

# YLLÄTTÄVIÄ HAVAINTOJA SUOMEN SUBSTANTIIVIEN FONOTAKTISESTA KOMPLEKSISUUDESTA?

**F**red Karlsson (2005) raportoi laskelmiaan suomen substantiivien fonotaktisesta kompleksisuudesta Kimmo Koskenniemen noin neljännesvuosisata sitten tietokoneistaman Tuomen (1972) aineiston perusteella. Kaikki kolme ansaitsevat suuren kiitoksen kvantitatiivisen tiedon tuottamisesta suomen äänneopin harrastajien käyttöön. Karlsson analysoi nimenomaan supisuomalaisen substantiivien ensitavun fonotaksia, ja hän pitää havaintojaan yllättävinä. Yritän seuraavassa osoittaa, että hänen tulostensa päälinjat ovat itse asiassa hyvin odotuksenmukaiset.

Karlssonin (2005) taulukko 1 (s. 68) osoittaa, että aineistoon sisältyy vain 29 yksitavuista substantiivikantaa, joista 24 on rakennetta CVV eikä yksikään rakennetta CV. Karlsson pitää tätä asiantilaa hämmästyttävänä, koska yleisen tunnusmerkisyyden teorian perusteella olisi ennustanut, että yksinkertaisempi CV-rakenne esiintyisi ja että se lisäksi esiintyisi useammin kuin CVV. Mutta, hän sitten toteaa, CV-substantiivit (ja -verbit) ovat kiellettyjä — tosin hän ei kerro miksi. Hän pitää sanaa *yö* poikkeuksena (s. 68), koska se on ainoa VV-substantiivi. Karlssonin ilmaisema hämmästyminen on hämmästyttävää, onhan hän aiemmin (1983: 215) todennut, että »huomionarvoista on, ettei avonaisilla sanaluokilla eli substantiiveilla, adjektiiveilla ja verbeillä ole edustajia CV-sanojen joukossa».

CV-tavu on kyllä typologisesti yleisin tavutyyppi, mutta tästä ei välttämättä seuraa, että CV-substantiivit olisivat yleisiä. Niinpä CV-substantiivit eivät ole mahdollisia suomen lisäksi esimerkiksi useimmissa germaanisissa kielissä — kun rakenteel-

la CV tarkoitetaan, että V on foneettisesti lyhyt. Toisaalta Satsuki Nakai (Edinburghin yliopistosta) kertoo, että japanisessa — joka on vokaalien suhteen selvä kvantiteettikieli — on paitsi substantiiveja, jotka koostuvat pelkästään lyhyestä vokaalista (esim. /i/ 'vatsa'), myös runsaasti CV-substantiiveja (esim. /ki/ 'puu'). Suomessa avoimen sanaluokan sanan minimaalisen perusmuodon täytyy sisältää vähintään kaksi soinnillista moraa, ja suomi eroaa tässä suhteessa selvästi sekä germaanisista kielistä että japanista; esimerkiksi ruotsin *hatt* ei sisältäne minkäänlaisen tulkinnan mukaan kahta soinnillista moraa. Lyhimmät suomea koskevan rajoituksen täyttävät sanat ovat *ui* ja *yö*; kumpikaan sana ei siis ole tämän rajoituksen suhteen poikkeus. Ei ole myöskään yhtään avoimen sanaluokan lainasanaa, joka ei täyttäisi minimivaatimusta (ks. Suomi 2004). Palaan tämän rajoituksen todennäköiseen motivaatioon tuonnempana.

Karlssonin (2005) tulosten mukaan kaksitavuisissa, vokaaliin päättyvissä sanoissa kaksimorainen ensitavu on vallitseva, ja ensitavultaan kolmimorainen rakenne CVVC.CV on melkein yhtä yleinen kuin rakenne CV.CV. Yksimorainen ensitavu ((C)V) esiintyy vain 16 prosentissa sanoista; konsonanttiin päättyvissä kaksitavuisissa sanoissa osuus näyttää olevan samanlainen, samoin kolmitavuisissa ja pitemmissä sanoissa sekä vihdoin myös koko sanastossa (s. 69). Koska universaalisti suositaan lyhyttä yksimoraista CV-tavua, Karlsson olisi odottanut prototyyppisen ensitavun olevan yksimorainen. (Sivumennen sanoen ainakin minun ja erään fennistin intuition

mukaan *a.ort.ta*, *ny.ans.si* ja *se.ans.si* ovat kolmitavuisia osoitetulla tavalla, jolloin ensitavu on yksimorainen eikä nelimorainen niin kuin Karlsson analysoi.)

