenä.

Halle (1962) ja erityisesti Chomsky (1964) kritisoivat IA-fonologiaa. Halle esimerkiksi huomasi, että venäjän kielessä soinnillisuusassimilaatio vaikutti sekä foneemeihin että allofoneihin. Chomsky taas vakuuttui siitä, että monet fonologian ongelmat voitiin kuvata yksinkertaisemmin ja siis yleisemmin, jos ajateltiin fonologisten sääntöjen voivan järjestyä siten, että ensin sovelletaan ensimmäistä sääntöä (morfofoneemiikasta foneemiin -säännöt) ja vasta tämän jälkeen toista (fonemiikasta fonetiikkaan -säännöt).

Chomsky ja Halle (1968) osoittivat, että esimerkiksi sanojen *writer* vs. *rider* fonemisaatio /rajtʰr/ vs. /rajdʰr/ rikkoo useimpia tunnettuja IA-fonemisaation periaatteita (invarianssi, lineaarisuus, biuniikkisuus), mutta että käyttämällä kahta peräkkäin sovellettua sääntöä oli mahdollista kuvata tämän kaltaiset tilanteet.

/rajt/	/rajt+ʰr/	/rajd/	/rajd+ʰr/	Fonologinen representaatio
ra:jd		ra:jdʰr		Pituussääntö
rajRʰr		ra:jRʰr		“flapping” sääntö
[rajt]	[rajRʰr]	[ra:jd]	[ra:jRʰr]	Foneettinen representaatio
'write'	'writer'	ride'	'rider'	

Tämän kaltaisista esimerkeistä Chomsky ja Halle (1968, 340-341) päättelivät, että ei ole olemassa yhtä foneemista tasoa, joka sijoittuisi Sapir'n fonologisen ja foneettisen tason väliin. Samalla he luopuivat mallissaan kontrastin (opposition) käsitteestä, joka olisi suoraan sovellettavissa teoriaan (kuten Sapir'lla). Kontrastin käsitteen tilalle Chomsky & Halle loivat teorian, joka pyrki kuvaamaan fonologisia ilmiöitä säännöillä. Teoria pyrki kuvaamaan sääntöjen, jotka muunsivat (morfo)fonologisen tason ilmiöt foneettisen tason ilmiöiksi, muodon ja substanssin. Chomsky ja Halle omaksuivat Roman Jakobsonin (1986—1982) binaaristen distinktiivisten piirteiden teorian. Fonologiset piirteet kuvattiin formaalisti matriiseina. Huomattakoon myös, että tässä ja seuraavissa säännöissä pituutta koskevat prosodiset (pidentyminen, difton-

gisoituminen) ja segmenttaaliset prosessit liittyvät tai sotkeutuvat — sen mukaan mistä näkökulmasta asiaa tarkastelee — yhdeksi kokonaisuudeksi.

Chomskyn & Hallen klassinen analyysi (1968, 178-190) englannin kielen vokaalivaihtelusta ja velaariden pehmenemisestä ("English Vowel Shift" ja "Velar Softening") kuvaa hyvin generativistien pyrkimystä kuvata kaikki fonologiset ilmiöt säännöillä.

[aj] vs. [i] (*divine, divin-ity*)

[ij] vs. [e] (*serene, seren-ity*),

[ej] vs. [æ] (*profane, profanity*)

Koska lyhyt vokaali ei vaihtelee sanoissa tyyppiä *rigid* → *rigidity* ja *final* → *finality*, Chomsky ja Halle päättelivät, että pitkän vokaalin täytyi olla piilevänä pohjamuotona sanoissa, joissa vokaalivaihtelua esiintyy. Vokaalin laatu olisi kuitenkin sama kuin lyhytvokaalisessa muodossa:

/i:/, /e:/, and /æ:/

Ensimmäiseksi vokaali lyhenee tiettyjen suffiksien edellä. Kaikki jäljelle jäävät pitkät vokaalit diftongisoituvat. Tämän jälkeen sovelletaan kahta sääntöä, jotka muokkaavat pitkien vokaalien ydintä, nukleusta. Ensimmäinen sääntö muuttaa ylävokaalin /i/ ja keskivokaalin /e/ [-korkea]-vokaaleiksi.

1) [korkea] → [-korkea]

Toinen sääntö muuttaa keski- ja alavokaalit [alow] [-alow] vokaaleiksi.

2) [a-matala] → [-a-matala]

/devi:n/	/sere:n/	/profæ:n/	
ij	ej	æj	diftongisaatio
ej	ij	-----	[ahigh] → [-ahigh]
æj	-----	ej	[alow] → [-alow]

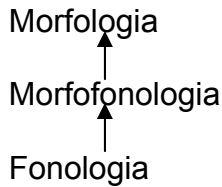
Chomskyn ja Hallen analyysieja moittittiin aikoinaan mm. liiasta abstraktiudesta ja siitä, että — jos teoria todella kuvaa puhujan kompetenssia — kielenpuhujalla ei voi olla tietoa yllä kuvatuista fonologisista prosesseista, jotka sisältävät kieli-historiallista tietoa, johon kielenpuhujalla ei ole pääsyä (esim. Kiparsky 1968). Standarditeoria ei myöskään kyennyt kuvaamaan tilanteita, joissa fonologista sääntöä ei sovelleta, vaan se ohitetaan (esim. Kissenbert 1970).

## 2.2 Fonologisesta ja morfologisesta komponentista

Keskeisiä kysymyksiä generatiivisessa kieliopissa kautta aikojen on ollut se, voivatko kieliopin eri osat toimia keskenään ja jos voivat, niin miten. Tähän kysymykseen on kaksi erilaista vastausta.

1) Klassisen neo-bloomfieldiläisen teorian mukaan morfologia ja fonologia ovat toisistaan riippumattomia (K. P. Mohanan 1995, 24). Tällaisessa teoriassa tehdään tiukka ero fonologisen ja morfologisen tason välille. Ilmiöt, joissa vaikuttavat sekä morfologiset että fonologiset säännöt, kuvataan osana morfofonologiaa. Morfofonologia taas kuvataan alueeksi, joka sijoittuu morfologian ja fonologian väliin. Morfofonologian käsite on kuitenkin uusbloomfieldiläisen teorian suurin kompastuskivi, koska morfofonologia jää täysin abstraktiksi allomorfien piirteiden, jotka eivät ole fonemaattisia eivätkä morfologisia, yhdistelyksi ja niiden kuvaukseksi ja vaille minkäänlaista empiiristä perustaa. Alla oleva kuvio 1 on ymmärrettävä niin, että löytämismenetelmään ("discovery analysis") perustuva fonologia "ruokkii" morfofonologiaa ja morfofonologinen taso "ruokkii" edelleen morfologista tasoa. Näin fonologisen tason läpäissyt representaatio joutuu morfofonologisten sääntöjen alaiseksi ja morfofonologinen representaatio morfologisten sääntöjen alle. Mallissa siis edetään äänneistä morfeemeihin kuten IA-fonologian esiteoreettisessa löytämisprosessissa.

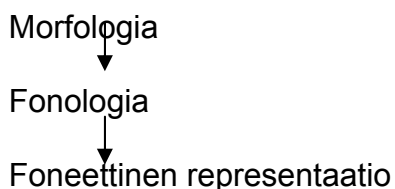




Kuvio 1. IA-fonologian mukainen tasokuvaus.

Edellä mainitun teoriaan voidaan olennaisena liittää Stampen (1972) ratkaisu, jonka mukaan on olemassa kahdenlaisia fonologisia sääntöjä (tai rajoituksia). Toiset säännöt vaikuttavat myös morfologisiin prosesseihin ja toiset vaikuttavat ainoastaan fonologisiin prosesseihin.

2) Klassisessa generatiivisessa fonologiassa ei hyväksytty jakoa morfofonologiseen ja fonologiseen tasoon, vaan oletettiin, että fonologia ja morfologia vaikuttivat toisiinsa. Tähän kuuluvat standarditeoria (Chomsky & Halle 1968) ja leksikaalinen fonologia (Pesetsky 1979, Kiparsky 1982b). Generatiivisessa fonologiassa suunta eri tasojen välillä on vastakkainen verrattuna klassiseen morfofonologiaan. Kun generatiivisessa kieliopissa keskeisellä sijalla on aina ollut tuotoksen generoiminen alla olevasta muodosta, on myös luontevaa, että generativistien piirissä ajateltiin morfologian edeltävän fonologista tasoa (Kuvio 2). Chomskylaisessa generatiivisessa kieliopissa syntaksi oli ”generoiva” komponentti ja semantiikka ja fonologia olivat tulkitsevia komponentteja, fonologinen komponentti tulkitsi foneettisesti (ns. foneettinen representaatio) syntaktisen pintarakenteen, johon oli sijoitettu leksikaaliset ja kieliopilliset morfeemit. Tätä ajattelutapaa edusti suomen kuvauksessa esimerkiksi Karlssonin *Suomen kielen äänne- ja muotorakenne* (1982).



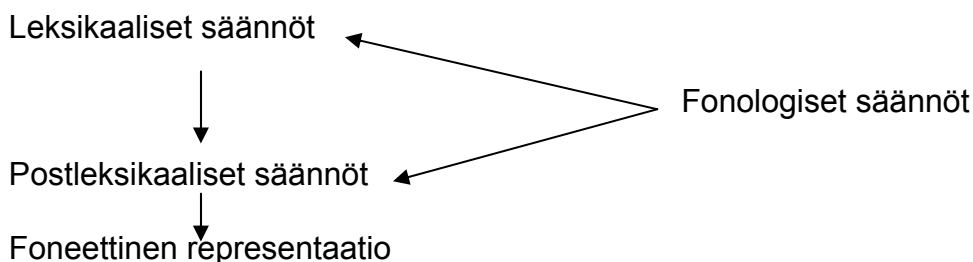
Kuvio 2. Klassisen generatiivisen fonologian mukainen tasokuvaus.

### 2.3 Kuinka morfologia ja fonologia vaikuttavat toisiinsa?

Jos hyväksytään ajatus, että morfologia ja fonologia vaikuttavat toisiinsa, on olemassa (vähintään) kaksi loogista vaihtoehtoa siitä, kuinka fonologinen ja morfologinen taso vaikuttavat toisiinsa.

1) Syklisen käsityksen mukaan osa fonologisista säännöistä tai rajoituksista on syklisiä. Tällä tarkoitetaan sitä, että osaa fonologisista säännöistä sovelletaan niin monta kertaa kuin mahdollista rakenteen eri tasoilla, kun taas osaa säännöistä sovellettiin syklin jälkeen. Sykliseen malliin perustuu valtaosa generatiivisen fonologian suuntauksista. Tiukasti syklisten mallien puutteena on, että syklisistä analyseista tulee helposti rakenteeltaan kompleksisia, kun yritetään keksiä, miksi sykliset sääntöjä sovelletaan tietyissä ympäristöissä. Toisaalta syklisissä malleissa on useita ”kielenulkoisia” mallintamiseen liittyviä vaikeuksia, jotka nekin liittyvät juuri syklisten sääntöjen soveltamiseen ja toisaalta soveltamatta jättämiseen.

2) Kerroksellisessa fonologiassa (tässä leksikaalinen fonologia) säännöt jakautuvat leksikaalisiin ja postleksikaalisiin sääntöihin (Kiparsky 1982c, 2003). Leksikaalisessa tasossa ovat leksikaalisten sääntöjen lisäksi morfologiset säännöt, joita sovelletaan ensin (ks. Kuvio 3 alla). Vasta tämän jälkeen voidaan soveltaa postleksikaalisia sääntöjä. Postleksikaaliselta tasolta ei voida palata enää leksikaalisiin sääntöihin, vaan on oletettava, että leksikaalisen ja postleksikaalisen tason välillä informaatio ei voi kulkea enää takaisin päin. Kerroksellisen fonologian suurimpia etuja on, että sillä voidaan kuvata sekä fonologisia että klassisen fonologian morfofonologiaa vastaavia ilmiöitä.

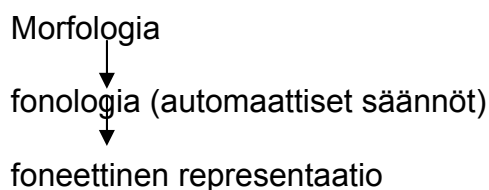


Kuvio 3. Kerroksellisen fonologian mukainen tasokuvaus.

Leksikaaliset säännöt kuuluvat leksikon sisään ja kuvaavat siis sanoihin kuuluvaa leksikaalista informaatiota. Nämä säännöt toimivat jokaisen affiksaatiokierron jälkeen ja ovat siten inherentisti syklisiä. Myös fonologiset säännöt ovat leksikaalisia, joten niitä sovelletaan jokaisessa syklissä. Postleksikaaliset säännöt eroavat edellisistä siinä, että niitä sovelletaan leksikaalisten ja fonologisten sääntöjen jälkeen.

## 2.4 Generatiivinen fonotaktiikka

Generatiivinen fonotaktiikka meni vielä klassista generatiivista fonologiaa kauemmaksi kieltäessään morfofonologisen tason käsitteen olemassaolon. Generatiivinen fonotaktiikka edusti äärimmäistä päätä generatiivisessa fonologiassa väittäessään, että mitkään fonologiset periaatteet tai säännöt eivät voi viitata morfologiseen rakenteeseen. Tämän ajattelun mukaan siis kaikki fonologiset prosessit, jotka ovat sensitiivisiä morfologiselle informaatiolle, ovat automaattisesti osa morfologiaa (K. P. Mohanan 1995, 38). Generatiivisessa fonotaktiikassa toisin sanoen kiellettiin morfofonologian olemassaolon. Jos tämä väite on oikea, niin kysymys siitä, miten fonologiset säännöt hyödyntävät informaatiota morfologisesta rakenteesta, ei ole adekvaatti.



Kuvio 4. Generatiivisen fonotaktiikan mukainen tasokuvaus.

## 2.5 Affiksaalinen morfologia

Affiksaalisen (Lieber 1980; Kiparsky 1983) morfologian mukaan ainoat sallitut morfologiset operaatiot ovat operaatioita, jotka yhdistävät affikseja ja vartaloita toisiinsa. Tämän hypoteesin uskottavuus riippuu kuitenkin siitä, miten affiksi määritellään. Tara Mohanan'in (1983) mukaan affiksi saadaan yksinkertaisimmin määritellyä, jos määritelmä jaetaan kolmeen tasoon (K. P. Mohananin 1995, 41 mukaan).

Piirteet: entiteetit, kuten [perfekti], [passiivi], [monikko]  
(universaaleja)

Morfeemit: abstrakteja kielikohtaisia morfologisia yksiköitä, kuten [ISTUA], [TAKKA]  
jne.

Formatiivit Fonologisen tason ilmiöt, jotka toimivat morfologisina yksiköinä. Esim. [-  
n], [-ssa], [-ttaa]

1970-luvun lopussa ja 1980-luvun alussa rajasymbolien ("Boundary Symbol") formaalinen käyttökelpoisuus kuitenkin kyseenalaistettiin. Useissa tutkimuksissa selvisi, että rajasymbolit olivat samalla kertaa liian vahvoja ja toisaalta liian heikkoja välineitä morfologisesti ohjautuvien fonologisten ilmiöiden formaaliin analyysiin (K. P. Mohanan 1995, 43). Rajasymboleille tilalle ehdotettiin kahta erilaista ratkaisua. Ensimmäinen ratkaisu oli korvata rajasymbolit puukuvaimella (esim. Rotenberg 1978; Aronoff ja Sridhar 1983). Toinen ratkaisu oli jakaa informaatio sekventiaalisen modulaarisuuden periaatteen mukaisesti (esim. Siegel 1974; Pesetsky 1979; Kiparsky 1982c; K. P. Mohanan 1982; Pulleyblank 1986).

Kiparskyn (1968-1973) mukaan fonologisia distinktioita neutralisoivia sääntöjä, joita sovelletaan täysin poikkeuksetta (automaattiset fonologiset säännöt), ei voida käyttää kaikkien erilaisten morfeemien esiintymisissä (Colen 1995, 73 mukaan). Tätä ehtoa kutsutaan generatiivisessa kirjallisuudessa nimellä "Alternation Condition". "Alternation Condition" (tässä AC) kieltää soveltamasta neutralisaatiosääntöjä kokonaan johtamattomissa ympäristöissä morfeemien sisällä (Cole 1995, 73). Kiparsky vastusti generatiivisen fonologian liikaa abstraktiutta ja AC:n liittykin tiukasti Kiparskyn propagoimaan leksikaaliseen fonologiaan. AC-analyysilla onkin selkeitä etuja kuvattaessa fonologisia ilmiöitä, jotka on helpommin kuvattavissa olettamalla niiden olevan leksikaalista alkuperää. AC oli (ja on ehkä vieläkin) tieteenteoreettisesti tehokas väline, koska sen avulla päästiin eroon liian abstrakteista pohjamuodoista. Esimerkiksi Chomsky ja Halle (1968, 51-55) analysoivat englannin kielen ilmiötä, jolle he antoivat nimen "Trisyllabic Shortening"

(kolmitavuisten sanojen lyhentyminen). Tämän analyysin mukaan englannin kolmetavuisissa sanoissa tavun pitkä painoton vokaali lyhenee tapauksissa *divinity*, *opacity*, *tabulate*, *derivative*. Tätä ilmiötä ei kuitenkaan ole johtamattomissa (monomorfeemisissa) sanoissa, kuten *Omaha*, *Ivory*, *nightingale*. SPE:n mukaan tämäntyyppiset sanat eivät ole enää kolmitavuisten sanojen lyhentymis-säännön alaisia. Kiparskyn (1982b) mukaan sanoissa tyyppiä *Pamela*, *camera*, *alibi*, *sycamore* ei ole mitään todisteita siitä, että ne olisivat SPE:n kuvaaman lain alaisia (Colen 1995, 74 mukaan). Pikemminkin on niin, että edellä mainitunlaiset sanat olivat kaksitulkitaisia, koska on yhdentekevää, merkitäänkö ne leksikkoon lyhyt vai pitkävokaalisina. Kiparskyn (1982b) mukaan onkin tärkeä määritellä, mitkä sanat täyttävät kolmitavuisten sanojen lyhentyminen -säännön ehdot ja mitkä eivät. Kiparskyn mukaan SPE:n analyysia oli muutettava konkreettisempaan suuntaan olettamalla, että osa tämänkaltaisista fonologisista distinktioista on kuvattava leksikaalisiksi. Näin päästäisiin lähelle konkreettista tietoa kielestä. AC:n mukaan pohjalla olevan muodon täytyy olla identtinen pintamuodon kanssa. Tällöin SPE:n säännöillä kuvattavat distinktiot ovatkin kuvattavissa leksikaalisiksi.

## 2.6 Leksikaalinen fonologia ja sykli

Leksikaalinen fonologia (LF) perustu(i)~(u) hypoteesiin, jonka mukaan osa fonologisista säännöistä toimii syklisesti. LF:ssä keskeiset periaatteet koskevat sitä, miten fonologian ja morfologian välinen vuorovaikutus toimii. Seuraavassa Kiparskyn leksikaalisen fonologian ydinkysymykset Mascarón (1976) muotoilemana (Colen 1995, 78 mukaan):

- 1) Miksi täytyisi olettaa, että on olemassa kahden tyyppisiä fonologisia sääntöjä, syklisesti toimivia ja ei-syklisiä?
- 2) Miksi syklisillä analyyseilla on niin kompleksinen muoto? Miksi niin pitäisi yleensä olla?
- 3) Millainen inherentti yhteys on syklisyyden ja sääntöjen järjestämisen välillä? Miten määritellä johdettu ympäristö?

Kiparskyn mukaan on olemassa syklisiä sääntöjä, jotka eivät ole neutraalistavia ja joita sovelletaan myös ei-johdetuissa (monomorfeemisissa) ympäristöissä. Tällaiset tapaukset ovat SCC:n periaatteiden vastaisia. On todisteita (englannin sanapaino, espanjan tavunmuodostus ja aspiraatio ja englannin velaarien pehmenneys), joiden mukaan 1) kaikki sykliset säännöt eivät ole neutralisoivia, 2) on olemassa pakollisia neutralisoivia sääntöjä, joita ei sovelleta joka sykllillä, 3) kaikki pakolliset neutralisoivat säännöt eivät ole SCC:n alaisia. Koska empiirinen evidenssi ei tukenut edellä kuvattunlaista syklisyyden käsitettä (esim. Mascaro 1976), on syklin käsite jouduttu muotoilemaan leksikaalisessa fonologiassa uudelleen.

Leksikaalisessa fonologiassa fonologiset säännöt seuraavat sananmuodostus- ja yhdyssanasääntöjä ja muita morfologisia sääntöjä (Cole 1995, 80). Leksikaalisia sääntöjä sovelletaan jokaisen syklin lopussa morfologisten sääntöjen jälkeen. Leksikaalisessa fonologiassa on syklisiä sääntöjä, jotka seuraavat sananmuodostus- ja yhdyssanasääntöjä eli juuri leksikaalisia sääntöjä. Kiparskyn (1982a, 46) mukaan jokainen leksikaalinen tapaus (entry) konstituoii fonologisen säännön (ns. identity mapping)  $\alpha \rightarrow \alpha$  jokaiselle leksikaaliselle tapaukselle (= leksikaalinen "entry") / $\alpha$ /. Tämä identtisyysääntö kilpailee kaikkien muiden sääntöjen kanssa kieliopin leksikaalisessa komponentissa (Cole 1995, 81). Esimerkiksi englannissa on identtisyysääntö, joka kilpailee Trisyllabic Shortening -säännön kanssa. Molempien sääntöjen syötös on sama, mutta vain jompaakumpaa sääntöä voidaan soveltaa (jos niiden tuotokset ovat erilaiset). Jos molemmat säännöt kuuluvat leksikaaliseen komponenttiin, ne molemmat ovat ns. "Elsewhere Condition" -prinsiipin alaisia.

"Elsewhere Condition" -periaatteen mukaan (kahden) säännön, joiden rakennekuvaus on päällekkäinen, mutta joiden tuotokset ovat erilaiset, on järjestäytyttävä disjunktiiivisesti. Kahdesta säännöstä spesifimpää sovelletaan ensin ja yleisempi sääntö seuraa erityisempää sääntöä.

"Elsewhere Condition" -prinsiipin mukaan leksikaalinen identtisyysääntö tulee aina sovelletuksi ennen leksikaalista fonologista sääntöä, koska se sisältää leksikaalista informaatiota (= spesifimpi). Leksikaalisessa fonologiassa osaa säännöistä ei koske "Elsewhere Condition" -prinsiippiä, koska nämä säännöt ovat leksikaalisen informaati-

tion pääsemättömissä (ne eivät toisin sanoen täytä leksikaalisia sovellusehtoja). Näitä sääntöjä kutsutaan postleksikaalisiksi säännöiksi. Koska postleksikaaliset säännöt eivät laukaise leksikaalisia sääntöjä (= eivät ole "Elsewhere"-prinsiipin alaisia), voivat postleksikaaliset säännöt vaikuttaa myös johtamattomissa ympäristöissä. Leksikaaliseen fonologiaan siis sisältyy oletus, että osaa säännöistä voidaan "a priori" pitää ei-syklisinä, koska niiden fonologisten sovellusehtojen kuvaus on monitasoista ja tasolta  $n+1$  ei voida palata tasolle  $n$ .

Kiparskyn (1982c, 47) mukaan rakennesääntöjä (sanapainosäännöt) ja tavusääntöjä ei koskaan ohiteta "Elsewhere Conditionin" perusteella, koska nämä säännöt eivät koskaan riko "Elsewhere Condition" -prinsiippiä vastaan (sovelletaan vain silloin, kun kahden säännön syötökset ovat identtisiä, mutta joiden tuotokset ovat erilaisia). Myöhemmin Kiparsky (1985, 91) oletti, että vain leksikaaliset säännöt, jotka tuottavat distinktiiviset tuotokset, rakentavat fonologisesti johdetun tason (= "derived string").

Kiparskyn (esim. 1982b) leksikaalinen fonologia sisältää leksikaalisen teorian kaikkien vahvimmat ja rajoittavimmat formuloinnit. Niistä seuraa kuitenkin vaikeuksia, ja Cole (1995, 83) luettelee seuraavat seikat, jotka ovat osoittautuneet ongelmalliseksi em-piirisen aineiston valossa.

- 1) SCC johdetaan "Elsewhere Condition" -prinsiipistä.
- 2) Kaikki leksikaaliset säännöt ovat syklisiä.
- 3) The "derived environment constraint" -prinsiippiä sovelletaan kaikkiin ja vain sykliisiin sääntöihin.
- 4) Syklisyys johtuu fonologian ja morfologian yhteisvaikutuksesta ("Cyclicity derives from the interleaving of phonology and morphology").

Mohananin & Mohananin mukaan (1984) SCC-prinsiippiä ei voida johtaa "Elsewhere Condition" -prinsiipistä (Cole 1995, 85 mukaan). Heidän mallissaan (analyysi malayalam-nimisestä kielestä) yksittäinen sääntö voi vaikuttaa sekä leksikaalisella että postleksikaalisella tasolla. Tässä mallissa ei ole eroteltu erikseen erillisiä sääntöjä leksikaalisille ja postleksikaalisille komponenteille. Mohanan & Mohanan ei myöskään anna eksplisiittistä selitystä sille, miksi ainoastaan leksikaaliset säännöt

ovat "Elsewhere Condition" –prinsiipin alaisia, ja voivat tämän prinsiipin vaikutuksesta järjestyä disjunkttiivisesti verrattuna postleksikaalisiin sääntöihin. Jos "Elsewhere"-prinsiippi vaikuttaisi yhtäläisesti sekä leksikaalisiin että postleksikaalisiin sääntöihin, seuraisi siitä se, että myös post-leksikaaliset säännöt olisivat johdettuja sanoja koskevien sääntöjen alaisia ("derived environment constraint") alaisia. Tällä olisi tietenkin haitallisia vaikutuksia, sillä jos postleksikaalista sääntöä tällöin sovellettaisiin johdettuihin (tai johtamattomiin) sanoihin, ohitettaisiin postleksikaaliset säännöt "Elsewhere Condition" –prinsiipin nojalla aivan kuin leksikaaliset säännöt.

Syötös	/hatanpää/
assimilaatiosääntö	hatanpää *ei voida soveltaa
tuotos	hat <u>a</u> mpää

Kiparsky itse hylkäsi (Kiparsky 1985) hypoteesin, jonka mukaan SCC on johdettavissa (derivable) oleva sääntö, ja päätyi sijoittamaan SCC universaalikielioppiin itsenäiseksi komponentiksi (Cole 1995, 84). Tämä höllensi entistä, liian tiukkaa formalisointia, mutta toisaalta se oli ilmiselvä heikennys verrattuna aikaisempaa formalisointiin, koska automaattisesti universaalikielioppiin sijoitettuna SCC operatiivinen status jäi aika olemattomaksi. (ks. myös Iverson & Wheeler 1988).

Suomen kielessä on selvä todiste sitä vastaan, että vain syklistet, leksikaaliset säännöt ovat johdettuja sanoja koskevan rajoituksen ("Derived Environment Constraint") alaisia (Kiparsky 1968, 1973). Kyseessä on sääntö, jota generatiivisessa fonologiassa kutsutaan spirantisaatiosäännöksi (= "Spirantization rule"). Tämän säännön mukaan /t/ vaihtuu /s/:ksi silloin, kun sananloppuinen e → i. Kiparskyn (1982a, 165-173) mukaan pohjamuotona on muoto /vete/. Tällöin muutos e → i ruokkii spirantisaatiosääntöä (sääntöä e → i on sovellettava ennen spiranttisääntöä). Spirantisaatiosääntöä ei voida soveltaa ennen e → i –muutosta, koska syötös /vete/ ei ole spirantisaatiosäännön soveltamiseen vaadittava ympäristö. Säännön t → s /\_i on kuitenkin oltava sääntö, jota voidaan soveltaa ainoastaan johdetuissa muodoissa. Tästä todisteena on se, että spirantisaatio ei toteudu johtamattomissa sanoissa, kuten *tila*, *äiti*. Leksikaalisen fonologian kriteereillä spirantisaatio on kuitenkin syklinen, leksikaalinen sääntö. Sääntö e → i taas ei voi olla postleksikaalinen sääntö,



koska sitä sovelletaan monomorfeemiseen sanaan ja sitä ei edellä mikään sääntö. Jos sitä sovellettaisiin syklisten sääntönä, se tuottaisi pohjamuodosta /vete/ muodon \*vesinä oikean muodon *vetenä* sijaan. Tätä kuvausta on kuitenkin laajasti kritisoitu. Todellisuudessa tätä sääntöä on pidettävä vahvasti leksikaalistuneena (kuten Karlsson 1982, 344).

Syötös:	/vete/
e → i	veti
t → s/ __i	vesi

## 2.7 Leksikaalisten piirteiden redundanssittomuus ja tunnusmerkkisyys

Fonologisten piirteiden spesifioimattomuus ("underspesification") perustuu ajatukseen, että fonologisten piirteiden ei tarvitse aina olla erikseen spesifioituja. Generatiivisen fonologian parissa spesifioimattomuus perustuu kahteen oletukseen, jotka ovat alkujaan Chomskylta ja Hallelta (1968):

1. Leksikaalinen minimaalisuus: Pohjamuotojen täytyy sisältää minimimäärä informaatiota, joka tarvitaan sanojen erottelemiseksi leksikossa. Tästä on hyvänä esimerkkinä Kalevi Wiikin (1975) esitys suomen kielen vokaalisoinnusta (ks. myös Kiparsky 1982a, 143-159).

2. Täysi spesifikaatio (= "Full Specification"): Fonologisen komponentin tuotoksen täytyy sisältää täydellisesti (tai vähintään maksimaalisesti) spesifioitu piirrematriisi.

Chomskyn ja Hallen mukaan fonologisen komponentin tehtävänä oli lisätä ei-distinktiiviset fonologiset piirteet tuotokseen, joka siis sisältää leksikossa ainoastaan minimimäärän tarvittavia distinktiivisiä piirteitä. Leksikaalisen minimaalisuuden periaatteen vuoksi leksikkomuodot voivat kantaa vain minimimäärän leksikaalista informaatiota. Täyden spesifikaation periaate taas määrää, että tuotoksen on oltava fonologisilta piirteiltään täysin spesifioitu. Näin ollen leksikon täytyy sisältää ns. redundanssisääntöjä, jotka lisäävät puuttuvat fonologiset piirteet tuotokseen. Tyypillisiä

redundanssisääntöjä ovat siis seuraavan kaltaiset säännöt, joiden mukaan vokaali sisältää piirteet [+kontinuantti] ja [+sonoranttinen] segmentit ovat myös soinnillisia.

V → [+Continuant]

[+son] → [+voice]

Jo siis fonologisessa teoriassa hyväksytään Chomskyn ja Hallen (1968) kaksi ydinperiaatetta (yllä), tästä seuraa automaattisesti se, että joudutaan käyttämään redundanssisääntöjen kaltaisia formalisointeja spesifioimattomuuden kuvauksessa. Toisin sanoen oletetaan, että tuotoksessa fonologisten piirteiden täytyy olla täysin spesifioituja.

*Kontekstista riippuvalla tunnusmerkkisyydellä* tarkoitetaan kielikohtaista tunnusmerkkisyyttä, jossa eri kielten välillä on eroavaisuuksia. Kontekstista riippuva tunnusmerkkisyys on myös sidottua tiettyihin sana-asemiin tai tietyn tyyppisen segmentin läheisyyteen. Esimerkiksi suomessa (ja monessa muussa kielessä) sanansisäinen nasaali on aina samapaikkainen kuin sitä seuraava konsonantti (\*nk, \*np, \*mt).

*Kontekstista riippumaton tunnusmerkkisyys* on universaalia. Se perustuu oletukseen (Steriade 1995,119), jonka mukaan kaikilla fonologisilla piirteillä on sekä tunnusmerkkiset että tunnusmerkittömät arvot, jotka ovat osa universaalia kielioppia (= UG). Tunnusmerkitön arvo viittaa normaaliin, neutraaliin kyseessä olevan artikulaatioelimen tilaan. Esimerkiksi piirteen [nasaalinen] tunnusmerkitön arvo on [-nasaalinen]. Näin on päätelty olevan siksi, että segmentit ovat tavallisesti oraalisia ja oraaliset segmentit ovat nasaalisia segmenttejä vapaammin esiintyviä. Niinpä kontekstiltään vapaa neutralisointisäännön voi kirjoittaa alla olevaan muotoon (ks. Steriade 1995, 119).

[ ] → [-Nasaali]

Yllä mainitun kaltaisten redundanssisääntöjen mukaan piirteiden spesifiointi toimisi johdoksellisesti. Ei ole myöskään selvää, että kaikki fonologiset piirteet olisivat yksinkertaisesti binaarisia.

*Asemasta riippuvainen neutralisaatio* on kolmas kielen syntagmaattista ennustettavuutta lisäävä piirre. Asemasta riippuva neutralisaatio tarkoittaa ilmiötä, jossa tietyt fonologiset piirteet neutralisoituvat tietyssä sana-asemassa. Asemasta riippuvainen neutralisaatio ei koske jotain yksittäistä tunnusmerkkistä piirrettä, vaan kyseessä on tietyssä sana-asemassa vaikuttavat prosodiset seikat. Asemasta riippuvan neutralisaation voi kuvata seuraavasti (Goldsmith 1990, 123; Steriade 1995, 120-121).

- 1) Asemasta riippuva neutralisaatio (suodatinversio) ("filter version"): \* $\alpha$ F asemassa X, missä X on määritelty prosodisesti tai morfologisesti.
- 2) Asemasta riippuva neutralisaatio (lisenssiversio) ("licensing version"): Piirteen  $\alpha$ F täytyy olla lisensoitu asemassa X, missä X on määritelty prosodisesti tai morfologisesti.

Suodatinversion mukaan asemasta riippuva neutralisaatio on ilmiö, jonka mukaan piirre  $\alpha$ F ei saa esiintyä tietyssä morfologisessa tai prosodisessa ympäristössä. Tämän teorian mukaan asemasta riippuva neutralisaatio on ymmärrettävä säännöksi tai rajoitukseksi, joka rajoittaa tietyn piirteen esiintymistä tietyssä asemassa. Jos käytetään formulointia kaksi, voi piirre  $\alpha$ F esiintyä vain ympäristössä, missä sen olemassaolo on "lisensoitu" so. sen ympäristö sallii piirteen  $\alpha$ F olla tietyssä asemassa. Lisenssiteorian mukaan esimerkiksi suomessa piirre [+labiaalinen] ei voi esiintyä koodassa, ellei tätä piirrettä ole lisensoitu seuraavan tavun alkukonsonanttiin. Niinpä suomessa ei esiinny (omaperäisissä) sanoissa konsonanttiyhtymiä kuten \**mt*, \**pk*, \**pt*, mutta kahden labiaalinen yhtymät *mp*, *pp*, *mm* ovat sallittuja.

Teoria nimeltä *radikaali spesifioimattomuus* on saanut alkunsa Paul Kiparskyn (1982b, 1985) leksikaalisessa fonologiassa. Radikaalissa spesifioimattomuudessa keskeisiä ovat lokaalisuuden ja invarianssi käsitteet. Kiparsky oli ensimmäinen, joka huomasi että lokaalisuuden vastaiset rikkomukset voivat olla seurausta fonologisista prosesseista, joissa segmentit, jotka on spesifioitu jonkun tietyn piirteen kannalta, voidaan ohittaa johtoprosesseissa. Kiparskyn (1985) pääargumentti oli, että redun-

danssisääntöjen sovelletaan vain tietyillä fonologisilla tasoilla. Nondistinktiiviset, ennustettavat piirteet lisättiin tuotokseen aina vasta postleksikaalisella tasolla (Steriaden 1995, 124 mukaan).

Kiparsky siis oletti samoin kuin Chomsky ja Halle (1968), että tunnusmerkkiteoria koostuu universaalista luettelosta, joka sisältää kaikki tunnusmerkkiset piirteet. Tunnusmerkkisyyttä vastaan voidaan kuitenkin rikkoa, kielissä voi olla segmenttejä, jotka universaalista näkökulmasta ovat vahvasti tunnusmerkkisiä. Esimerkiksi piirre \*[+sonorantti, -soinnillinen] on vahvasti tunnusmerkkinen piirre, mutta on olemassa kieliä, joissa tällaisia segmenttejä kuitenkin esiintyy. Kiparsky oletti, että on olemassa rakenteensuojeluprinsiippi ("Structure Preservation") niminen periaate, joka kieltää leksikaaliset säännöt, jotka tuottavat sellaisia rakenteita, jotka eivät ole kielessä sallittuja pohjamuotoja (Steriaden 1995, 125 mukaan).

## 2.8 Mikä on fonologinen läpinäkyvyys?

SPE:n aikaisessa fonologiassa keskeinen idea oli serialismi: sääntöjä sovellettiin sarjassa siten, että ensi sovellettiin sääntöä 1, jonka tuotos toimi syötöksenä säännölle 2. Klassiseksi muodostunut opaakin fonologisen prosessin muotoilu on Kiparskylta (1973, 79): Fonologinen prosessi P muotoa

$A \rightarrow B/C\_D$  on opaakki, jos pintamuodossa on jompikumpi seuraavista.

a) A ympäristössä  $C\_D$ .

b) B muussa ympäristössä kuin  $C\_D$  silloin kun B on johdettu prosessilla P.

McCarthy (1999, 2) termeillä ilmaistuna kohta a) ei ole pintamuodoltaan tosi ("non-surface-true") ja kohdan b) tapaukset eivät ilmene pintamuodossa ("non-surface-apparent"). McCarthy muotoilemana Kiparskyn (1973) fonologisen opaakkisuuden määritelmä on seuraavanlainen (tekijän suomentamana):

1) Lingvistiksi merkittävät yleistykset eivät useinkaan ole pintamuodoltaan tosia. Tällä tarkoitetaan sitä, että yleistys Y:llä on tärkeä rooli kielessä K, mutta kielessä K on pintamuotoja, jotka rikkovat yleistystä Y vastaan. Serialismi selittää tämän siten,

että yleistys Y on voimissaan ainoastaan yhdellä derivaation tasolla. Myöhemmät derivaation tasot peittävät tämän ilmiön Y, ja voivat olla jopa sen vastaisia.

2) Kielellisesti merkittävät yleistyksset eivät ole useinkaan pintamuodossa ilmeneviä ("not-surface-apparent"). Yleistys Y heijastuu pintamuotoon P, mutta ehdot jotka aiheuttivat Y:n soveltamisen, eivät ole nähtävissä pintamuodossa P. Serialismin mukainen selitys on, että yleistyksen Y vaatimat ehdot ovat relevantteja ainoastaan tasolla missä Y on voimissaan. Myöhemmät tasot voivat peittää yleistyksen Y tarvitsemat ehdot tai myöhemmillä tasoilla voivat vaikuttaa ehdot, jotka ovat jopa vastakkaiset kuin Y:n ehdot.

Esimerkiksi suomessa suffiksaalisessa astevaihtelussa on omistusliitteen edellä heikko aste, vaikka "pintamuodossa" astevaihtelun ehdot eivät toteudukaan (umpitavu → heikko aste). Generatiivisen fonologian puitteissa tämä voidaan analysoida seuraavasti. Omistusliitteen edellä on heikko aste, koska derivaatiossa tasolla 2 tavu on umpinainen. Tasolla 3 astevaihtelusääntö ei enää ole voimassa ja heikko aste jää voimaan.

Taso 1:	peruna+i	a → o/___i (imperf., mon.)
	/perunoi/	
Taso 2:	perunoi+den	
	/perunoiden/	
Taso 3:	perunoiden+si	(n → ∅)
	/perunoidesi/	

### 3. Deklaratiivinen fonologia

#### 3.1 Deklaratiivisen fonologian perusoletukset

Deklaratiivinen fonologia (= DF) on paremminkin tutkimussuuntaus, kuin yksi yhtenäinen teoreettinen suuntaus. DF ei perustu transformaatioihin eikä siinä generoida pintamuotoja pohjamuodoista käsin. DF on siis pintamuoto-orientoitunut malli, jossa kielellinen representaatio ja kielenkuvaus erotetaan tiukasti toisistaan. DF:ssä rajoitukset eivät perustu kielellisiin universaaleihin, vaan ne muodostavat kielellisen ilmiön osittaisen kuvauksen. Samoin kuin uudemmissa generatiivisissa teorioissa (OT), pintamuotoa koskevat rajoitukset ovat normi, ja pohjamuotoja koskevat rajoitukset ovat poikkeuksia (Scobbie, Coleman & Bird 1996, 685). DF:ssä rajoitukset eivät kilpaile keskenään, vaan jokaisen rajoituksen on toteuduttava. Deklaratiivinen fonologia on siis pintamuoto-orientoitunut malli, jossa ”pintamuodon” representaatio käsitetään essentiaalisesti, eli foneettisesti.

Deklaratiivisessa fonologiassa keskeistä on pyrkimys mahdollisimman yksinkertaiseen pintamuoto-orientoituneeseen kuvaukseen. Niinpä generatiivisesta fonologiasta tutut ilmiöt, kuten esimerkiksi assimilaatiosäännöt ja leksikaaliset säännöt, säännöt, jotka estävät toisia sääntöjä toteutumasta, eivät ole hyväksytyjä. Näihin sääntöjen interaktio-ongelmiin on siis otettu äärimmäinen kanta (Scobbie, Coleman & Bird 1996, 685): Kaikki fonologiset säännöt ovat rajoituksia, joten ainoastaan rajoitusten interaktio joudutaan kuvaamaan. Rajoitukset eivät toimi yhdessä erilaisten elementtien, kuten leksikon ja sääntöryhmien, kanssa. Lisäksi kaikki rajoitukset ovat tasavertaisia, niitä sovelletaan samalla tavalla ja kaikki rajoitukset täytetään.

Otan esimerkin. Perinteisesti syvämuodon ja pintamuodon suhde on fonologisesti kuvattu siten, että seuraavat kaksi ehtoa täytyy:

- 1) Sekä pohja- että pintamuodot ovat tarpeellisia.
- 2) Pinta- ja pohjamuoto ovat erilaisia.

Deklaratiivisessa fonologiassa pintamuotojen ja ”alla olevien” piilevien muotojen suhde on mahdollisimman yksinkertainen ja kuvauksessa piilevät muodot ovat pintarakenteen osittaisia kuvauksia (Scobbie, Coleman & Bird 1996, 697). Tämä voitaisiin ilmaista siten, että piilevät muodot koostuvat distinktiivisistä pintamuodon representaatioista. Paremmin DF:aan sopiikin kuvaus, jonka mukaan tällaiset rakenteet ovat yksinkertaisesti vähemmän rajoitettuja (enemmän disjunktivisia) pintamuodon kuvauksia.

Deklaratiivisessa fonologiassa universaalikielioppi koostuu vain niistä rajoituksista, jotka ovat tosia kaikissa kielissä. Universaaleja rajoituksia vastaan ei voida rikkoa (kuten OT:ssa) vaan ne kuvataan implikaatioina tai disjunktioina (Scobbie, Coleman & Bird 1996, 700). Universaalit ovat siis parametrisoituja.

### 3.2 Leksikko deklaratiivisessa fonologiassa

Deklaratiivisessa fonologiassa leksikko täytyy jakaa useisiin arbitraarisiin (ala)osioihin, jotka eroavat toisistaan sen mukaan, miten tiettyjä fonologisia sääntöjä sovelletaan kuhunkin osioon. Nämä leksikon osat sisältävät ”ylimääräistä” informaatiota leksikon osista vain siinä mielessä, että jonkin rajoituksen strukturaalinen kuvaus voi sisältää informaatiota siitä leksikon osasta, mihin rajoitusta sovelletaan. Esimerkiksi voidaan rakentaa kuvaus leksikon osasta, missä ei ole soinnillisuuteen perustuvaa kontrastia obstruenttien välillä. Tähän sisältyy automaattisesti ajatus, että jos muita kuvauksia ei esiinny, on muissa osissa leksikkoa voimassa oppositio soinnillisten ja soinnittomien obstruenttien välillä.

Af     obstruentti(f)  $\cap$  leksikon osa a(f)  $\rightarrow$  soinniton(f)

Leksikon osaan a kuuluvat morfeemit ovat rajoitetumpia ja spesifimmin kuvattuja kuin morfeemit, jotka eivät kuulu tähän osaan leksikkoa ja jotka eivät sisällä tätä rajoitusta. Deklaratiiviseen hypoteesiin kuuluu ajatus, että leksikon sisäinen jakautuminen alaryhmiin on arbitraarista ja kategoriat vaihtelevat yksilöstä toiseen ainakin kuhunkin kategoriaan kuuluvien sanojen määrän, kompleksisuuden, opittavuuden ja sanaston laajuuden suhteen (Scobbie, Coleman & Bird 1996, 697).

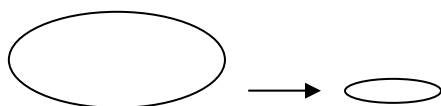
Epäproduktiivinen, leksikaalistunut vaihtelu voidaan oppia vasta aikuisiällä ja tämä vaihtelu talletetaan leksikkoon ilman pintamuotoon vaikuttavia seuraamuksia.

Koska deklaratiiivinen fonologia on yksitasoinen malli, ei leksikkoon voida olettaa muotoja, jotka eivät ole sallittuja pintamuodossa. DF:ssa tiukka rakenteensäilytys ("Structure Preserving") täytyy aina olla voimassa toisin kuin leksikaalisessa fonologiassa (esim. Kiparsky 1982b), jossa rakennetta suojelevat rajoitukset tai säännöt eivät toimi enää pintamuodossa, ja näin postleksikaaliset säännöt voivat toimia esteet-tömästi ja peittää leksikaaliset prosessit alleen.

### 3.3 Rajoitukset deklaratiiivisessa fonologiassa

Deklaratiivisen fonologian teoreettinen pohja perustuu Johnsonin (1988) ja Kaplanin & Bresnanin (1982) distinktiolle lingvististen objektien ja näiden formaalin teoreettisen kuvauksen välillä (Scobbie, Coleman & Bird 1996, 686).

Kieli                      Formaali kuvaus



Kuvio 5. Deklaratiivinen fonologia.

Formaaleissa kuvauksissa käytetään generatiivisesta fonologiasta tuttuja sääntöjä tai sääntöjen kuvausten representaatioita, kuten puudiagrammeja (Scobbie, Coleman & Bird 1996, 685).

Koska DF:ssa kielellinen kuvaus perustuu kielellisen objektin formaaliin kuvaukseen, käytetään siinä formaalista logiikasta tunnettuja konnektoreita, jotka on kehitetty formaaliin kuvaukseen. Tätä kuvaustapaa käytetään kaikissa kuvauksissa fonologisista säännöistä leksikkoon. Kuvauksista koostuu kielen osittainen kuvaus. Itsenäiset osittaiset kuvaukset voivat viitata samaan objektiin, ja näin ollen osittaisista kuvauksista voi koostua tarkempi kuvaus. Tämä tarkoittaa, että deklaratiiivisessa



fonologiassa ei tehdä eroa "Structure-building" ja "Structure-checking" -sääntöjen välille. Kielen osittainen kuvaus sisältää seuraavat osatekijät (Scobbie, Coleman & Bird 1996, 688):

- a) kaikki universaalit rajoitukset.
- b) kaikki kielikohtaiset rajoitukset.
- c) kaikki morfeemikohtaiset rajoitukset, jotka viittaavat kyseessä oleviin morfeemeihin.

Deklaratiivisessa fonologiassa leksikaalinen aines muodostaa osan kielen osittaisesta kuvauksesta. Morfeemit ovat rajoituksia, koska kaikki representaatiot, jotka sisältävät leksikaalista ainesta, lisensoidaan osaksi kyseisen leksikaalisen osan vaikutuksessa. Koska DF on pikemminkin yleisempi teoreettinen rakennelma kuin yksittäinen teoria, siihen ei sisälly vaatimuksia a priori kuten että lisensoiminen tai fonologinen hallitseminen ("Government") ovat fonologisia primitiivejä. Generatiivinen fonologia SPE:stä lähtien operoi symboleilla ja symbolien välisillä operaatioilla. Sama ilmiö on uusimmassakin generatiivisessa teoriassa (OT). Rajoitukset ja leksikaalinen aines nähdään ontologisesti erilaisina. Morfeemeja sovelletaan representaation osana.

*Vajaamäärittely* ("Underspecification") on kuvattavissa yksinkertaisesti DF:n puitteissa. Koska kaikki kielenkuvaus on osittaista, on leksikaalisen informaation ja fonologisen säännön yhteispeli (ainakin formaaliselta kannalta) helposti selitettävissä. Vajaamäärittelyssä on kyse siitä, että köyhempään tai puutteelliseen informaatioon lisätään leksikaalista informaatiota (Scobbie, Coleman & Bird 1996, 690). Esimerkiksi kohta a on kuvaus, jonka mukaan korkea takavokaali voi olla joko pyöreä tai ei-pyöreä leksikossa. Kohta b kuvaa rajoitusta, jossa tietyssä ympäristössä vokaali on pyöreä.

- a) Leksikko: E...  $\text{back}(x) \cap \text{high}(x)$
- b) Sääntö: A...  $\text{back}(x) \cap \text{high}(x) \dots \rightarrow \text{round}(x)$

Kuvaukset a ja b muodostavat kuvauksen, jonka mukaan vokaali on takainen, pyöreä ja korkea. Pelkkä kuvaus a väittää, että kyseessä voi olla joko pyöreä tai ei-pyöreä vokaali. Kuvaus b ei taas voi olla vokaalin, joka on korkea, takainen, ei-pyöreä osittainen kuvaus. Kuvauksen a mukaan vokaali on vajaamääritetty. Deklaratiivisessa fonologiassa vajaamäärittely on siis eksklusiivinen disjunktio, joka on osa universaalia kielioppia (Scobbie, Coleman & Bird 1996, 691).

Rajoitusten yhteistyö –ongelma ("the interaction problem") on yksinkertaistettu deklaraatiivisessa fonologiassa, koska ainoastaan rajoitusten interaktio otetaan huomioon. Näin on siksi, että myös leksikaalinen informaatio käsitetään rajoituksiksi. Tämä on DF:n kiistaton etu muihin rajoituksiin perustuviin malleihin nähden, koska DF:ssä leksikaalinen informaatio ja rajoitukset ovat ontologisesti samanlaisia (siis toisin kuin OT:ssa). Kaikkia rajoituksia täytyy soveltaa (niiden täytyy olla tyydyttyjä). Näin ollen rajoitusten järjestys ei voi vaikuttaa kieliopillisuuteen.

Leksikaalisen informaation täytyy edustua täydellisesti. Niinpä kyseessä on absoluuttinen uskollisuus leksikon ja "pintamuodon" välillä. Produktiiviset fonologiset vaihtelut analysoidaan käyttämällä osittaisia leksikaalisia kuvauksia. Destruktiivisia operaatioita tai transformaatiota ei sallita (esim. piirteitä muuttavat säännöt ja segmenttejä poistavat säännöt ovat kiellettyjä) (Scobbie, Coleman & Bird 1996, 696). Esimerkiksi segmenttien poistot kuvataan segmentin vaihteluna nollan kanssa ja epen-teesit nollan vaihteluna segmentin kanssa. Jos siis on olemassa esim. morfeemi x, joka vaihtelee nollan kanssa, on morfeemin x leksikaalisessa muodossa disjunktio  $x \cup \emptyset$ . Tämä on kuvattavissa sulkumerkeillä. Jos siis on olemassa lekseemi  $A(x)B$ , jossa A ja B ovat pakolliset elementit ja sulkumerkeillä merkitty (x) on vaihtoehtoinen elementti, joka voi puuttua. Lekseemi  $A(x)B$  voidaan ilmoittaa myös disjunktiona seuraavalla tavalla (1). Tässä kuvauksessa ilmoitetaan, että tällä rajoituksella on kaksi vaihtoehtoista reaalistumaa, AB ja AxB. Edelleen disjunktioita 1 voidaan tarkentaa toisella tarkemmalla rajoituksella (2), joka kieltää x:n esiintymisen jossain ympäristössä.

- 1)  $AB \cup AxB$
- 2)  $xB'$

Epenteesit kuvataan siis samaan tapaan kuin segmentin poistot (= vaihtelut nollan kanssa). Rajoitus 1 kuvaa vaihtelevan morfeemin ympäristön (A ja B). Rajoitus B on varsinainen epenteesisääntö. Jos siis on olemassa spesifimpi sääntö (B'), joka saa epenteesin ehdot täytetyksi, lisätään elementti x.

- 1) AB                                      (A<B)  
 2) AB' → Ax B'                        (A<<x<<B)

Deklaratiiviseen hypoteesiin kuuluu piirre, että rajoitusten soveltamista ei välttämättä aloiteta leksikaalisesta informaatiosta käsin. Yksittäiset rajoitukset liitetään yhteen Boolean konektoreilla (Scobbie, Coleman & Bird 1996, 692).

Syötöksiin perustuvat suhteet ovat DF:ssa automaattisia, koska automaattiset kuvaukset yhdistetään toisiinsa ja näistä syntyvät konjunktiot ovat transitiivisia ja symmetrisiä (Scobbie, Coleman & Bird 1996, 693). Niinpä a ja b ovat ekvivalentteja. Sekä a että b ovat transitiivisessa (tai syötössuhteessa) c:hen, jonka mukaan c on tosi.

- a) (velaari → frikatiivi)  $\cap$  (frikatiivi → soinniton)  
 b) (frikatiivi → soinniton)  $\cap$  (velaari → soinniton)  
 c) velaari → soinniton

Vastasyötös ("counterfeeding") on deklaratiivisessa fonologiassa mahdoton, koska rajoitusten muodostamat konjunktiot ovat symmetrisiä ja transitiivisia. Generatiivisessa fonologiassa vastasyötös-ongelmat ovat ikuisuuskyseminen, joihin ei ole löytynyt ratkaisua rajoitusperusteiseen fonologiaan siirtymisen myötäkään. Yksinkertainen vastasyötösongelma löytyy suomen kielestä. Astevaihtelussa geminaatat vaihtelevat yksinäisklusiilin kanssa ja yksinäisklusiilit heikkenevät edelleen tai niiden tilalle tulee nolla.

*tt : t : d*

*kk : k : Ø*

*pp : p : v*

Chomskyn ja Hallen (1968) mukainen kuvaus olisi siis seuraavanlainen. Samanlainen kuvaus on tietenkin mahdollinen kaikissa vastaavanlaisissa kuvauksissa, joissa ei sallita sääntöjen samanaikaista soveltamista. Näiden sääntöjen soveltaminen järjestyksessä  $b < a$  estää sen, ettei geminaatta /tt/ heikkene suoraan /d/:ksi (*katto < katon > kadon*).

a)  $tt \rightarrow t/A\_B$

b)  $t \rightarrow d/A\_B$

Deklaratiivisessa fonologiassa segmentin leksikaalinen representaatio sisältää sen vaihtoehdot. Koska geminaatta /tt/ vaihtelee pintamuodossa yksinäisklusiilin /t/ kanssa, täytyy osittaisen kuvauksen sisältää molemmat vaihtoehdot. Scobbien, Colemanin & Birdin (1996, 694) mukaan kyseessä ei kuitenkaan ole heikko, suppletiivinen konjunktio, koska kyseessä on vajaamääritys: /tt/ : /t/ vaihtelu koskee fonologista pituutta, kun taas /t/ : /d/ vaihtelussa on kyse myös ääntöpaikan vaihtumisesta (ks. Suomi 1980).

a)  $tt \cup t$

b)  $t \cup d$

Deklaratiivinen fonologia soveltaa "Elsewhere Condition" (EC) periaatetta sääntöjen sisäisen järjestäytymisen kuvauksessa. Generatiivisessa fonologiassa EC:n määritelmä on seuraava.

"Elsewhere Condition": Kahdesta säännöstä, joilla on erilainen tuotos, sovelletaan ensin sääntöä, joka on spesifimpi.

DF:ssa "Elsewhere"-prinsipiillä tarkoitetaan konjunktioita, jossa kuvaus rakentuu yleisemmästä säännöstä ja spesifisemmän säännön strukturaalisesta negaatiosta (Scobbie, Coleman & Bird 1996, 692). Logiikan termein tämä rakenne on siis konjunktio, jonka osat ovat:

1) yleisempi sääntö (so. "Elsewhere"-sääntö)

2) spesifimmän säännön negaatio

EC toimii deklarativisessa fonologiassa seuraavalla tavalla. Ensimmäisen säännön (sääntö a) mukaan takavokaali on pyöreä. Säännön b mukaan takainen alavokaali ei ole pyöreä. Säännön a' takainen ei-pyöreä vokaali on pyöreä. Säännöt a ja b ovat siis "Elsewhere"-suhteessa toisiinsa, koska niiden tulokset ovat erilaiset. Näistä a on varsinainen "Elsewhere"-tapaus, koska b on a:ta spesifimpi. Näin ollen a:sta voidaan kirjoittaa "eksplisiittinen" yleinen sääntö a'.

a) takainen  $\rightarrow$  pyöreä

b) takainen  $\cap$  ala  $\rightarrow$  pyöreä

a') takainen  $\cap$  ala  $\rightarrow$  pyöreä

Edelleen a') voidaan derivoida seuraavasti.

takainen  $\cap$  (takainen  $\cap$  ala)  $\rightarrow$  pyöreä

takainen  $\cap$  (takainen  $\cap$  ala)  $\rightarrow$  pyöreä

takainen  $\cap$  ala  $\rightarrow$  pyöreä

Muotoilulla a on teoreettinen etu a':han nähden, koska a eksplikoi takaisuuden ja pyöreäyden yhteyttä. Vaikka b on spesifimpi kuin a, voidaan yleistykset a vielä tehdä. Sääntöjen a ja b soveltaminen ei siis riko deklarativista hypoteesia vastaan.

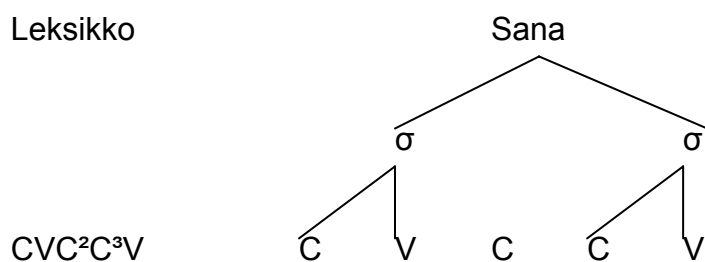
## 4. Autosegmentaalista fonologiaa

### 4.1 Tavunmuodostus formaalista näkökulmasta

Seuraavassa luvussa käsittelemme tavua ja erilaisia tavunmuodostusta koskevia teorioita. Tavun käsitteen keskeisyydestä lukemaan ja kirjoittamaan opittaessa todistaa se, että lukemaan ja kirjoittamaan oppivat lapset ovat tutkimusten mukaan paremmin tietoisia tavun käsitteestä kuin foneemisista yksiköistä (esim. H. Lyytinen et al. 2004b; U. Richardson et al. 2004). Tavunmuodostusta kielessä on mahdollista kuvata kahdella formaalilla mallilla (Blevins 1995, 223). Ensimmäinen malli koostuu kahdesta syklistä, joista ensimmäinen kulkee oikealta vasemmalle ja toinen vasemmalta oikealle. Ensimmäisessä syklistä tavun vokaalit ja vokaalien vasemalla puolella olevat konsonantit liitetään tavutasoon. Tätä sykliä kutsutaan ydintavutukseksi. Toinen sykli kulkee vasemmalta oikealle ja tässä syklistä oikealla puolella olevat konsonantit (=koodakonsonantit) liitetään tavutasoon.

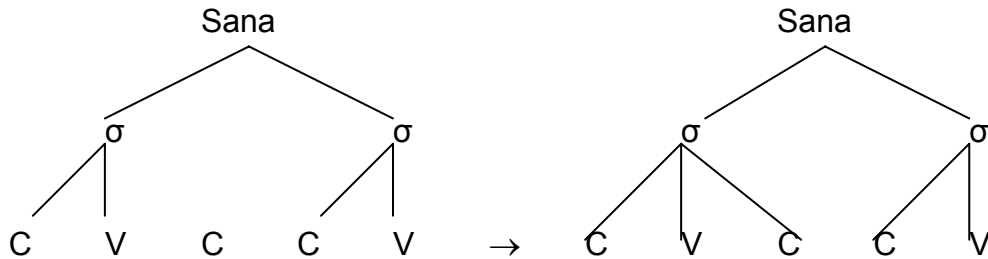
#### 1) Ydintavutus (Core syllabification)

Oikealta vasemmalle; luo (C)V-sekvenssit



Kuvio 6. Ydintavutus oikealta vasemmalle.

## 2. Toinen sykli vasemmalta oikealle:



Kuvio 7. Ydintavutus vasemmalta oikealle.

Edellisen kaksiosaisen mallin etuja ovat ainakin seuraavat:

Kuvaa hyvin kielen systeemiä sikäli, että suomessa tavun alussa voi olla vain yksi konsonantti. Onset-aseman ulkopuolelle jäävät segmentit liitetään automaattisesti koodaan.

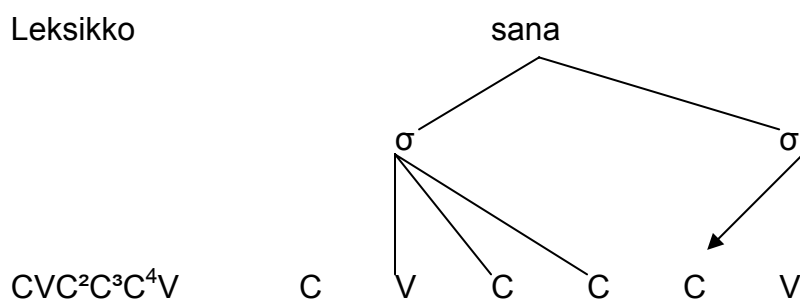
Haittoja: Onko oikeutettua formalisoida edellisen kaltaisia kaksisuuntaisia operaatioita? Sallitaanhan esimerkiksi suomessa lainasanoissa sananalkuisia konsonanttiyhtymiä. Kaksisuuntainen malli ei ole ainakaan psykologisessa mielessä validi: eihän voida olettaa, että leksikosta haettu sana tavutetaan mielessä kaksisuunnaisesti lineaariseen tapaan.

Formaalina mallina tavutuksen kaksisuuntainen kuvaus kuitenkin toimii kaikissa ympäristöissä. Ensimmäinen sykli on oikealta vasemmalle ja toinen sykli vasemmalta oikealle. Kohdassa d on sana tyyppiä *kiuas*, jossa ei ole toisen tavun tavunalkuista konsonanttia. Tämän tyyppisten sanojen takia tarvitaan lisäksi sääntöä, jonka mukaan diftongit kuuluvat samaan tavuun.

Leksikko		1. sykli	2. sykli
a) CVCV	>	[CV][CV]	-----
b) CVCCV	>	[CV]C[CV]	[[CV]C][CV]
c) CVVCV	>	[CV][V][CV]	[[CV]V][CV]
d) CVVVC	>	[CV]VVC	[CV]V[VC]

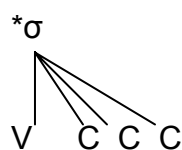
Toinen malli, millä voidaan kuvata tavutusta, on yksisuuntainen malli, jossa tavutus etenee pelkästään vasemmalta oikealle. Tässä mallissa siis konsonantit liitetään tavutasoon koodassa olevista konsonanteista lähtien. Koodan ulkopuolelle jäävät konsonantit muodostavat seuraavan tavun tavunalkuisen konsonantin. Jos tavunmuodostus on tämän mallin mukainen operaatio, on oletettava, että tavunmuodostuksessa on mukana sääntöjä, jotka määrittävät sekä koodan maksimaalisen rakenteen että tavunalkuisen konsonantin aseman: tavunalkuinen konsonanttihan on pakollinen, jos edellisessä tavussa on kooda.

Lisäksi tämä malli implikoi, että on olemassa erillinen "koodan vahvistamistendessi" eli (painollisen) tavun kooda rakennetaan ensin. Tällöin lienee oletettava, että on olemassa sääntö(jä), jotka määrittävät maksimaalisen tavun laajuuden; ilman näitä tavutus ei näin onnistuisi.



Kuvio 8. Yksisuuntainen malli.

Sääntö:



(Ei tavuja, joissa on kolme nukleuksen jälkeistä konsonanttia)



Kuvio 9. Tavun laajuuden rajoitus.

Nukleus:	CC[V]CCC[V]
Kooda:	CC[V][CC]C[V]
Onset:	[CC][[V][CC]][C][V]

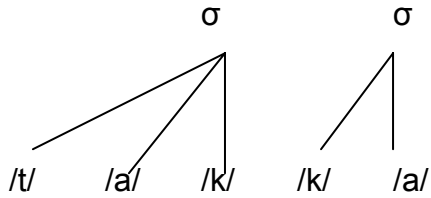
Oletan kuitenkin tässä työssä Niemeä & Bergiä (2000, 211) seuraten, että suomessa ei ole olemassa erillistä prosessia tai prosesseja, joiden tarkoitus on vahvistaa koodaa. Niemen & Bergin mukaan kooda luodaan pikemminkin segmenteistä, jotka jäävät tavunalkuisen konsonantin ulkopuolelle. Koska suomessa ei sallita sanan sisällä kuin yksi tavunalkuinen konsonantti, voidaan tavun muodostusta kuvata formaalista näkökulmasta yksinkertaisimmin kahtena operaationa, joista ensimmäinen kulkee oikealta vasemmalle ja toinen vasemmalta oikealle.

Kirjoitusprosessia kuvattaessa on otettava huomioon, että puhetta ja kirjoitusta ei ainakaan hypoteesin tasolla kuvata grafeemeiksi muutetuksi puheeksi. Lähtökohtana pidän oletusta, jonka mukaan puhe on prosodinen ilmiö, kun taas kirjoitusta on ainakin hypoteettisesti pidettävä lineaarisesti järjestettynä grafeemijonona. Näin ollen kirjoittamaan oppiva joutuu sisällyttämään oppimisprosessiin myös segmentaalista tasoa laajempaa informaatiota.

#### 4.2 Erilaisia tavun hierarkkisen kuvauksen malleja

Tavunkuvauksen formaaleissa kuvauksissa mallit eroavat toisistaan siinä, miten ne suhtautuvat tavutasoa alempiin tasoihin. Malleja on neljä kappaletta. Yksinkertaisin tavumalli on ns. "flat scture" –malli, jonka mukaan tavutason alapuolella ei ole mitään tasoa (Blevins 1995, 212). Tässä mallissa fonologinen kuvaus rakentuu siis sana- ja tavutasoista ja näiden alapuolella olevasta fonologisesta tasosta.

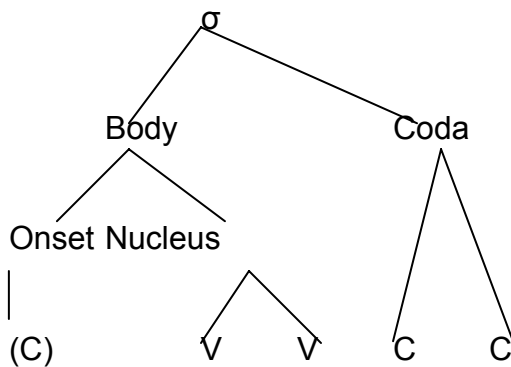
1) Tavussa ei ole tavutasoa alempia rakenteita (flat structure)



Kuvio 10. "Flat structure".

Malleilla 2 ja 3 on yhteistä se, että molemmissa tavutasoa alemmat yksiköt ovat binaarisesti haarautuvia. Tätä on pidettävä ainakin formalismin näkökulmasta etuna, koska tämä helpottaa tavun kuvausta (ks. alla kuvat 11 ja 12).

2) Binaarinen jako: Body:  $\sigma \rightarrow \text{Body Coda}$ ; Body  $\rightarrow \text{Onset Nucleus}$  (McCarthy 1979)



Kuvio 11. Binaarinen jako 1.

Lyhyt tavu: haarautumaton tavu

Pitkä tavu: 1) haarautuva tavu

2) haarautuva nukleus

Ylipitkä tavu: 1) haarautuva kooda

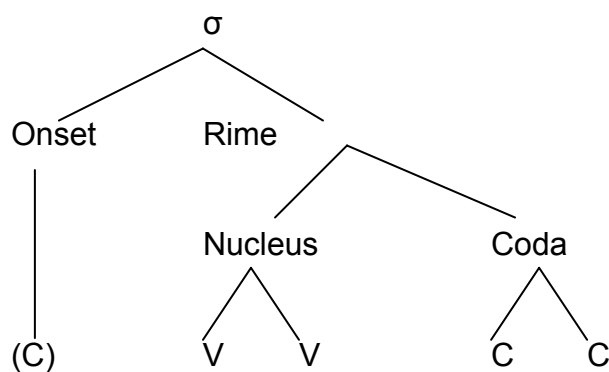
2) haarautuva nukleus + kooda

Tässä mallissa Bodyn haarautuminen ei kerro mitään tavun pituudesta. Sikäli tämä malli sisältää redundanssia. Malli myös implikoi sitä, että tavunalkuisen konsonantin ja nukleuksen relaatio on kiinteämpi kuin nukleuksen ja koodan keskinäinen suhde. Koska kooda on sidottu suoraan tavutasoon, tekee malli selkeän eroa pitkän ja ylipitkän tavun välille, koska pitkässä tavussa ei ole koskaan haarautuvaa koodaa.

Taulukko 1. Tavun pituuden kuvaus Body-mallissa.

	Haarautuva tavu	Haarautuva body	Haarautuva kooda	Haarautuva nukleus
Lyhyt tavu	- !	+ -		
Pitkä tavu	+ -	+ -	- !	+ -
Ylipitkä tavu	+	+ -	+ -	+ -

3) Binaarinen jako:  $\sigma \rightarrow$  Onset Rime; Rime  $\rightarrow$  Nucleus Coda (riimimalli)



Kuvio 12. Binaarinen jako 2.

Lyhyt tavu: haarautumaton riimi

Pitkä tavu: 1) haarautuva riimi

2) haarautuva nukleus

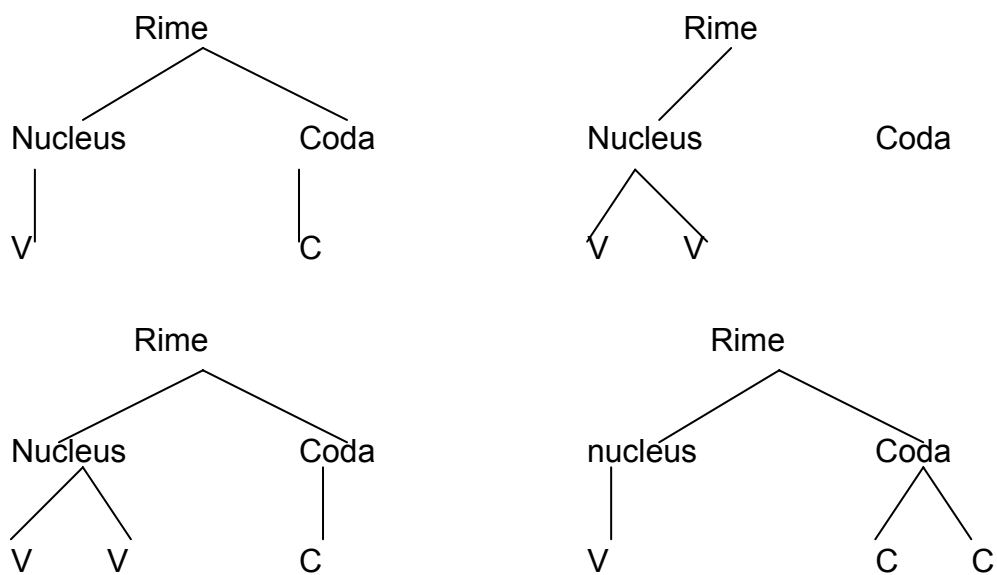
Ylipitkä tavu: 1) haarautuva nukleus + kooda

2) haarautuva kooda

Tämä malli on ehkä käytetyin tavun osien kuvauksen malli. Pitkän ja ylipitkän tavun ehdottomasti erottava piirre on se, että pitkässä tavussa ei voi olla haarautuvaa koodaa.

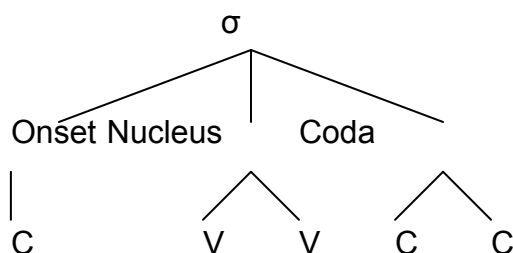
Taulukko 2. Tavun pituuden kuvaus riimimallissa.

	Haarautuva tavu	Haarautuva riimi	Haarautuva nukleus	Haarautuva kooda
Lyhyt tavu	+ -	- !		
Pitkä tavu	+ -	+ -	+ -	- !
ylipitkä tavu	+ -	+	+ -	+ -



Kuvio 13. Pitkän ja ylipitkän tavun merkinnän ekonomiaa.

#### 4) Kolmihaarainen malli: $\sigma \rightarrow$ Onset Nucleus Coda



Kuvio 14. Kolmihaarainen malli.

Lyhyt tavu: haarautumaton nukleus

Pitkä tavu: 1) haarautuva nukleus

2) kooda

Ylipitkä tavu: 1) haarautuva nukleus + kooda

2) haarautuva kooda

Taulukko 3. Tavun pituuden kuvaus kolmihaaraisessa mallissa.

	Haarautuva tavu	Haarautuva nukleus	Haarautuva kooda
Lyhyt tavu	+ -	- !	
Pitkä tavu	+ -	+ -	- !
ylipitkä tavu	+ -	+ -	+ -

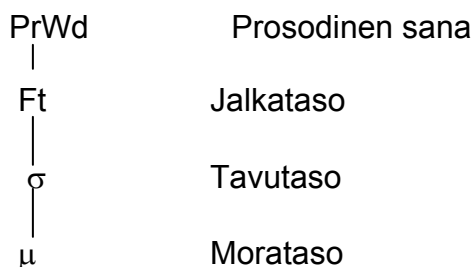
Edellä esitetystä ilmenee selvästi, että tavun hierarkinen rakenne on avoin erilaisille tulkinoille, eikä ole löydettävissä yhtä mallia, joka olisi ylivoimainen muihin malleihin vertailtaessa.

### 4.3 Morateoria ja sen soveltaminen geminaattoihin

Olen edellä käsitellyt fonologian autonomisuutta, fonologian suhdetta leksikkoon ja morfologiaa. Esimerkkinä olen käyttänyt klusiilien astevaihtelua, joka koskee tavujen ja tavunvälisten klusiilien pituutta. Tiivistäen kieliopissa pituus näyttää kytkeytyvän sekä fonologiseen että morfologiseen tietoon, ja voi olettaa, että näin on myös pituuden prosessoinnissa niin lukemisessa kuin kirjoittamisessa. Siirryn nyt tarkastelemaan tarkemmin fonologian monitasoisuutta ja segmentaalisen fonologian asemaa ja pituuden kuvausta monitasoisessa fonologiassa. Autosegmentaalinen, epälineaarinen fonologia koostuu useista päällekkäisistä tasoista. Pitkä ja lyhyt segmentti tulkitaan identtisiksi segmentaalisella tasolla, mutta ne eroavat toisistaan metrisellä

tasolla. Autosegmentaalisen fonologian lähtökohtana on morateoria, missä pitkä ja lyhyt segmentti ovat yksimoraisia piilevällä tasolla (underlying), mutta pitkä konsonantti edustuu kahtena lyhyenä morana metrisellä tasolla. Lyhyt konsonantti on piilevä nollamorainen, mutta voi olla kielikohtaisesti yksimorainen. Prosodiset kategoriat ovat hierarkkisessa suhteessa toisiinsa. Prosodisessa hierarkiassa alempien konstituenttien täytyy olla sidottuja ylempään tasoon (so. morat tavutasoon, tavutaso jalkatasoon ja jalkataso prosodiseen sanaan), mutta ylempien konstituenttien ei tarvitse kuulua alempiin tasoihin.

Prosodinen hierarkia Kagerin (1999, 146) mukaan:



Kuvio 15. Prosodinen hierarkia.

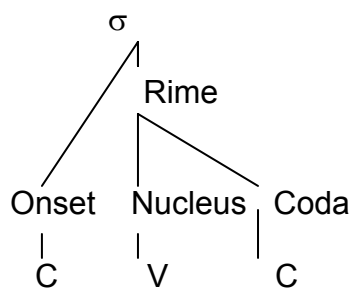


Kuvio 16. Ajoitusyksiköt autosegmentaalisisessa teoriassa.

Geminaatan yhdeksi pitkäksi segmentiksi tulkitsemisen puolesta puhuu se, että jos geminaatta tulkitaan koostuvaksi kahdesta lyhyestä segmentistä, on se rikkomus OCP:tä (ei rinnakkaisia identtisiä segmenttejä) vastaan (Harrikari 2000, 3). Monosegmenttien tulkinnan puolesta todistaa myös se, että svaavokaali ei voi hajottaa pitkää konsonanttia eikä samapaikkaisia konsonanttiyhtymiä (Harrikari 2000, 54). Näin ollen ne siis koostuisivat yhdestä jakamattomasta segmentistä.

Yksi eniten käytetyistä tavun kuvauksen malleista on Piken & Piken (1947, Hayesin (1989, 253) mukaan) malli. Mallissa tavutaso jakaantuu kahtia tavunalkuiseen

konsonanttiin (onset) ja riimiin (eng. rimeen). Rime jakaantuu edelleen kahteen osaan: nukleukseen ja koodaan (ks. esimerkki alla). Näiden alapuolella on skeletaallinen taso, joka taas liittyy sen alapuolella olevaan segmentaaliseen tasoon. Tavun-alkuinen konsonantti kuuluu aina onsetsiin ja sitä seuraava vokaali tai vokaalit nukleukseen. Suomessa nukleus koostuu joko yhdestä lyhyestä vokaalista tai kahdesta lyhyestä vokaalista tai yhdestä pitkästä vokaalista. Nukleuksen jälkeiset konsonantit luetaan koodaan kuuluviksi. Näin ollen koodaan kuuluvat kaikki segmentit, jotka ovat vähemmän sonorisia kuin vokaalit.

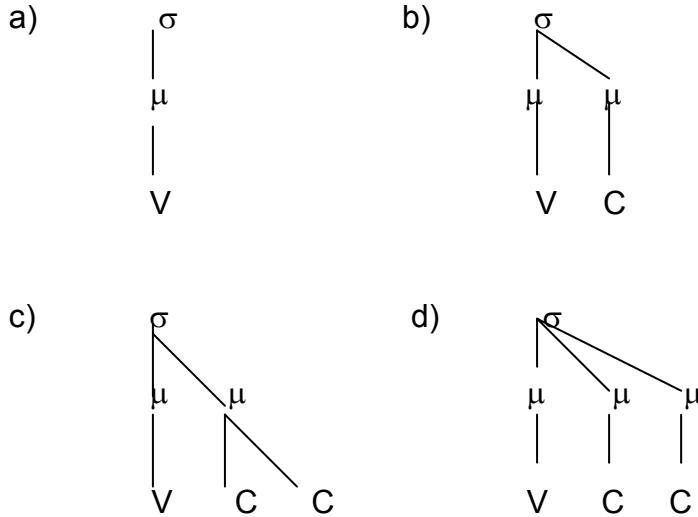


Kuvio 17. Moraan perustuva tavumalli.

Moraan perustuva tavumalli on sen sijaan Piken & Piken mallia yksinkertaisempi, koska se perustuu vain kolmeen tasoon: tavutasoon  $\sigma$ , moratasoon  $\mu$  ja segmentaaliseen tasoon (De Lacy 1997, 13). Kielet eroavat toisistaan siinä, lasketaanko kooda morasta koostuvaksi vai ei. Suomessa kooda on morainen, koska kooda vaikuttaa tavujen pituuteen ja koodassa olevan konsonantin pituus luo fonologisia oppositioita (esimerkiksi *kato* – *katto*). Sekä Piken & Piken mallissa että moramallissa tavunalkuinen konsonantti (onset) liittyy suoraan tavutasoon, koska nukleuksen edellä oleva konsonantti ei vaikuta tavun pituuteen. Lyhyt tavu koostuu yhdestä morasta, kuten alla olevasta esimerkistä a käy ilmi. Pitkä tavu taas on vähintään kaksimorainen (esimerkki b). Ylipitkissä tavuissa ei ole selvää, koostuvatko ne kahdesta morasta (esimerkki c) vai kolmesta morasta (esimerkki d).

Seuraavista esimerkeistä selviää, että lyhin mahdollinen tavu koostuu yhdestä morasta ja yhdestä assosiaatiosta moratasosta tavutasoon. Pitkässä tavussa assosiaatioita tavutasoon on kaksi. Ylipitkissä tavussa assosiaatioita tavutasoon mallissa c on kaksi ja mallissa d kolme. Perinteisesti morateorian mukaan tavuja on aino-

astaan kahta laatua; yksimoraisia lyhyitä tavuja sekä kaksimoraisia pitkiä tavuja (Hayes 1995, 52).