En voi pitää Karlssonin tuloksia mitenkään yllättävinä. Ensiksikin, kuten edellä jo totesin, CV-tavun universaalista yleisyydestä ei voi tehdä johtopäätöksiä CV-sanojen yleisyydestä. Edelleen ei ole aina selvää, mitä CV-tavulla maailmalla kulloinkin tarkoitetaan. Arvioin hiljan erään käsikirjoituksen, jossa se, mitä tässä kirjoituksessa tarkoitetaan suomen CVV-tavulla, olikin analysoitu (pitkän vokaalin sisältäväksi) CV-tavuksi; ehkä suomen CVV-tavut kirjautuvat CV-tavuiksi joissakin typologisissa katsauksissa. Toiseksi, koska suomi on kvantiteettikieli, jossa jokainen vokaalifoneemi on aina mora, on odotuksenmukaista, että ensitavu on usein vähintään kaksimorainen — kun foneemianalysissa noudatetaan mielestäni ainoaa järkevää ratkaisua eli Karlssonin (1969) ehdottamaa tulkintaa, jossa kontrastiivisesti pitkät monoftongit ja konsonantit sekä diftongit tulkitaan kahden foneemin jonoiksi. Kolmanneksi, koska sanansisäiset CC- ja CCC-jonot 1. ja 2. tavun rajalla ovat hyvin yleisiä ja näistä konsonanteista muut paitsi jonon viimeinen ovat ensitavuun kuuluvia moria, on myös tästä syystä odotuksenmukaista, että vähintään kaksimoraiset ensitavut ovat hyvin yleisiä. Karlsson (1983: 116) toteaa kyseisessä asemassa esiintyvistä omaperäisistä CC-jonoista, että »useimpia on niin runsaasti, ettei leksikaalisia frekvenssejä kannata laskea», ja suuri enemmistö hänen esimerkksanoistaan on nimenomaan substantiiveja. Neljänneksi, koska suomen kymmenestä niin sanotusta perustavutyyppistä vain kaksi on yksimoraisia (ks. esim. Karlsson 1983: 133) ja koska näistä tavutyyppi V on lisäksi tunnetusti melko harvinainen, ei ole yllätys, että suurin osa sananalkuisista tavuista on monimoraisia. Viidenneksi, kaikki perus-

tavutyyppit »eivät esiinny missä tahansa sana-asemassa. Sanan ensimmäinen tavu on rakenteellisesti eriytynein. Se on ainoa paikka, jossa kaikki perustavutyyppit voivat esiintyä» (mts. 137). Toisaalta on hyvin tiedossa, että CV-tavut ovat yleisiä myös muualla kuin ensitavussa; Karlssonin (mts. 217) mukaan vapaiden morfeemien suosituin kanoninen tyyppi on »nimenomaan CV-loppuinen» kaksitavuinen rakenne. Tästä seuraa, että useat nimenomaan monimoraiset tavutyyppit esiintyvät vain ensitavussa. Tämäkin asiantila ennustaa, että sananalkuisia monimoraisia tavuja täytyy olla runsaasti.

Toisaalta ei CVCV-sanoja suinkaan ole yllättävän vähän, toisin kuin Karlsson (2005) antaa ymmärtää; kun tässä kappaleessa puhun sanoista, tarkoitan siis Karlssonin yksityiskohtaisemmin tarkastelemissa, vokaaliin päättyviä kaksitavuisia substantiivikantoja. Supisuomalaisen sanan alussa voi esiintyä 11 konsonanttia (/p t k s h m n l r v j/), jolloin — sivuuttaen tietyt konsonantin ja vokaalin yhdistämisrajoitukset — erilaisia sananalkuisia CV-tavuja voi olla kaikkiaan  $11 \times 8 = 88$  kpl. Supisuomalaisten CVVCV-sanojen ensitavussa voi esiintyä 26 VV-jonoa (kaksoisvokaalit ja diftongit) eli erilaisten CVV-tavujen määräksi saadaan  $11 \times 26 = 286$  kpl. Supisuomalaisten CVCCV-sanojen mahdollisia ensitavuja on  $11 \times 8 \times 10 = 880$  kpl; sanansisäisen CC-jonon ensimmäisenä jäsenenä voivat edellä mainituista konsonanteista esiintyä muut kuin /v j/, ja lisäksi /ŋ/. Nämä vertailut saavat riittää, ja lukujen valossa voisi hyvinkin olettaa, että CVCCV-sanoja olisi kymmenkertainen määrä CVCV-sanoihin verrattuna ja CVVCV-sanojakin yli kolminkertainen määrä CVCV-sanoihin verrattuna. Näin ei kuitenkaan ole laita, vaan CVCCV-sanoja (1 795 kpl) on vain alle kaksi ja puoli kertaa enemmän kuin CVCV-sanoja (756 kpl), ja CVVCV-sanoja (950 kpl) vain alle