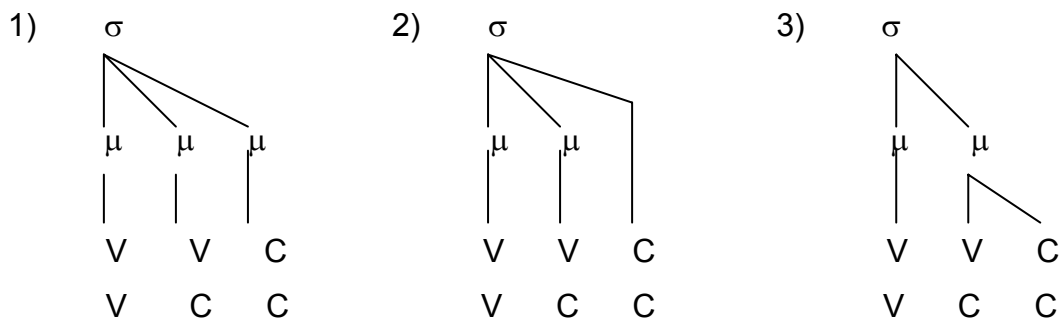
Kuvio 18. Kolmimorainen malli.

Perinteisesti ylipitkät tavut on tulkittu fennistiikan parissa kolmimoraisiksi (esim. Karlsson 1982, 134). Ongelmana tässä on se, että moran käsite on monimutkainen, koska on oletettava kolmentyyppisiä tavuja: lyhyet tavut ovat yksimoraisia ja pitkät tavut ovat kaksi- tai kolmimoraisia. Jako kaksi- ja kolmimoraisiin tavuihin on siinä mielessä turha, että pitkät tavut tulkitaan suomessa aina pitkiksi riippumatta siitä, ovatko ne kaksi- vai kolmimoraisia. Lisäksi ylipitkät tavut voivat esiintyä omaperäisissä sanoissa vain ensimmäisessä pääpainollisessa tavussa (Karlsson 1982, 115).

Loogisesti mahdollisia ovat seuraavat mallit (ks. kuvio 19). Mallissa 1 ylipitkä tavu koostuu kolmesta morasta, kun taas malleissa 2 ja 3 on kaksi moraa. Malli 3 eroaa malleista 1 ja 2 siinä, että mallissa 3 on yksi assosiaatio vähemmän tavutasoon, mutta yksi assosiaatio edellisiä enemmän moratasoon. Malli 1 on oikea, jos ylipitkät tavut todella ovat kolmemoraisia. Jos ylipitkät tavut tulkitaan kaksimoraisiksi, on valinta tehtävä mallien 2 ja 3 välillä. Malli 2 on relevantti silloin, jos tavunloppuisen konsonantin sallitaan liittyvän suoraan segmentaaliseen tasoon. Jos jokainen taso liittyy ainoastaan yhtä pykälää alempaan tasoon, on malli 3 relevantti.



Suosituin vaihtoehto lienee malli 3 (De Lacy 1997, 15). Mallien 1 ja 2 etuna on se, että niissä tavun pituus on mallia 3 selkeämmin tavutason ominaisuus. Suomessa ensimmäinen mora voi olla vain vokaali. Toinen mora voi olla joko vokaali tai konsonantti. Toisen moran jälkeinen kolmas segmentti voi olla ainoastaan konsonantti. Jos toisen moran jälkeinen segmentti on geminaatan ensimmäinen osa, voi se muutamaa poikkeusta lukuunottamatta olla vain soinniton klusiili. Sonoranttisen konsonantin jäljessä toisen moran jälkeinen segmentti voi olla myös sibilantti *s* (esim. *pörssi*, *vinssi*). Universaalien vokaalien fonotaksia koskevan rajoituksen mukaan tavussa voi olla korkeintaan kaksi vokaalia.



Kuvio 19. Erilaisia tapoja kuvata ylipitkä tavu.

On myös mahdollista olettaa, että "Strick Layer Hypothesis" (esim. Selkirk 1984) on oikea. "Strick Layer Hypothesis'n" mukaan noodilla *X* on pääsy vain tasolle *X* ja tasolle *X-1* (De Lacy 1997, 15). Näin ollen tavutasolla  $\sigma$  on pääsy vain moratasoon  $\mu$  ja mora-tasolla on pääsy segmenttaaliselle tasolle, mutta tavutaso ei voi kontrolloida segmen-taalista tasoa, koska tavutaso on segmenttaalista tasoa kaksi noodia ylempänä. Tällöin malli 3 olisi oikea.

Toinen mahdollinen malli on Hayesin (1995) teorian mukainen. Hayesin mukaan mora täytyy olla tunnistettavissa fonologisessa tuotoksessa (De Lacyn 1997, 15 mukaan). Yksinkertaisin tapa täyttää tämä vaatimus on olettaa, että morataso voi dominoida ainoastaan yksittäisiä moria. Hayes (1995, 53) esittää, että tavu on pitkä, jos tavutasossa on kaksi assosiaatiota moratasoon. Hayesin (1995) mukaan tavut jakaantuvat pitkiin ja lyhyisiin tavuihin yksinkertaisesti siten, että lyhyet tavut ovat

yksimoraisia ja pitkät tavut kaksimoraisia (De Lacyn 1997, 16 mukaan). Jos Hayes on oikeassa, on malli 3 relevantein.

Kun vertaillaan edellä olevia puukuvaimia selviää nopeasti, ettei ole helppo löytää empiiristä näyttöä siitä, mikä esitellyistä malleista on adekvaatti ja onnistuu parhaiten kuvaamaan kielellistä todellisuutta. Koska fonologiset oppositiot koostuvat binaarisista piirteistä, on todennäköistä, että kielisysteemi tekee eroa ainoastaan ensimmäisen ja toisen moran välille siten, että yksimorainen tavu on lyhyt ja kaksimorainen pitkä. Silloin on yhdentekevää, muodostaako ylipitkän tavun viimeinen segmentti kolmannen moran vai ei. Merkitystä on enemmän sillä, tulkitaanko viimeinen segmentti tavutason vai moratason ominaisuudeksi, sillä jos segmentti liittyy suoraan tavutasoon, on sillä merkitystä myös tavun prominenssiin.

Suomessa pitkän vokaalin jälkeinen geminaatta voi olla vain soinniton obstruentti (ks. Taulukko 5). Syynä tähän voi olla se, että kieli koettaa välttää tilanteita, joissa ensimmäinen pääpainollinen tavu on ylipitkä ja toinenkin tavu on pitkä obliikvi-muodoissa (esim. *\*taakkan*). Tätä sääntöä vastaan rikkoo pitkäkonsonantisista kolmimoraisista sanoista vain kolme prosenttia (esim. sana *huussi*). Poikkeuksena ovat sanat, joissa toinen tavu on umpinainen ja geminaatta on heikkoasteinen muoto (esimerkiksi sanat *kaarros*, *piirros*, *käännös*) ja käänteisessä astevaihtelussa olevat e'-loppuiset sanat kuten *kiille'*, *viere'* ja *piirre'*. Näin ollen on todennäköistä, että toisen moran jälkeinen geminaatta voi olla vain pitkä soinniton obstruentti, jos sen edellä on pitkä vokaali tai vokaaliyhtymä. Jos ylipitkät tavut tulkitaan automaattisesti kolmimoraisiksi, on vaikea keksiä syytä, miksi sonoranttinen geminaattakonsonantti ei käy pitkän vokaalin tai pitkän vokaaliaineksen jäljessä. Jos taas ylipitkät tavut tulkitaan kaksimoraisiksi, on tilanne selkeämpi.

Taulukko 4. CVVC<sup>2</sup>.C<sup>2</sup>V-tyypin nominit Käänteissanakirjan (Tuomi 1980 mukaan).

/CVVC.CV/	N		%	
muu kuin klusiili	91	2*	20,8	0,6
klusiili	346	346*	79,2	99,4
Yhteensä	437	348*	100	100

\* Ilman e'-nomineja.

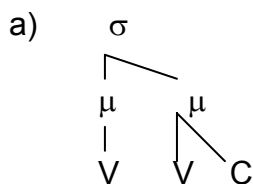
On mahdollista, että suomessa ei sallita aitoja kolmemoraisia tavuja kuin tietyin ehdoin: geminaatta ei voi muodostaa kolmatta moraa. Suomessa mahdollisia ovat tietenkin ylipitkät sonoranttiin loppuvat tavut, jotka muodostuvat yksitavuisesta sanasta ja sijapäätteestä (esim. *jään*, *pään*, *maan*). Sonoranttinen geminaatta on mahdollinen päätteissä ja omistusliitteissä (esim. *maamme*, *päällä*), mutta ei nominatiiveissa. Jos edellä mainittu pitää paikkansa, on moran määritelmä yksinkertainen: Sonorantit ovat suomessa aina moraisia. Tällöin tulee yksinkertaisesti selitetyksi se, miksi suomessa ei juurikaan ole ylipitkiä tavuja, joissa viimeinen segmentti on sonoranttisen geminaatan toinen osa. Jos siis soinnilliset geminaatat sallittaisiin ylipitkissä tavuissa, johtaisi tämä siihen, että ylipitkät tavut tulkittaisiin automaattisesti kolmi-moraisiksi. Toisaalta ylipitkien tavujen tulkitseminen kaksimoraiseksi aiheuttaa sen, että segmenttien ja morien määrä eivät enää vastaa toisiaan.

Suomen kielen minimisana on kaksimorainen CVV-tyyppi (*maa*, *pää*, *luu* etc.). Kaksimoraisia CVC-tyyppisiä sanoja ei suomen kielessä ole (lukuunottamatta pronomineja ja partikkeleita). Harrikarin (2000) mielestä tämä on osoitus siitä, että tyyppin CVV tavut ovat pidempiä kuin umpitavut CVC. Suomen kielen minimisana koostuu siis kaksimoraisesta pitkästä tavusta tai kahdesta yksimoraisesta lyhyestä tavusta (*maku*, *latu*, *muta*). Hayesin (1995, 71) mukaan moraisen trokee jalka voi koostua joko yhdestä pitkästä tavusta tai kahdesta lyhyestä tavusta, kuten suomen kielen minimisanat. Toisaalta on vahvoja perusteita, joiden mukaan suomi voidaan luokitella syllabiseksi trokeeksi. Vahvin peruste tälle on se, että suomessa sekä pääpainollinen että sivupainollinen tavu voivat olla lyhyitä, yksimoraisia tavuja ((esim. (*ma.ku*).ja, (*ba.naa*)(*ne.ja*)). Syllabinen trokee on juuri edellisten esimerkkien tapaan insensi-

tiivinen tavun pituudelle. Toisaalta suomessa voi sivupaino siirtyä neljännelle tavulle, jos kolmas tavu on lyhyt esimerkiksi tapauksissa *((ha.rak).ko.(jaan.kin)* tai *((ha.rak).(ko.jaan).kin)*.

Näyttää siltä, että suomessa ei ole kolmiportaista pituusoppositiota, jossa tavut jakautuisivat kolmeen ryhmään; ylipitkiin, pitkiin ja lyhyisiin tavuihin. Ylipitkät tavut eivät Karlssonin (1982, 115) mukaan esiinny omaperäisissä sanoissa muualla kuin ensimmäisessä pääpainollisessa tavussa. Niinpä ylipitkät tavut näyttävät liittyvän paremminkin jalkatason kuin tavutason ominaisuuksiin. Toisaalta suomessa vierasperäisissä sanoissa ylipitkä tavu voi olla myös muualla kuin ensimmäisessä tavussa, kuten esimerkiksi sanoissa *kohortti*, *mahorkka*, *Escortti*. Tämän perusteella ylipitkiä tavuja ei vältetä ensimmäisen tavun ulkopuolellakaan.

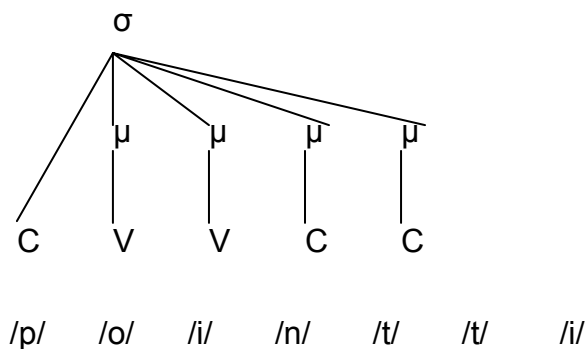
Jos siis suomessa ylipitkät tavut ovat prominentimpia kuin muut pitkät tavut ja suomessa on rajoitus, jonka mukaan ylipitkä tavu on pääpainollinen, johtaa se seuraavaan loogiseen johtopäätökseen: pääpaino voi sijoittua lainasanoissa myös toiselle tavulle, jos toinen tavu on ylipitkä. Tämän mukaan suomessa olisivat mahdollisia seuraavalla tavalla painotetut sanat, kuten *\*ma.(hork.ka)*, *\*ko.(hort.ti)*. Koska suomessa on kiinteä pääpaino ensimmäisellä tavulla, on oletettava, että kieli ei tee eroa pitkien ja ylipitkien tavujen välillä. Tällöin ylipitkien tavujen oikea tulkinta on seuraavanlainen (kuvio 20). Ylipitkien tavujen tavuloppuinen konsonantti liittyy toiseen moraan.



Kuvio 20. Ylipitkän tavun kuvaus haarautuvana morana.

Morateoria eroaa kaikista edellä esitetyistä malleista siinä, että se ei erota tavun sisäisiä yksiköitä toisistaan muuten kuin niiden moraisuuden suhteen. Morateoria on edellä esittelyjä tavumalleja yksinkertaisempi, ja siten myös sisältää vähemmän redundanssia. Morateoria on kuitenkin ongelmallinen kuvattaessa kieliä kuten suomi,

joissa on lyhyiden ja pitkien tavujen lisäksi ylipitkiä tavuja. Ilman mitään sovelluksia mora kuvaa tällaisissa kielissä vain tavun pituutta, eikä sekä tavun pituutta että sen painollisuutta. Tavallisesti morateoriassa tavut jaetaan lyhyisiin tavuihin (= yksimorainen) ja pitkiin tavuihin (= kaksimorainen). Joskus morateorian piirissä hyväksytään kolmimoraiset tavut viittaamaan ylipitkiin tavuihin. (Näin tekee mm. Karlsson 1982, 134). Jos kolmimoraiset tavut hyväksytään, ei ole mitään syytä, miksi ei hyväksyttäisi myöskin (suomessa marginaalisia) neljämoraisia tavuja. Tällöin mora-teoria jää kuitenkin puhtaasti deskriptiiviseksi.



Kuvio 21. Nelimorainen tavu.

Lyhyt tavu: 1 mora

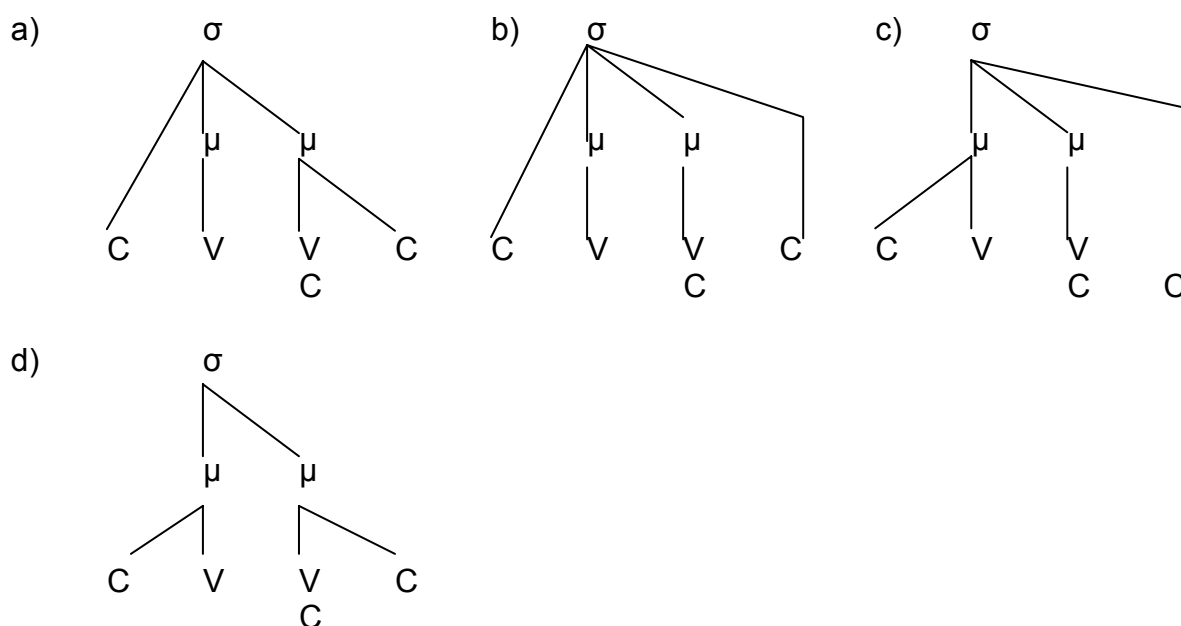
Pitkä tavu: 2 moraa

ylipitkä tavu: 3 moraa (tai jopa neljä)

Jos moraa kuitenkin käytetään määritelmänsä mukaisesti viittaamaan **sekä tavun pituuteen että tavun painollisuuteen**, joudutaan segmentaalisen rakenteen ja morarakenteen välillä tekemään väistämättä kompromisseja (ellei haluta olettaa, että suomessa on prominenssiltaan kolmen tyyppisiä tavuja: lyhyitä, pitkiä ja ylipitkiä. Yksinkertaisin tapa lienee kuvata pitkien tavujen lisäksi myös ylipitkät tavut kaksimoraisina. Tällöin ylipitkän tavun tavunloppuinen konsonanttiaines liitetään joko edelliseen moraan tai suoraan tavutasoon. Jos viimeinen segmentti liitetään moratasoon, implikoi tämä sitä, että kahdella samaan moraan liittyvällä segmentillä on yhteisiä piirteitä (a). Tämä pitää paikkansa ainakin siltä osin, että prominenssiin liittyvä toisen moran pidennys jakautuu näiden kahden moran kesken. Toisin sanoen

toisen moran pidennys ei pidennä tavun ensimmäisen vokaalin jälkeistä segmenttiä, jos sitä seuraa vielä toinen konsonantti samassa tavussa (*karkki, lompsa*). Samoin geminaatta on merkittävästi (16 %) lyhyempi pitkän vokaaliaineksen jälkeen kuin lyhyen vokaalin jälkeen (Lehtonen 1970, 105-138; Karlsson 1982, 72).

Jos taas viimeinen segmentti liitetään suoraan tavutasoon, on tavun ylipituus tulkittava suoraan tavutason ilmiöksi samaan tapaan kuin tavun alkukonsonantti. Aina-kaan puhtaasti formaaleista syistä tämä malli ei ole houkutteleva, koska se lisää yhden uuden assosiaation tavutasoon. Tälle on vaikea keksiä formaaleja perusteluja varsinkin, jos tavun alkukonsonantti liitetään myös tavu suoraan tavutasoon (b). Jos tavun alkukonsonantti liitetään tavun ensimmäiseen moraan, on helpompi hyväksyä suora assosiaatio segmenttaaliselta tasolta tavutasoon (c). Valinta mallien a ja c välillä merkitsee sitä, että joudutaan päättämään, määrääkö tavun pituuden assosiaatioiden määrä tavutasoon (malli c) vai moratasoon (malli a). Neljäs teoriassa mahdollinen malli olisi malli d, jossa sekä tavunalkuinen konsonantti että ylipitkän tavun tavunloppuinen konsonantti liittyisi moratasoon. Malli d tekee tavutasolla selkeästi eron yksi ja kaksimoraisen (lyhyen ja pitkän) tavun välille, mutta on mutkikas kuvattaessa mora- ja segmenttitasoa.



Kuvio 22. Ylipitkät tavut.

Taulukko 5. Ylipitkät tavut (malli a).

	Assosiaatioita tavutasoon	Assosiaatioita moratasoon
lyhyt tavu	1(-2)	1
pitkä tavu	2(-3)	2
ylipitkä tavu	2(-3)	3

Taulukko 6. Ylipitkät tavut (malli c).

	Assosiaatioita tavutasoon	Assosiaatioita moratasoon
lyhyt tavu	1	1(-2)
pitkä tavu	2	2(-3)
ylipitkä tavu	3	2(-3)

Ei liene olemassa yksiselitteistä empiiristä keinoa selvittää, liitetäänkö tavunalkuinen konsonantti tavutasoon vai ensimmäiseen moraan. Kummallakin näkemyksellä näyttää olevan kannatuksensa. Mallia valittaessa on kiinnitettävä huomiota ainakin seuraaviin seikkoihin.

A) Tavun alkukonsonantti on tavutason piirre.

- Tavun alkukonsonantin kesto riippuu suoraan tavutason piirteistä. Tavun alkukonsonantti on pidempi VV-jonon kuin yhden vokaali edessä (Suomi 2006, 492). Toisaalta ainoastaan sanassa pääpainollisen tavun alkukonsonantti on muita sanassa olevia lyhyitä konsonanteja pidempi (Suomi 2006, 492). Kuitenkin sanan alkukonsonantti on Suomen (2005) aineistossa ainoa segmentti, jonka pituus ei vaihtele merkittävästi vertailuissa sanatyyppien CV.CV ja CVC.CV välillä.

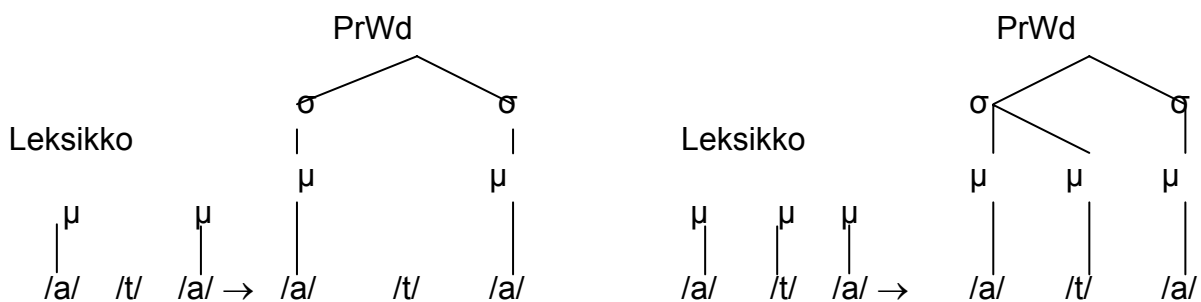
B) Tavun alkukonsonantti kuuluu ensimmäisen moraan.

- Alkukonsonantti jakaa ensimmäisen moran kanssa yhteisiä piirteitä esim. aksenttiin liittyvän segmenttien keston pidentymisen muodossa (aksentista ks. Suomi 2006, 229).

- Alkukonsonantti ja ensimmäinen mora muodostavat elimellisen yhteyden, jota ei voi rikkoa. Tätä on vaikea empiirisesti todistaa, koska konsonantti määritelmänsä mukaisesti kuuluu aina sonantin yhteyteen ( $C, V \rightarrow CV$ ;  $V, C, V \rightarrow V.CV$ ).

Koska kysymystä alkukonsonantin asemasta tavutasoon nähden ei voida ainakaan tämän työn puitteissa empiirisesti ratkaista, liitän tavunalkukonsonantin suoraan tavutasoon. Ongelma ei myöskään ole keskeinen tässä työssä.

Morateoriassa tavunmuodostus on yksinkertaisempi prosessi kuin muissa edellä esitetyissä autosegmentaalisissa malleissa. Morateoriassa oletetaan, että tavut muodostetaan moraisista segmenteistä alkaen, koska tällöin tavunalkuinen konsonantti eroaa geminaatasta yksinkertaisesti siten, että geminaatta on jo leksikossa morainen. Morataso liitetään tavutasoon (kuvio 23).



Kuvio 23. Tavunmuodostus morateoriassa.

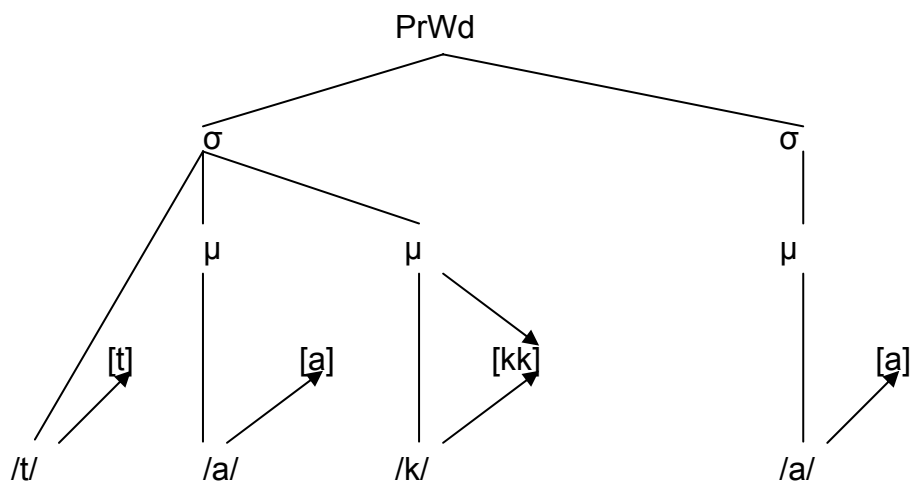
#### 4.4. Pituuden koodaamisongelmat kirjoituksessa fonologisessa mallissa

Tarkastelen tässä luvussa pituuden koodaamista autosegmentaalisessa fonologiassa morateorian kannalta. Lähtökohtana morateoriassa on, että moraiset segmentit ovat myös leksikossa moraisia. Näin ollen puheessa kvantiteettidikotomia rakentuu sen mukaan, onko segmentti morainen vai ei. Suomessa kvantiteetin merkintä on juuri se asema, jossa tehdään luultavasti eniten kirjoitusvirheitä. Jos puheessa pitkä kvantiteetti eroaa lyhyestä kvantiteetista autosegmentaalisella tasolla, on oletettava, että informaatiota autosegmentaaliselta tasolta tarvitaan myös kirjoitusprosessissa.



## Hypoteesi A

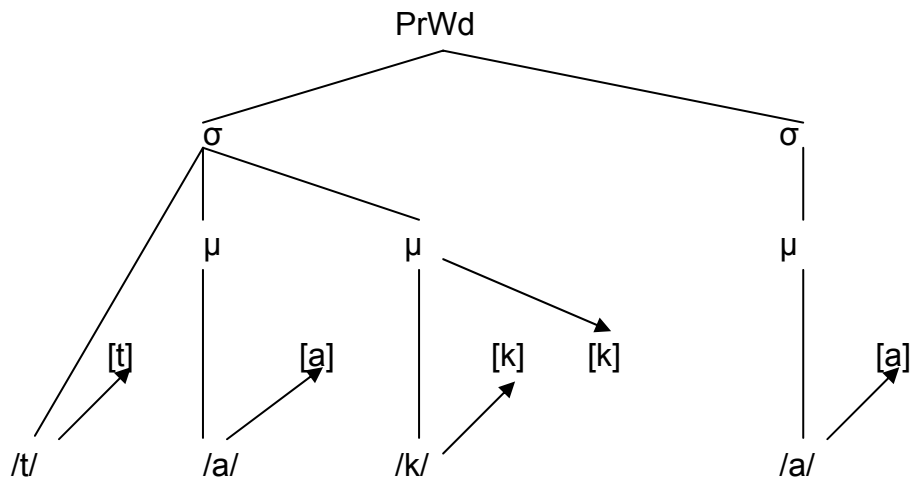
Jos kirjoitusta kuvataan puheen koodauksena ja samalla hyväksytään autosegmentaalinen malli, on ajateltava, että lyhyet segmentit koodataan ”suoraan” segmentaaliselta tasolta, mutta geminaattojen ja pitkien vokaalien koodauksessa on käytettävä informaatiota moratasolta (tai yleisemmin ajoitustasolta). Morainen konsonantti (= geminaatta) koodataan kaksoisgrafeemiksi sekä segmentaalisen tason että moratason informaation avulla.



Kuvio 24. Kirjoitus ja autosegmentaalinen malli A.

## Hypoteesi B

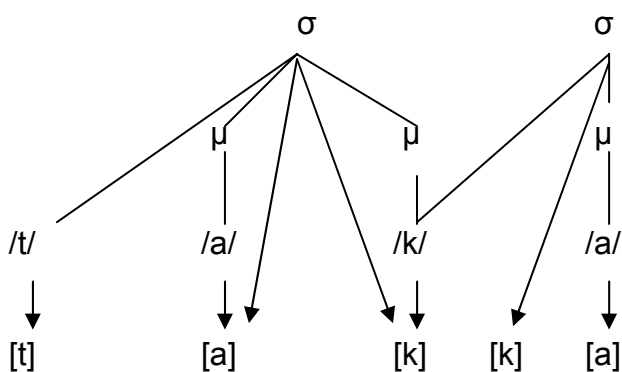
Tämän hypoteesin mukaan kirjoituksessa kaksoisgrafeemi ei ole kiinteä yksikkö, vaan kaksoisgrafeemin molemmat osat merkitään itsenäisesti. Geminaatan ensimmäinen osa koodataan segmentaalisen informaation avulla ja toinen osa moratasolta saadun informaation avulla.



Kuvio 25. Kirjoitus ja autosegmentaalinen malli B.

### Hypoteesi C

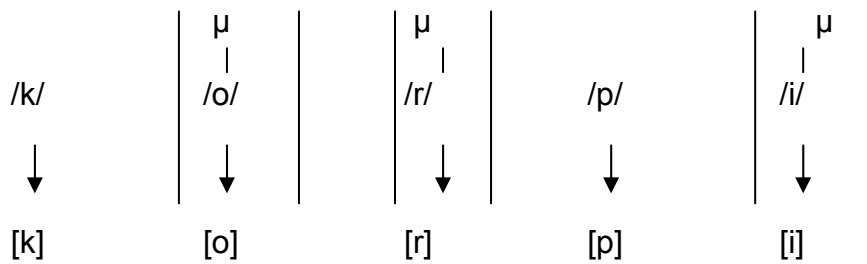
Kolmannessa mahdollisessa hypoteesissa kirjoitusprosessissa käytetään hyväksi tavutason informaatiota. Tällöin on oletettava, että tavu on keskeisin sanaa pienempi yksikkö, joka on tarpeellinen myös kirjoitusprosessissa. Empirian tasolla tämä näkyy siten, että kirjoitusprosessissa tavujen muodostaminen on keskeinen prosessi. Voi olla, että tällöin joudutaan oletamaan, että grafeemitkin yhdistyvät tavutasoon.



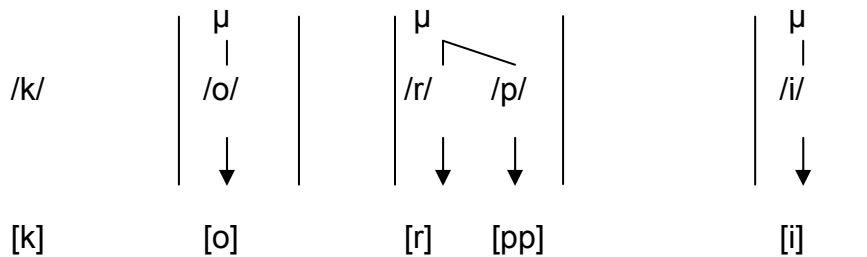
Kuvio 26. Kirjoitus ja autosegmentaalinen malli B.

Malleille on yhteistä se, että informaatio ajoitustasolta on välttämätöntä, ja että pelkältä segmentaaliselta tasolta koodaaminen ei ole mahdollista. Tämä implikoi sitä, että moratasoon sisältyvä prosodinen informaatio on välttämätöntä kirjoitus-

prosessissa. Morataso sisältää mm. prominenssiin liittyvää kestoinformaatiota. Jos morataso sisältää kestoinformaatiota, on yksinkertaisinta olettaa, että ylipitkät tavut ovat kaksimoraisia. Näin on siksi, että pitkässä tavussa toisen moran muodostava segmentti on pidennetty verrattuna muihin segmentteihin. Jos taas tavu on ylipitkä, ei toisen moran muodostavaa segmenttiä pidennetä, vaan kesto näyttää jakautuvan soinnillisen segmentin ja sitä seuraavan segmentin välille. Mora voisi jakautua kahden segmentin välille.



/korpi/

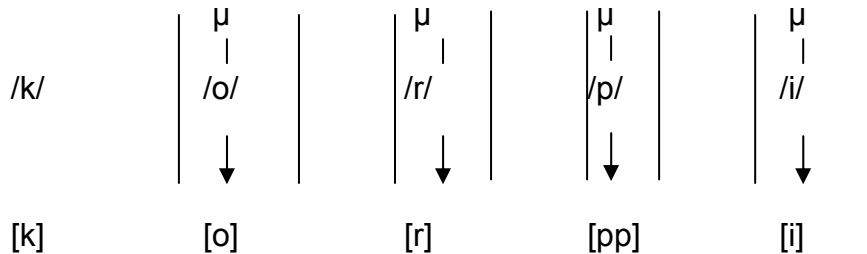


/korppi/

Kuvio 27. Mora ajoitusyksikkönä 1.

Jos ylipitkät tavut kuvataan kaksimoraisiksi, jää selitettäväksi, miten toisen moran pidennys jää pois silloin, kun kolmas mora on tavunloppuinen. Tähän on kaksi mahdollista selitystä. Hypoteettisesti on mahdollista, että tavutaso kontrolloi segmenttien kestoja. Näin ei kuitenkaan näytä olevan, sillä suomessa ei ole tavun isokroniaa (Suomi & Ylitalo 2003, 39). Toinen mahdollisuus on, että ilmiön syy on segmenttaalisella tasolla. Tämä selitys taas ei ole kovinkaan käyttökelpoinen siksi, että siinä ylemmille tasoille kuuluvaa informaatiota siirretään segmenttaaliselle tasolle kuuluvaksi.

Kuvattuna kolmimoraisena:



Kuvio 28. Mora ajoitusyksikkönä 2.

#### 4.5 Äänteiden kestoista

On ollut pitkään tiedossa, että geminaatta ja pitkä vokaali ovat kestoiltaan noin kaksi kertaa vastaavia lyhyitä segmenttejä pidempiä. Tässä työssäni nojaudun etupäässä Lehtosen (1970) esittämiin äänteiden kestoja kuvaaviin lukuihin. Olen myös laskenut konsonanttien kestoja kirjoitustestissä käyttämäni epäsanoina. Lehtonen vertaili toisiinsa erilaisia sanarakenteita, esimerkiksi rakenteita CVCV ja CVCVV, siten että kulloinkin vertailtavina olevat rakenteet koostuivat samoista foneemisegmenteistä ja erosivat toisistaan vain kvantiteetin suhteen; VV oli aina kaksoisvokaali ja CC eräin poikkeuksin aina kaksoiskonsonantti. Rakenteen CVCV edustajia olivat esimerkiksi sanat /tapa/, /lika/, /ripi/ ja /sata/, ja rakenteen CVCVV edustajia esimerkiksi sanat /tapaa/, /likaa/, /ripii/ ja /sataa/. Näin voitiin kontrolloida äänteiden ominaiskestoerojen mahdollisesti vääristävät vaikutukset itse tutkittavaan ilmiöön, kvantiteettivastakohdan toteutumiseen kestoerojen avulla. Tämän asetelman vuoksi Lehtonen ei kuitenkaan verrannut — eikä olisi voinut luotettavasti verrata — sanarakenteita toisiinsa muutoin kuin pareittain, aina kahta rakennetta kerrallaan. Niinpä vertailussa CVCV–CVCVV olivat mukana tietyt CVCV-sanat, vertailussa CVCV–CVCCV toiset CVCV-sanat. Lehtonen ei myöskään juuri tehnyt siinä mielessä syntagmaattisia vertailuja, että hän olisi esimerkiksi systemaattisesti verrannut keskenään vokaalien kestoja sanan eri tavuaseissa. Lehtonen ei kaikkiaan suhteuttanut pääaineistostaan saamia tuloksia millään tavoin sanapainoon. Lehtosen mittaamat

sanat esiintyivät lausekehyksissään sellaisissa asemissa, että niille lankesi lievä aksentti ("a weak sentence stress") (Lehtonen 1970, 45; Suomi 2006, 484).

Lehtosen (1970) työssä geminaattojen kestot on mitattu ainoastaan ympäristössä /pa\_a/. Lehtosen laskemia lukuja tarkasteltaessa on otettava huomioon, että soinnillisissa konsonanteissa geminaattaa pidentää se, että geminaatan alkuosa muodostaa sanan toisen pidennetyn moran. Viimeiseen (Taulukko 7) sarakkeeseen olen laskenut Lehtosen (1970) aineistosta, kuinka monta prosenttia yksinäiskonsonantti on geminaatan kestosta ympäristössä /pa\_a/. Luvuista ilmenee, että soinnillisissa konsonanteissa geminaatta on suhteessa yksinäiskonsonanttiin kestoltaan lyhyempi kuin sama suhde soinnittomilla konsonanteilla. Tosin ero ei ole kovinkaan suuri. Suomessa sanansisäisen geminaatan kesto vaihtelee rajusti sen mukaan, sisältääkö geminaattaa seuraava tavu pitkän vai lyhyen vokaaliaineksen. Lehtosen (1970, 71) aineiston perusteella näyttää siltä, että geminaatta on pisin pitkän vokaalin edellä (CVV), seuraavaksi pisin lyhyessä tavussa (CV) ja kaikkein lyhin umpitavun edellä (CVC).

Taulukko 7. Sanansisäisten konsonanttien ominaiskestot (ms.).

	/pa_a/	/pa_an/	/pa_aa/	geminaatta	C : CC %
/p/	102	99	118	191	53 %
/t/	94	94	108	197	<b>48 %</b>
/k/	98	102	112	205	48 %
/s/	94	82	101	190	49 %
/h/	79	76	85	----	---
/m/	67	73	79	142	47 %
/n/	56	55	65	134	<b>42 %</b>
/ŋ/	---	---	---	146	---
/l/	48	52	52	124	39 %
/r/	45	50	61	125	36 %
/j/	83	88	100	---	---
/v/	57	60	69	---	---
/d/	51	53	62	---	---

Edelleen Lehtosen (1970, 71) laskemista kestoista nähdään, että soinnillisten ja soinnittomien geminaattojen ja yksinäiskonsonanttien suhteellisten kestojen välillä on eroja. Soinnilliset konsonantit ovat keskimäärin 8,3 % suhteellisesti pidempiä geminaattoina kuin yksinäiskonsonantteina suhteessa soinnittomiin konsonantteihin. Geminaatan kesto vaihtelee huomattavasti yksinäiskonsonanttia enemmän. Syynä lienee ainakin osittain se, että geminaatta ei pidentyessään sekoita kesto-suhteita samaan tapaan kuin yksinäiskonsonantti: yksinäiskonsonanttihan voi sekaantua geminaattaan, jos sen kesto lisätään liikaa. Geminaatan keston vaikuttanee myös se, että lyhyen vokaalin jälkeen geminaatan alkuosa muodostaa sanan toisen moran, jota pidennetään prosodisista syistä (Suomi 2002, 5). Suomen (2002, 5) mukaan pitkä pääpainollinen vokaali on noin 2,6 kertaa pidempi kuin pääpainollinen yksinäisvokaali. Samoin lyhyen pääpainollisen tavun jälkeisen toisen tavun vokaali (CV.CV) on 1,6 kertaa pidempi kuin toisen vokaali sanatyypeissä CVV.CV ja CVC.CV. Edelleen sanatyyppissä CVV.CV pitkä vokaali on 2,45 kertaa pidempi kuin tyyppissä

CV.CVV. Samoin sanan toisen moran muodostava konsonantti on sanatyypissä CVC.CV pidentynyt esimerkiksi sanassa /kulta/, mutta ei sanoissa kuten /maksu/, joissa sanan toisen moran muodostava konsonantti on soinniton. Karlssonin (1983: 152) mukaan ”kun M2 on konsonantti, resonanteilla on selvempi taipumus pidentyä kuin obstruenteilla”. Suomen ja Ylitalon (2003, 1) mukaan lievästi aksentoiduissa sanoissa, joissa M2 oli resonantti, se piteni merkitsevästi (68 %), mutta kun M2 oli obstruentti, pidentymistä ei tapahtunut.

Koska tässä käyttämissäni Lehtosen (1970) laskemissa luvuissa ei ole eroteltu eri konsonantteja, ovat luvut sikäli epätarkkoja, että niistä ei voida päätellä, kuinka paljon pidempi soinnillinen geminaatta on silloin, kun se muodostaa sanan toisen moran.

Taulukko 8. Yksinäiskonsonantin ja geminaatan suhteelliset pituudet konsonanteittain.

	/pa_a/	Gem. /pa_a/	Geminaatta	CC : C %
/p/	102	191	1,9	46,6
/t/	94	197	<b>2,1</b>	<b>52,2</b>
/k/	98	205	<b>2,1</b>	<b>52,2</b>
/s/	94	190	2,0	50,5
Obstr. ka.	97	196	2,0	50,5
/m/	67	142	2,1	47,2
/n/	56	134	<b>2,4</b>	<b>52,8</b>
//	48	124	2,6	61,3
/r/	45	125	2,8	64,0
Soinn. ka.	54	131	2,4	58,8

Konsonanttien ääntöpaikka näyttää vaikuttavan konsonanttien pituuteen. Labiaalit ovat Lehtosen (1970, 71) aineiston mukaan kestoiltaan pidempiä kuin dentaalit ja velaarit.

Taulukko 9. Konsonanttien ominaiskestot ääntöpaikan mukaan.