▷

kolmannes enemmän kuin CVCV-sanoja. Kun jaetaan Karlssonin raportoimat näitä rakenteita edustavien sanojen lukumäärät kussakin rakenteessa mahdollisten erilais-  
ten ensitavujen määrällä, saadaan numeerinen arvio siitä, kuinka tehokkaasti eri rakenteet ovat ensitavujensa osalta käytössä. CVCV-sanojen osalta arvioksi saadaan  $756/88 = 8,6$ ; rakenteen CVVCV osalta  $950/286 = 3,3$ ; ja rakenteen CVCCV osalta  $1795/880 = 2,0$ . Toisin sanoen CVCV-sanoja on liki yhdeksän kertaa enemmän kuin on mahdollisia erilaisia CV-tavuja sanan alussa, kun taas CVCCV-sanoja on vain kaksi kertaa enemmän kuin on mahdollisia sananalkuisia CVC-tavuja. Näistä luvuista voidaan edelleen päätellä, että tarkastelluissa rakenteissa satunnaisten aukkojen suhteellinen osuus sanojen teoreettisesta maksimimäärästä on pienin rakenteessa CVCV ja suurin rakenteessa CVCCV. Itse asiassa siis rakenne CVCV on mainituista sanarakenteista suosituin.

Karlsson (2005) toteaa (s. 68) kaksitavuisten sanojen osalta, että pyrkimys välttää kaksoisvokaaleja ja diftongeja toisessa tavussa on hyvin vahva; näitä esiintyy aineiston noin 6 000 supisuomalaisen, johtamattoman kaksitavuisen substantiivin joukossa vain seitsemän. Näyttää lisäksi siltä, että rajoitus on erityisen vahva silloin, kun ensitavu on lyhyt; aineistoon ei sisälly yhtään supisuomalaista substantiivia, jonka rakenne olisi muotoa (C)VVCV. Karlsson ei jälkimmäistä rajoitusta totea, ja se mitä sanon seuraavassa, ei siis ole kommentti hänen johtopäätöksinsä eikä se päde sanoihin, joissa ensitavu on pitkä. Olen hiljan hahmotellut suomen prosodisen järjestelmän osa-alueiden keskinäisiä suhteita (Suomi 2005), ja mainitsen nyt vain joitakin tässä yhteydessä tärkeitä pääkohtia. Lause-aksentti toteutuu yleensä tonaalisena nousu-laskuna siten, että nousu tapahtuu sanan ensimmäisen ja lasku toisen moran aikana.

Aksentti toteutuu yksimuotoisena tonaalisena kuviona sanarakenteesta riippumatta. Jotta tämä yksimuotoisuus olisi mahdollista, segmenttikestojen täytyy vaihdella sanarakenteiden sisällä ja niiden välillä. Rakenteessa CVCV(C) tonaalinen nousu tapahtuu ensitavun aikana ja lasku toisen tavun aikana, ja jotta tonaalinen kuvio olisi samanlainen kuin muissa rakenteissa, toisen tavun yksinäisvokaalin keston täytyy olla hyvin pitkä. Tästä seuraa, että rakenteen CVCVV(X) toisen tavun vokaalin kesto ei ole kovin paljon suurempi kuin rakenteen CVCV(C) toisen tavun vokaalin kesto. Tässä valossa ei ole yllättävää, että rakennetta CVCVV(X) ei suosita: se ei eroa riittävästi selvästi suositusta rakenteesta CVCV(C).

Miksi avoimen sananluokan sanan täytyy sisältää vähintään kaksi soinnillista moraa? Todennäköisesti siksi, että aksentti vaatii kunnolla toteutuakseen vähintään kaksi soinnillista moraa: ensimmäisen moran aikana tapahtuu tonaalinen nousu ja myöhemmän moran aikana lasku. Avointen sanaluokkien sanoja aksentoidaan hyvin usein, toisin kuin suljettujen sanaluokkien sanoja. Toki myös yksimoraisia sanoja (esimerkiksi pronomineja) aksentoidaan, mutta aksentti toteutuu tällöin monin eri tavoin ja toisin kuin pitemmissä sanoissa (Suomi painossa).