Ääntöpaikka	Segmentti	/pa_a/	gem. /pa_a/	Geminaatta	CC : C %
Labiaali	/p/	102	191	1,9	46,6
	/m/	67	142	2,1	47,2
Dentaali	/t/	94	197	2,1	52,2
	/s/	94	190	2,0	50,5
	/n/	56	134	2,4	52,8
	/l/	48	124	2,6	61,3
	/r/	45	125	2,8	64,0
Velaari	/k/	98	205	2,1	52,2
	/ŋ/	---	146	---	

Jos pääpainollinen tavu on lyhyt, on toisen tavun lyhyt vokaali puolipitkä (Lehtonen 1970, 127-129). Tätä hyvin tunnettua ilmiötä voidaan Kari Suomen (2002, 2004) termein kutsua sanan toisen moran pidennykseksi (ks yllä).

Taulukko 10. Segmenttien kestot CVCV- ja CVCCV-sanatyypeissä.

	C	V	C	V
CVCV	72	65	78	102
CVCCV	70	80	146	47

Edelleen sanan ensimmäinen vokaali (= ensimmäinen mora) on suunnilleen yhtä pitkä riippumatta toisen tavun rakenteesta (Lehtonen 1970, 126). Myös kestoerot toisen tavun alkukonsonantissa ovat pieniä.



Taulukko 11. Umpitavuisuuden vaikutus toisen tavun alkukonsonanttiin.

	<b>C</b>	<b>V</b>	<b>C</b>	<b>V</b>	<b>C</b>
CVCV	72	65	<u>78</u>	102	-----
CVVC	67	62	<u>70</u>	100	74
CVCVV	71	65	<u>89</u>	144	-----
CVVCVC	73	67	<u>83</u>	154	59

Sanan ensimmäisen vokaalin kesto ei näytä vaihtelevan myöskään silloin, kun pääpainollinen tavu loppuu geminaattaan. Tällöin V1 on hieman pidempi kuin V1 on avotavussa. Toisen tavun rakenne ei vaikuta tässäkään tapauksessa tavun keston. Geminaattakonsonantti on hieman pidempi, kun sitä seuraa pitkä vokaali verrattuna geminaattaan, jota seuraa lyhyt vokaali.

Taulukko 12. Geminaatta avo- ja umpitavun edellä.

	<b>C</b>	<b>V</b>	<b>C</b>	<b>V</b>	<b>C</b>
CVCCV	70	80	146	47	-----
CVCCVC	80	87	134	55	88
CVCCVV	66	78	188	131	-----
CVCCVVC	73	77	192	123	60

Toisen tavun tavunalkuinen konsonantti on hieman pidempi kestoiltaan silloin, kun sitä seuraa pitkä vokaali verrattuna samaan sanahahmoon, jossa sanansisäistä lyhyttä konsonanttia seuraa lyhyt vokaali.

Taulukko 13. Toisen tavun alkukonsonantin kesto.

	<b>C</b>	<b>V</b>	<b>C</b>	<b>V</b>	<b>C</b>
CVVCV	75	145	67	46	-----
CVVCVC	82	148	58	57	85
CVVCVV	72	168	85	133	-----
CVVCVVC	77	175	87	140	59

Myöskin ylipitkissä tavuissa sanansisäinen geminaatta on pidempi kestoiltaan silloin, kun sitä seuraa pitkä vokaali. Myös tässä ryhmässä V2:n kesto vaikuttaa pääpainollisen tavun vokaalin keston.

Taulukko 14. Geminaatan pituus ylipitkissä tavuissa.

	<b>C</b>	<b>V</b>	<b>C</b>	<b>V</b>	<b>C</b>
CVVCCV	80	130	145	42	-----
CVVCCVC	76	149	129	41	78
CVVCCVV	75	138	181	132	-----
CVVCCVVC	68	136	185	123	59

Kuten edellä mainituista sanansisäisten segmenttien kestoista selviää, on sanansisäinen pitkä obstruentti keskimäärin kaksi kertaa pidempi kuin lyhyt sanansisäinen obstruentti (yksinäiskonsonantin kesto on 50,5 prosenttia geminaatan kestoista). Soinnillisissa konsonanteissa geminaatta on keskimäärin 2,4 kertaa pidempi kuin yksinäiskonsonantti (soinnillinen pitkä konsonantti 52,8 prosenttia pitempi kuin yksinäiskonsonantti). Tämän perusteella kirjoittamaan oppivan lapsen olisi helpompi oppia ensin erottamaan soinnilliset geminaatat yksinäiskonsonanteista kuin erottaa pitkän ja lyhyen klusiilin toisistaan, koska pitkän ja lyhyen segmentin ero on soinnillisissa konsonanteissa suurempi kuin klusiileissa.

Käyttämässäni kirjoitustestin epä sanoissa (luku 7) Praatilla ([www.praat.org](http://www.praat.org); Boersma & Weenink 2006) segmentoimani sanansisäisten konsonanttien kestot ovat

seuraavanlaiset. Soinnittomien klusiilien keston laskettiin umpivaiheen lisäksi konsonanttien purkausvaiheet. Lehtonen (1970) taas laski klusiilien purkausvaiheet umpivaiheen jälkeen osaksi seuraavaa vokaalia. Segmentoinnin kriteereillä ei ole tässä merkitystä, koska kahden eri tutkimuksen tuloksia ei kuitenkaan voida verrata suoraan toisiinsa. Lisäksi puhenopeus vaihtelee eri tutkimuksien välillä. Tämän vuoksi suhteelliset kestot ovat absoluuttisia kestoja tärkeämpiä. Tässä segmentoidut epäsanat on äänitetty yhdeltä puhujalta. Puhuja luki listasta kunkin epäsanon kolme kertaa. Näistä valittiin kuulon perusteella paras. Huomiota tässä herättää puheen hitaus, jolla on ehkä tekemistä sen kanssa, että puhuja pyrki lausumaan yksittäin esiintyvät sanat mahdollisimman huolellisesti.

Lyhyen vokaalin jälkeinen pitkä klusiili on noin 2,4 kertaa vastaavaa yksinäisklusiilia pidempi. Luvut vastaavat suurin piirtein Lehtosen (1970, 71) vastaavia kestoja. Pitkän vokaalin jälkeinen nasaaligeminaatta (lyhyen ja pitkän suhde 2,2) on suhteessa vastaavaan yksinäisklusiiliin hieman lyhyempi kuin lyhyen vokaalin jälkeinen nasaaligeminaatta (lyhyen ja pitkän suhde 2,4).

Taulukko 15. Lyhyen klusiilin kesto toisen tavun alussa.

	<b>C<sup>1</sup></b>	<b>V<sup>1</sup></b>	<b>C<sup>2</sup></b>	<b>V<sup>2</sup></b>	<b>C<sup>3</sup></b>	<b>V<sup>3</sup></b>
/lutiipo/	0,077	0,085	0,126	0,103	0,134	0,089
/sudemmi/	0,110	0,071	0,111	0,110	0,244	0,066
/lutiippo/	0,046	0,065	0,141	0,234	0,128	100
/sutippo/	0,124	0,079	0,108	0,084	0,353	0,91
/suteemi/	0,119	0,088	0,122	0,134	0,088	0,076
/sutiippo/	0,111	0,075	0,127	0,183	0,287	0,089
Keskisarvo	-----	-----	<u>0,123</u>	-----	-----	-----

Taulukko 16. Pitkän klusiilin kesto toisen tavun alussa.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/nuttiippo/	0,066	0,106	0,368	0,208	0,136	0,111
/ruttemmi/	0,049	0,051	0,270	0,104	0,237	0,105
/ruttiippo/	0,053	0,094	0,316	0,140	0,257	0,094
/ruttippo/	0,103	0,090	0,242	0,055	0,298	0,073
/nuttippo/	0,049	0,086	0,281	0,048	0,141	0,100
/ruttemi/	0,047	0,075	0,281	0,078	0,080	0,0105
Keskiarvo	-----	-----	<u>0,293</u>	-----	-----	-----

Vastaavasti ensimmäisen ja toisen tavun rajalla pitkä nasaali on noin 2,7 kertaa pidempi kuin lyhyt nasaali. Tämä ero on suurempi kuin Lehtosen (1970, 71) luvuista laskettu suhde. Lehtosella ero soinnillisen lyhyen ja pitkän konsonantin välillä on 1 : 2,4.

Taulukko 17. Lyhyen nasaalin kesto toisen tavun alussa.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/sunippo/	0,105	0,085	0,056	0,084	0,415	0,075
/sunipo/	0,139	0,126	0,070	0,101	0,134	0,097
Keskiarvo	-----	0,106	0,063	-----	-----	-----
C:CC	-----	-----	0,063	-----	-----	-----

Taulukko 18. Pitkän nasaalin kesto toisen tavun alussa.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/runnipo/	0,085	0,131	0,165	0,077	0,125	0,101
/runniippo/	0,121	0,124	0,176	0,058	0,308	0,075
Keskiarvo	-----	0,128	0,171	-----	-----	-----
C:CC	-----	-----	<u>0,171</u>	-----	-----	-----

Lyhyt klusiili on pitkän vokaalin jäljessä keskimäärin 103 millisekuntia. Tämä luku on samansuuntainen kuin Lehtosella (1970, 71). Klusiiligeminaatta taas on keskimäärin 0,229 millisekuntia. Lyhyen ja pitkän klusiilin suhde on epäsanoissa 1 : 2,2.

Taulukko 19. Lyhyen klusiilin kesto toisen tavun alussa pitkän vokaalin jälkeen.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/reetipo/	0,100	0,225	0,097	0,059	0,116	0,80
/ruutemmi/	0,042	0,230	0,103	0,096	0,186	0,80
/ruutemi/	0,064	0,218	0,108	0,079	0,093	0,099
/neetippo/	0,117	0,225	0,104	0,060	0,296	0,106
Keskiarvo	-----	-----	<u>0,103</u>	-----	-----	-----

Taulukko 20. Klusiiligeminaatan keston pitkän vokaalin jälkeen.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/seettipo/	0,140	0,170	0,236	0,66	0,109	0,083
/reettippo/	0,073	0,180	<u>0,213</u>	0,052	0,262	0,070
/nuuttemi/	0,070	0,188	<u>0,247</u>	0,090	0,077	0,082
/nuuttemmi/	0,079	0,188	<u>0,218</u>	0,090	0,205	0,095
Keskiarvo	-----	-----	<u>0,229</u>	-----	-----	-----

Lyhyt nasaali on pitkän vokaalin jäljessä 69 millisekuntia ja nasaaligeminaatta vastaavassa asemassa keskimäärin 147 millisekuntia pitkä. Geminaatta on siis lyhyttä nasaalia suunnilleen 2,1 kertaa pidempi. Pitkän vokaalin jälkeinen nasaaligeminaatta on näin ollen suhteellisesti lyhyempi kuin lyhyen vokaalin jäljessä (2,1 vs. 2,7).

Taulukko 21. Lyhyen nasaalin kesto pitkän vokaalin jäljessä.

	<b>C<sup>1</sup></b>	<b>V<sup>1</sup></b>	<b>C<sup>2</sup></b>	<b>V<sup>2</sup></b>	<b>C<sup>3</sup></b>	<b>V<sup>3</sup></b>
/tuunippo/	0,025	0,242	0,069	0,053	0,285	0,087
/ruunipo/	0,073	0,267	0,069	0,069	0,122	0,101
Keskiarvo	-----	-----	<u>0,69</u>	-----	-----	-----

Taulukko 22. Nasaaligeminaatan kesto pitkän vokaalin jälkeen.

	<b>C<sup>1</sup></b>	<b>V<sup>1</sup></b>	<b>C<sup>2</sup></b>	<b>V<sup>2</sup></b>	<b>C<sup>3</sup></b>	<b>V<sup>3</sup></b>
/tuunnipo/	0,136	0,121	0,139	0,084	0,106	0,071
/ruunnippo/	0,063	0,211	0,154	0,062	0,257	0,087
C:CC	-----	-----	<u>0,147</u>	-----	-----	-----

Toisen ja kolmannen tavun rajalla olevien klusiilien kestot eroavat huomasti ensimmäisen ja toisen tavun rajalla olevista konsonanteista, joiden keskimääräinen pituus on geminaatoissa 0,293 ms. ja lyhyissä klusiileissa 0,123. Pitkä klusiili on tässä asemassa noin 2,5 kertaa pidempi kuin vastaava lyhyt klusiili. Toisen ja kolmannen tavun rajalla yksinäiskonsonantin kesto on keskimäärin 0,124 ms., mutta pitkän klusiilin 0,314 ms.

Taulukko 23. Lyhyt kolmannen tavun tavunalkuinen klusiili.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/lutipol/	0,077	0,085	0,126	0,103	0,134	0,089
/lutiipol/	0,046	0,065	0,141	0,234	0,128	100
/seettipol/	0,140	0,170	0,236	0,66	0,109	0,083
/ruunipol/	0,073	0,267	0,069	0,069	0,122	0,101
/sunipol/	0,139	0,126	0,070	0,101	0,134	0,097
/runnipol/	0,085	0,131	0,165	0,077	0,125	0,101
/tuunnipol/	0,136	0,121	0,139	0,084	0,106	0,071
/reettipol/	0,100	0,225	0,097	0,059	0,116	0,80
/nuttipol/	0,049	0,086	0,281	0,048	0,141	0,100
Keskisarvo	-----	-----	-----	-----	<u>0,124</u>	-----

Taulukko 24. Klusiiligeminaatta toisen ja kolmannen tavun rajalla.

	C <sup>1</sup>	V <sup>1</sup>	C <sup>2</sup>	V <sup>2</sup>	C <sup>3</sup>	V <sup>3</sup>
/ruttipol/	0,103	0,090	0,242	0,055	0,298	0,073
/reettipol/	0,073	0,180	0,213	0,052	0,262	0,070
/tuunipol/	0,025	0,242	0,069	0,053	0,285	0,087
/ruttipol/	0,103	0,090	0,242	0,055	0,298	0,073
/sutipol/	0,124	0,079	0,108	0,084	0,353	0,91
/neetipol/	0,117	0,225	0,104	0,060	0,296	0,106
/sunipol/	0,105	0,085	0,056	0,084	0,415	0,075
/runnipol/	0,121	0,124	0,176	0,058	0,308	0,075
Keskisarvo	-----	-----	-----	-----	<u>0,314</u>	-----

Nasaaleissa suunta on samankaltainen kuin toisen ja kolmannen tavun rajaisissa klusiiligeminaatoissa. Pitkä nasaali on 2,6 kertaa lyhyttä nasaalia pidempi. Syynä tähän voi olla se, että epäsanat luettiin listasta ilman kehyslauseetta. Tällöin on mahdollista, että lukija pidentää toisen ja kolmannen tavun rajaista pitkää konsonanttia enemmän kuin ”normaalissa” puheessa.

Taulukko 25. Kolmannen tavun tavunalkuinen nasaali.

	<b>C<sup>1</sup></b>	<b>V<sup>1</sup></b>	<b>C<sup>2</sup></b>	<b>V<sup>2</sup></b>	<b>C<sup>3</sup></b>	<b>V<sup>3</sup></b>
/nuuttemi/	0,070	0,188	<u>0,247</u>	0,090	0,077	0,082
/ruttemi/	0,047	0,075	0,281	0,078	0,080	0,0105
/ruutemi/	0,064	0,218	0,108	0,079	0,093	0,099
/sutemi/	0,119	0,088	0,122	0,134	0,088	0,076
Keskiarvo	-----	-----	-----	-----	<u>0,085</u>	-----

Taulukko 26. Toisen ja kolmannen tavun tavunrajainen nasaaligeminaatta.

	<b>C<sup>1</sup></b>	<b>V<sup>1</sup></b>	<b>C<sup>2</sup></b>	<b>V<sup>2</sup></b>	<b>C<sup>3</sup></b>	<b>V<sup>3</sup></b>
/sutemmi/	0,110	0,071	0,111	0,110	0,244	0,066
/nuuttemmi/	0,079	0,188	<u>0,218</u>	0,090	0,205	0,095
/ruutemmi/	0,042	0,230	0,103	0,096	0,186	0,80
/ruttemmi/	0,049	0,051	0,270	0,104	0,237	0,105
Keskiarvo	-----	-----	-----	-----	<u>0,218</u>	-----



## 5. Astevaihtelun tarkastelua eri kuvausmenetelmin

### 5.1. Astevaihtelu morfofonologian ongelmana

Kvantitatiivinen astevaihtelu on suomen kielessä tyypin *maa : maita* vaihtelun lisäksi ainoa morfofonologinen ilmiö, joka vaikuttaa segmentin pituuteen. Astevaihtelu jaetaan perinteisesti kvalitatiiviseen ja kvantitatiiviseen astevaihteluun. Kvantitatiivinen astevaihtelu voidaan lisäksi jakaa radikaaliin astevaihteluun (*takka : takan*) ja monitavujen erikoisvaihteluun (*harakka : harakan*) (Karlsson 1982, 335). Radikaalin astevaihtelun ja monitavujen erikoisvaihtelun erottaminen toisistaan on oikeutettua silloin, jos näiden kahden katsotaan eroavan toisistaan astevaihteluympäristön perusteella. Tällöin korostetaan sitä, että radikaalissa astevaihtelussa on heikko aste ainoastaan lyhyestä vokaalista ja konsonantista koostuvan umpitavun edessä. Monitavujen erikoisvaihtelussa heikko aste voi olla myös *i*-loppuisen diftongin edellä (esim. *harakka : harakoita*). Jos taas tavun painoa pidetään määrävänä tekijänä, ei ole tarpeen erottaa monitavujen vaihtelua ja radikaalia astevaihtelua toisistaan. Tällöin riittää, että postuloidaan sääntö ”heikko aste pitkän tavun edellä”.

Edellä kuvatun kaltaisista lyhyen ja pitkän tavun suhteisiin vaikuttavista ilmiöistä mainitsee myös Lauri Kettunen (1949). Tosin Kettusen kuvaamissa ilmiöissä sivupainollinen tavu lyhentää edellistä painotonta tavua, ei lisää painollisen tavun vahvuutta. Lauri Kettunen (1949, 69) on kiinnittänyt huomiota siihen, että heikkoa astetta tavataan suomen murteissa asemassa, joissa seuraava tavu on ollut tarpeeksi vahvasti sivupainollinen aiheuttaakseen heikon asteen edelliseen tavuun (esim. *sairaloinen, rapakoinen*). Samaa on Kettusen mukaan sanottava kaakkoishämäläis-kannakselaisesta taivutustyyppistä *pu.nain : pu.naist : pu.na.seks*, jossa vahvasti sivupainollinen tavu on lyhentänyt toista tavua. Kettusen mukaan (1949, 72): ”--- mutta kun jo suomesta edellä esittämäni näytteet osoittavat, miten sivupaino on kyennyt aikaansaamaan kvantiteetin lyhenemistä edellisessä tavussa, ei ole syytä olettaa muuta myöskään ensi ja toisen tavun rajalla tapahtuneesta klusiilinheikennyksestä.”. Seuraavassa oletan, että suomen metrinen mitta on morainen trokee. Seuraavista esimerkeistä ilmenee, että vähintään kolmemoraiseksi tulkit-

tava pitkä tavu lyhentää edellistä tavua. Apokopeemuodot ovat tulkittavissa joko kolmemoraisiksi tai jopa neljämoraisiksi. On kuitenkin oletettavaa, että kolmimorainen tulkinta on oikeutetumpi ja riittää tekemään selväksi eron *i*-diftongillisten muotojen ja diftongittomien muotojen välillä.

3	4	4
μ <u>uuu</u>	μ <u>uuuu</u>	μ <u>u</u> <u>uuu</u>
(pu. <u>nain</u> )	: (pu. <u>naist</u> )	: (pu. <u>na</u> ). <u>(seks)</u>

tai:

3	3	3
μ <u>uuu</u>	μ <u>uuu</u>	μ <u>u</u> <u>uu</u>
(pu. <u>nain</u> )	: (pu. <u>naist</u> )	: (pu. <u>na</u> ). <u>(seks)</u>

Kuvio 29. Painottoman ja sivupainollisen tavun pituusvaihtelut.

Ravilan (1951, 298) mukaan taas astevaihtelun syntyä selvitettäessä on korostettava tavun avonaisuuden merkitystä pikemminkin kuin tavun umpinaisuutta. Ravilan (ed.) mukaan ”*Esisuomalaiset geminaattaklusiilit kk, tt, pp ovat kantasuomessa heikentyneet kaikissa asemissa, paitsi painollisen ja painottoman tavun rajalla avotavun edessä. Kaikissa niissä asemissa, joissa geminaattaklusiilit ovat heikentyneet, ovat myös yksinäisklusiilit heikentyneet. Yksinäistenuisklusiilit, mikäli eivät ole joutuneet kokonaan toisenlaisten äänten muutosten alaisiksi, ovat säilyneet heikentymättöminä kaikissa niissä asemissa, joissa geminaattaklusiilikin on säilynyt, sekä kaikissa niissä asemissa, joissa geminaattaklusiili ei ole esiintynyt.*”

Yleinen klusiilien astevaihtelusääntö muotoillaan suomen kielen muoto-opeissa usein tyyliin ”heikko aste umpitavun edessä”.

CC → C./\_\_VC

Tämä geminaattaklusiileja koskeva sääntö ei kuitenkaan ole puhtaasti fonologinen sääntö, koska omistusliitteet eivät laukaise astevaihtelua. Umpitavusääntö on ollut

toimiva kielihistoriallisesti, koska kantasuomessa sanan toisessa, painottomassa tavussa ei voinut olla pitkää vokaalia. Nykysuomessa umpitavusääntöä joudutaan tarkentamaan ja rajaamaan esimerkiksi muotoilulla ”umpitavussa lyhyen vokaalin edessä”. Avoin kysymys on, pitäisikö myös vahvalle asteelle muotoilla oma sääntö ”vahva aste avotavun edessä”. Näyttää kuitenkin siltä, ettei tälle säännölle ole tarvetta eikä edellytyksiä, koska vahvan asteen täytyy olla leksikkoon merkittyä. Tästä yksinkertaisena osoituksena on, että suomessa on lyhyen ja pitkän klusiilin oppositio (esim. *kato* vs. *katto*), ja sanat täytyy pitää leksikossa erillään. Ainakin historiallisesta näkökulmasta suomessa on ollut ketjuvaihtelua (*kk* : *k* :  $\emptyset$ ; *pp* : *p* : *v*; *tt* : *t* :  $\delta \sim v \sim \emptyset$ ). Näin ollen fonologis-leksikaalinen kontrasti ulottui myös astevaihteluun. Kolmanneksi heikko aste johtimen edellä ei vahvene jouduttuaan avotavuun (esim. sanassa *lepikko* : *lepikossa*, *\*leppikossa*). Samalla tapaan heikko aste ei vahvene juuressa tyypissä *säädytön* : *säädyttömän*, vaan heikko aste säilyy läpi taivutusparadigman. Fonologinen avotavusääntö olisi toki teoreettisesti mahdollinen monitavujen erikoisvaihtelun kuvauksessa, koska suomesta puuttuu sanat tyyppiä *\*haraka*, *\*ullako*.

Astevaihtelun laukaisijana on perinteisesti pidetty tavun umpinaisuutta. Karlssonin (1982, 323) mukaan radikaalissa astevaihtelussa heikko aste on umpitavun edessä. Karlsson (1982, 336) tosin mainitsee, että tavun painollisuutta voi yhtä hyvin pitää astevaihtelun kriteerinä monitavujen vaihtelussa. Kuten on tunnettua, ainoastaan umpitavu laukaisee radikaalin astevaihtelun. Monitavuissa heikko aste voi olla valinnaisesti myös *i*-loppuisen diftongin edellä.

*takka* : *takan* : *takkaa* : *takkoina*

*harakka* : *harakan* : *harakkaa* : *harakkoina* ~ *harakoina*

Terho Itkonen (1957, 267) toteaa näitä käsitellessään, että ”*Eri tutkimuksissa joskus nähdyt viittaukset siihen, että i-loppuinen diftongi olisi jossain murteissa käsitetty tavua sulkevaksi, sisältää pikemminkin vain asiantilan toteamuksen kuin selityksen*”. Jos astevaihtelu laukaisevaksi tekijäksi tulkitaan Itkosen tapaan paino, selittävät samat tekijät sekä radikaalin astevaihtelun että monitavujen vaihtelun. Tällöin astevaihtelu on sensitiivinen vain tavun pituuden suhteen.

Käänteisellä astevaihtelulla tarkoitetaan puolestaan vaihtelua, jossa nominatiivin heikko aste vahvenee muissa sijoissa, esim. genetiivissä ja obliikvisijoissa (*laude: lauteet, varvas: varpaat*). Käänteistä astevaihtelua voi pitää selvästi leksikaalistuneena. Tämä näkyy siinä, että lainasanat ja muut uudet sanat eivät mukaudu käänteiseen vaihteluun. Käänteisen astevaihtelun kuvauksen ongelmana on se, että heikkoasteisessa muodossa tulisi odottaa suffiksillista muotoa (Karlsson 1982, 332). Käänteisessä astevaihtelussa olevissa sanoissa heikko aste on kuitenkin nominatiivissa ja vahva aste esimerkiksi genetiivissä ja ulkopaikallissijoissa. Karlssonin (1982, 332) mukaan suomen nominatiivi on kuitenkin tulkittava päätteettömäksi. Karlsson esittää edelleen, että käänteisen astevaihtelun kuvaus on sukua muille vartalonmuodostussäännöille, kuten vahvan vokaalivartalon muodostamiselle (*raide* → *raitee-*).

Karlssonin (1982, 197) mukaan sanan erityyppiset vartalot eivät ole keskenään hierarkkisesti samanarvoisia, ja nominatiivi on tulkittava niistä ensisijaiseksi. Tätä puoltaa myös se, että nominatiivi on yleisin sijamuoto ja siten helposti yleistettävä esimerkiksi kielenomaksumisessa. Lapsi muodostaa nominatiivista virheellisiä muotoja (esimerkiksi *käsit* 'kädet'), ja tämä on tulkittavissa todisteeksi siitä, että lapsi käyttää nominatiivia taivutuksen lähtömuotona myös sellaisissa tapauksissa, joihin kuuluu jokin muu taivutusvartalo. Tämäntyyppisiä muotoja lapsi ei ole oppinut ympäristöstään, vaan hän on muodostanut ne itse yleistämällä nominatiivin lähtömuodoksi myös vartaloihin, joihin se ei kuulu.

Käänteisessä astevaihtelussa olevissa sanoissa sananloppuinen *-e'* pitenee obliikvi-muodoissa. Tämä saattaa toimia (ikonisena) merkkinä käänteisestä astevaihtelusta (esim. *lähde': lähteen*). *e'*-loppuisissa nomineissa on olemassa käänteisessä vaihtelussa olevan *e'*-tyypin lisäksi myös *e'*-loppuinen taivutustyyppi, jossa ei ole vokaalinpidennystä, eikä astevaihtelua (esim. *nalle: nallen*). Mikäli pidentyvää loppu-*e*:tä pidetään merkkinä astevaihtelusta, niin voi ajatella, että jos vokaalinpidennystä ja astevaihtelua ei esiinny, on tapahtunut paradigmaattinen siirtymä *e'*-tyypistä tunnusmerkittömään *e*-tyyppiin. Karlsson (1982, 364) toteaa, että paradigmaattisia

suhteita esiintyy sekä sanamuotojen sisällä että eri paradigmojen välillä. Tätä klas-  
sista paradigman mallia Karlsson kutsuu pintaparadigmaksi.

Pintaparadigmoissa morfologiset ja fonologiset säännöt liittyvät makrosäännöiksi,  
jotka ovat paradigmaattisessa suhteessa useisiin pintaparadigmoihin. Karlsson  
(1982, 364) esittää, että pintaparadigmojen reaalisuudesta on useita näyttöjä esimer-  
kiksi kielen omaksumisessa. Lapsi voi mukauttaa uuden sanan joko *häive'*- tai *nalle*-  
sanojen taivutusparadigmaan. Koska *nalle*-tyypin sanoissa ei ole vokaalinpidennystä,  
siihen mukautettavat sanat eivät sisällä tietoa käänteisestä astevaihtelusta.

Toisaalta taas sanojen paradigmat ovat pintavaikutussyhteyksien vuoksi sekaantuneet  
paikoin siten, että on olemassa sanoja, jotka ovat loppukahdenteisia (kuten  
*kolmek:in*), vaikeivät olekaan tyypillisiä *e'*-sanoja. Suomen kielessä on useita  
kymmeniä sekaparadigmoja, joissa tämän kaltaista sekaantumista on tapahtunut  
(esim. *kannel : kantelen ~ kantele : kanteleen, manner : manteren ~ mantereen*). Se  
on osoitus siitä, että tällaisissa paradigmoissa on horjuntaa, joka voi osaltaan vaikut-  
taa käänteisen astevaihtelun epäproduktiivistumiseen. Nykysuomen *e'*-sanat ovat  
hyvä esimerkki siitä, miten diakroniset suhteet ovat usein sotkeutuneet ja moni-  
mutkaistuneet niin, ettei niitä synkronisesta näkökulmasta kyetä enää yksiselitteisesti  
erottelemaan toisistaan.

Hypoteettisesti voi ajatella, että käänteisen astevaihtelun lähtömuotona ei olekaan  
yksikön nominatiivi, vaan jokin muu sija, josta käänteinen vaihtelu tulisi yksin-  
kertaisemmin johdetuksi. Monikon nominatiivi (esim. *lauteet*) tuntuu olevan monessa  
tapauksessa tämäntyyppisten sanojen lähtömuoto, ja ne näyttävätkin esiintyvän  
usein monikkomuotoisina (vrt. esim. Leivo 1984). Tämänkaltaista tulkintaa tukee  
lisäksi se, että useita käänteisessä astevaihtelussa olevia sanoja käytetään primaa-  
risti monikossa (esim. *lauteet, varpaat, pyyteet*). Tämä on myös käänteisen aste-  
vaihtelun yksinkertaisin ja relevantein selitys, koska siinä vaihtelunalainen konso-  
nantti tai konsonanttiyhtymä käyttäytyy samalla tavalla kuin normaalissa aste-  
vaihtelussa. Ei ole myöskään mitään perusteita olettaa *a priori*, etteikö käänteisen  
vaihtelun kaltaisilla harvinaisilla morfologisilla ilmiöillä voisi olla edellisen mukaista  
tulkintaa. Kyseisen tulkinnan voisi ajatella saavan tukea muista vartalonmuo-



Syötös: /harakka+i+ta/

Juuritaso: há.rak.ká (juuritason sanapaino)  
(hára).(koi.ta)

Sanataso :

Postleks. taso:

Tuotos: /harakoita/

Syötös: /harakka+i+ta/

Juuritaso: há.rak.ka  
(ha.rak).ko.ja

Sanataso : (ha.rak).(ko.ja) (sanatason paino)

Postleks. taso:

Tuotos: /harakkoja/

Kiparskyn (2003) ratkaisu on elegantti, koska se pystyy selittämään variaation monitavuisten sanojen taivutuksessa prosodisilla ja fonologisilla tekijöillä. Kaksitavuisissa sanoissa variaatiota ei esiinny, koska suomen kiinteä sanapaino estää variaation. Seuraavassa luvussa käsittelen Kiparskyn mallia tarkemmin.

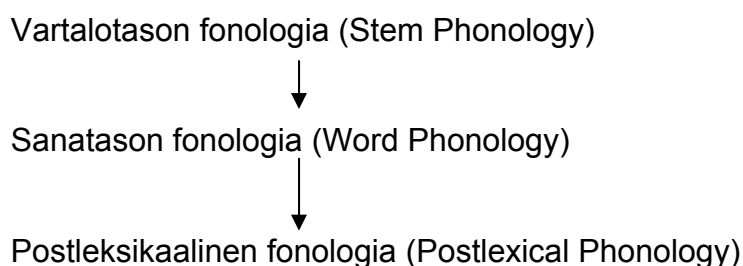
## 5.2 Leksikaalinen fonologia ja astevaihtelu

Leksikaalisessa fonologiassa kielioppi jaetaan kahtia leksikaaliseen ja postleksikaaliseen moduuliin. Leksikaalinen fonologia (esim. Pesetsky 1979; Kiparsky 1982) sisältää seuraavat väitteet kieliopin rakenteesta (K. P. Mohanan 1995, 59):

- a) Vain leksikaaliset säännöt voivat viitata sanansisäisiin rakenteisiin.
- b) Vain postleksikaaliset säännöt voivat vaikuttaa sanojen läpi.
- c) Vain leksikaaliset säännöt voivat olla syklisiä.
- d) Leksikaaliset säännöt ovat rakennetta suojaavia ("Structure preserving").
- e) Vain leksikaalisissa säännöissä voi olla leksikaalisia poikkeuksia.
- f) Leksikaaliset säännöt sovelletaan ennen postleksikaalisia sääntöjä.

Kohdissa a ja b on leksikaalisen fonologian ydinajatus. Kohta c on osittain epäselvä, ja sen toimivuudesta eivät leksikaalisen fonologian eri haarat ole päässeet yhteisymmärrykseen. Kohta e on ongelmallinen siksi, että vaikka yleensä näyttää siltä, että leksikkoon sidottu informaatio sisältää poikkeuksia enemmän kuin puhtaasti fonologiset säännöt, on sitä vaikea nostaa formaalin kielenkuvauksen tasolle. Sama kritiikki pätee myös viimeiseen kohtaan.

Uusin sovellus leksikaalisesta fonologiasta lienee Kiparskyn (2003) kuvaus suomen monikon genetiiveistä. LPM-OT:ssa (leksikaalisessa optimaalifonologiassa) fonologiset ja morfologiset syklit jakautuvat kolmeen itsenäiseen osaan, joissa kaikissa sekä uskollisuus- että tunnusmerkkisyysrajoitukset voivat järjestäytyä toisistaan riippumattomasti (esim. Kiparsky 2003, 109). Jokainen taso ”ruokkii” seuraavaa tasoa siten, että vartalotason tuotos on sanatason syötös ja sanatason tuotos on edelleen postleksikaalisen tason syötös.



Kuvio 30. Leksikaalinen optimaaliteoria.

Koska fonologiset syklit seuraavat toisiaan, on mahdollista, että seuraava sykli peittää edellisessä syklissä tapahtuneet fonologiset (tai morfologiset) prosessit. Seuraavassa on esimerkki siitä, miten postleksikaalinen fonologia ”kätkee” sanatason fonologiset prosessit alleen näkymättömiin (ks. myös Anttila 1995 monikon genetiiveistä). Se on samalla esimerkki siitä, miten muotojen valinta ja fonologisten ja foneettisten segmenttien pituus ja kesto (geminaatta /jj/ Suomessa ei ole oppositiossa /j/:n kanssa) riippuu monimutkaisista morfofonologisista rajoituksista ja niiden järjestyksestä. Esimerkki siis kuvaa myös, kuinka samat säännöt ja rajoitukset koskevat niin ”klassisen” IA-fonologian foneemeja kuin (allo)foneja. Tarkastelen tässä myös



morfonologiaan kuuluvan astevaihtelun ja tavun pituuden välistä suhdetta optimaalisuusteorian valossa.

Seuraavassa taulukossa 27 sanatason rajoitushierarkia \*L' >> \*LAPSE generoi voitta-jaksi kandidaatin 1 (harakkoja), joka rikkoo sanatasolla alempana hierarkiassa olevaa rajoitusta \*LAPSE vastaan. Kandidaatti 2 (ha.rak).(ko.ja) häviää, koska se rikkoo sanatasolla rajoitusta \*L' vastaan. (taulukosta 27 on tietenkin jätetty pois rajoitukset, jotka eivät ole adekvaatteja tässä esimerkissä.)

Taulukko 27. Sanatason rajoitukset.

/harakka-i-ta/	*L'	*LAPSE	
1)→(ha.rak).ko.ja		*	
2)(ha.rak).(ko.ja)	*		

\*L': Ei lyhyitä painollisia tavuja.

\*LAPSE : Painoton tavu on painollisen tavun vieressä.

Postleksikaalisella tasolla taas sanatason tuotos toimii syötöksenä (harakkoja). Postleksikaalisella tasolla rajoitus \*LAPSE on hierarkiassa korkeammalla kuin rajoitukset \*H/O, \*O' ja \*L'. Tästä on todisteena yleisgeminatio tapauksessa (ha.rak).(ko.jaan). Tästä seuraa se kauaskantoinen johtopäätös, että postleksikaalisella tasolla rajoitus \*LAPSE on aina suomessa adekvaatti voittajaa generoitaessa.

Taulukko 28. Postleksikaalisen tason rajoitukset.

/(ha.rak).ko.ja/	*LAPSE	*CV'.CVV	*CVC.CVV
(ha.rak).ko.jaan	*!		
(ha.rak).(ko.jaan)		*	
→ (ha.rak).(ko.jaan)			*

\*LAPSE: Painoton tavu on painollisen tavun vieressä.

\*CV'.CVV: Ei rakenteita, jossa on lyhyt painollinen tavu ja pitkä painollinen tavu.

\*CVC'.CVV: Ei kahta peräkkäistä pitkää tavu.

Postleksikaalisen tason painorajoituksissa rajoitus \*LAPSE on siis ”kiinteästi” prominenssirajoitusten \*H/O, \*O' ja \*L' yläpuolella. Rajoitus \*LAPSE on tulkittava seuraavasti ”painoton tavu on painollisen vieressä”.

SANATASO:(FtBIN >> LEFT-HEAD >> ALIGN-LEFT >>\*CLASH >> \*L', \*LAPSE)

POSLEKSIKAALINEN TASO: \*LAPSE >> \*L' >> \*CV'CVV >> \*CVC'CVV

Kuvio 31. Sanatason ja postleksikaalisen tason painohierarkia.

Kuten yllä olevasta Kuvioista 5 ilmenee, ovat rajoitukset \*L', \*LAPSE sanatasolla vapaassa vaihtelussa. Jos hierarkia on \*L' >> \*LAPSE, voittaa kandidaatti (ha.ra).(koi.ta), koska kandidaatti (ha.rak).(ko.ja) rikkoo rajoitusta \*L' vastaan. Kandidaatti (ha.rak).ko.ja rikkoo rajoitusta \*LAPSE vastaan ja kandidaatti (ha.ra).koi.ta rajoitusta \*LAPSE vastaan.

Ongelmana on, lisätäänkö omistusliite sanatasolla vai postleksikaalisella tasolla. Jos omistusliitteet liitetään sanatasolla, on oletettava, että rajoituksen \*LAPSE yläpuolella on rajoitus, joka kieltää sanatasolla painolliset lyhyet tavut. Merkitsen tätä rajoitusta tässä rajoituksella \*L' ”ei lyhyitä painollisia tavuja”. Voittaja on siis kandidaatti (ha.ra).(koi.taan).

Taulukko 29. Sanataso.

/harakka-i-ta-an/	*L'	*LAPSE
→(ha.rak).ko.jaan		*
(ha.ra).(koi.taan)		

\*L': Ei lyhyitä painollisia tavuja.

\*LAPSE: Painoton tavu on painollisen tavun vieressä.

Postleksikaalisella tasolla painorajoituksen \*LAPSE on oltava hierarkiassa rajoituksen \*L' yläpuolella, koska muuten yleisgeminatio ei olisi mahdollinen (yleisgeminatio vaatii lyhyen painollisen tavun ja sitä seuraavassa tavussa on oltava pitkä vokaali). Astevaihtelu on siis sanatason ilmiö, koska muuten ei voida selittää sitä, miksi ei ole heikkoasteisia geminoituneita muotoja tyyppiä *\*harakojjaan*. Näin ollen dissimilaatiota koskevat rajoitukset \*HH, \*LL ovat postleksikaalisella tasolla hierarkiassa alhaalla eivätkä vaikuta enää postleksikaalisen tason ilmiöihin.

Taulukko 30. Postleksikaalinen taso.

/(ha.rak).ko.jaan/	*LAPSE	*L'	*CV'.CVV	*CV'C.CVV
(ha.rak).ko.jaan	*			
(ha.rak).(ko.jaan)		*	*	
(ha.rak).koj.jaan	*			
→ (ha.rak).(koj.jaan)				*

\*LAPSE: Painoton tavu on painollisen tavun vieressä.

\*L': Ei lyhyitä painollisia tavuja.

\*CV'.CVV: Ei rakenteita, jossa on lyhyt painollinen tavu ja pitkä painollinen tavu.

\*CVC'.CVV: Ei kahta peräkkäistä pitkää tavu.

Tässä tapauksessa mahdollinen hierarkia postleksikaalisella tasolla on siis seuraavanlainen:

\*LAPSE >> \*L' >> \*CV'.CVV >> \*CV'C.CVV >> \*HH

### 5.3 Astevaihtelun kuvaus DF:ssa

Deklaratiivinen fonologia on konkreettista fonologiaa siinä suhteessa, että fonologiset vaihtelut kuvataan disjunktioina. Av:ssa heikkoa astetta ei voida kuvata säännöllä, joka johtaa heikon asteen vahvasta asteesta kuten fonologisissa teorioissa, jotka sallivat johto-operaatiot (esim. generatiivinen fonologia). Ajatellaan, että seuraavat ehdot riittävät kuvaamaan astevaihtelua.

- 1) geminaattojen vaihtelu on produktiivista.
- 2) kvalitatiivinen vaihtelu on epäproduktiivista leksikaalista vaihtelua.

Lisäksi voidaan olettaa olevan seuraavanlaisia morfofonologisia ympäristöjä (X tarkoittaa mitä tahansa segmenttiä tai segmenttiryhmää; K = klusiili):

- 1) XK.KV
- 2) XK.KVV(C)
- 3) X.KV
- 4) X.KVC
- 5) X.KVV(C)

Malli A (Kvalitatiivinen ja kvantitatiivinen vaihtelu säännönmukaisina)

Joudutaan kirjoittamaan yksi sääntö, jonka mukaan geminaatta tai yksinäisklusiili edustuu yksinäisklusiilina umpitavun edessä. Kvalitatiivinen AV merkitään tarkempana disjunktiona, koska heikkoasteisia konsonantteja on useita. Ongelmaksi tulee, miten kuvata kvalitatiivisen astevaihtelun ulkopuolella olevia sanoja. Ainoa vaihtoehto on olettaa, että nämä sanat on merkitty erikseen leksikkoon vaihtelemattomina (sääntö c).

- a) p U v/ \_\_VC  
 t U d/ \_\_VC  
 k U Ø/ \_\_VC  
 k U v/[+lab V]\_\_ [+lab V]C
- b) [+gem] U [-gem]/ \_\_VC
- c) [-gem] U [-gem]/ \_\_VC

Malli B (Kvalitatiivinen AV rajoituksin merkittynä)

Koska deklaratiiivinen fonologia on ennustava teoria, voi olettaa, että kvalitatiivisen astevaihtelun ei tarvitse olla leksikkoon merkittyä: valtaosa suomen yksinäisklusiilisista sanoista on kuitenkin edelleen kvalitatiivisessa astevaihtelussa. Geminaatta on merkitty erillisenä piirteenä [+gem].

- a) [+gem] U [-gem]/ \_\_VC
- b) p U v/ \_\_VC  
 t U d/ \_\_VC  
 k U Ø/ \_\_VC  
 k U v/[+lab V]\_\_ [+lab V]C

Malli C (Geminaattaklusiilit lyhyinä klusiileina)

Ainakin teorian tasolla voidaan ajatella seuraavanlaistakin kuvausta. Leksikkoon geminaatat on merkitty yksinäiskonsonanteina, jotka pitenevät geminaatoiksi avotavussa. Kvalitatiivinen vaihtelu on automaattista, mutta osa leksikosta on merkitty vaihtelun ulkopuolelle (sääntö c).

- a) [-gem] U [+gem]/\_\_V(VX)
- b) p U v/\_\_VC  
t U d/\_\_VC  
k U Ø/\_\_VC  
k U v/[+lab V]\_\_ [+lab V]C
- c) [-gem] U [-gem]/\_\_V(VX)

## 6. Geminaattojen funktionaalisesta kuormituksesta

### 6.1 Ensimmäisen jalan rakenteiden käyttöfrekvenssit

Tunnusmerkkisyyden käsite on alun perin peräisin Prahan koulukunnassa vaikuttaneilta fonologeilta, N. S. Trubetzkoylta ja Roman Jakobsonilta. Teoriassa fonologiset oppositiot jakoutuivat multilateraalsiin ja bilateraalsiin oppositioihin (Lauerma 1993, 3). Trubetzkoy'n ajattelussa tunnusmerkkisyys kytkeytyy fonologisten oppositioiden neutraalistumiseen, joten tunnusmerkkisyys jää vain osaa foneemioppositioista koskettavaksi kielikohtaiseksi ilmiöksi (emt., 4). Jakobson epäili kuitenkin jo 1930-luvulla multilateraalisten oppositioiden olemassaoloa. Fonologiassa alettiin purkaa tai järjestää multilateraalisia oppositioita binaarisiksi oppositioiksi. Jakobson (1941) huomasi lapsen kieltä ja afasiaa tutkiessaan, että samat distinktioltaan selvimmät piirteet, jotka ilmaantuivat lapsen kieleen ensimmäisinä, hävisivät myös afaatikoilta viimeisenä. Jakobson oivalsi, että puheen oppiminen on selkeän hierarkkista: ensimmäiseksi kehittyvät distinktioltaan selvimmät oppositiot ja monimutkaisemmat oppositiot kehittyvät aste asteelta yksinkertaisimpien oppositioiden pohjalta. (Lauerman 1993, 5 mukaan).

Haspelmathin (2006, 16-21) mukaan kielellinen tunnusmerkkisyys olisi korvattava muilla käsitteillä, kuten tuottamisen ja prosessoinnin helppoudella, käytön korkeafrekvenssisyydellä, distinktisyydellä ja selittävyydellä. Tuottamisen helppouden kautta tunnusmerkkisyysteorian tunnusmerkkiset rakenteet. Tuottamisen helppouden kautta voidaan myös selittää paradigmojen samankaltaisuus. Korkeafrekvenssisyys on yksi Haspelmathin (2006, 16) tunnusmerkkisyyden korvaajiksi ehdottamista parametreista. Korkeafrekvenssisten muotojen säilymistä pidetään myös yhtenä kielen muutoksen universaaleista, joka selittyy Bybeen (1985, 119) mukaan suurtaajuuteen liittyvällä leksikaalisella vahvuudella. Suurtaajuinen kategoria muistetaan paremmin, koska suurempi esiintymäfrekvenssi kasvattaa sen leksikaalista vahvuutta. Lopulta jos kategorioiden distinktio häviää, säilyy suuritaajuisempi kategoria.

Frekvenssi voidaan jakaa kahteen pääluokkaan, käyttöfrekvenssiin ja tyyppi-  
frekvenssiin. Haspelmathin (2006, 16) mukaan käyttöfrekvenssi on yksi tärkeimmistä  
tunnusmerkkisyyden käsitteen korvaajista, koska perinteisesti tunnusmerkkisinä pide-  
tyt rakenteet ovat myös kielessä harvemmin käytettyjä. Frekventtisyyden toinen laji,  
tyyppifrekvenssi, tarkoittaa leksikossa taivutustyyppiin kuuluvien sanojen määrää.  
Tyyppifrekvenssi vaikuttaa morfologiseen prosessointiin siten, että pientaajuiset  
muodot prosessoidaan pikemmin sääntöjen kuin ulkomuistin avulla ja suurtaajuiset  
ulkomuistin perusteella (Bybee 2001, 28). Korkea tyyppifrekvenssi voidaan yhdistää  
produktiivisuuteen, mutta tämä sidos ei ole mutkaton ja tarvitsee tuekseen todisteita.  
Konnektionistiset mallit perustuvat juuri tyyppi- ja käyttöfrekvenssin yhteisvaiku-  
tukseen.

Tässä luvussa esittelemäni sanatyyppejä koskevat käyttöfrekvenssitiedot on laskettu  
Suomen Kuvalehden (SK) vuosikerrasta 1999. Tämän lisäksi olen laskenut osittain  
samat tiedot Pentti Holapan teoksesta *Kansliapäällikkö*, joka on myös sähköisenä  
aineistona. Edellä mainitut aineistot kuuluvat osana laajempaa sähköistä aineistoa  
Tieteen tietotekniikan keskuksen Kielipankin suomenkieliseen aineistoon  
(<http://www.csc.fi/tutkimus/alat/kielitiede>). Frekvenssitiedot on koottu datasta, jonka  
laajuus on noin 57 000 kirjoitettua sanaa. Näistä aineistoista on laskettu työssäni  
käyttämäni käyttöfrekvenssiluvut. Käyttöfrekvenssin lisäksi olen laskenut *Nyky-  
suomen sanakirjan* (NS) yhdistämättömistä sanoista (noin 65 000 sanaa) sana-  
tyyppien ns. tyyppifrekvenssit, joihin palaan seuraavassa luvussa.

Yleensä fonologiassa ei toimita tilastollisin perustein vaan vanha strukturalistinen  
sääntö on ”kerran oppositio aina oppositio”. Mutta kun tarkastellaan suoritusta, luke-  
mista ja kirjoitusta, tällöin tilastolliset asiat kuten ilmiön yleisyys nousevat tärkeiksi.  
Luvuissa on mukana myös kaksitavuisia sanoja pidemmät sanat, mutta näistä on  
otettu mukaan laskuihin ainoastaan ensimmäisen tavun ja toisen tavun rakenne  
toisen tavun ensimmäiseen vokaaliin asti. Näin on tehty siksi, että suuria teksti-  
aineistoja käsiteltäessä oli teknisesti erittäin vaikeaa etsiä frekvenssilukuja ensitavua  
kauempaa sanasta, koska ensi- ja toisen tavun rajaa kauempana Unix-pohjainen  
haku ei pysty erittelemään, mitkä sanamuodot ovat kyseessä.



Kuten alla olevista luvuista näkyy, yleisin kirjoitetun kielen kaksitavuinen sanarakenne on CV.CV(X), jonka yleisyyteen vaikuttanee se, että klusiilit ovat juoksevassa tekstissä usein heikossa asteessa.. Toiseksi yleisin jalkatyyppi on CVCCV, joista suurin osa on rakenteeltaan tyyppiä CVC<sup>2</sup>C<sup>3</sup>V(X). Rakenne, jossa on pitkä avotavu ensimmäisenä tavuna (CVV.CV(X)) on kolmanneksi yleisin tyyppi. Harvinaisimpia ovat ylipitkät tavut (yhteensä 10,3 % kaikista tavuista). Aineistosta selviää, että lyhyemmät rakenteet ovat yleisempiä kuin pitkät rakenteet. Toisin sanoen lyhyet ja pitkät pääpainolliset tavut ovat yleisempiä kuin ylipitkät tavut.

Taulukko 31. Ensimmäisen tavun ja toisen tavun rakenne toisen tavun ensimmäiseen vokaaliin asti (SK).

Tyyppi	N	%
(C)V CV	21 313	37,2
(C)VC CV	18 529	32,3
(C)VV CV	11 581	20,2
(C)VVC CV	4 615	8,1
(C)VCC CV	1 279	2,2
Yhteensä	57 317	100

Seuraavassa on vertailun vuoksi laskettu edellä mainitut samat pääpainollisen tavun frekvenssityypit Pentti Holapan teoksesta *Kansliapäällikkö* (1996). Sanatyyppien frekvenssit ovat yhteneviä asiatyylisen tekstin (Suomen Kuvalehti) kanssa. Kaiken kaikkiaan geminaattoja on Holapan tekstissä kaikista sanatyypeistä laskien 15,4 prosenttia. Holapan teoksen valitsin vertailuaineistoksi, koska se oli helposti saatavilla sähköisessä muodossa Tieteen tietotekniikan keskuksen aineistoista ([www.csc.fi](http://www.csc.fi)).

Taulukko 32. Ensimmäisen tavun ja toisen tavun rakenne toisen tavun ensimmäiseen vokaaliin asti (Holappa).

Tyyppi	N	%
(C)V CV	15 729	37,8
(C)VC CV	13 330	32,0
(C)VV CV	8 420	20,2
(C)VVC CV	3 635	8,7
(C)VCC CV	509	1,2
Yhteensä	41 623	100

Geminaattoja on edellä mainituissa rakenteissa seuraavasti (Suomen Kuvalehti 1999). Geminaattojen yhteenlaskettu osuus kaikissa sanatyypeissä on 14,6 %. Yleisimpiä geminaattoja ovat dentaaligeminaatat *///* ja *tt/*. *///*:n yleisyys selittynee ainakin osaksi sillä, että *///* kuuluu adessiivin ja ablatiivin sijapäätteeseen (vrt. Taulukko 33 alla) eli ne esiintyvät tietyssä morfologisessa asemassa. Toisaalta, frekvenssiltään dominoivien kaksitavuisten vartaloiden jälkeen ne esiintyvät juuri poiskatkaistussa osassa, kolmannen tavun rajalla, ja ensitavun lyhyen vokaalin jälkeen ei ole juuri koskaan suffiksia, pitkän vokaalin jälkeen joskus.

Taulukko 33. Geminaattojen määrät ja suhteelliset osuudet eri asemissa 1. ja 2. tavun rajalla (SK).

Yhtymä	(C)VCCV...	(C)VVCCV...	(C)VCCCV...
{kk}	245	632	349
{pp}	163	161	53
{tt}	<u>1471</u>	<u>1120</u>	182
{ss}	246	150	149
{mm}	290	20	-----
{nn}	869	336	-----
{ll}	<u>1283</u>	<u>402</u>	-----
{rr}	200	42	-----
			-----
Yhteensä	4767	2863	733
%	25,7 %	62,0	57,3 %

Tarkasteltaessa tarkemmin juoksevan tekstin sanarakenteita huomataan, että valtaosassa sanoista (78,4 %) on joko toisen tavun tavunalkuinen yksinäiskonsonantti tai tavunrajainen kahden konsonantin yhdistelmä. Geminaatalliset sanarakenteet tyyppiä (C)VCCV muodostavat 12,2 prosenttia kaikista ensimmäisen ja toisen tavun välisistä yhtymistä. Edelleen ylipitkissä tavuissa olevat geminaatat ovat kaikkein harvinaisimpia.

Taulukko 34. Geminaattojen ja yksinäiskonsonanttien määrät ja frekvenssit ensimmäisen ja toisen tavun rajalla (SK).

Grafeemi	(C)V <u>CCV</u>	(C)Vx <u>CCV</u>	CV(X).CV	Yhteensä
{k/	245 (3,9 %)	981 (15,6 %)	5072 (80,5 %)	6298
{p}	163 (6,9 %)	214 (9,1 %)	1976 (84,0 %)	2353
{t}	1471 (17,0 %)	1302 (15,0 %)	5890 (68,0 %)	8663
{s}	246 (5,8 %)	299 (7,1 %)	3680 (87,1 %)	4225
{m}	290 (9,3 %)	20 (0,6 %)	2797 (90,0 %)	3107
{n}	869 (22,7 %)	336 (8,8 %)	2628 (68,6 %)	3833
{l}	1283 (18,2 %)	402 (5,7 %)	5359 (76,1 %)	7044
{r}	200 (5,9 %)	42 (1,2 %)	3168 (93,0 %)	3410
Yhteensä	4767	3596	30 570	38 993
%	12,2	9,2	78,4	100

Ensimmäisen tavun tavutyypit jakaantuvat kokonaisuudessaan seuraavasti. Yleisin pääpainollisen tavun tyyppi on lyhyt tavu (C)V (37,2 %) ja toiseksi yleisin tavutyyppi (C)VC1 (24,2 %). Kolme yleisin tavutyyppiä (C)V, (C)VC1 ja (C)VV kattavat kaikista tavutyypeistä 81,4 prosenttia. Geminaatat on seuraavassa taulukossa alleviivattu.

Taulukko 35. Ensimmäisen ja toisen tavun rajalla olevat yhtymien määrät järjestettynä niiden käyttöfrekvenssin mukaan (SK).

Tyyppi	N	%
(C)V CV	21 313	37,2
(C)VC <sup>1</sup>  C <sup>2</sup> V	13 762	24,0
(C)VV CV	11 581	20,2
(C)V <u>C</u> <sup>1</sup>   <u>C</u> V	4 767	8,3
(C)VVC <sup>1</sup>   <u>C</u> V	2 863	5,0
(C)VVC <sup>1</sup>  C <sup>2</sup> V	1 752	3,1
(C)VCC <sup>1</sup>   <u>C</u> V	733	1,3
(C)VCC <sup>1</sup>  C <sup>2</sup> V	546	1,0
Yhteensä	57 317	100

Geminaattoja oli konjunktioissa ja partikkeleissa siten, että klusiiligeminaatoista konjunktioita tai partikkeleita (*että, jotta, vaikka*) oli 30,8 %, ja soinnillisista geminaatoista partikkelien osuus oli 13,4 %. Samoin geminaatta-[s]:n osuus on 28,8 prosenttia. Jos partikkelit ja konjunktiot vähennetään kaikkien geminaattojen määrästä, jää klusiiligeminaattojen osuudeksi kaikkiin juoksevan tekstin sanoihin suhteutettuna noin 5,3 prosenttia. Yhteensä geminaattojen osuudeksi jää noin 11,2 prosenttia.

## 6.2 Pitkien ensitavun ja toisen tavun vokaalien käyttöfrekvenssi

Pitkistä vokaaleista pääpainollisen tavun pitkät vokaalit ovat melko frekventtejä: 20,6 prosenttia ensitavun vokaaleista on pitkiä vokaaleja. Tämä näkyy myös luvussa 7 esittelemäni kirjoitustestin tuloksissa siten, että painolliset pitkät vokaalit hallitaan kaikista pitkistä segmenteistä parhaiten.

Taulukko 36. Pääpainollisen tavun pitkät vokaalit ja niiden käyttöfrekvenssi (SK).

Sekvenssi	Yhteensä
(C)aa	1363
(C)ee	147
(C)ii	1665
(C)oo	117
(C)uu	1308
(C)yy	159
(C)ää	936
(C)öö	4
Yhteensä	5699 (20,6 %)
(C)V	21 910 (79,4 %)

Juoksevan tekstin toisessa tavussa pitkät vokaalit ovat harvinaisia. 99,9 prosenttia aineis-tossa olevista kaksitavuisista sanoista, joissa oli toisessa tavussa pitkä vokaali, oli yksikön partitiiveja (*takkaa*, *pakkaa* etc.). 6,3 prosentissa kolmitavuisista sanoista on pitkä toisen tavun vokaali (ks. Taulukko 37).

Taulukko 37. Vokaalien kaksoisgrafeemien määrät ja niiden käyttöfrekvenssi toisessa tavussa juoksevassa tekstissä (SK).

sekvenssi	yhteensä
S.Caa.CX	626
S.Cee.CX	387
S.Cii.CX	298
S.Coo.CX	250
S.Cuu.CX	242
S.Cyy.CX	37
S.Cää.CX	134
S.Cöö.CX	4
yhteensä	1978 (6,3 %)
S.CV.CX	29 203 (93,7 %)

S = ensimmäinen tavu.

X = mikä tahansa segmentti tai useampia segmenttejä.

(CV + CVV = 31 181)

Useamman pitkän kvantiteetin yhdistelmiä on juoksevassa tekstissä niukalti. Taulukossa 38 on laskettu juoksevasta tekstistä (Suomen Kuvalehti 1999) sanat, joissa on kahden pitkän segmentin yhtymä (siis esim. tyytit *taakka, saakka, tyyppi* jne.).

Taulukko 38. Ensimmäisen tavun pitkän vokaalin ja tavarajan geminaatan yhdistelmien määrät juoksevassa tekstissä (SK).

Sekvenssi	Yhteensä
(C)VVkkV...*	103
(C)VVppV...	62
(C)VVttV...	411
(C)VVssV...	55
(C)VVmmV...*	10
(C)VVnnV...	152
(C)VVIIV...	122
(C)VVrrV...	41
Yhteensä	956

Sanamuotoja (\* 4 saamme), (\*16 saakka) oli näistä 4 ja 16 kappaletta.

Juoksevassa tekstissä on vain vähän, 3,2 prosentissa kaikista yhtymistä, erilaisia yhden ja kahden pitkän kvantiteetin yhdistelmiä ensimmäisen ja toisen tavun rajalla. Seuraavissa luvuissa on ilmoitettu ainoastaan ne sanat, jotka koostuvat lyhyestä tai pitkästä vokaalista ja lyhyestä tai pitkästä konsonantista. Kahden pitkän segmentin yhtymät ovat selvästi harvinaisin sanatyyppi (3,2 % kaikista sanatyypeistä).

Taulukko 39. Pitkien ja lyhyiden kvantiteettien yhdistelmät (SK).

Sanatyyppi	N	%
(C)V.CV	21 910	74,0
(C)VC.CV	4 767	16,1
(C)VV.CV	1 978	6,7
(C)VVC.CV	956	3,2
Yhteensä	29 611	100

Eri sanatyyppien käyttöfrekvenssin jälkeen siirryn seuraavassa luvussa 6.3 käsittelemään geminaattojen ja vastaavien yksinäiskonsonanttien tyyppifrekvenssiä.



### 6.3 Geminaattojen tyypifrekvenssi

Pitkien ja lyhyiden segmenttien frekvenssitietoja olen laskenut myös *Nykysuomen sanakirjan* (NS) yhdistämättömistä sanoista. *Nykysuomen sanakirjan* aineisto on saatavilla sähköisessä muodossa Kotimaisten Kielten tutkimuskeskuksen aineistoista ([www.kotus.fi/aineistot](http://www.kotus.fi/aineistot)). Sanakirja-aineistosta olen laskenut pitkien ja lyhyiden segmenttien yleisyystietoja siksi, että eri sanatyypin leksikaalisella frekvenssillä voi olla merkitystä lukemaan oppimisessa sanatyypin käyttöfrekvenssin lisäksi. *Nykysuomen sanakirjan* yhdistämättömistä sanoista laskettuna kaksitavuisia sanatyyppiä on seuraavasti. Laskuissa on mukana kaikki sanaluokat (Taulukko 40). Sanakirjasta laskettuna yleisin pääpainollisen tavun rakenne on konsonanttiyhtymään loppuva pitkä tavu (esim. *rinta*, *kanta*, *sompa*). CV.CV-tyyppi on vasta sijalla kolme.

Taulukko 40. Sanatyypin tyypifrekvenssit (NS).

Tyyppi	N	%	Geminaattoja	Geminaattoja %
(C)VCCV	2 503	35,4	733	32,1
(C)VVCV	1 381	19,5	-----	-----
(C)VCV	1 111	15,7	-----	-----
(C)VVCCV	962	13,6	464	50,6
(C)VCCCV	649	9,2	493	79,0
(C)VSCV	473	6,7	-----	-----
Yhteensä	7 079	100	1 690	23,9

Yleisimmät ensimmäisen ja toisen tavun rajalla olevien geminaattojen ja vastaavien toisen tavun tavunalkuisten yksinäiskonsonanttien muodostamat tyypit on NS:sta laskettu siten, että sanansisäinen konsonantti on joko yksinäiskonsonantti tai geminaatta. Klusiilit muodostavat selkeästi soinnillisista konsonanteista eroavan luokan sanatyypin yleisyyden suhteen. Klusiileissa yleisimpiä ovat geminaatat kun taas

soinnilliset konsonantit esiintyvät klusiileja yleisemmin yksinäiskonsonantteina (mukaan kaikki sanaluokat). Klusiileista geminaattoja on 55,3 % ja soinnillisista konsonanteista 37,5 prosenttia.

Taulukko 41. Yksittäiskonsonanttien ja niiden geminaattojen määrät ja tyyppifrekvenssit NS:n sanoissa.

Klusiilit	%	Soinnilliset konsonantit	%
1. (C)VSKKV 456	21,9	1. (C)VVCV 591	56,6
2. (C)VVCCV 346	16,6	2. (C)VCV 422	40,4
3. (C)VCCV 352	16,9	3. (C)VCCV 301	28,8
4. (C)VSCV 323	15,5	4. (C)VVCCV 91 (2*)	8,7
5. (C)VVCV 313	15,0	-----	-----
6. (C)VCV 296	14,2	-----	-----
Yhteensä 2 086	100	1 045	100

S = soinnillinen konsonantti

\*Muut kuin e'-nominit

Klusiilien kvantiteettisuhteet ovat sellaiset, että lyhytkonsonanttisia sanoja on kaikista sanoista 87,0 prosenttia. Pitkä vokaali on ensitavussa 22,0 prosentissa sanoista. Geminaatallisia kaikista alla olevista sanatyypeistä on vain 13,0 prosentissa sanoista. 9,9 prosenttia sanasisäisistä klusiileista on lyhyen vokaalin jälkeen geminaattoja. Pitkän vokaalin jälkeisiä geminaattoja on siis vain 3,0 prosentissa sanoista. Kaiken kaikkiaan lyhyitä klusiileja on 87,0 prosenttia ja geminaattoja ainoastaan 13,0 prosenttia. On kiintoisaa, että soinnillisia geminaattoja on prosentuaalisesti enemmän kuin klusiiligeminaattoja (17,6 %). Syynä lienee, että klusiilit ovat juoksevassa tekstissä usein heikkoasteisessa muodossa, mikä vähentää niiden määrää. Pitkän vokaalin jälkeiset soinnilliset geminaatat ovat harvinaisia (2,0 %). Ne ovat joko heikkoasteisia muotoja, kuten *suunta* : *suunnan* tai partikkeleita (*kiinni*, *tuonne* etc.).

Taulukko 42. Sanatyyppien tyyppifrekvenssit *Nykysuomen sanakirjan* mukaan (sanansisäiset klusiilit).

	/k/	/p/	/t/	N	%
(C)V.CV	5 072	1976	5 890	12 938	68,3
(C)VV.CV	1 192	271	2 057	3 520	18,6
(C)VC.CV	245	163	1 471	1 879	9,9
(C)VVC.CV	103	62	411	576	3,0
Yhteensä				18 931	100

Samoin kuin klusiileissa sanatyyppi CVCV on yleisin myös soinnillisissa konsonanteissa. Ylipitkissä tavuissa soinnilliset geminaatat ovat vain yhden prosenttiyksikön harvinaisempia kuin klusiiligeminaatat ylipitkissä tavuissa (3,0 % vs. 2,0 %). Geminaatat lyhyen vokaalin jälkeen näyttävät olevan yllättäen soinnillisissa konsonanteissa hieman yleisempiä kuin klusiileissa (15,6 ja 9,9 prosenttia).

Taulukko 43. Sanatyyppien, joissa on sanansisäiset soinnilliset konsonantit, tyyppifrekvenssit *Nykysuomen sanakirjan* mukaan.

	/m/	/n/	/l/	/r/	N	%
(C)V.CV	1316	1551	4163	2148	9178	54,2
(C)VV.CV	1481	1077	1196	1020	4774	28,2
(C)VC.CV	290	869	1283	200	2642	15,6
(C)VVC.CV	10	152	122	41	325	2,0
Yhteensä					16 919	100

Kaikista *Nykysuomen sanakirjan* yhdistämättömistä sanoista (yhteensä noin 72 000 sanaa) laskettuna yleisin pääpainollisen tavun rakenne on pitkä umpitavu (C)VC (37,7 % kaikista sanoista), ja toiseksi yleisin tavutyyppi on lyhyt tavu (C)V (29,2 prosenttia kaikista sanoista laskettuna). Pitkiä pääpainollisia vokaaleja on 9,4 prosentissa sanoista ja geminaattoja 14,6 prosentissa. Geminaattojen osuus kaikkiin sanatyypeihin suhteutettuna on siis 14,6 prosenttia.

Taulukko 44. Pääpainolliset tavujen määrät NS:n mukaan.

Tyyppi	Yhteensä		Pitkä vokaali		Geminaatta	
(C)VC.CX	24 503	37,7 %	-----		5 516	22,5 %
(C)V.CX	19 004	29,2 %	-----		-----	
(C)VV.CX	12 941	19,9 %	4107	31,7 %	-----	
(C)VVC.CX	5 784	8,9 %	1 997	34,5 %	2 543	44,0 %
(C)VCC.CX	2 847	4,4 %	-----		1 433	50,3 %
Yhteensä	65 079	100 %	6 104	9,4 %	9 492	14,6 %

Seuraavissa luvuissa ovat mukana kaikki *Nykysuomen sanakirjan* yhdistämättömät perussanat. Laskuissa mukana olevat sanat ovat siis vähintään kolmitavuisia, koska sanat on haettu tietokannasta siten, että sananloppuinen merkki /X/ voi olla mikä tahansa merkki tai mitä tahansa useita merkkejä. Toiseksi viimeisessä sarakkeessa on merkitty sulkuihin rakenne X.CVCC.CVX, joka on marginaalinen. Valtaosa sanoista muotoa X.CVVC.CVX on verbejä, kuten *haaveilla* : *haaveilen*. Yleisimmät tavutyytit toisessa tavussa ovat lyhyt toinen tavu (37,8 % kaikista vähintään kolmitavuisista sanoista), pitkä umpitavu toisessa tavussa (39,4 %). Ylipitkiä rakenteita toisessa tavussa on 10,0 prosentissa sanarakenteista. Pelkästään nomineista lasketuna ylipitkiä toisia tavuja on ainoastaan 0,7 prosentissa sanoista.

Taulukko 45. Yleisimmät sanarakenteet vähintään kolmitavuisissa sanoissa järjestettynä pääpainollisen tavun yleisyyden mukaan sekä toisen tavun rakenteen yleisyyden mukaan (NS).

1. tavu	S.CV.CVX	S.CVC.CVX	S.CVV.CVX	S.CVVC.CV	Yhteensä
1) CV.	6 072	7 057	2 036	1 457 (125)	16 747
2) CVC.	6 122	6 127	2 286	1 586 (120)	16 241
3) CVV.	3 260	4 240	992	833 (21)	9 346
4) CVVC.	1 650	974	545	379 (14)	3 562
5) VC.	1 321	1 219	468	294 (55)	3 357
6) V.	1 161	1 397	277	239 (44)	3 118
7) CVCC.	717	318	288	153 (17)	1 493
8) VV.	248	283	104	98 (6)	739
9) VCC.	178	65	29	16 (6)	294
10) VVC.	122	54	34	24 (1)	235
Yhteensä	20 851	21 734	7059	5 079 (409)	55 132
Yhteensä %	37,8	39,4	12,8	10,0	100

X = mikä tahansa segmentti ja S = ensimmäinen tavu

Sanatyyppien, joissa esiintyy geminaatta, osuus juoksevassa tekstissä on (käyttöfrekvenssi, Suomen Kuvalehti) yhteensä 11,2 prosenttia. Sanatyyppin (C)VC.CV... sanoista geminaatallisia on 8,3 prosenttia suhteutettuna kaikkiin sanatyyppisiin. Ylipitkissä tavuissa geminaattoja oli sanatyyppissä (C)VCC.CV... 1,3 prosenttia kaikista sanoista ja tyyppissä (C)VVC.CV 5,0 prosenttia kaikista sanoista. *Nykysuomen sanakirjan* yhdistämättömistä sanoista laskettuna geminaattoja oli kaiken kaikkiaan 14,6 prosenttia. Sanatyyppissä (C)VC.CV... geminaattoja oli 22,5 prosenttia kaikista tämän sanatyyppin sanoista ja tyyppissä (C)VCC.CV... 50,3 prosenttia. Pitkän vokaaliaineksen jälkeen geminaattoja oli 44,0 prosenttia tämän sanatyyppin sanoista. (C)VC.CV...-sanat muodostivat kaikista NS:n yhdistämättömistä sanoista 37,7 prosenttia. Ylipitkien tavujen osuus kaikista sanoista on NS:ssä 13,3 prosenttia (juoksevassa tekstissä 6,3 prosenttia). Tämän perusteella harvinaisimpia

tavutyyppejä on leksikossa suhteellisesti laskettuna enemmän kuin juoksevassa tekstissä (käyttöfrekvenssi).

#### 6.4 Kolmitavuisten sanatyyppien tyypifrekvenssi

Seuraavaksi siirryn käsittelemään erikseen kolmitavuisten sanojen tyypifrekvenssiä. Käsittelem kolmitavuiset sanat erikseen siksi, että työni empiirisen osuuden epäsanat ovat juuri kolmitavuisia. Tähän paneudutaan paremmin luvussa 7. Seuraavissa laskelmissa ovat mukana kaikki sanaluokat, mutta ainoastaan vain kaikki CV-loppuiset sanat. Nämä luvut on laskettu *Nykysuomen sanakirjan* yhdistämättömistä sanoista. Ne muodostavat valtaosan kolmitavuisista sanoista. Muiden kuin CV-loppuisten sanojen ottaminen mukaan ei myöskään muuttaisi sanatyyppien frekvenssisuhteita. Yleisimmät toisen painottoman tavun rakenteet kolmitavuisissa sanoissa näyttävät seuraavalta. Lyhyt toinen painoton tavu näyttää olevan yleisin tavutyyppi. Ylipitkät tavut ovat pienin ryhmä, vaikka segmenttien yhdistymismahdollisuuden ovat laajimmat ylipitkissä tavuissa. Kaiken kaikkiaan yleisin toisen tavun rakenne on pitkä tavu (CVC tai CVV), joita on kaikista pääpainottomista tavuista 46,1 prosenttia. Toiseksi yleisin tyyppi on lyhyt toinen tavu (CV 38,3 %). Ylipitkiä tavuja on 15,7 prosenttia.

Taulukko 46. Pääpainottoman tavun rakenteiden määrät ja tyypifrekvenssit kolmitavuisissa sanoissa (NS).

Tyyppi	N	%
S.CV.CV	5 009	38,3
S.CVC.CV	4 285	32,7
S.CVVC.CV	2 048	15,7
S.CVV.CV	1 741	13,3
Yhteensä	13 082	100

S = pääpainollinen tavu.

Seuraavaksi taulukossa 47 esitän yleisimmät kolmitavuiset sanatyypit NS:n yhdistämättömistä sanoista. Suluissa on muiden kuin e'-nominien määrät. Luvuista ilmenee selvästi, että suomessa ei ensimmäisen ja toisen tavun rajaa kauempana ole

juurikaan muita geminaattoja kuin geminaattaklusiileja. Myös soinnilliset geminaattakonsonantit ovat harvinaisia (esim. *metalli*). Kolmitavuisten CV-loppuisten nominien rakenteet on koostettu taulukkoon 47.

Taulukko 47. Konsonanttien kvantiteettioppositio 2. ja 3. tavun rajalla (ei e'-nomineja) (NS).

	{k}	{t}	{p}	{s}	{m}	{n}	{l/}	{r}
S.CVC.CV	585	184	6	49	3	9	34	1
S.CV.CV	14	12	2	3	<u>251</u>	<u>458</u>	<u>684</u>	<u>766</u>
S.CVVC.CV	<u>93</u>	<u>140</u>	<u>9</u>	5	0	3	7	0
S.CVV.CV	15	11	1	<u>221</u>	<u>130</u>	<u>197</u>	<u>519</u>	<u>128</u>
yhteensä	707	347	18	278	384	667	1244	895
-VC.CV	678	324	15	54	3	12	41	1
-VC.CV %	<u>96 %</u>	<u>93 %</u>	<u>83 %</u>	<u>19 %</u>	1 %	2 %	3 %	---
-V.CV	29	23	3	224	381	655	1203	894

Ensimmäisen ja toisen tavun rajan ulkopuolella yksinäisklusiilit ovat harvinaisia, ja klusiilit ovat enimmäkseen geminaattoja. Jos yksinäisklusiileja esiintyy, ovat ne jonkin morfologisen operaation tuotosta (esim. käänteisessä astevaihtelussa olevien nominien heikkoasteisia nominatiiveja /*kaavake*/, /*juoksute*/). Soinnilliset konsonantit taas ovat tässä asemassa enimmäkseen yksinäiskonsonantteja ja geminaatat ovat harvinaisia. Kaiken kaikkiaan voidaan sanoa, että konsonanttien kvantiteetti on ensimmäisen ja toisen tavun rajaa kauempana sanassa vahvasti morfologisesti ohjautuvaa.

### 6.5 Kaksitavuisten nominien kvantiteetin funktionaalinen kuormitus

Tässä luvussa tutkin sanojen frekvenssin perusteella sitä, kuinka paljon sanoja erotetaan toisistaan kvantiteetin avulla ensimmäisen ja toisen tavun rajalla. Tyyppejä CVCCV ja CVCV, joissa on sanansisäinen klusiili joko geminaattana tai yksinäis-

konsonanttina, on yhteensä 647 sanaa. Sanoissa, joissa on sanansisäinen klusiili, on geminaattoja ja yksinäiskonsonantteja suurin piirtein saman verran (45,6 ja 54,4 prosenttia). Luvut on laskettu *Nykysuomen sanakirjan* yhdistämättömistä sanoista.

Taulukko 48. Yksinäisklusiilien ja geminaattojen määrät (NS).

	Yksinäisklusiili	Geminaatta	Yhteensä
Yhteensä	295	352	647
%	45,6	54,4	100

Minimiparit on määritelty siten, että minimipariksi lasketaan kukin sanakirja esiintymä kertaalleen. Esimerkiksi sanat */kato/* ja */katto/* lasketaan yhdeksi minimipariksi riippumatta siitä, että fonologiselta hahmolta homonyymisiä esiintymiä olisi enemmänkin. Seuraavassa taulukossa 49 on esitelty kaksitavuiset sanat, joissa sanansisäisen klusiilin kvantiteetti on erottavana tekijänä kaksitavuisessa sanassa. Minimiparien prosenttiluvut on laskettu rakenteiden kokonaismäärästä, eli ne osoittavat, että kvantiteettiin perustuvat minimiparit ovat harvinaisia. Minimipareja CVCV-CVCCV on 9,7 prosenttia kaikista CVCV ja CVCCV tyyppien sanoista (yhteensä 647 sanaa). Minimipareja tyyppijä CVX.CV-CVXC.CV on 5,8 prosenttia näiden sanatyyppien määrästä. Tässä yhteydessä sanansisäinen X tarkoittaa konsonanttia.

Taulukko 49. Sanansisäisen klusiilin pituuteen perustuvat minimiparit (NS).

Sanatyyppit	Kvantiteetti-minimipareja	%	Sanatyyppijä yhteensä
CVCV – CVCCV	63	9,7	647
CVSCV – CVSCCV	41	5,3	776
CVVCV – CVVCCV	38	6,6	577
Yhteensä	142	7,1	2000

S = soinnillinen konsonantti (*n, m, l, r*).



Sibilantin /s/ tyypifrekvenssi (Taulukko 50) eroaa klusiileista siinä, että tyypifrekvenssien perusteella suurin osa sanoista on lyhytkonsonanttisia (64,4 %).

Taulukko 50. Sanasisäinen /s/ yksinäiskonsonanttina ja geminaattana (NS).

	Yksinäiskonsonantti	Geminaatta	Yhteensä
Yhteensä	271	150	421
%	64,4	35,6	100

Kvantiteettiin perustuvia minimipareja on tyypissä /s/ vs. /ss/ sama määrä kuin klusiileissa (7,1 %) (ks. Taulukko 51 alla).

Taulukko 51. /s/ vs. /ss/ perustuvat minimiparit (NS).

Sanatyyppi	Kvantiteetti-minimipareja	%	Sanatyyppiä yhteensä
CVCV – CVCCV	18	10,7	168
CVSCV – CVSCCV	7	5,2	134
CVVCV – CVVCCV	5	4,2	119
Yhteensä	30	7,1	421

S = soinnillinen konsonantti n, m, l, r.

Soinnillisissa konsonanteissa lyhyet konsonantit ovat yleisempiä kuin geminaatat. Kielihistoriallisesti tämä johtuu siitä, että varhaiskantasuomessa geminaattoina saattoivat esiintyä ainoastaan klusiilit. Soinnilliset geminaatat ovat kehittyneet myöhemmin esim. ääntenmuutosten seurauksena (\**ln* > *ll*). Soinnillisissa konsonanteissa 77,1 prosenttia sanoista on lyhytkonsonanttisia.

Taulukko 52. Soinnilliset yksinäiskonsonantit ja geminaatat (NS).

	Yksinäiskonsonantti	Geminaatta	Yhteensä
Yhteensä	1013	301	1314
%	77,1	22,9	100

*Nykysuomen sanakirjan* yhdistämättömistä sanoista lasketut kvantiteettiminimiparien määrät ja suhteelliset osuudet, mitkä on esitetty taulukossa 53. Prosenttiluvut on laskettu kaikista CV(X)CV - CV(X)CCV-sanoista, joissa on joko geminaatta tai soinnillinen konsonantti, joka voi esiintyä geminaattana. Soinnillisissa konsonanteissa (*n, m, l, r*) minimipareja on 8, 2 prosenttia kaikista sanoista.

Taulukko 53. Sanansisäisen soinnillisen konsonantin pituuteen perustuvat minimiparit (NS).

Sanatyyppi	Kvantiteetti-minimipareja	%	Sanatyyppiä yhteensä
CVCV – CVCCV	59	8,2	723
CVVCV – CVVCCV	0	0	91 (2*)
Yhteensä	59	8,2	814

\* muita kuin e'-nomineja.

Alla olevassa taulukossa 54 on laskettu kaikista *Nykysuomen sanakirjan* kaksitavuisista nomineista kvantiteettiminimiparit. Ensimmäisessä sarakkeessa ("kaikki") on kaikkien geminaattojen ja niiden sanansisäisten yksinäiskonsonanttien määrä, jotka voivat esiintyä geminaattoina ympäristössä CV(X)CV – CV(X)CCV.