Karlsson (2005) kysyy lopuksi, miksi mutkikkaita fonotaktisia rakenteita (monimoraisia alkutavuja) suositaan yksinkertaisempien (yksimoraisten alkutavujen) kustannuksella. Olen antanut omat selitykseni edellä. Karlsson tarjoaa kolme mahdollista syytä. Kolmas näistä on morfofonologinen: hänen mukaansa uudet sanat ja lainat suosivat kvantitatiivista astevaihtelua kvantitatiivisen kustannuksella, ja edellinen on mahdollinen vain vähintään kaksimoraisissa tavuissa, esimerkiksi *rokki* : *rokin*; siis tarkemmin sanottuna vahva aste esiintyy

vähintään kaksimoraisessa tavussa. Tämä selitys ei kuitenkaan ole vakuuttava. Ensimmäikään ei ole ilmeistä, että uudet ja lainasanat suosivat nimenomaan monimoraisia alkutavuja; lainasanojen osalta moralku riippuu sitä paitsi pitkälti siitä, minkälainen sana on rakenteeltaan lainanantajakielessä, joten suosimisesta puhuminen voi olla harhaanjohtavaa. On näet runsaasti ensitavultaan yksimoraisia uusia ja lainasanoja, joissa ei ole minkäänlaista morfoloogista vaihtelua, ja tämä sananmuodostuskeino näyttää olevan hyvin produktiivinen: *blogi* : *blogin*, *hype* : *hopen*, *laku* : *lakun*, *mega* : *megan*, *toto* : *toton* jne. Kvantitatiivisen astevaihtelun suosiminen kvalitatiivisen kustannuksella ei siis näytä riittävän selittämään monimoraisien ensitavujen ylivoimaisuutta. Toiseksi, ja tämä peruste on edellistä painavampi, on äärimmäisen vaikeata nähdä, miten uusien ja lainasanojen käyttäytyminen voisi selittää taannehtivasti kielen vanhojen sanojen rakennetta. Käyttäytyypä *rokki* nykyään miten tahansa, miten se voisi selittää jo kauan ennen Elvistä muotoutuneen sanaston rakenteellisia ominaisuuksia? *Rokki* sisältyy ainakin Tuomen (1972) toiseen painokseen vuodelta 1980, mutta se ja sen kaltaiset uudet sanat eivät riitä selittämään niitä tuloksia, jotka koko aineistosta on saatavissa. ■

KARI SUOMI

*Suomen kielen, informaatiotutkimuksen ja logopedian laitos (SILO)*

*PL 1000*

*90014 Oulun yliopisto*

Sähköposti: *kari.suomi@oulu.fi*

## LÄHTEET

- KARLSSON, FRED 1969: Suomen yleiskielen segmentaalifoneemien paradigma. – *Virittäjä* 73 s. 351–362.
- 1983: *Suomen kielen äänne- ja muotorakenne*. Helsinki: WSOY.
- 2005: Phonotactic complexity of Finnish nouns. – A. Arppe, L. Carlson, K. Lindén, J. Piitulainen, M. Suominen, M. Vainio, H. Westerlund & A. Yli-Jyrä (toim.), *Inquiries into words, constraints and contexts. Festschrift in the honour of Kimmo Koskenniemi on his 60th birthday* s. 65–70. Preprint. (Elektroninen julkaisu: <http://csli-publications.stanford.edu/site/ONLN.html>. Myös: [www.ling.helsinki.fi/~fkarlsson/kimmo\\_festschrift\\_Karlsson.pdf](http://www.ling.helsinki.fi/~fkarlsson/kimmo_festschrift_Karlsson.pdf).)
- SUOMI, KARI 2004: Moraic patterns in Finnish. – M. S. Peltola & J. Tuomainen (toim.), *Studies in speech communication* s. 87–97. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 72. Turku: Turun yliopisto.
- 2005: Suomen kielen prominenssien foneettisesta toteutumisesta. – *Virittäjä* 109 s. 221–243.
- painossa: On the tonal and temporal domains of accent in Finnish. – *Journal of Phonetics*.
- TUOMI, TUOMO 1972: *Suomen kielen käänteissanakirja*. Suomalaisen Kirjallisuuden Seuran Toimituksia 274. Helsinki: Suomalaisen Kirjallisuuden Seura.