*Nykysuomen sanakirjassa* minimiparit näyttävät olevan yleisimpiä sanatyypeissä CVCV – CVCCV, joissa niitä on 9,5 prosenttia (Taulukko 54) ja harvinaisimpia sanapareissa CVVCV – CVVCCV (5,8 %) ja CVCCV – CVCCCV (6,0 %). Jos minimiparien määrä suhteutetaan kaikkiin kaksitavuisiin nomineihin, jää minimiparien määrä erittäin pieneksi (Taulukko 54). Kaiken kaikkiaan konsonanttien kvantiteet-

timinimipareja on vain 3,6 prosenttia kaikista kaksitavuisista sanoista. Vertailukohtana voi pitää sitä, että homonyymejä on NS:n kaksitavuisissa sanoissa yli kaksi kertaa enemmän kuin kvantiteettiminipareja: CVCV 84; CVCCV 200; CVVCV 91; CVVCCV 46; CVSCV 83; CVCCCV 56 (yhteensä 560).

Laskutavasta riippumatta sanansisäinen kvantiteetti on harvinainen merkityksiä erottavana tekijänä. Kirjoituksen oppimisessa tämä saattaa näkyä siinä, että kvantiteetin merkitseminen opitaan myöhemmin kuin muissa asemissa olevien grafeemien merkintä.

Taulukko 54. Minimipareja sanatyypeittäin (NS).

Sanatyyppi	N	Minimiparit	%
CVCV – CVCCV	1527	145	9,5
CVVCV – CVVCCV	659	38	5,8
CVCCV – CVCCCV	779	47	6,0
Yhteensä	2965	230	<u>7,8</u>

## 6.6 Minimiparit ja niiden funktionaalinen kuormitus

Kaikkiin NS:n yhdistämättömiin sanoihin verrattaessa käy helposti ilmi, että kvantiteettiin perustuvat minimiparit ovat kaksitavuisissakin nomineissa perifeerinen ilmiö. Minimipareja on kaikkiin kaksitavuisiin sanoihin suhteutettuna vain 3,2 prosentissa sanoista (Taulukko 55).

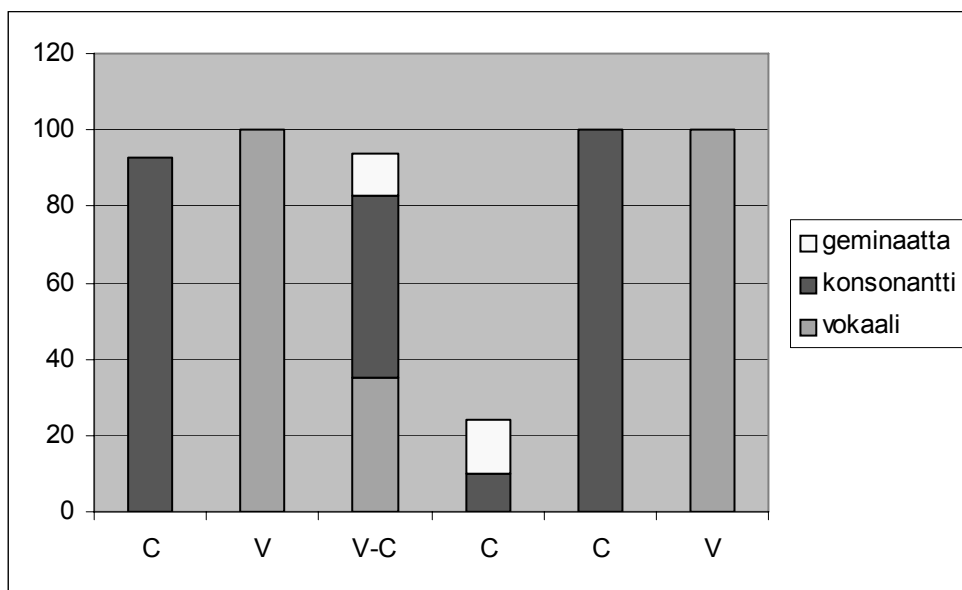
Taulukko 55. Minimiparit suhteutettuina kaikkiin kaksitavuisiin sanoihin ( $S = n, m, l, r$ ) (NS).

Sanatyyppi	N	Minimiparit	%
CVCV – CVCCV	3286	145	4,4
CVVCV – CVVCCV	2203	38	1,7
CVSCV -CVCCCV	905	47	5,2
Yhteensä	6394	203	<u>3,2</u>

Jos siis kirjoittamaan opeteltaessa sanatyypin frekvenssillä on merkitystä (ks. Uppstad 2005, 40), oppii lapsi merkitsemään kirjoitukseen merkityksiä erottelevat tekijät sen mukaan, kuinka frekventti ilmiö kyseessä oleva tekijä on. Moderni kielitiede on Saussuren jälkeen keskittynyt kielen rakenteellisiin kontrasteihin, joka tarkoittaa merkityksiä erottelevien oppositioiden yksiselitteistä korostamista. Vaihtoehtoinen malli nostaa suhteellisen frekventtiyden rakenteellisten kontrastien yläpuolelle. Frekvenssinäkemyksen mukaan pelkkään strukturalistiseen opposition käsitteeseen nojaava kielitiede ei pysty selittämään esimerkiksi sitä, miten nämä oppositiot rakentuvat lapsen oppiessa kieltä. Frekvenssiteorioiden mukaan kirjoituksen tehtävä on funktionaalinen; kirjoittaja pyrkii ilmaisemaan merkityksiä ja kirjoitus on työkalu tähän tarkoitukseen. Koska kvantiteetti erottelee merkityksiä kielessä harvoin, näkyy tämä myös kirjoittamaan oppimisessa: kvantiteetin merkitseminen horjuu.

Seuraavaksi esittelen tarkemmin eri segmenttien funktionaalista kuormitusta leksikossa tyyppifrekvenssien pohjalta. Tätä varten olen laskenut eri segmenttien yleisyyden *Nykysuomen sanakirjasta*. Sananalkuinen konsonantti on 93 %:ssa kaksitavuisista sanoista. Ensimmäisen moran muodostama vokaali on kaikissa sanoissa

(100 %) samoin kuin toisen tavun alkukonsonantti ja toisen tavun vokaali. Sanan toisen moran muodostaa 48 %:ssa sanoista konsonantti, 35 %:ssa vokaali ja 11 %:ssa geminaatan alkuosa. Kolmas mora on yhteensä 24 prosentissa sanoista (geminaatan alkuosa 14 %, konsonanttiyhtymän alkuosa 10 %). Geminaatallisia muotoja on yhteensä 25,6 prosentissa kaksitavuisista nomineista. Toisen moran muodostama geminaattoja on 11,1 prosentissa sanoista ja kolmannen moran muodostavia geminaattoja 14,5 prosentissa. Frekvenssitietojen pohjalta voidaan rakentaa seuraavanlainen kuvaus suomen kaksitavuisista sanoista (N = 6 606).



Kuvio 32. Kaksitavuisien sanojen rakenteen kuvaus mora-asemittain.

Seuraava kuvio 33 kuvaa sitä, kuinka monessa prosentissa sanoista on kyseiset grafeemit. Esimerkiksi kohdassa 1 on kuvattu grafeemiyhdistelmä (tavun alkukonsonantti + vokaali) on 93 prosentissa kaksitavuisista nomineista. Kohdassa kaksi kuvattu umpitavu on 23,5 prosentissa sanoista. Toinen vokaali seuraa ensimmäistä vokaalia 19,5 prosentissa sanoista. Geminaattoja on 11,1 %:ssa (lyhyen vokaalin jälkeen) kaikista sanoista ja ylipitkissä tavuissa 7,0 prosentissa kahden vokaaligrafeemin jälkeen ja 7,4 prosentissa kaksitavuisista sanoista.

	1)	2)	3)
(93 %)	1. mora	2. mora	3. mora
C	V	-----	-----
C	V	C (23,5 %)	-----
C	V	V (19,5 %)	-----
C	V	<u>C.C</u> (11,1 %)	-----
C	V	C	<u>C.C</u> (7,4 %)
C	V	V	<u>C.C</u> (7,0 %)
C	V	V	C.C (6,9 %)
C	V	C	C.C (2,0 %)

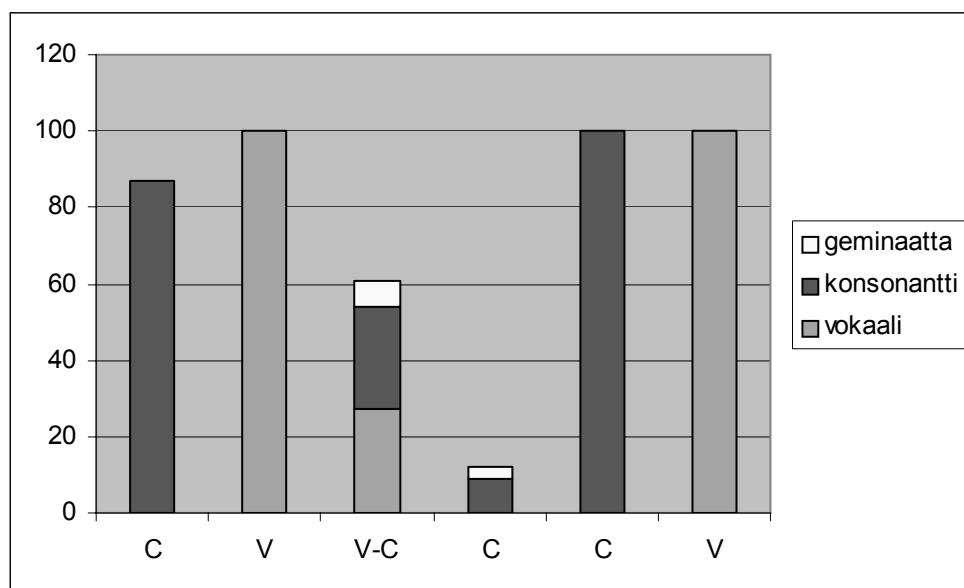
Kuvio 33. Grafeemien suhteellinen yleisyys morittain merkittynä.

Pääpainollisen tavun rakenteet morina kaikista *Nykysuomen sanakirjan* yhdistämättömistä sanoista (N = 68 502) laskettuna on koottu kuvioon 34. Eri tavutyypin yleisyys on kutakuinkin sama kuin pelkästään kaksitavuisista sanoista (vrt. Kuvio 33) laskettuna. Poikkeamana kaksitavuisiin sanoihin on geminaattojen suhteellinen osuus kaikissa NS:n yhdistämättömissä sanoissa pienempi. Tämä johtunee siitä, että kaksitavuisia pidemmissä sanoissa on sanan toinen tavu usein umpitavu, jonka edellä geminaatta lyhenee yksinäiskonsonantiksi (kvantitatiivisessa astevaihtelussa). Geminaatalliset ylipitkät tavut ovat kaksitavuisissa sanoissa huomattavasti yleisempiä kuin NS:n koko sanastossa (CVVCC 7,0 % vs. 3,5 % ja CVCCC 7,4 % vs. 2,0 %).

(87 %)	1. mora	2. mora	3. mora
C	V (27,7 %)	-----	-----
C	V	C (23,1 %)	-----
C	V	V (17,5 %)	-----
C	V	<u>C.C</u> (7,1 %)	-----
C	V	C	<u>C.C</u> (4,5 %)
C	V	V	<u>C.C</u> (3,5 %)
C	V	V	C.C (2,0 %)
C	V	C	C.C (1,7 %)

Kuvio 34. Grafeemien suhteellinen yleisyys morittain merkittynä (NS:n kaikki sanat).

Alla olevassa kuviossa 35 on esitetty eri segmenttiyhtymien yleisyys graafisessa muodossa. Laskuissa on kaikki *Nykysuomen sanakirjan* yhdistämättömät sanat.



Kuvio 35. Pääpainolliset tavut NS:n yhdistämättömissä sanoissa (prosentteina).

Kuten edellä mainituista tyyppifrekvenssitiedoista ilmenee, ei pitkä kvantiteetti ole kovin frekventti ilmiö. Geminaatat muodostavat vain pienen osan varsinkin ylipitkien tavujen tavunloppuisista konsonanteista. Kaksitavuisissa sanoissa geminaattoja on vain 14,4 prosenttia ja NS:n kaikista yhdistämättömistä sanoista laskettuna ainoastaan 5,5 prosenttia.

## 6.7 Minimiparien frekvenssi

Seuraavaksi tarkastelen minimiparien määrää juoksevassa tekstissä. Teen tämän siten, että lasken kaikkien *Nykysuomen sanakirjan* kaksitavuisien minimiparien yleisyyden aineistosta, joka koostuu noin 180 000 000 sanasta (tarkkaan 179 556 341 sanaa). Tämä jättimäinen data koostuu Kielipankin ([www.csc.fi/kielipankki](http://www.csc.fi/kielipankki)) kaikista suomenkielisistä aineistoista, joihin on pääsy www-pohjaisella korpustyökaluohjelmalla nimeltä *www-lemmie* (<http://www.csc.fi/english/research/software/www-lemmie>).

Yhdeksi minimipariksi on laskettu esiintymät, joissa esiintyy minimiparien molemmat osat (esim. */katto/* – */kato/*, */tuli/* – */tulli/*). Taulukkoon 56 on kerätty kaikki ne minimiparit, joiden kumpikin osa esiintyy Kielipankin aineistossa enemmän kuin 10. kertaa.

Harvinaisimpia tässä aineistossa ovat minimiparit, jotka rakentuvat ylipitkästä tavusta. Yhteensä minimipareja, jonka kumpikin jäsen esiintyy Kielipankin aineistossa 10 kertaa tai enemmän on 97 kappaletta. Taulukoissa 56 ja 57 olevat skeletaalit tarkoittavat minimiparityyppejä siten, että tyyppin CVCV - CVCCV minimiparit ovat tyyppiä */kuka/* - */kukka/*, tyyppin CVVCV - CVVCCV tyyppiä */viita/* - */viitta/* ja CVCCV - CVCCCV tyyppiä */kunta/* - */kuntta/*.



Taulukko 56. Niiden minimiparisanatyyppien esiintymät, jotka esiintyivät Kielipankin suomenkielisessä aineistossa enemmän kuin kymmenen kertaa (10 >).

	<b>CVCCV-CVCCV</b>	<b>CVVCV-CVCCV</b>	<b>CVCCV-CVCCCV</b>	<b>Yhteensä</b>
/k/	19	2	3	24
/t/	7	9	2	18
/p/	5	0	2	7
/m/	5	0	-----	5
/n/	6	0	-----	6
/l/	20	0	-----	20
/r/	6	0	-----	6
/s/	11	0	-----	11
yhteensä	<b>79</b>	<b>11</b>	<b>7</b>	<b>97</b>

Minimiparien määrä vähenee edelleen, jos listataan vain ne minimiparit, joiden kumpaakin osaa esiintyy sanomalehtiaineistossa vähintään 100 kertaa. Näin seulottuna ylipitkän tavun varaan rakentuvat minimiparit kutistuvat lähes olemattomiin (11 kpl). Yhteensä kvantiteettiminimipareja on 47 kappaletta.

Taulukko 57. Niiden minimiparisanatyyppien esiintymät, jotka esiintyivät Kielipankin suomenkielisessä aineistossa enemmän kuin sata kertaa (100 >).

	<b>CVCV-CVCCV</b>	<b>CVVCV-CVVCCV</b>	<b>CVCCV-CVCCCV</b>	<b>Yhteensä</b>
/k/	10	1	1	12
/t/	4	8	0	12
/p/	1	0	1	2
/m/	2	0	-----	2
/n/	4	0	-----	4
/l/	11	0	-----	11
/r/	1	0	-----	1
/s/	3	0	-----	3
yhteensä	<b>36</b>	<b>9</b>	<b>2</b>	<u><b>47</b></u>

Kaiken kaikkiaan kvantiteettiminimipareja ei näytä juurikaan esiintyvän juoksevassa tekstissä, vaikka niitä näyttää olevan jonkin verran leksikossa (tässä NS). Siksi ei ole oletettavaa, että lukemaan ja kirjoittamaan oppiva lapsi pystyisi oppimaan kvantiteetin merkitsemisen minimiparien kautta (esim. /kisa/ vs. /kissa/ = /vesa/ vs. X, X = /vessa/). Minimiparien avulla oppiminen on tietenkin mahdollista, mutta se ei liene onnistu datan (syötöksen) kautta, koska minimiparien käyttöfrekvenssi ei tähän riittäne.

### 6.8 Koonta sanatyyppien frekvenssien tarkastelusta

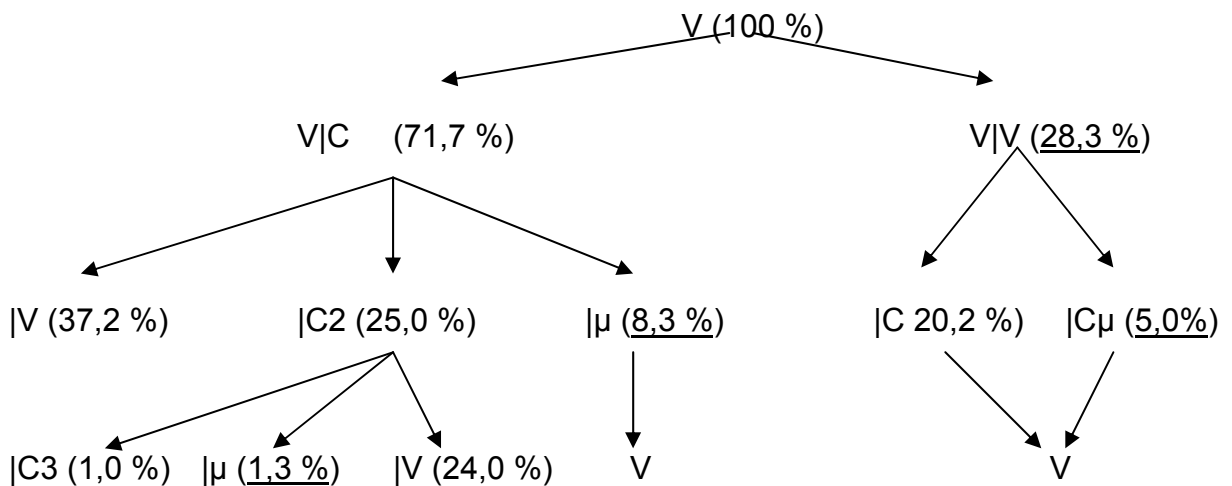
Jos sanatyypit kirjoitetussa tekstissä järjestetään käyttöfrekvenssin perusteella, saadaan seuraavanlainen hierarkia. Kaikissa tavuissa on vähintään yksi vokaali.

- 1) Sanoista 71,7 prosentissa lyhyttä ensitavun vokaalia (yhtä vokaaligrafeemia) seuraa yksi konsonantti (sanatyypit (C)V.CV, (C)V $\bar{C}$ .CV, (C)V $\bar{C}$ C.CV).

1) 28,3 prosentissa kaksitavuisista sanoista on ensitavussa kaksi vokaalia ((C)VV.CV, (C)VVC.CV).

2) Lyhyttä ensitavun vokaalia (yhtä vokaaligrafeemia) ja konsonanttia (C)VC| seuraa toinen ei-identtinen konsonanttigrafeemi 25,0 % sanoista ja vokaali 37,2 prosentissa sanoista. Identtinen konsonanttigrafeemi (gemaatta) seuraa 8,3 %:ssa tapauksista.

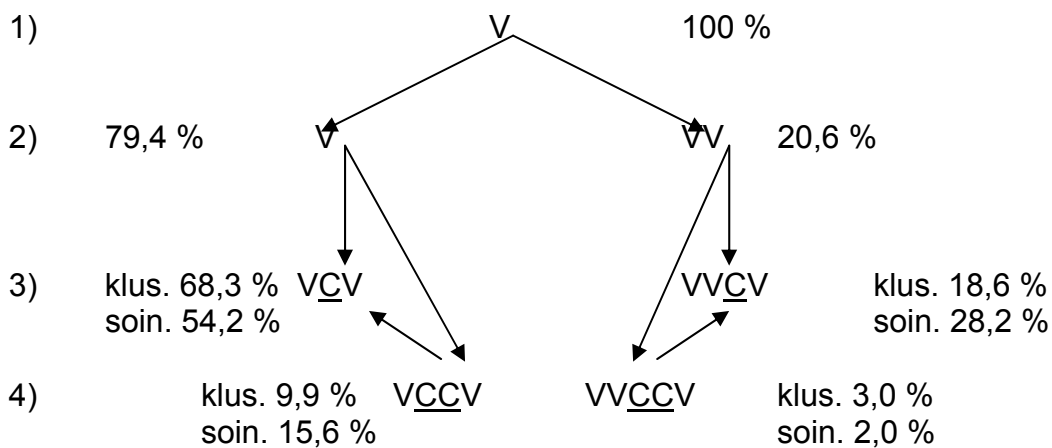
3) Ensitavun vokaaligrafeemia seuraa toinen vokaaligrafeemi 28,3 prosentissa tapauksista. 20,2 prosentissa tavuista kahta vokaaligrafeemia seuraa yksi konsonantti ja 5,0 prosentissa gemaatta (kaksi identtistä konsonanttigrafeemia).



Kuvio 36. Sanatyyppien yleisyys tasoittain juoksevassa tekstissä (SK).

Jos kuvauksessa keskitytään pelkästään segmenttien pituuteen, saadaan kuvion 37 tapainen hierarkia. Jokaisessa sanassa on (odotuksen mukaisesti) vähintään yksi vokaali (taso 1). 20,6 prosentissa on kaksi peräkkäistä vokaaligrafeemia. 79,4 prosentissa yhtä vokaaligrafeemia seuraa konsonantti. Jos konsonantti on klusiili (k, p, t), on se yksinäisigrafeemi 86,9 prosentin todennäköisyydellä ja gemaatta 12,9 prosentin todennäköisyydellä. Kahden identtisen vokaaligrafeemin jäljessä klusiili on gemaatta vain 3,0 prosentissa juoksevan tekstin sanoista, joissa on sanansisäinen klusiili. Yhden vokaalin jälkeen klusiiligemaatta on 9,9 prosentissa.

Lyhyttä vokaalia seuraava soinnillinen konsonantti on geminaatta 15,6 prosentissa sanoista ja yksinäiskonsonantti 54,2 prosentissa sanoista. Kahden identtisen vokaaligrafeemin jälkeen soinnillinen konsonantti on geminaatta vain 2,0 prosenttisesti ja yksinäiskonsonantti 98,0 prosenttisesti. Kaiken kaikkiaan konsonanttien pitkän ja lyhyen kvantiteetin suhde on lyhyen vokaalin jälkeen 17 % vs. 83 % ja kahden identtisen vokaaligrafeemin jälkeen 90,2 % vs. 9,8 %.



Kuvio 37. Vokaalien ja konsonanttien kvantiteettityyppien jakauma niiden yleisyyden perusteella juoksevassa tekstissä (SK).

Päätelmät:

1) Kirjoittamaan oppimisen näkökulmasta geminaatat ovat ”täydellisen” oikeinkirjoitustaidon omaksumisessa viimeinen opittava, koska altistus niiden oppimiseksi on pienempi kuin yksinäiskonsonanttien kohdalla. Geminaattoja on vähän kun tarkastellaan juoksevan testin käyttöfrekvenssejä (Taulukko 34) ja sanakirjatekstin tyypifrekvenssejä (Taulukko 40). Siksi geminaatat ovat ortografian pullonkaula.

2) Geminaatan merkitseminen yksinäiskonsonantilla voi olla yksinkertainen yleistyminen siitäkin, että sanan sisällä tavu alkaa suomessa melkein aina yhdellä konsonantilla.

3) Geminaatan merkitseminen yhdellä konsonantilla ei välttämättä johda ”kielen vastaisiin” sanamuotoihin kuten esimerkiksi yhden vokaaligrafeemin tai yhden konsonanttigrafeemin poisjättäminen johtaisi, esim. *takka* > *taka*; *tuli* > *\*tli*.

4) Koska geminaatan ja yksinäiskonsonantin muodostamia minimipareja on niin sanatyypeissä kuin juoksevassa tekstissä hyvin vähän, kirjoittamaan oppijan on vaikea muodostaa aitoja esimerkkejä siitä, miten kvantiteetti erottaa eri sanahahmoja ja merkityksiä kirjoituksessa toisistaan.

5) Koska kirjoitettu kieli ei ole puheen tapaan prosodinen ilmiö, etenkin pitkän kvantiteetin muuttaminen segmentaaliseksi esitykseksi ja sen merkitseminen kirjoituksessa on vaikeampaa kuin sen tuottaminen ja signaloiminen puheessa tavu- ja sanarakenteen eri asemissa. Tämä ei sulje pois mahdollisuutta, että myös pituuden hahmottaminen puheessa on vaikeata etenkin dyslektikoilla (ks. esim. Richardson et al. 2003; Leppänen et al. 2002). Puheen tuottaminen on erittäin paralleelia prosessointia. Lisäksi se on ilmiö, jossa esimerkiksi semanttinen, leksikaalinen ja morfologinen ennakointi on tärkeässä osassa. Leksikaalisen, morfologisen ja fonologisen tiedon sovittaminen yhteen on ollut formaalisten kielioppien tekijöidenkin kesto-ongelma. Geminaatta ja yksinäiskonsonantti vaihtelevat myös samassa leksikaalisessa yksikössä (astevaihtelu), ja ne ovat puheessa vapaassa vaihtelussa esimerkiksi joissakin lainasanoissa (*int(t)ernetti*) (ks. lukua 6.9) ja dialekteissa ja sosiolekteissa, joissa esiintyy geminaatiota.

6) Mikään ei takaa kuitenkaan sitä, että kirjoitus**prosessissa** kirjoittaja ei erottaisi lyhyttä ja pitkää kvantiteettia toisistaan vaan että kvantiteettivirheet ovat kirjaimen poisjättämiä. Monitorointi voi keskittyä kirjoittamisessa ensisijaisesti kirjainten kvaliteettiin, ja yksikin grafeemi täyttää tällöin monitorointiehdot. Kirjoitusprosessista on todisteita ainoastaan siitä, että lukihäiriöstä kärsivät merkitsevät tavallista useammin kvantiteettia kahden identtisen grafeemin sijasta yhdellä grafeemilla.

Edellä käsitellyn frekventtiyden jälkeen siirryn käsittelemään tunnusmerkkisyyden teoriaa ja sen soveltamista piituden käsitteeseen ja lukemisen ja kirjoittamisen oppimiseen.

## 6.9 Frekvenssi ja tunnusmerkkisyyden teoria

Tunnusmerkkisyyden käsite on peräisin Prahan koulukunnasta, jossa tunnusmerkkisyys kytkeytyi aluksi fonologisten oppositioiden neutraalistumiseen (laajemmin ks. esim. Lauerma 1993, 5). Chomskyn ja Hallen generatiivisessa fonologiassa syvä-rakenteen segmentit katsottiin joko tunnusmerkkisiksi tai tunnusmerkittömiksi. Tunnusmerkkisyys ei ollut generatiivisessa fonologiassa enää sidoksissa fonologisten piirteiden neutraalistumiseen, vaan tunnusmerkkisyyden lähtökohtana on segmenttien universaali, ”luonnollinen” tunnusmerkittömyys (Lauerma 1993, 5). Esimerkiksi soinnittomat klusiilit ovat tunnusmerkittämiä ja vokaaleihin kuuluu tunnusmerkittömänä piirre [+soinnillinen]. Tarkastelen seuraavassa tunnusmerkkisyyden kriteerejä Lassin (1985) esittämänä (Lauerman 1993, 6 mukaan). Lassilla tunnusmerkkiset yksiköt

- 1) ovat maailman kielissä tunnusmerkittämiä harvinaisempia
- 2) eivät esiinny oppositioparin edustajana opposition neutraalistuttua
- 3) ovat tekstifrekvenssiltään pienempiä
- 4) ilmaantuvat lapsen puheeseen myöhemmin ja katoavat afaatikon puheesta aikaisemmin
- 5) kadotessaan sulautuvat tunnusmerkittömään yksikköön
- 6) ovat kielen kehityksen kannalta epävakaampia, herkempiä muutokselle
- 7) implikoivat vastaavan tunnusmerkittömän yksikön esiintymisen.

Näyttää siltä, että suomenkin kielessä sanatyyppeiden frekvenssi ja niiden kompleksisuus odotuksenmukaisesti osuvat yhteen: kompleksiset rakenteet ovat vähiten frekventtejä sanatyyppejä. Ainakin silloin, jos kompleksinen määritellään siten, että pituudeltaan kompleksisia ovat ne sanahahmot, joissa kvantiteettioppositio on muutenkin jonkin yksikön äärimmäisellä laidalla (ks. Steriade 1997, 2001). Tällöin ylipit-

kissä tavuissa olevat geminaatat ovat kompleksisia, koska ne sijaitsevat kaukana tavun ytimeistä. Samalla logiikalla vähiten tunnusmerkisiä ja vähiten kompleksisia ovat lyhyistä tavuista koostuvat sekvenssit (CV.CV). Asia ei ole kuitenkaan näin yksinkertainen, sillä on oletettavaa, että pituuden suhteen kompleksisia ovat kaikki ne asemat sanassa, joissa kvantiteettioppositioita yleensä voi esiintyä. Näin ollen kompleksinen on esimerkiksi sanahahmo CVV.CV, koska se on kvantiteettioppositiossa CVVC.CV-tyyppin kanssa. Tällöin kielellinen kompleksisuus määriteltäisiin seuraavaan tapaan. Jokainen sanatyyppi on yhtä kompleksinen, koska kaikissa alla olevissa sanatyypeissä on saman verran oppositiopareja.

Sanatyyppi	oppositiossa	kompleksisuusarvo
(C)V.CV	CVVCV, CVCCV	2
(C)VV.CV	CVCV, CVVCCV	2
(C)VC.CV	CVCV, CVVCCV	2
(C)VVC.CV	CVCCV, CVVCV	2

Kuvio 38. Fonologinen kompleksisuus 1.

Kompleksisuus voidaan määritellä myös siten, että CVVCV on kompleksinen ja tunnusmerkkinen CVCV:n suhteen, jonka numeroarvo on 0, mutta CVVCCV on tunnusmerkkinen CVVCV:henkin verrattuna, ja se saisi tunnusmerkkisyyden osalta arvon 2. Tällöin kompleksinen määritellään hypoteettisen universaalien kielellisen tunnusmerkkisyyden teorian kautta. Sanatyyppi CVCV on tunnusmerkitön. CVVCV saa kompleksisuusarvon 1, koska siinä on kvantiteettioppositio suhteessa tunnusmerkittömään sanaan CVCV. Sanatyyppi CVVCCV olisi kaikkein kompleksisin, koska siinä on kaksi kvantiteettioppositiota.

Sanatyyppi	Kompleksisuusarvo
(C)V.CV	0
(C)VV.CV	1
(C)VC.CV	1
CVVC.CV	2

Kuvio 39. Fonologinen kompleksisuus 2.

Vaikka asia olisikin näin, ei tämä oppositioiden määrään perustuva kompleksisuus näy kirjoittamisessa, koska kvantiteettivirheet merkitään yleisesti yhdellä kirjaimella kahden grafeemin sijaan. Tällöin ei ole mitään todisteita siitä, etteikö myös CVV.CV-hahmo olisi kielellisesti kompleksinen. Tämän puolesta todistaa se, että puheessa pitkien tavujen jälkeinen lyhyt klusiili pitenee usein vierassanoissa (ks. esim. Karlsson 1982, 153-154) (esim. *kookos* > *kookkos*; */balkan/* ”Balkan” > */balkkan/* ”Balkkan”). Koska tämä ilmiö on niin yleinen, ei voida ajatella, että obstruentin pidennys tässä asemassa sinällään lisäisi kielellistä kompleksisuutta. Toinen mahdollisuus on, että tämä vaihtelu on luonteeltaan foneettista. Kaukana tavun ytimeistä fonologiset oppositiot eivät olisi vakiintuneita<sup>1</sup>.

Voidaankin päätellä, että kielellisen, fonologisen kompleksisuuden lisäksi on olemassa grafortaktinen (visuaalinen) kompleksisuus, joka ilmenee siinä, että ne kvantiteettioppositiot, jotka ovat kirjoitetussa kielessä harvinaisia, tuottavat kirjoittamaan oppiville lapsille eniten vaikeuksia. Kirjoitustehtävän tuloksissa tämä näkyy siinä, että vaikka pitkät klusiilit olivat vaikeita sekä lyhyen että pitkän vokaalin jäljessä, kaikkein vaikein asema oli pitkän vokaalin jälkeinen pitkä nasaali. Koska oppositio lyhyen ja pitkän nasaalin (pitkä ja lyhyen soinnillisen konsonantin) välillä on tässä asemassa harvinainen (ks. Taulukot 34 ja 43), on se myös vaikeinta oppia merkitsemään.

<sup>1</sup> Toisenlaiseen ratkaisuun on päädytty Turusen (2003) ja Torvelaisen (2007) kielenomaksumista käsittelevissä väitöskirjoissa, joissa pitkää painollista tavua on pidetty tunnusmerkittömänä.



Seuraavaksi käsittelen kielellistä tunnusmerkkisyyttä ja tunnusmerkkisyyden muita erilaisia kriteerejä. Jos tunnusmerkittömiksi määritetään ne kontrastit, jotka ovat kielessä yleisiä, ja taas tunnusmerkkisiksi kontrastit, jotka ovat kielessä harvinaisia, saadaan seuraavanlainen (kuvio 40) tunnusmerkkisyysjatkumo juoksevan tekstin kvantiteettidistinktioiden kannalta katsottuna. Tässä käytetty Suomen Kuvalehden (1999) sisältää 35850 sanaa, joissa kvantiteettidistinktio on mahdollinen ensitavun vokaalissa, ensimmäisen ja toisen tavun rajalla olevassa konsonantissa tai molemmissa. Jos siis sanatyyppien käyttöfrekvenssi määrittää niiden tunnusmerkkisyyden, ovat tunnusmerkkisimpiä pitkän vokaalin jälkeiset geminaatat tyyppiä *taakka*, *tyyppi* jne.. Toiseksi tunnusmerkkisimpiä ovat lyhyen vokaalin jälkeiset geminaatat tyyppiä *takka*. Tunnusmerkittömiä ovat avotavut ja lyhyet konsonantit.

Alempaan kuvioon 41 on merkitty sanatyyppien yleisyys *Nykysuomen sanakirjan* mukaan (N=32 327). Myös ”leksikossa” (tyyppifrekvenssi, NS) ylipitkät tavut ovat kaikkein tunnusmerkkisimpiä. Toiseksi tunnusmerkkisimpiä ovat klusiiligeminaatat ja heti näiden jälkeen soinnilliset geminaatat. Lyhytkonsonanttiset muodot ovat vähiten tunnusmerkkisiä, kuten sopii odottaakin.

Kuvio 40. Tunnusmerkkisyysjatkumo juoksevasta tekstistä.

*CVV <u>CC</u> V, *CVV <u>CC</u> V > *CV <u>CC</u> V > *CV <u>CC</u> V > *CVV <u>C</u> V > *CVV <u>C</u> V > *CV <u>C</u> V >						
(n, m, l, r)	(k, p, t)	(k, p, t)	(n, m, l, r)	(k, p, t)	(n, m, l, r)	(n, m, l, r)
f: 325	576	1 879	2 642	3 520	4 774	9 178
0,9 %	1,6 %	5,2 %	7,4 %	9,8 %	13,3 %	25,6 %
*CV <u>C</u> V						
(k, p, t)						
12 938						
36,1 %						

Kuvio 41. Tunnusmerkkisyysjatkumo *Nykysuomen sanakirjassa*.

*CVV <u>CCV</u>	*CVV <u>CCV</u> >	*CV <u>CCV</u> >	*CV <u>CCV</u> >	*CVV <u>CV</u> >	*CVV <u>CV</u> >	*CV <u>CV</u> >
(n, m, l, r),	(k, p, t)	(k, p, t)	(n, m, l, r)	(k, p, t)	(n, m, l, r)	(k, p, t)
f: 750	1 744	1 745	3 400	3 965	5 223	6 388
2,3 %	5,4 %	5,4 %	10,5 %	12,3 %	16,2 %	19,8 %
*CV <u>CV</u>						
(n, m, l, r)						
9 112						
28,2 %						

Juoksevassa tekstissä (Kuvio 40) lyhyitä konsonantteja on ensimmäisen ja toisen tavun rajalla 84,9 prosenttia ja geminaattoja 15,1 prosenttia. Lyhyitä klusiileja on ensimmäisen ja toisen tavun rajalla 87,0 prosenttia ja pitkiä klusiileja tällöin 13,0 prosenttia. *Nykysuomen sanakirjasta* laskettuna lyhyitä konsonantteja on samassa asemassa 76,4 prosenttia ja geminaattoja 23,6 prosenttia. Erot juoksevan tekstin ja NS:n välillä selittyvät sillä, että NS:ssä on enemmän klusiiligeminaattoja. Tämä onkin odotettavaa, sillä leksikossa nominit ovat yksikön nominatiivissa, ja tällöinhän nomineissa on vahva aste, vaikka esim e'-nomineissa tilanne on kuitenkin toisin päin, mutta e'-nomineja ei ole niin paljon, että ne vaikuttaisivat radikaalisti kvantiteettisuhteisiin (ks. Taulukoita 41 ja 47).

Näyttääkin siltä, että tunnusmerkkisyyshierarkia frekvenssin perusteella muodostuu erilaiseksi leksikkomuotojen ja juoksevan tekstin frekvenssilukujen välillä. Tämä käytö- ja tyyppifrekvenssin ero saattaa olla osasyynä siihen, että klusiilien kvantiteetti on soinnillisia konsonantteja vaikeampi merkitä. Kirjoittamaan oppiva lapsi hakee mentaalileksikosta tarvittavan sanan, jonka kvantiteetti lemma-muodossa voi kuitenkin olla erilainen kuin leksikkumuotoinen sana.

## 7. Fonologisen teorian kokeellinen soveltaminen

### 7.1 Koeasetelma ja menetelmät

Kirjoitustestiin osallistui 230 toisella ja kolmannella luokalla olevaa lasta (110 2.-luokkalaista ja 120 3.-luokkalaista). Kukin ärsyke esitettiin kuulokkeista kaksi kertaa 75 desibelin voimakkuudella. Ärsykkeiden välissä oli 2 sekunnin hiljaisuus. Vastausaikaa annettiin lapsille niin kauan kuin he tarvitsivat. Kokeen vetäjä katsoi, koska kaikki lapset olivat laittaneet kynän pöydälle merkiksi siitä, että olivat ratkaisseet osion. Ohjeena lapsia pyydettiin kirjoittamaan kuulemansa paperille. Lapsille oli kerrottu, että kuullut sanat voivat olla tuntemattomia hölynpölynsanoja. Testissä käytettiin 28:aa epäsanaa. Epäsanojen oikeinkirjoituksessa kiinnitettiin huomiota ensisijaisesti kvantiteetin merkitsemiseen. Jos vastaus oli mahdoton tulkita, se merkittiin hylätyksi. Koska alkuperäistä listaa jouduttiin lyhentämään käytännön syistä, jäi osa sana-asemista pienemmälle huomiolle. Tämä voi olla ongelma vastauksia tulkittaessa, mutta tuloksia lienee kuitenkin pidettävä ainakin suuntaa antavina. Vähiten epäsanoina oli konsonantti /n/:n ja geminaatta /nn/:n sisältäviä epäsanoina (alleviivattu), näitä kumpaakin on kussakin sana-asemassa ainoastaan kaksi kappaletta, kun taas lyhyen klusiilin /t/ ja geminaatan /tt/ sisältäviä sanoja on sana-asemasta riippuen neljästä kahdeksaan kappaletta.

Lyhyen ja pitkän kvantiteetin oikeinkirjoitusta tutkittiin seuraavissa sana-asemissa. Sana-aseman jälkeen on merkitty niiden sanojen määrä, joissa oli kyseinen segmentti mainitussa sana-asemassa. Tämän jälkeen on esitelty ne epäsanat, joissa on kyseinen segmentti.

Pääpainollinen lyhyt vokaali (16 epäsanaa)

*ruttippo, suturemi, lutiipo, ruttiiippo, lutipo, sutippo, ruttemi, sunippo, sunipo, sutemi, sutiippo, runnipo, nuttiiippo, ruttemmi, runnippo, nuttippo*

2) Pääpainollinen pitkä vokaali (12 epäsanaa)

*reettippo, tuunippo, nuuttemi, seettippo, neetippo, ruunipo, nuuttemmi, ruutemi, ruutemmi, tuunnipo, reetipo, ruunnippo*

3) Toisen tavun tavunalkuinen konsonantti lyhyen pääpainollisen vokaalin jälkeen (konsonantit /n/ = 2 ja /t/ = 6).

*sunippo, sunipo*

*sutemmi, lutiipo, lutipo, sutipppo, sutemi, sutiippo*

4) Toisen tavun tavunalkuinen konsonantti pitkän pääpainollisen vokaalin jälkeen (konsonantit /n/ = 2 ja /t/ = 4).

*tuunippo, ruunipo*

*neetipppo, ruutemi, ruutemmi, reetipo*

5) Ensimmäisen ja toisen tavun rajalla olevat geminaatat (/nn/ ja /tt/) lyhyen pääpainollisen vokaalin jälkeen. (/nn/ = 2 /tt/ = 6). Geminaatta klusiilin /tt/ sisältävistä sanoista ovat asteriksilla \* merkityt sanat kielenvastaisia sikäli, että niissä on vahva aste (geminaatta) umpitavun edellä. Myöskin epäsanat /*ruttiippo*/ on rakenteeltaan kielelle vieras.

*runnipo, runnippo, ruttippo\*, ruttiippo, ruttemi, nuttiipo, ruttemmi\*, nuttippo*

6) Ensimmäisen ja toisen tavun rajalla olevat geminaatat (/nn/ ja /tt/) pitkän vokaalin jälkeen. (/nn/ = 2 /tt/ = 4). Kielenvastaiset vahva-asteiset muodot on merkitty edelleen asteriksilla \*.

*tuunnippo, ruunnippo*

*reettippo\*, nuuttemi, seettippo, nuuttemmi\**

7) Toisen tavun lyhyt vokaali (= 24)

*ruttippo, reettippo, sutemmi, tuunippo, nuuttemi, seettippo, lutipo, sutipppo, ruttemi, neetipppo, ruunipo, nuuttemmi, sunippo, sunipo, ruutemi, sutemi, ruutemmi, runnipo, tuunnippo, reetipo, ruttemmi, runnippo, nuttippo, ruunnippo*

8) Toisen tavun pitkä vokaali (= 4)

*lutiipo, ruttiippo, sutiippo, nuttiipo*

9) Lyhyen toisen tavun vokaalin jälkeinen kolmannen tavun alkukonsonantti (/m/ = 4 /p/ = 8). Tässä on huomattava, että yksinäisklusiileja ei juuri löydy ensimmäisen ja toisen tavun rajaa kauempana, vaan klusiilit esiintyvät geminaattoina (esim. *harakka, ullako* jne.).

*nuuttemi, ruttemi, ruutemi, sutemi*

*seettipo, lutipo, ruunipo, sunipo, runnipo, tuunnipo, reetipo, nuttupo*

10) Toisen tavun lyhyen vokaalin jälkeinen geminaatta (/mm/ = 4 /pp/ = 8)

*sutemmi, nuuttemmi, ruutemmi, ruttemmi*

*ruttippo, reettippo, tuunippo, sutippo, neetippo, sunippo, runnippo, ruunnippo*

## 7.2 Pituuden grafeeminen merkitseminen sana-asemittain

Esittelen seuraavassa kirjoitustestin virheanalyysin tulokset grafeemiasema kerrallaan alkaen sanan (pääpainollisen tavun) ensimmäisestä vokaalista. (Taulukoissa esitetään virheiden prosentuaaliset osuudet sekä se, kuinka paljon samoja virheitä oli eri lapsilla.) Sitten tarkastelen sanatason vaikutusta kirjoitusvirheisiin. Lopuksi esittelen kirjoitusprosessin mallini, jonka on tarkoitus auttaa hahmottamaan kirjoitusprosessin etenemistä kielitieteen näkökulmasta.

**Pääpainollisen tavun lyhyen vokaalin** merkitsemisessä tehtiin virheitä hyvin vähän. Koko aineistossa virheitä oli yhteensä 14, eli yhteensä 5,6 prosenttia lapsista teki virheen: 12 lasta teki yhden ja ainoastaan yksi lapsi 230:stä teki kaksi kvantiteetti-virhettä lyhyen pääpainollisen tavun vokaalin merkitsemisessä.

Taulukko 58. Pääpainollisen tavun lyhyen vokaalin merkitseminen 1.

Virheitä N	Vastaajia	%
0	217	94,3
1	12	5,2
2	1	0,4
Yhteensä	230	100

Yhteensä virheitä oli 14, ja vastauksia kaikkiaan 3680, eli vääriä vastauksia oli 0,4 prosenttia kaikista vastauksista. Hylättyjä vastauksia oli 5,3 prosenttia. Lyhyt pääpainollinen vokaali oli testissä 16 epäsanassa, joten jaettu summa on virheiden osalta  $14 / 16 = 0,875$ .

Taulukko 59. Pääpainollisen tavun lyhyen vokaalin merkitseminen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
14	3472	194	3680
0,4 %	94,3 %	5,3 %	100 %

**Pääpainollisen tavun pitkä vokaali** osoittautui lyhyen vokaalin tavoin helpoksi merkitä (Taulukot 60 ja 61). Tosin oikein vastanneiden lasten prosenttiosuus oli näissä hieman pienempi, 90 prosenttia. Lisäksi yksi lapsi merkitsi pitkän vokaalin yhdellä grafeemilla peräti 10 kertaa. Ero pääpainollisen tavun lyhyissä ja pitkissä vokaaleissa tehtyjen virheiden osalta on tilastollisesti merkittävä todennäköisyydellä 0,030 (df 229,  $t = -2,184$ ), Wilcoxon 0,000 ( $Z = -3,608$ ).

Taulukko 60. Pääpainollisen tavun pitkän vokaalin merkitseminen 1.

Virheitä N	Vastaajia	%
0	207	90,0
1	18	7,8
2	4	1,7
10	1	0,4
Yhteensä	230	100

Virheitä pääpainollisen tavun pitkissä vokaaleissa oli yhteensä 1,3 prosenttia, ja oikeita vastauksia oli 90,0 prosenttia. Pääpainollisia pitkiä vokaaleja oli 12 epäsa-  
nassa, joten jaettu summa on 3, eli lapset tekivät keskimäärin 3 virhettä sanaa kohti.

Taulukko 61. Pääpainollisen tavun pitkän vokaalin merkitseminen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
36	2484	240	2760
1,3	90	8,7	100

**Yksinäisklusiilin merkintä** horjui jonkin verran toisen tavun alussa, mutta oikeita vastauksia on kuitenkin prosentuaalisesti enemmän kuin pääpainollisen tavun pitkän vokaalin merkitsemisessä (Taulukot 62 ja 63). 14 lasta merkitsi kerran lyhyen klusiilin pitkänä. Sitä lukuisimmat virheet ovat harvinaisia. Jaettu summa on 4. Pitkän vokaalin jälkeisessä lyhyessä klusiilissa jaettu summa on 3,5. Ero virheiden välillä sanatyypeissä CV.tV – CVV.tV ei ole tilastollisesti merkittävä (0,644, df 229, t 0,462) Wilcoxon 0,684 (Z -0,407).

Taulukko 62. Lyhyen klusiilin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä 1.

Virheitä N	Vastaajia	%
0	212	92,2
1	14	6,1
2	3	1,3
4	1	0,4
Yhteensä	230	100

Lyhyen pääpainollisen vokaalin jälkeisissä toisen tavun tavunalkuisissa klusiileissa virheitä oli yhteensä 1,7 prosentissa vastauksista. Hylättyjä vastauksia oli 6,1 prosenttia.

Taulukko 63. Lyhyen klusiilin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
24	1272	84	1380
1,7 %	92,2 %	6,1 %	100 %

Pitkän vokaalin jälkeisen lyhyen klusiilin oikeinkirjoitus hallittiin hyvin. Kahdeksan vastaajaa teki yhden virheen. Lisäksi yksi oppilas teki kaksi virhettä ja yksi neljä virhettä. Virheitä oli kaiken kaikkiaan ainoastaan 1,7 prosentissa kaikista vastauksista (ks. Taulukko 64).



Taulukko 64. Lyhyen klusiilin merkitseminen toisen tavun alussa pitkän vokaalin jäljessä 1.

Virheitä N	Vastaajia	%
0	220	95,7
1	8	3,5
2	1	0,4
4	1	0,4
Yhteensä	230	100

Taulukko 65. Lyhyen klusiilin merkitseminen toisen tavun alussa pitkän vokaalin jäljessä 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
14	880	26	920
1,5	95,7	2,8	100

**Lyhyt soinnillinen konsonantti** näyttää olevan lyhyttä klusiilia helpompi merkitä. Ainakin nasaalin (n) merkintä näyttää olevan helppoa, sillä sen merkinnässä ei tehty lainkaan virheitä. Samoin pitkän vokaalin jälkeen nasaalin merkintä näyttää olevan melko yksiselitteistä. Jaettu summa on siis lyhyen vokaalin jälkeen 0 ja pitkän vokaalin jälkeen 1.

Taulukko 66. Nasaalin merkitseminen toisen tavun alussa lyhyen vokaalin jäljessä.

Virheitä N	Vastaajia	%
0	230	100
Yhteensä	230	100

Taulukko 67. Nasaalin merkitseminen toisen tavun alussa pitkän vokaalin jälkeen 1.

Virheitä N	Vastaajia	%
0	228	99,1
1	2	0,9
Yhteensä	230	100

Oikeita vastauksia oli lyhyen nasaalin merkitsemisessä pitkän pääpainollisen vokaalin jälkeen 99,1 prosenttia. Lyhyttä kvantiteettia merkittiin pitkällä ainoastaan 0,4 prosentissa kaikista vastauksista.

Taulukko 68. Nasaalin merkitseminen toisen tavun alussa pitkän vokaalin jälkeen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
2	456	2	460
0,4 %	99,1 %	0,4 %	100

**Geminaattojen merkintä** oli yksinäiskonsonanttien oikeinkirjoitukseen verrattuna horjuvaa. Etenkin geminaattaklusiilien merkinnässä ilmeni virheitä. Lyhyen pääpainollisen vokaalin jälkeisistä geminaatoista merkitsi kaikki oikein vain 75,2 prosenttia lapsista. Lisäksi monet lapset tekivät useita virheitä tässä asemassa (ks. Taulukko 69 alla). Jaettu summa on lyhyen vokaalin jälkeisissä klusiiligeminaatoissa 17,7. Vastavissa nasaaligeminaatoissa jaettu summa on 8.

Taulukko 69. Klusiiligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jäljessä 1.

Virheitä N	Vastaajia	%
0	173	75,2
1	34	14,8
2	12	5,2
3	4	1,7
4	2	0,9
5	2	0,9
6	3	1,3
Yhteensä	230	100

Lyhyen pääpainollisen vokaalin jälkeisissä klusiiligeminaatoissa tehtiin virheitä 7,7 prosentissa kaikista vastauksista. Hylättyjä vastauksia oli paljon, 17,1 prosenttia. Tähän voi olla syynä se, että kun klusiiligeminaatan merkitsemisessä oli epävarmuutta, näkyi tämä epävarmuus myös siten, että muualla sanassa tehtiin myös useita virheitä. Tämän vuoksi vastaus oli hylättävä.

Taulukko 70. Klusiiligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jäljessä 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
106	1038	236	1380
7,7 %	75,2 %	17,1	100 %

**Nasaaligeminaatan merkintä** taas oli klusiiligeminaattaa helpompaa. Vastauksista oli 94,8 prosenttia oikein (Taulukot 71 ja 72). Erot virheitä tehneiden lasten määrässä sanatyypin CVttV ja CVnnV merkitsemisessä ovat tilastollisesti merkittäviä todennäköisyydellä 0,001 (df 229, t 3,450) Wilcoxon 0,000 (Z -6,034).

Taulukko 71. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jälkeen 1.

Virheitä N	Vastaajia	%
0	218	94,8
1	8	3,5
2	4	1,7
Yhteensä	230	100

Lyhyen pääpainollisen vokaalin jälkeisissä nasaaligeminaatoissa virheitä oli 3,5 prosentissa vastauksista. Hylättyjä vastauksia oli tässä vain 1,7 prosenttia.

Taulukko 72. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla lyhyen vokaalin jälkeen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
16	436	8	460
3,5 %	94,8	1,7	100

**Pitkän vokaalin jälkeen nasaaligeminaatan merkitseminen** oli kaikista sana- asemista vaikeinta. Kaikista vastauksista oli oikein ainoastaan puolet, 50 prosenttia. Koska epäsanaja oli ainoastaan kaksi, ei kahta virhettä enempää virheitä voitu tehdä. Jaettu summa nasaaleissa on 78,5 ja klusiiligeminaatoissa 26. Ero pitkän vokaalin jälkeisten klusiili- ja nasaaligeminaattojen merkitsemisen välillä on tilastollisesti merkittävä todennäköisyydellä 0,000 (df 229 T -9,704), Wilcoxon 0,000 (Z -8,183). Vaikka näitä epäsanaja oli vain kaksi, on tulos todennäköisesti validi, koska vastaajia oli 230. Vastauksia kertyy kahdestakin epäsanasta yhteensä  $2 \times 230 = 460$  kappaletta.

Taulukko 73. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla pitkän vokaalin jälkeen 1.

Virheitä N	Vastaajia	%
0	115	50,0
1	73	31,7
2	42	18,3
Yhteensä	230	100

Pitkän vokaalin jälkeisistä nasaaligeminaattojen oikeinkirjoituksessa oli virheitä 34,1 prosentissa vastauksista. Oikeita vastauksia oli vain puolet, 50 prosenttia (ks. Taulukko 74).

Taulukko 74. Nasaaligeminaattojen merkitseminen 1. ja 2. tavun rajalla pitkän vokaalin jälkeen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
157	230	73	460
34,1 %	50,0 %	15,9	100 %

**Pitkän vokaalin jälkeisissä klusiiligeminaatoissa** virheitä ei sen sijaan tehty paljonkaan enemmän kuin lyhyen vokaalin jälkeisissä klusiiligeminaatoissa. Oikein vastanneiden prosentuaalinen osuus oli 76,5 prosenttia, kun sen lyhyen vokaalin jälkeisissä klusiiligeminaatoissa oli 75,2 prosenttia. Ero näiden kahden välillä ei ole tilastollisesti lainkaan merkitsevä (0,872 df = 229, T 0,161), kun taas ero sanatyyppeihin CVVnnV ja CVVttV vastanneiden lasten välillä ovat tilastollisesti merkittäviä todennäköisyydellä 0,000 (df 229, t -9,704) Wilcoxon 0,000 (Z -8,183).

Taulukko 75. Klusiiligeminaattojen merkitseminen pitkän vokaalin jälkeen 1.

Virheitä N	Vastaajia	%
0	176	76,5
1	27	11,7
2	10	4,3
3	11	4,8
4	6	2,6
Yhteensä	230	100

Pitkän vokaalin jälkeisistä klusiiligeminaatoista oli oikein 76,5 prosenttia ja väärin 10,2.

Taulukko 76. Klusiiligeminaattojen merkitseminen pitkän vokaalin jälkeen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
94	704	122	920
10,2 %	76,5 %	13,3 %	100 %

**Toisen tavun yksinäisvokaali** merkittiin säännöllisesti yhdellä grafeemilla. Ainoastaan viisi lasta 230:stä merkitsi yhden kerran lyhyen vokaalin kahdella grafeemilla. Jaettu summa on 0,21.

Taulukko 77. Toisen tavun yksinäisvokaalin merkitseminen 1.

Virheitä N	Vastaajia	%
0	225	97,8
1	5	2,2
Yhteensä	230	100

Toisen tavun lyhyissä vokaaleissa virheitä oli erittäin vähän, 0,09 prosenttia, kun taas oikeita vastauksia oli 97,8 prosenttia (Taulukko 78).

Taulukko 78. Toisen tavun yksinäisvokaalin merkitseminen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
5	5400	115	5520
0,09 %	97,8	2,1	100

**Toisen tavun pitkän vokaalin merkintä** sen sijaan oli yksinäisvokaalin merkintää hieman mutkikkaampaa. Oikein vastanneita oli lapsista 87,4 prosenttia. Pitkää vokaalia merkitsi yhdellä grafeemilla yhteensä 28 oppilasta. Toisen tavun pitkissä vokaaleissa jaettu summa on 10,5. Ero lyhyen ja pitkän toisen tavun vokaalin merkitsemisessä on tilastollisesti merkittävä todennäköisyydellä 0,000 (df 229, T -4,895) Wilcoxon 0,000 (Z -4,836).

Taulukko 79. Toisen tavun pitkän vokaalin merkitseminen 1.

Virheitä N	Vastaajia	%
0	201	87,4
1	17	7,4
2	9	3,9
3	1	0,4
4	1	0,4
Yhteensä	230	100

Toisen tavun pitkistä vokaaleista merkittiin oikein 87,4 prosenttia ja väärin 4,6 prosenttia. Toisen tavun pitkän vokaalin oikeinkirjoitus näyttää siis olevan hieman vaikeampaa kuin toisen tavun lyhyen vokaalin merkitseminen.

Taulukko 80. Toisen tavun pitkän vokaalin merkitseminen 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
42	804	74	920
4,6 %	87,4 %	8,0 %	100 %

**Yksinäisklusiilia** merkittiin ensimmäisen ja toisen tavun rajalla yhdellä grafeemilla 92,6 prosenttisesti oikein. Tosin 17 oppilasta merkitsi lyhyen klusiilin kahdella grafeemilla ainakin kerran (ks. Taulukko 81 alla). Lyhyen nasaalin /m/ merkitsi oikein 95,7 prosentissa vastaajista. Jaettu summa klusiileissa on 3,75 ja nasaaleissa 2,75.

Taulukko 81. Lyhyen klusiilin merkitseminen toisen ja kolmannen tavun rajalla 1.

Virheitä N	Vastaajia	%
0	213	92,6
1	13	5,7
2	1	0,4
4	1	0,4
5	1	0,4
6	1	0,4
Yhteensä	230	100

Kolmannen tavun tavunalkuisen yksinäisklusiilin oikeinkirjoituksessa virheitä oli yhtä paljon kuin lyhyen klusiilin merkitsemisessä toisen tavun alussa (1,7 prosenttia vastauksista oli väärin).



Taulukko 82. Lyhyen klusiilin merkitseminen toisen ja kolmannen tavun rajalla 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
30	1704	70	1804
1,7 %	94,5 %	3,9 %	100 %

Kolmannen tavun tavunalkuisen nasaalin merkitsi oikein 95,7 prosenttia vastaajista, kun kolmannen tavun tavunalkuinen klusiili merkittiin oikein 92,6 prosentissa vastauksia. Ero klusiilien ja nasaalien merkitsemisessä ei ole tilastollisesti merkittävä (0,410, df 229, t 0,825) (Wilcoxon 0,836, Z -0,207).

Taulukko 83. Lyhyen nasaalin merkitseminen toisen ja kolmannen tavun rajalla 1.

Virheitä N	Vastaajia	%
0	220	95,7
1	9	3,9
2	1	0,4
Yhteensä	230	100

Kolmannen tavun tavunalkuinen nasaali merkittiin oikein 95,7 prosenttisesti. Vääriä vastauksia oli ainoastaan 1,2 prosenttia kaikista vastauksista.

Taulukko 84. Lyhyen nasaalin merkitseminen toisen ja kolmannen tavun rajalla 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
11	880	29	920
1,2 %	95,7 %	3,2 %	100 %

**Toisen ja kolmannen tavun rajalla klusiiligeminaatat** merkitsi oikein 80,9 prosenttia vastaajista, mikä on hieman enemmän kuin klusiiligeminaatoissa

ensimmäisen ja toisen tavun rajalla (75,2 %). Jaettu summa on klusiiligeminaatoissa 10,5, ja nasaaleissa 7,75. Erot klusiilien ja nasaalien merkitsemisessä eivät ole tilastollisesti merkittäviä (0,246, df 229, t 1,164) Wilcoxon 0,341 Z -0,952).

Taulukko 85. Klusiiligeminaatan merkitseminen 2. ja 3. tavun rajalla 1.

Virheitä N	Vastaajia	%
0	186	80,9
1	26	11,3
2	9	3,9
3	2	0,9
4	3	1,3
5	2	0,9
6	2	0,9
Yhteensä	230	100

Kaiken kaikkiaan toisen- ja kolmannen tavun rajalla olevista klusiiligeminaatoista kirjoitettiin oikein 80,9 prosenttia ja kvantiteettivirhe oli 9,1 prosentissa vastauksista (Taulukko 86). Hylättyjä vastauksia oli 14,6 prosenttia, mikä on suuri määrä.

Taulukko 86. Klusiiligeminaatan merkitseminen 2. ja 3. tavun rajalla 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
84	1488	268	1840
4,6 %	80,9 %	14,6 %	100 %

Toisen- ja kolmannen tavun rajalla olevat nasaaligeminaatat merkitsi oikein 208 lasta 230:stä. Prosentuaalisesti tämä on 90,4 prosenttia.

Taulukko 87. Nasaaligeminaattojen merkitseminen 2. ja 3. tavun rajalla 1.

Virheitä N	Vastaajia	%
0	208	90,4
1	18	7,8
2	1	0,4
3	1	0,4
4	2	0,9
Yhteensä	230	100

Toisen- ja kolmannen tavun rajalla olevista nasaaligeminaatoista oli oikein 90,4 prosenttia ja kvantiteettivirhe oli 3,4 prosentissa vastauksista. Virheitä oli hieman vähemmän kuin vastaavissa klusiiligeminaatoissa (3,4 vs. 4,6 %).

Taulukko 88. Nasaaligeminaattojen merkitseminen 2. ja 3. tavun rajalla 2.

Virheitä	Oikeita vastauksia	Hylätty	Vastauksia yhteensä
31	832	57	920
3,4 %	90,4 %	6,2 %	100 %

Seuraavassa alaluvussa kokoaan oikeinkirjoitustestin tulokset ja tarkastelen niitä frekvenssilukujen ja autosegmentaalisen mallin avulla.

### 7.3 Tiivistys oikeinkirjoitustestistä

Alla olevaan taulukkoon 89 on laskettu kaikista muuttujista (eri sana-asetat) jaetut summat. Jaettu summa on laskettu siten, että jokaisessa grafeemiasemassa esiintyvät virheet on jaettu niiden sanojen määrällä, joissa kyseessä oleva virhe esiintyy. Esimerkiksi pitkän vokaalin jälkeinen nasaaligeminaatta esiintyi kahdessa sanassa. Tällöin jaettu summa on laskettu jakamalla kyseisissä sanoissa esiintyvien virheiden määrä kahdella ( $157 / 2 = 78,5$ ). Tämä luku siis kertoo virheiden

suhteellisen määrän per epäsana. Eniten virheitä esiintyy siis pitkän vokaalin jälkeisissä nasaaleissa (jaettu summa 78,5) ja toiseksi eniten pitkän vokaalin jälkeisissä klusiiligeminaatoissa (jaettu summa 26). Taulukossa 89 oikealla oleva sarake kertoo, paljonko virheitä tehtiin prosentteina suhteutettuna niiden sanojen määrään, joissa oli kyseinen äänne. Seuraavaksi pyrin esittämään kvantiteetin merkitsemisessä esiintyneet virheet tavutyyppin mukaan. Lyhyissä pääpainollisissa vokaaleissa ei tehty juurikaan virheitä (0,4 %), mutta pitkissä vokaaleissa virheitä oli hieman enemmän (1,3 prosentissa vastauksista). Geminaatat osoittautuivat odotetusti yksinäiskonsonantteja vaikeammiksi. Kun yksinäisklusiileissa tehtiin virheitä 24 kappaletta (cv.Tv) ja 14 kappaletta (cvv.Tv) eli yhteensä 38, oli virheitä geminaatoissa 165 kappaletta. Ylipitkät tavut osoittautuivat kaikista hankalimmiksi. Suluissa olevat luvut kertovat virheiden määrän testissä eri segmenttien kohdalla. Prosentit perustuvat jaettuun summaan (so. ne on vakioitu sanojen määrän mukaan).

Näyttää siltä, että suurin osa virheistä kasaantuu ylipitkiin tavuihin, eli virheiden määrä riippuu tavutason muuttujista: tavun pituudesta, tavutyyppin yleisyydestä ja tavutyyppin kompleksisuudesta. Tämän puolesta todistaa se, että ylipitkissä tavuissa tehtiin enemmän virheitä kuin toisen ja kolmannen tavun rajalla olevassa geminaatassa (11,3 vs. 4,6 %).

Näyttääkin siltä, että virheluokat korreloivat sanahahmojen frekvenssiin juoksevassa tekstissä. Nasaaligeminaatat ylipitkässä tavussa ovat juoksevassa tekstissä harvinaisin näistä tavutyypeistä (frekvenssi 0,9 %). Toiseksi harvinaisin on klusiiligeminaatat (1,6 prosenttia). Toisen ja kolmannen tavun rajalla olevien geminaattojen yleisyyttä juoksevasta tekstistä ei ole laskettu laskentateknisten vaikeuksien takia. Pitkiä vokaaleja on 6,3 prosenttia kaikista toisen tavun vokaaleista.

Sanahahmoparien CVttV ja X.CVppV välillä on tilastollisesti merkitsevä ero (0,001 df 229, t 3,407) Wilcoxon 0,000 (Z -3,702). Tämä todistaa sen puolesta, että ensi- ja toisen tavun rajaa kauempana olevat klusiiligeminaatat ovat helpompi kirjoittaa kuin ensi- ja toisen tavun rajalla olevat. Ero selittyy siis sana-asemalla eli sanatason muuttujilla. Syynä tähän voi olla se, että yksinäisklusiileja ei esiinny omaperäisessä

sanastossa muualla kuin ensimmäisen ja toisen tavun rajalla. Nasaalien välillä sana-hahmoissa CVn.nV ja X.CVm.mV ei ole tilastollista eroa (0,889, df 229, t 0,140) Wilcoxon 0,821 (Z -0,226). Tässä sana-asema ei näytä vaikuttavan.

Yksinäiskonsonanttien merkintään sana-asema ei liioin vaikuta. Toisen tavun tavun-alkuisien klusiilien ja kolmannen tavun tavunalkuisten klusiileiden välillä ei myöskään ole tilastollisesti merkitsevää eroa (0,582, df 229, t -0,552) (Wilcoxon 0,157 df 229, Z -1,415). Nasaalien välillä on tilastollinen ero toisen tavun tavunalkuisten ja kolmannen tavun tavunalkuisten nasaalien välillä (T-testi 0,002, df 229, t -3,108) (Wilcoxon 0,002, Z -3,051). Tämä johtuu siitä, että toisen tavun tavunalkuisissa nasaaleissa lyhyen vokaalin jälkeen ei tehty yhtään virhettä, mutta kolmannen tavun tavun-alkuisissa konsonanteissa virheitä oli 1,2 prosentissa.

Taulukko 89. Tiivistys oikeinkirjoitustestin tuloksista.

Grafeemiasema	Jaettu summa	%
cV.cv...	0,8	0,4
cVV.cv...	3	1,3
cv.Tv...	4	1,7
cv.Nv...	0	0
cvv.Tv...	3,5	1,5
cvv.Nv	1	0,4
cvTTv...	17,7	<b>7,7</b>
cvNNv...	8	<b>3,5</b>
cvvTTv...	26	<b>11,3</b>
cvvNNv...	78,5	<b>34,1</b>
x.cV.cv	0,2	0,1
x.cVV.cv	10,5	4,6
x.cv.Pv	3,8	1,7
x.cv.Mv	2,8	1,2
x.cvPPv	10,5	<b>4,6</b>
x.cvMMv	7,8	<b>3,4</b>

Virheitä tehtiin sana-aseman mukaan eniten seuraavissa paikoissa (ks. Taulukko 90 alla). Taulukon ensimmäisessä sarakkeessa on sanahahmo, jossa kyseinen segmentti on. Taulukon viimeiseen sarakkeeseen on merkitty kyseisen segmentin yleisyys ensimmäisen sarakkeen sanatyypissä. Frekvenssi on laskettu juoksevasta tekstistä (ks. luku 5). Kohdassa 4 on neljä eri kohtaa (a, b, c ja d) siksi, että näiden välillä ei ole virheiden yleisyyden mukaan tilastollista eroa.

Taulukko 90. Virheelliset segmentit ja niiden sana-asemat järjestettynä virheiden määrän perusteella.

Sanahahmo	Segmenttityyppi	Virheitä	Frekvenssi
1) CVV <u>n</u> .nV	nasaaligeminaatta	34,1 %	0,9 %
2) CVV <u>t</u> .tV	klusiiligeminaatta	11,3 %	1,6 %
3) CV <u>t</u> .tV	klusiiligeminaatta	7,7 %	5,2 %
4a) X.CV <u>p</u> .pV	klusiiligeminaatta	4,6 %	-----
4b) X.CV <u>V</u> .CV	painoton pitkä vokaali	4,6 %	6,3 %
4c) CV <u>n</u> .nV	nasaaligeminaatta	3,5 %	15,6 %
4d) X.CV <u>m</u> .mV	nasaaligeminaatta	3,4 %	-----

Seuraavat tilastolliset erot eri sana-aseman mukaan ovat tilastollisesti merkitseviä:

Ero ylipitkien tavujen nasaaligeminaattojen ja klusiiligeminaattojen (1 vs. 2) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,000 (df 229 T  $-9,704$ ); Wilcoxon 0,000 (Z  $-8,183$ ). Nasaaligeminaatat olivat helpompia merkitä (3,5 vs. 4,6 %).

Ero ylipitkien tavujen nasaaligeminaattojen ja klusiiligeminaattojen (2 vs. 3) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,000 (df 229 T  $-9,704$ ); Wilcoxon 0,000 (Z  $-8,183$ ). Klusiiligeminaatat olivat helpompia merkitä (11,3 vs. 34,1 %).

Ero ylipitkien tavujen klusiiligeminaattojen ja pitkien tavujen klusiiligeminaattojen (2 vs. 3) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,002 (df 229,  $T -3,079$ ); Wilcoxon 0,002, ( $Z -3,056$ ). Lyhyen tavun jälkeiset klusiiligeminaatat olivat helpompia merkitä (7,7 vs. 11,3 %).

Ero klusiiligeminaattojen ja toisen ja kolmannen tavun rajalla olevien klusiiligeminaattojen (3 vs. 4a) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,001 (df 229,  $t 3,407$ ) Wilcoxon 0,000 ( $Z -3,702$ ). Toisen ja kolmannen tavun rajalla olevat klusiiligeminaatat olivat helpompia (4,6 vs. 7,7 %).

Ero ensitavun ja toisen tavun rajalla olevien klusiiligeminaattojen ja toisen tavun pitkien vokaalien välillä (3 vs. 4b) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,009 (df 228,  $t 2,618$ ), Wilcoxon 0,000 ( $Z -4,120$ ). Toisen tavun pitkä vokaali oli helpompi merkitä (4,6 vs. 7,7 %).

Ero ensi- ja toisen tavun rajaisten klusiiligeminaattojen ja samassa asemassa olevien nasaaligeminaattojen (3 vs. 4c) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,001 (df 229  $T 3,450$ ), Wilcoxon (0,000,  $Z -3,971$ ).

Ero ensi- ja toisen tavun tavunrajaisten klusiiligeminaattojen ja toisen ja kolmannen tavun tavunrajaisten nasaaligeminaattojen (3 vs. 4d) merkitsemisessä on tilastollisesti merkitsevä todennäköisyydellä 0,001 (df 229,  $T 3,414$ ); Wilcoxon 0,001, ( $Z -3,382$ ). Nasaaligeminaatat olivat helpompia merkitä (3,4 vs. 7,7 %).

Erot kohdan 4 sisällä eivät ole tilastollisesti merkittäviä. Ero toisen ja kolmannen tavun rajalla olevien klusiiligeminaattojen ja toisen tavun pitkien vokaalien (4a vs. 4b) merkitsemisessä ei ole tilastollisesti merkittävä todennäköisyydellä 0,743 (df 228,  $T 0,328$ ); Wilcoxon 0,685, ( $Z -0,406$ ). Myöskään ero toisen tavun pitkien vokaalien ja ensi- ja toisen tavun tavunrajaisten nasaaligeminaattojen (4b ja 4c) merkitsemisessä ei ole tilastollisesti merkitsevä todennäköisyydellä 0,234 (df 228,  $T 1,194$ ); Wilcoxon 0,249, ( $Z -1,152$ ). Samoin ero toisen ja kolmannen tavun rajalla olevien klusiiligeminaattojen ja ensi- ja toisen tavun tavunrajaisten nasaaligeminaattojen merkit-

semisessä (4a vs. 4c) ei ole tilastollisesti merkittävä todennäköisyydellä 0,246 (df 229, T 1,164); Wilcoxon 0,341 (Z -0,952).

Kolmitavuisten epäsanon oikeinkirjoitustestin tulosten mukaan pituusoppositio on helpointa merkitä pääpainollisen tavun vokaaleissa (virheiden laskemisessa on käytetty jaettua summaa). Pääpainollisen tavun pitkistä ja lyhyistä vokaaleista oli merkitty oikein 98,3 prosenttia (1,7 % väärin). Toiseksi helpoin asema on nasaalin pituusoppositio lyhyen pääpainollisen vokaalin jälkeen. Kolmanneksi helpoin asema on nasaalin pituusoppositio toisen ja kolmannen tavun rajalla. Edelleen pitkä ja lyhyt klusiili on helpompi merkitä toisen ja kolmannen tavun rajalla kuin ensimmäisen ja toisen tavun rajalla. Tämä on yllättävää sikäli, että pituusoppositio luulisi olevan helpoin hallita pääpainollisen tavun jälkeen. Kaikista vaikeinta on opposition hallinta nasaaleissa pitkän vokaalin jälkeen: näissä virheitä oli yhteensä 34,5 prosentissa. Tämä luku on kuitenkin harhaanjohtava siltä osin, että pitkän pääpainollisen vokaalin jälkeisen lyhyen nasaalin oikeinkirjoitus on helppoa (0,4 % vääriä vastauksia).

Taulukko 91. Pituusopposition hallinta sanan eri asemissa.

Sana-asema	Virheitä yhteensä	Virheitä yhteensä %
1) (C)V vs. (C)VV	3,8	1,7
2) (C)V.nV vs. (C)Vn.nV	8	3,5
3) S.CV.mV vs. S.CVm.mV	10,5	4,6
4) S.Cv vs. S.Cvv	10,7	4,7
5) S.CV.pV vs. S.CVp.pV	14,3	6,3
6) (C)V.tV vs. (C)Vt.tV	21,7	9,4
7) (C)VV.tV vs. (C)VVt.tV	29,5	12,8
8) (C)VV.nV vs. (C)VVn.nV	79,5	34,5

S = mikä tahansa pääpainollinen tavu.

Näyttää siltä, että soinnillisten segmenttien oikeinkirjoitus oli helpompaa kuin klusiilien (Taulukko 92). Lyhyt nasaali /n/ merkittiin väärin geminaattana ainoastaan 0,4 prosentissa vastauksia (virheitä 2). Lyhyttä kolmannen tavun tavunalkuista nasaalia



merkittiin kaksoiskonsonantilla 1,2 prosentissa tapauksista. Lyhyen vokaalin jälkeisistä nasaaligeminaatoista merkittiin yksinäiskonsonantilla vain 3,5 prosentissa pääpainollisen tavun ja toisen tavun rajalla olevista tapauksista. Toisen ja kolmannen tavun rajaisista nasaaligeminaatoista 3,4 prosenttia merkittiin virheellisesti yksinäiskonsonantilla. Klusiileissakin lyhyt klusiili merkitään pääsääntöisesti yhdellä konsonantilla. Ainoastaan 1,7 prosenttia lyhyen pääpainollisen vokaalin ja 1,5 prosenttia pitkän pääpainollisen vokaalin jälkeisistä toisen tavun alkukonsonanteista kirjoitettiin kahdella grafeemilla. Kolmannen tavun alkukonsonantteina olleista klusiileista 1,7 prosenttia merkittiin virheellisesti geminaattana. Klusiiligeminaatoista lyhyen pääpainollisen vokaalin jälkeen merkittiin virheellisesti 7,7 prosentissa vastauksista ja pitkän pääpainollisen vokaalin jälkeen 11,3 prosentissa. Toisen ja kolmannen tavun välisissä geminaatoissa virheitä oli vähemmän, 4,6 prosenttia. Pitkän vokaalin jälkeiset nasaaligeminaatat erottuvat omaksi ryhmäkseen: niistä 34,1 prosenttia kirjoitettiin väärin yksinäiskonsonantilla.

Taulukko 92. Vastaukset segmentin tyyppin mukaan laskettuna.

Grafeemiasema	Jaettu summa	%
x.Nv...,x.Mv	3,8	1,6
x.Tv..., x.Pv	11,3	4,9
xN.Nv..., xm.mv	15,8	6,9
xT.Tv..., xP.Pv	48,3	23,6
cvvN.Nv...	78,5	34,1

Tulosten perusteella klusiilien merkitseminen oikein näyttää olevan vaikeampaa kuin soinnillisten konsonanttien. Lyhyen vokaalien jälkeisten klusiiligeminaattojen merkitseminen on huomattavasti vaikeampaa kuin lyhyen vokaalin jälkeisten nasaaligeminaattojen (23,6 vs. 6,9 %). Samoin yksinäisklusiilien merkinnässäkin on horjuntaa enemmän kuin yksinäisnasaalien merkinnässä (4,9 vs. 1,6 %). Kuitenkin kaikkein vaikein asema kirjoitusprosessin kannalta on nasaaligeminaattojen merkitseminen pitkän vokaalin jälkeen. Näissä pitkäa nasaalia merkittiin yksinäiskonsonantilla 34,1 prosentissa vastauksista. Syynä lienee se, että sanastossa ei ole muutamaa poikkeusta lukuunottamatta sanoja, joissa on ylipitkissä tavuissa soinnil-

linen geminaatta. Poikkeuksen tähän muodostavat heikkoasteiset muodot, kuten *e'*-nominat (*jääne, kiille*) ja umpitavun edellä olevat soinnilliset geminaatat (*käännös*).

## 8. Lopuksi

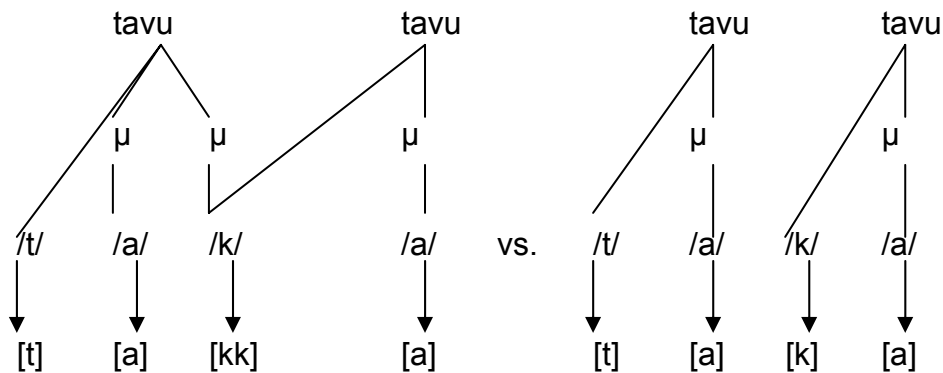
Olen edellä tarkastellut kvantiteettierojen fonologiaa ja osoittanut, että fonologisesti se voidaan kuvata parhaiten autosegmentaalisen fonologian avulla. Olen myös osoittanut, miten kirjoituksen kuvaamisessa yksinkertainen IA-malliin perustuva segmentaalinen fonologinen esitys ei vastaa nykykäsitystä fonologiasta eikä sen avulla voi kuvata fonologisesti kvantiteettieroja.

Kvantiteetin merkitseminen kirjoituksessa eroaa muiden segmenttien koodaamisesta siinä, että pitkä kvantiteetti sisältää segmentaalisen tason ja kahden autosegmentaalisen tason, sana- ja tavutason informaatiota. Kuten kirjoitustestin virheanalyysi osoitti, autosegmentaalisten tasojen informaation kompleksisuus myös ennustaa myös pituuden merkitsemistä kirjoituksessa. Lyhyt vokaali on yksimorainen ja tavunalkuinen konsonantti liittyy joko tavutasoon tai moratasoon, tässä kuvauksessa tavutasoon. Niiden koodaaminen kirjoituksessa ei tuota vaikeuksia.

Pitkä vokaali on kaksimorainen, fonologisessa mallissa se liittyy assosiaatioviivoilla kahteen moraan eli puheen ajoitusyksikköön. Fonologisen teorian pohjalta voidaan olettaa, että pitkän konsonantin merkitseminen on vaikeampaa kuin lyhyen, koska geminaatta liittyy sekä moratasoon että tavutasoon ja sen merkitsemisessä tarvitaan siis kolmen tason informaatiota. Näin pitkän eli geminaattakonsonantin koodaaminen vaatii enemmän ”prosessointiresursseja” esimerkiksi lyhytkestoisen muistin kapasiteettia kuin yksinäiskonsonantin koodaaminen. Ja näin näyttää olevankin. Aktuaalisessa prosessoinnissa on tietenkin mahdollista, että kirjoittaja joutuu myös yksinäiskonsonanttia kirjoittaessaan miettimään, voisiko tai pitäisikö valita yksinäiskonsonantti vai geminaatta tai kummasta on kyse. Virheanalyysin perusteella tämä ei kuitenkaan vaikuta olennaiselta pituuden koodaamisessa, jossa ongelma on nimenomaan pitkän koodaaminen. Edelleen aktuaalisessa kirjoitusprosessissa on ehkä mukana myös leksikaalista ja morfologista tietoa esimerkiksi pituuden ennakoinnissa astevaihtelun osalta ja tuotoksen monitoroinnissa. Tässä rajoitin kuitenkin vain fonologisen kuvauksen ennustavuuteen pitkän ja lyhyen kvantiteetin koodaamisessa.

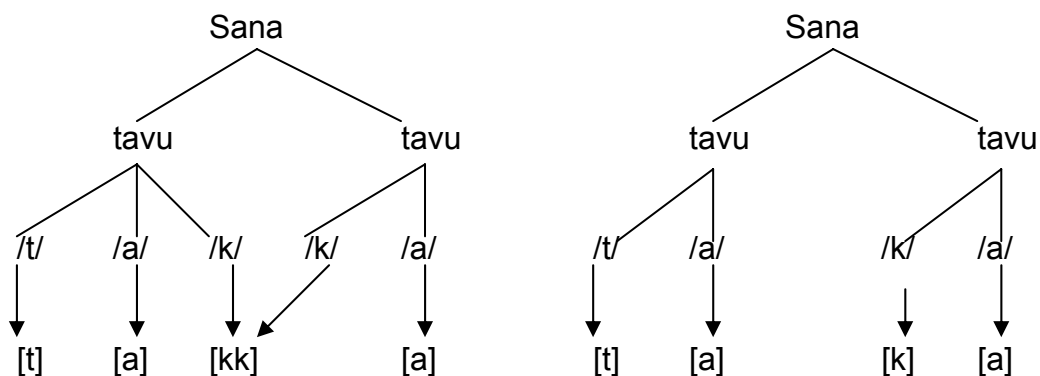
Myös frekvenssitieto pitkien konsonanttien harvinaisuudesta tukee konsonanttien pituuden kompleksisuutta eli tässä monitasoisuuden tulkintaa. Pitkä konsonantti on lyhyeen verrattuna harvinaisempi niin tyyppi- kuin käyttöfrekvenssin osalta.

Seuraavassa mallinnan tämän fonologisen tiedon morateoriaa käyttäen ja ilman morateoriaa. Morakäsitettä hyödyntäen pitkä ja lyhyt geminaatta kuvataan seuraavasti:



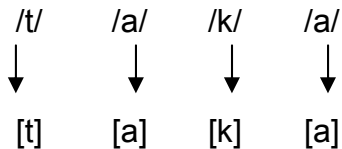
Kuvio 42. Pitkä ja lyhyt konsonantti kuvattuna morateoriaa käyttäen.

Jos sama asia halutaan kuvata ilman morateoriaa, niin voidaan käyttää ns. Flat-Structure-mallia, jossa ei oleteta segmentaalisen ja tavutason lisäksi olevan muita sanatasoa alempia tasoja. Tässä mallissa siis segmentit liittyvät suoraan tavutasoon ja tavutasot sanatasoon.



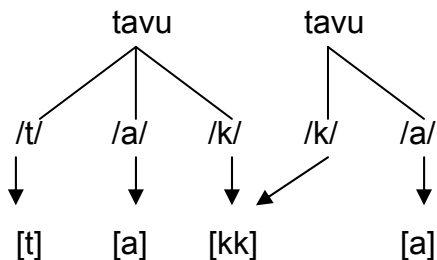
Kuvio 43. Pitkä ja lyhyt konsonantti kuvattuna ilman morateoriaa.

Kirjoitettaessa joudutaan kvantiteettia merkittäessä pitämään mielessä (ilmeisesti lyhytkestoisessa muistissa) useiden segmenttien kokonaisuus. Sanassa /taka/ jokainen segmentti edustaa ”pelkästään” itseään eli jokainen eri äänne voidaan muuttaa kirjaimiksi.



Kuvio 44. Grafeemin ja segmentin vastaavuus lyhyissä konsonanteissa.

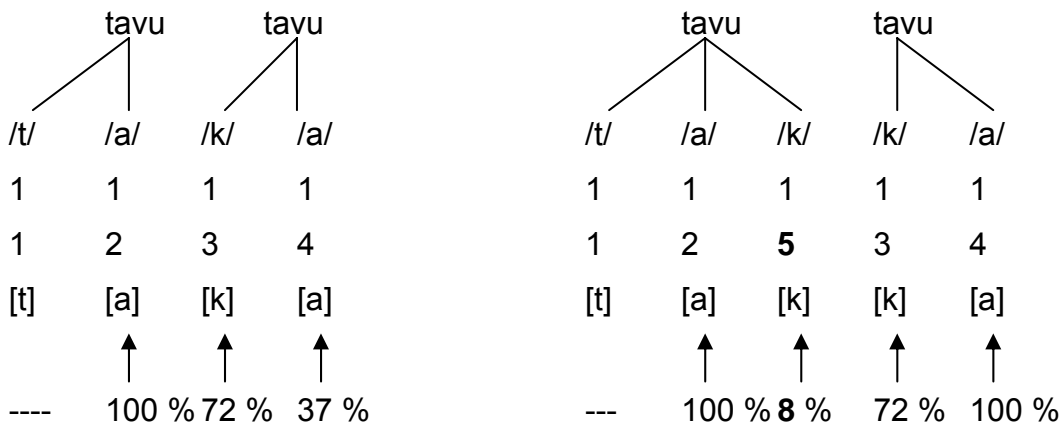
Pitkää kvantiteettia merkittäessä joudutaan teoriasta riippumatta käyttämään hyväksi useiden eri tasojen informaatiota. Ensinnäkin geminaatta kuuluu kahteen eri tavuun: tällöin joudutaan käyttämään tavun rajan ylittävää informaatiota. Toiseksi koska konsonantti ei ole koskaan tavua muodostava, joudutaan pitkää kvantiteettia prosessoitaessa sisällyttämään samaan prosessiin myös (ainakin) seuraavan tavun vokaali.



Kuvio 45. Grafeemin ja segmentin vastaavuus pitkissä konsonanteissa.

Edellä kuvattua kirjoitusprosessia voidaan havainnollistaa seuraavalla tavalla. Kuvitellaan, että sanan jokaisella segmentillä on työmuistin toiminnan yksikköinä ominaisarvo 1. Sanassa /taka/ tämä tekee yhteensä 4 yksikköä. Sanan pituuden kasvaessa kasvaa luonnollisesti edellä mainittu työmuistia rasittavien yksiköiden summakin. Kuvitellaan edelleen, että kielessä ei olisi kvantiteettioppositioita. Tällöin prosessissa sanansisäisen konsonantin kohdalla yksiköiden summa olisi 3. Pituus-

oppositio mutkistaa tilannetta huomattavasti. Jotta konsonantin pituus voidaan prosessoida, täytyy myös seuraava tavu olla samassa prosessissa mukana. Tällöin geminaatta saa vähintään arvon 5 (jos toinen tavu on lyhyt), koska prosessointiin tulee mukaan assosiaatioiden luominen sana- ja tavutasolle. Jos pituusdistinktio jätetään kirjoituksessa prosessoimatta, selvittää kolmen yksikön prosessoinnilla. Kun malliin yhdistetään juoksevan tekstin frekvenssiluvut, on helppo nähdä, miksi kvantiteetti on kirjoituksen oppimisessa viimeiseksi opittava asia.



Kuvio 46. Tavurakenne ja frekvenssi.

Mallini avulla on helppo ymmärtää, miksi lukemaan ja kirjoittamaan oppivat lapset tekevät eniten virheitä juuri geminaattojen merkitsemisessä, mutta eivät tee juuri lainkaan virheitä lyhyen konsonantin merkitsemisessä. Alla on koottu vielä lyhyen ja pitkän konsonantin merkitsemisessä tehdyt virheet. Yksinäisklusiilien kirjoittamisessa tehtiin virheitä lyhyen vokaalin jälkeen 7,8 prosentissa epäsanoista. Lyhyen ja pitkän vokaalin jälkeisissä yksinäisklusiileissa tehtyjen virheiden välillä ei kuitenkaan ole tilastollisesti merkittävää eroa (0,644, df 229, t 0,462). Lyhyen vokaalin jälkeisissä geminaatoissa virheitä tehtiin sitä vastoin runsaasti, 24,8 prosentissa sanoista. Pitkän vokaalin jälkeisissä geminaatoissa virheitä oli 23,4 prosentissa. Ero näiden välillä on tilastollisesti merkittävä (0,002, df 229, t -3,079). Lyhyissä nasaaleissa ei tehty virheitä juuri lainkaan. Ainoastaan pitkän vokaalin jälkeen oli kaksi virhettä. Toisen ja kolmannen tavun rajalla virheitä oli yksinäisklusiileissa 7,3 prosentissa ja nasaaleissa

(/m/) 4,3 prosentissa. Virheiden määrä on samansuuntainen kuin ensimmäisen ja toisen tavun rajallakin.

Erot yksinäiskonsonantin ja geminaatan merkitsemisen välillä ovat siis tilastollisesti merkittäviä kaikissa sana-asemissa. Koska yksinäiskonsonantteja ei merkitä geminaatoiksi, pituuden merkitsemistä ei voi redusoida foneemien lineaariseen representaatioon. Tilanne olisi toinen, jos lyhyttäkin segmenttiä merkittäisiin kirjoitusprosessissa säännöllisesti pitkänä kahdella grafeemilla. Tätä voisi pitää aitona horjuntana lyhyen ja pitkän kvantiteetin välillä. Näiden empiiristen tulosten valossa näyttääkin siltä, että kyseessä on pikemminkin **tendensi jättää kvantiteetti huomiotta**, koska sen merkitseminen vaatii epäsanaympäristössä paljon lyhytkestoisen muistin resursseja, ja merkitä kirjoitusprosessissa vain äänteen, erityisesti konsonanttien, kvaliteetti yhdellä grafeemilla.

Tässä on kyse prosessista, ja silloin oikea termi on "fonologisen informaation linearisointi grafeemiseksi esitykseksi kirjoitusprosessissa". Työssäni mallinnetaan siis tämän prosessin luonnetta formaalisti ja eksplisiittisesti fonologian teorioista lähtien (vs. taksonomisesta ja spesifioimattomasta "teoriasta" lähtien). Lähtökohtaisesti fonologinen informaatio on monitasoista, ja siis se tulee ottaa huomioon mallia laadittaessa. Tässä myös oletetaan, että kirjoitus etenee grafeemeittain vasemmalta oikealle, vaikka ei välttämättä niin, että tuottamisyksiköt ovat grafeemeja.

Tietysti on myös mahdollista sekin, että puhe "käännetään" suoraan kirjoitukseksi joillakin (primitiivisillä) assosiaatioilla tai leksikaalisten yksiköiden kirjoittamista koskevilla muistisäännöillä, hieman kuin konnektionistisesti puheen ja kirjoituksen yksiköitä yhdistelemällä. Tai sitten lienee mahdollista sekin, että fonologista tietoa ei käytetä ollenkaan (vaan se esim. syntyy kirjoitusprosessin myötä). Saattaahan tällaistakin tapahtua aktuaalisessa kirjoittamisessa aina ja se tietysti sotkee "selviä ja suorita" vastaavuuksia fonologisen mallintamisen ja sen empiirisiä tuotoksia kuvaavan vaikutuksen välillä eli se vähentää fonologisen mallin ennustavuutta. Tämä ei kylläkään päde tutkituissa pseudosanoissa, jotka olivat lapsille uusia ja joilla ei ole valmista assosiaatiota kirjoitukseen vaan joissa kuultu on muutettava fonologiseksi ja grafee-

miseksi esitykseksi. Mallit ovat kuitenkin malleja, niillä ei koskaan ole yksi-yhteen vastaavuutta todellisuuteen.

Virheanalyysin tuloksissa korostuu myös frekvenssinäkökulma. Nasaaligeminaattoja ei ole ylipitkissä tavuissa juurikaan muualla kuin johdetuissa ympäristöissä (esim. *suunta : suunnan; kaartaa : kaarran*), siksi niiden merkinnässä tehtiin noin 34 prosenttia virheitä. Ylipitkien tavujen klusiiligeminaatat eivät osoittautuneet vaikeammiksi kuin klusiiligeminaatat lyhyen vokaalin edessäkään (ei tilastollisesti merkittävää eroa pitkiin tavuihin verrattuna). Ylemmällä rivillä on ilmoitettu Wilcoxonin epäparametrisen testin tulos ja alemmalla rivillä T-testin tulos.

	virheitä (%)	ero tilastollisesti merkitsevä		
Vt	7,8	on	0,000	(Z -4,817)
Vtt	24,8	on	0,000	(t -4,866, df 229)
Vn	0	on	0,001	(Z -3,176)
Vnn	5,2	on	0,001	(t -3,337, df 229)
VVt	4,3	on	0,000	(Z -5,638)
VVtt	23,4	on	0,000	(t -5,963, df 229)
VVn	0,9	on	0,000	(Z -9,587)
VVnn	50,0	on	0,000	(t -13,334, df 229)
Vm	4,3	on	0,018	(Z -2,360)
Vmm	9,5	on	0,017	(t -2,415 df 229)
Vp	7,3	on	0,000	(Z -3,722)
Vpp	19,2	on	0,002	(t -3,213, df 229)

Lyhyiden klusiilien ja nasaalien välillä on tilastollisesti merkittävä ero ympäristössä CV.CV (T-testi 0,000, df 229, t 3,811) (Wilcoxon 0,000 Z -3,944). Nasaalit ovat



helpompia kirjoittaa kuin klusiilit, sekä yksinäiskonsonanteina että geminaattoina. Ero lyhyiden konsonanttien välillä pitkän vokaalin jälkeen ei ole tilastollisesti merkittävä (T-testi 0,096, df 229,  $t$  1,673) (Wilcoxon 0,123,  $Z$  -1,544).

Tuloksia voidaan tulkita myös ärsykkeiden foneettiselta kannalta. Klusiilit ovat vaikeampia kirjoittaa oikein kuin nasaalit (mahdollisesti tämä koskee yhtä lailla kaikkia soinnillisia konsonantteja  $m$ ,  $n$ ,  $l$ ,  $r$ ), koska klusiileissa on inkluusion ja okluusion välillä tauko. Klusiilit kirjoitetaan yhtä hyvin olipa kysymyksessä yksinäisklusiili tai geminaatta lyhyen tai pitkän vokaalin jälkeen. Vaikeaa on selittää, miksi pitkän vokaalin jälkeen nasaaligeminaatta on hyvin vaikea. Nasaaligeminaatta on pitkän vokaalin jälkeen lyhyempi kuin lyhyen vokaalin jäljessä (ks. kappale 4.5). Toinen selitys, joka ei ole ristiriidassa ensimmäisen kanssa, on se, että kahden peräkkäisen soinnillisen foneemin kvantiteetti on vaikea prosessoida.

Kaiken kaikkiaan olen tässä työssä pyrkinyt etsimään kvantiteetin koodamisen vaikeuksien syytä etupäässä autosegmentaalista fonologiasta, ilmiön frekventtiydestä ja ilmiön foneettisesta puolesta käsin. Pituuden foneettinen puoli tosin jää työssäni hyvin vähälle käsittelylle, koska sitä on käsitelty perusteellisesti mm. Richardson (mm. 1998, 2003, 2004). Pitkän kvantiteetin monimutkaisuus autosegmentaalisen fonologian kannalta sekä pitkän kvantiteetin lyhyttä kvantiteettia pienempi käyttöfrekvenssi tukevat toisiaan. Pienempi frekvenssi kertoo myös fonologisesta kompleksisuudesta suhteessa lyhyeen kvantiteettiin. Samoin on ylipitkissä tavuissa geminaatan ja yksinäiskonsonantin (CVV.CV vs. CVVC.CV) välinen kestoero pienempi kuin ero lyhyen ja pitkän välillä sanatyypeissä CV.CV vs. CVC.CV.

Tavun käsitteen kannalta on helppo ymmärtää, miksi juuri tavunloppuiset segmentit ovat kirjoituksessa kaikkein vaikeimpia. Oletan Niemeä ja Bergiä (2000, 211) seuraten, että suomessa ei ole erillistä koodan vahvistustendenssiä, jonka mukaan tavun loppuun kasataan niin paljon segmenttejä kuin mahdollista, vaan kooda rakentuu niistä segmenteistä, jotka jäävät seuraavan tavun tavunalkuisen konsonantin ulkopuolelle. Jos siis tavulla on merkitystä kirjoitusprosessissa (ainakin

kirjoitusprosessin alkuvaiheessa), on sanaa muodostettaessa kiinnitettävä huomio ensiksi tavunalkuisiin konsonantteihin (taso 1). Tavunloppuiset konsonantit liitetään tavuun vasta tasoilla 2 ja 3. On sinällään yhdentekevää, muodostavatko sananloppuiset konsonantit kaksi tasoa (tasot 2 ja 3) vai kuvataanko tavunloppuiset konsonantit aina tasolla 2. Joka tapauksessa on selvää, että kirjoitusprosessissa segmenttejä kooda- tessa tavunloppuisiin konsonantteihin kiinnitetään huomiota vasta sitten, kun rajoitus ”tavun alussa on yksi konsonantti” on täytetty.

Leksikko		Taso 1		Taso 2		Taso 3
CVCV	>	[CV].[CV]				
CVCCV	>	[CV]C.[CV]	>	[[CV]C].[CV]		
CVCCCV	>	[CV]CC.[CV]	>	[[CV]C]C.[CV]	>	[[[CV]C]C].[CV]

Sananalkuiset konsonantit liittyvät ensimmäiseen tavuun automaattisesti, koska ei ole mahdollista, että ne voisivat liittyä edellisen tavun loppuun. (Suomen kielessä on kahdella konsonantilla alkavia sanoja noin 3 prosenttia kaikista sanoista). Sanan- sisäinen geminaatta on hankala myös siksi, että jotta geminaatta voitaisiin koodata oikein, edellyttää koodaus, että kun rajoitus ”tavun alussa on yksi konsonantti” on täytetty, on samaan segmenttiin palattava myös tasolla 2 (tai 3). Näin on tietenkin siksi, että geminaatta kuuluu myös edellisen tavun koodaan.

Kielenkehityksen alkuvaiheessa lapset ovat paremmin tietoisia sanan ja tavun käsitteestä kuin foneemin käsitteestä (ks. esim. Lyytinen et al. 2004b ja Richardson et al. 2004). Tällöin tavulla voidaan olettaa olevan suuri merkitys laajempien yksiköiden segmen- toinnissa, esimerkiksi juuri tavutasolta segmenttaaliselle tasolle. Jos oletetaan, että segmentointi tapahtuisi edellä esitetyn mallin vastaisesti muodostamalla ensin tavun kooda, ja vasta tämän jälkeen seuraavan tavun tavunalkuinen konsonantti, mutkistuu tavunmuodostus entisestään. Tällöin tavutus kulkisi vasemmalta oikealle lineaarisen fonologian mallin mukaisesti. Toimiakseen tämä malli vaatii kuitenkin autosegmentaalista informaatiota, koska muuten tavutusprosessi ei voi saada tietoa siitä, kuinka paljon segmenttejä tavun loppuun

voidaan kasata. Mahdollinen on tietenkin malli, jossa kooda ensin täytetään maksimaalisesti, ja tämän jälkeen siitä irrotetaan viimeinen segmentti seuraavan tavun alkukonsonantiksi.

Leksikko		Taso 1		Taso 2		Taso 3
CVCV	>	[CVC]V	>	[CV][CV]		
CVCCV	>	[CVCC]V	>	[CVC][CV]		
CVCCCV	>	[CVCC].CV	>	[[CVC]C]CV	>	[[[CVC]C]CV]

Tämäntyyppisen tavutusprosessin etuna on se, että suomessa sanansisäinen tavunalkuinen konsonantti on aina yksinäiskonsonantti: tällöin tavunmuodostus olisi tältä osin automaattista. Ylipitkissä tavuissa, joissa on kaksi tavunloppuista konsonanttia, toimii myös sääntö, jonka mukaan tavun lopussa voi olla korkeintaan kaksi konsonanttia. Tällöin kolmas konsonantti jää automaattisesti seuraavan tavun alkuun. Olkoonpa tavutusprosessin kulku kumpi hyvänsä, niin tavunloppuiset konsonantit (ja geminaatat) ovat tavutasonkin kannalta sanan monimutkaisin asema.

Tutkielmani alussa käsittelem Uppstadin & Tonnessenin (2005, 60) käsityksiä fonologian ontologiasta kirjoituksen näkökulmasta. Toivon työni osoittaneen, että mikään käsitys Uppstadin ja Tonnessenin viiden kohdan listassa ei tunnu olevan riittävä kuvattaessa fonologisen perusyksikön ja grafeemin suhdetta. Ensimmäisen käsityksen mukaan fonologia on lingvistinen ilmiö, kirjoitettu kieli ei ole. Tämä käsitys implikoi, että leksikkoon on pääsy ainoastaan puheen kautta ja että fonologinen rakenne on suoraan relevantti kirjoitettuun kieleen nähden, mutta ei toisin päin. Tämän käsityksen heikkous on siinä, että ei voida todistaa, että fonologia todella olisi kirjoituksen mukaista. Todennäköisesti niin ei ole. Totta kuitenkin on, että kirjoitettu kieli ei voi sisältää joitain puheelle ominaisia prosodisia piirteitä. Toisen käsityksen mukaan fonologia on luonnollinen, kirjoitettu kieli on artefakti. Tämä käsitys implikoi sitä, että puhe on sisäsyntyistä, kirjoitus on artefakti ja puhe kuuluu ihmisenä olemiseen, mutta kirjoitettu kieli on ulkoista. Tämä hypoteesi ei kuitenkaan tuo mitään uutta kirjoituksen tutkimukseen käytännön tasolla. Käsitys kirjoitetusta kielestä artefak-

tina on liian yleinen toteama. Kolmannen käsityksen mukaan fonologia tarkoittaa segmentaalista, lineaarista rakennetta. Käsitys implikoi että prosodia on perifeerista, se ei ole fonologisen rakenteen osa ja äänteelliset yksiköt ovat lineaarisesti järjestyneitä. Tämän toivon osoittaneeni vääräksi, koska kielen fonologista rakennetta ja fonologista pituutta ei pystytä kuvaamaan lineaarisena rakenteena. Neljännen mallin mukaan fonologia on platonista. Tämä käsitys implikoi, että fonologia on abstraktia, se on kaukainen ilmiö verrattuna foneettiseen, akustiseen ääntämykseen. Fonologiselle tasolle täytyy olla pääsy lukemaan opittaessa ja fonologiaan pääsee ainoastaan fonologisten yksiköiden tunnistamisen tietoisuuden kautta. On vaikea todistaa, että puheessa todella käytetään kirjoittamisessa käytettävien yksiköiden kaltaisia yksiköitä. Voi olla hyvinkin niin, että fonologinen tietoisuus saavutetaan vasta kirjaimia opittaessa. Loogisempaa on kuitenkin olettaa, että ihmisellä on jonkin tasoinen tietämys tavua pienemmistä yksiköistä, mutta tämän tietämys ei ole suoraan rinnastettavissa kirjoituksen kielelliseen esitykseen. Osoitin kuitenkin, että pituuden osalta monitasoinen abstrakti malli ennustaa myös kirjoitusongelmia. Viidennen käsityksen mukaan fonologia on hyvin representationaalista. Käsitys implikoi että on olemassa spesifinen mentaalinen taso, mihin fonologinen rakenne voidaan totuudenmukaisesti sijoittaa psykologisen reaalisuuden ja kommunikatiivisen relevanssin vuoksi ja representaatio on kaukainen ja kvalitatiivisesti erilainen kuin puhe. Tämänkin oikeaksi todistaminen on vaikeaa. Ainakaan formaali fonologinen teoria ei voi operoida foneettisen representaation kaltaisilla yksiköillä, tämän toivon osoittaneeni luvussa kaksi. On runsaasti kielen ÄÄNNERAKENTEEN ilmiöitä, mukaan lukien fonologinen pituus, joiden kuvaaminen on mahdotonta, jos oletetaan fonologisten yksiköiden olevan pelkästään foneettisia representaatiota.

## Lähteet

- ANTTILA, ARTO 1995: Deriving variation from Grammar: A Study of Finnish Genitives. Stanford University. ROA-63. <http://roa.rutgers.edu/files/63-0000/63-0000-ANTTILA-0-0.PDF>.
- ARONOFF, MARK — SRIDHAR S. N. 1983: Morphological levels in English and Kannada; or Atarazing Reagan. – J. Richardson, M. Marks & A. Chukerman (toim.), CLS 19. *Parasession on the Interplay of Phonology, Morphology, and Syntax* s. 16–35. Chicago: CLS.
- BERG, THOMAS — NIEMI, JUSSI 2000: Syllabification in Finnish and German: Onset filling vs. onset maximization. *Journal of Phonetics* 28(2) s. 187–216.
- BLEVINS; JULIETTE 1995: The Syllable on Phonological Theory. – John A. Goldsmith (toim.), *Handbook of Phonological Theory* s. 206–244. Oxford: Basil Blackwell.
- BYBEE, JOAN L. 1985: Morphology: A study of the relation between meaning and form. *Typological studies of language* 9. Philadelphia: Benjamins.
- BYBEE, JOAN L. 1988a: Morphology as lexical organization. – M. Hammond & M. Noonan (toim.), *Theoretical morphology* s. 119–141. San Diego: Academic Press.
- BYBEE, JOAN L. 1988b: The diachronic dimension in explanation. – John A. Hawkins 1988 (toim.), *Explaining language universals* s. 350–379. Oxford: Basil Blackwell.
- BYBEE, JOAN 2001: Phonology and Language Use. *Cambridge Studies in Linguistics* 94: Cambridge University Press.

- CHOMSKY, NOAM 1964: Current Issues in Linguistic Theory. – J. A. Fodor & J. J. Katz (toim.), *The Structure of Language* s. 50–118. Readings in the Philosophy of Language. Englewood Cliffs: Prentice Hall.
- CHOMSKY, NOAM — HALLE, MORRIS 1968: The Sound Pattern of English. New York: Harper & Row.
- COLE, JENNIFER 1995: The cycle in phonological grammar. – J. Goldsmith (toim.), *Handbook of Phonological Theory* s. 70–113. Cambridge, MA: Basil Blackwell.
- DE LACY, PAUL 1997: Prosodic Categorisation. University of Auckland. ROA-236. Rutgers Optimality Archive. <http://ling.rutgers.edu/~delacy/files/delacy-1997-prosodic%20categorization.pdf>.
- FLEMMING, EDWARD 2005: Speech perception and phonological contrast (pre-publication version). <http://web.mit.edu/flemming/www/paper/Percphon03.pdf>.  
– Julkaistu myöhemmin samana vuonna D. Pisoni & R. Remez (toim.), *The Handbook of Speech Perception*. Oxford: Basil Blackwell.
- FLEMMING, EDWARD 2006: The role of distinctiveness constraints in phonology. Käsikirjoitus. <http://web.mit.edu/flemming/www/paper/inventories.pdf>.
- FORD, ALAN — SINGH, RAJENDRA 1983: On the status of morphophonology. – J. Richardson, M. Marks & A. Chukerman (toim.), *CLS 19: Papers from Parasession on the Interplay of Phonology, Morphology, and Syntax* s. 63–78. Chigaco: CLS.
- FORD, ALAN — SINGH, RAJENDRA 1985: Towards a non-paradigmatic morphology. – M. Niepokuj, M. VanClay, V. Nikiforidou & D. Feder (toim.), *Papers from BLS 11* s. 87–95. Berkeley: BLS.

- FOX, ANTHONY 2000: Prosodic Features and Prosodic Structure. The Phonology of Suprasegmentals. Oxford: Oxford University Press.
- FUDGE, E. C. 1967: The nature of phonological prime. *Journal of Linguistics* 3. s. 1-36.
- GOLDSMITH, JOHN 1976: Autosegmental phonology. Väitöskirja. MIT. New York. Garland Press.
- GOLDSMITH, JOHN 1990: Autosegmental and Metrical Phonology. Oxford: Basil Blackwell.
- HALLE, MORRIS 1962: Phonology in Generative Grammar. *Word* 18 s. 54–72.
- HARRIKARI, HELI 2000: Segmental Length in Finnish – Studies within a Constraint-based Approach. *Helsingin yliopiston yleisen kielitieteen laitoksen julkaisuja no. 33*. Helsinki: Helsingin yliopisto.
- HASPELMATH, MARTIN 2006: Against markedness (and what to replace it with). *Journal of Linguistics* 42 s. 25–70.
- HAYES, BRUCE 1989: Compensatory Lengthening in moraic phonology. *Linguistic Inquiry* 20. s. 253-306.
- HAYES, BRUCE 1995: Metric Stress Theory. Principles and Case studies. Chicago: The University of Chicago Press.
- HOUSEHOLDER, FRED 1965: On some recent claims in phonological theory. *Journal of Linguistics* 1. s. 13-34.
- ITKONEN, TERHO 1957: Mellakoihin vai mellakkoihin? Yleiskielemme eräiden taivutushorjuvuuksien taustaa. *Virittäjä* 61 s. 259–286.

- IVERSON, GREGORY — WHEELER, DEIRDRE 1988: Blocking and the Elsewhere Condition. – M. Hammond & M. Noonan (toim.), *Theoretical Morphology: Approaches in Modern Linguistics* s. 325–338. San Diego: Academic Press.
- JAKOBSON, ROMAN 1948: Russian Conjugation. *Word* 4. s. 155-167.
- JOHNSON, MARK 1988: *Attribute-value logic and the theory of grammar*, Stanford: CSLI. Stanford University.
- KAGER, RENE 1999: *Optimality Theory*. Cambridge Textbook in Linguistics. Cambridge: Cambridge University Press.
- KAPLAN, R. M. — BRESNAN, JOAN 1982: *Lexical-Functional Grammar*. — Joan Bresnan (toim.) *The mental representation of grammatical relations*. Cambridge, Mass.: The MIT Press.
- KARLSSON, FRED 1982: Suomen kielen äänne- ja muotorakenne. Porvoo-Helsinki-Juva: WSOY.
- KETTUNEN, LAURI 1949: Astevaihtelun salaisuus. *Virittäjä* 53 s. 69–75.
- KIPARSKY, PAUL 1968: How Abstract Is Phonology? – O. Fujimura (toim.), *Three Dimensions of Linguistic Theory* s. 1–136. Tokyo: Taikusha.
- KIPARSKY, PAUL 1973: "Elsewhere" in phonology. – S. Anderson & P. Kiparsky (toim.), *A Festschrift for Morris Halle* s. 93–106. New York: Holt Rinehard and Winston.
- KIPARSKY, PAUL 1982a: *Explanation in Phonology*. Dordrecht: Foris.



- KIPARSKY, PAUL 1982b: From cyclic phonology to lexical phonology. – H. van der Hulst & N. Smith (toim.), *The Structure of Phonological Representations* Volume 1 s. 131–175. Dordrecht: Foris.
- KIPARSKY, PAUL 1982c: Lexical phonology and morphology. – I. S. Yang (toim.), *Linguistics in the Morning Calm* Volume 2 s. 3–91. Seoul: Hanshin.
- KIPARSKY, PAUL 1983: Word formation and the lexicon. — F. Ingemann (toim.). *Proceedings of the Mid-America Linguistics Conference*. s. 3-29. Lawrence: University of Kansas.
- KIPARSKY, PAUL 1985: Some consequences of lexical phonology. *Phonology Yearbook* Volume 2 s. 85–138.
- KIPARSKY, PAUL 2003: Finnish Noun Inflection. – Diane Nelson & Satu Manninen (toim.), *Generative Approaches to Finnic and Saami Linguistics* s. 109–161. Stanford: CSLI Publications. Saatavilla [www-muodossa](http://www.muodossa) <http://www.stanford.edu/~kiparsky/Papers/finnish.article.pdf>.
- KISSENBERT, CHARLES 1970: On the Functional Unity of Phonological Rules. *Linguistic Inquiry* 1 s. 291–306.
- LAMB, SIDNEY 1966: Prolegomena to a theory of phonology. *Language* 42. s. 536-573.
- LAUERMA, PETRI 1993: Vatjan vokaalisointu. *Suomalais-ugrilaisen Seuran toimituksia* 214. Helsinki: Suomalais-ugrilainen Seura.
- LEBEN, WILLIAM RONALD 1973: Suprasegmental Phonology. New York & London: Garland.

- LEHTONEN, JAAKKO 1970: Aspects of Quantity in Standard Finnish. *Studia Philologia Jyväskyläensia* 6. Jyväskylä: University of Jyväskylä.
- LEIWO, MATTI 1984: Käänteisen astevaihtelun lingvistiikkaa ja psykolingvistiikkaa. – Kari Sajavaara, Jorma Tommola & Matti Leiwo (toim.), *Psykolingvistiisiä kirjoituksia* s. 87–100. Suomen soveltavan kielitieteen yhdistyksen (AFinIA) julkaisuja 37. Jyväskylä: Suomen soveltavan kielitieteen yhdistys.
- LEPPÄNEN, PAAVO H. T — RICHARDSON, ULLA — PIHKO, ELINA — EKLUND, KENNETH M. — GUTTORM, TOMI K. — ARO, MIKKO — LYYTINEN, HEIKKI 2002: Brain Responses to Changes in Speech Sound Durations Differ Between Infants With and Without Familiar Risk for Dyslexia. *Developmental Neuropsychology* 22 (1) s. 407–422.
- LIEBER, ROCHELLE 1980: On the organization of the lexicon. Painamaton väitöskirja. MIT.
- LYYTINEN, HEIKKI — ARO, MIKKO — EKLUND, KENNETH — ERSKINE, JANE — GUTTORM, TOMI — LAAKSO, MARJA-LEENA — LEPPÄNEN, PAAVO H. T. — LYYTINEN, PAULA — POIKKEUS, ANNA-MAIJA — RICHARDSON, ULLA — TORPPA, MINNA 2004a: The Development of Children at Familiar Risk for Dyslexia: Birth to Early School Age. *Annals of Dyslexia* 54 (2) s. 184–220.
- LYYTINEN, HEIKKI — AHONEN, T. — EKLUND K. — GUTTORM, T. — KULJU, P. — LAAKSO, M.-L. — LEIWO, M. — LEPPÄNEN, P. — LYYTINEN, P. — POIKKEUS, A.-M. — RICHARDSON U. — TORPPA, M. — VIHOLAINEN, H. 2004b: Early Development of Children at Familiar Risk for Dyslexia – Follow-up from Birth to School Age. *Dyslexia* 3 (10) s. 146–178.
- MASCARO, JOAN 1976: Catalan phonology and the phonological cycle. Painamaton väitöskirja. MIT.

- MCCARTHY, JOHN 1979: On stress and the syllabification. *Linguistic Inquiry* 10 s. 443–466.
- MCCARTHY, JOHN 1999: Sympathy and Phonological Opacity. *Phonology* 16 s. 331–399. Saatavilla www-muodossa <http://people.umass.edu/jjmccart/sympathy.pdf>.
- MOHANAN, K. P. 1982: Lexical phonology. Painamaton väitöskirja. MIT.
- MOHANAN, K. P. — MOHANAN, TARA 1984: Lexical phonology of the consonant system in Malayalam. *Linguistic Inquiry* 15. s.575-602.
- MOHANAN, K. P. 1995: Organization of Grammar. – J. Goldsmith (toim.), *The Handbook of Phonological Theory* s. 24–69. Oxford: Basil Blackwell.
- MOHANAN, TARA 1983: The structure of the melody. Painamaton väitöskirja. MIT.
- PESETSKY, DAVID 1979: Russian morphology and lexical theory. Painamaton väitöskirja. MIT.
- PULLEYBLANK, DOUGLAS 1986: Rule Application on a Non-cyclic Stratum. *Linguistic Inquiry* 17(3) s. 573–580.
- RAVILA, PAAVO 1951: Astevaihtelun arvoitus. *Virittäjä* 55 s. 292–300.
- RICHARDSON, ULLA 1998: Familial dyslexia and sound duration in the quantity distinctions of Finnish infants and adults. *Studia philologica Jyväskyläensia* 44. Jyväskylä: University of Jyväskylä.
- RICHARDSON, ULLA — LEPPÄNEN, PAAVO H. T. — LEIWO, MATTI — LYYTINEN, HEIKKI 2003: Speech Perception of Infants With High familiar Risk

for Dyslexia Differ at the Age of 6 Months. *Developmental Neuropsychology* 23(3) s. 385–397.

RICHARDSON, ULLA — THOMPSON, JENNIFER M. — SCOTT, SOPHIE K. — GOSWAMI, USHA 2004: Auditory Processing Skills and Phonological Representation in Dyslexic Children. *Dyslexia* 10 (3) s. 215–233.

ROTENBERG, JOEL 1978: The syntax of phonology. Painamaton väitöskirja. MIT.

SCOBIE, JAMES M. — COLEMAN, JOHN S. — BIRD, STEVEN 1996: Key aspects of Declarative Phonology. – Jacques Durand & Bernard Laks (toim.), *Current Trends in Phonology: Models and Methods* Volume 2 s. 685–709. European Studies Research Institute (ESRI). Manchester: University of Salford. Saatavilla www-muodossa <http://www.qmuc.ac.uk/ssrc/pubs/scob962.pdf>.

SELKIRK, ELISABETH 1984: Phonology and Syntax. The Relation between Sound and Structure. Cambridge: MIT Press.

SIEGEL, DOROTHY 1974: Topics in English morphology. Painamaton väitöskirja. MIT.

SINGH, RAJENDRA 1987: Well-formedness conditions and phonological theory. – W. U. Dressler, H. C. Luschyzy, O. E. Pfeiffer & J. R. Rennison (toim.), *Phonologica* 1984 s. 273–85. Cambridge: Cambridge University Press.

STAMPE, DAVID 1972: How I Spent My Summer Vacation. New York: Garland.

STERIADE, DONCA 1995: Markedness and Underspecification in the Study of Segments. – J. Goldsmith (toim.), *A Handbook of Phonological Theory* s. 114–174. Oxford: Basil Blackwell.

- STERIADE, DONCA 1997: Phonetics in phonology: the case of laryngeal neutralization. Julkaisematon käsikirjoitus. UCLA. <http://web.mit.edu/linguistics/people/faculty/steriade/publications.html>.
- STERIADE, DONCA. 2001: The Phonology of Perceptibility Effects: the P-map and its consequences for constraint organization. Julkaisematon käsikirjoitus. UCLA. <http://web.mit.edu/linguistics/people/faculty/steriade/publications.html>.
- SUOMI, KARI 1980: Voicing in English and Finnish stops. *Turun yliopiston Suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 10*. Turku: Turun yliopisto.
- SUOMI, KARI — TOIVANEN JUHANI — YLITALO, RIIKKA 2001: On distinguishing stress and accent in Finnish. Dept. of Linguistics and Phonetics, *Working Papers 49* s. 152–155. Lund: Lund University.
- SUOMI, KARI 2002: Word stress and duration in Finnish. *Proceedings from Fonetik 2002*. TMH-QPSR Vol. 44 s. 73–76. Kungliga Tekniska Högskolan. [http://www.speech.kth.se/prod/publications/files/qpsr/2002/2002\\_44\\_1\\_073-076.pdf](http://www.speech.kth.se/prod/publications/files/qpsr/2002/2002_44_1_073-076.pdf)
- SUOMI, KARI — YLITALO, RIIKKA 2003: Syllable weight and segmental durations in Finnish. *Reports in Phonetics Umeå University 9* s. 37–40. Umeå: Umeå University. <http://www.ling.umu.se/fonetik2003/pdf/037.pdf>.
- SUOMI, KARI — YLITALO, RIIKKA 2004: On durational correlates of word stress in Finnish. *Journal of Phonetics* Volume 32 (1) s. 35–63.
- SUOMI, KARI 2005: Temporal conspiracies for a tonal end: Segmental durations and accentual f<sub>0</sub> movement in a quantity language. *Journal of Phonetics* Volume 33 (3) s. 291–309.

SUOMI, KARI 2006: Suomen segmenttikestojen määräytymisestä. *Virittäjä* 4 s. 483–503.

TORVELAINEN, PÄIVI 2007: Kaksivuotiaiden lasten fonologisen kehityksen variaatio : puheen ymmärrettävyyden sekä sananmuotojen tavoittelun ja tuottamisen tarkastelu. *Jyväskylä studies in humanities* 73. Jyväskylä: Jyväskylän yliopisto.

TURUNEN, PIRJO 2003: Production of word structures : a constraint-based study of 2;6 year old Finnish children at-risk for dyslexia and their controls. *Jyväskylä studies in languages* 54. Jyväskylä: Jyväskylän yliopisto.

UPPSTAD, PER HENNING 2005: Language and Literacy. Some fundamental issues in research on reading and writing. Lund university. Department of Linguistics and Phonetics. Lund: Lund University.

UPPSTAD, PER HENNING — TONNESSEN, FINN EGIL 2005: The Status of the Concept of 'Phoneme' in Psycholinguistics. — Per Henning Uppstad, *Language and Literacy. Some fundamental issues in research on reading and writing.* s. 51-76. Lund university. Department of Linguistics and Phonetics. Lund: Lund University.

WIIK, KALEVI 1975: Vokaalisoinnun ongelmia. *Turun yliopiston fonetiikan laitoksen julkaisuja* 14. Turku: Turun yliopisto.

YLIVAKKURI, VALMA 1976: Onko suomen kielen kvalitatiivinen astevaihtelu epäproduktiivinen jäännös. *Sananjalka* 18 s. 53–69. Turku: Suomen kielen seura.