

Suomen tavuista

Kalevi Wiik

Käsittelen tässä artikkelissa suomen yleiskielen tavuja. Pysin selvittämään suomen tavujen rakennetta ja vastaamaan mm. kysymykseen, kuinka monta erilaista tavua suomen kielessä esiintyy.

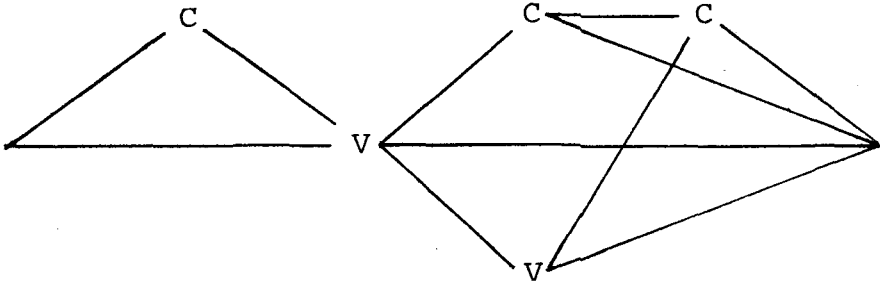
Esitykseni perustuu varsin rajalliseen suomen yleiskielen määritelmään. Tähän kieleen kuuluviksi lasken ainoastaan »supisuomalaiset» sanat; näin määriteltyyn »supisuomeen» eivät kuulu esim. sivistyssanat. Tästä kieli-
muodosta voidaan väittää esim., että sen tavu ei voi alkaa kahdella konsonantilla ja että sen konsonanttien inventaariin ei kuulu konsonantteja *b*, *g*, *f* ja *š*.

Käsittelen tavuja ensin siten, että en kiinnitä huomiota yksityisten vokaalien ja yksityisten konsonanttien välisiin eroihin. Merkitsen kaikkia vokaaleja symbolilla V ja kaikkia konsonantteja symbolilla C. Näin suomeen saadaan kymmenen tavutyyppeä. Tyypit ovat seuraavat:

| tavutyyppi: | esimerkkejä: |
|-------------|-------------------------------|
| (1) V | <i>i-lo, ko-e</i> |
| (2) CV | <i>ta-lo</i> |
| (3) VC | <i>il-ta, te-os</i> |
| (4) CVC | <i>ras-kas</i> |
| (5) VV | <i>ai-ta, aa-mu, ko-me-aa</i> |
| (6) CVV | <i>kai-ta, sa-noi</i> |
| (7) VVC | <i>ait-ta</i> |
| (8) CVVC | <i>taas-kaan</i> |
| (9) VCC | <i>irs-tas, int-tää</i> |
| (10) CVCC | <i>kars-ki, kant-ta</i> |

Tavutyypin rakenne voidaan esittää kuvan 1 mukaisena Markovin ketjuna. Kaavion saa läpäistä viivoja pitkin vasemmalta oikealle; tuloksena on kymmenen erilaista tavutyyppeä. Kaaviosta käy ilmi, että suomen tavuissa on neljä positiota. Näistä toinen on aina pakko täyttää vokaalilla. Nimitän tätä vokaalia ydinvokaaliksi. Muut positiot voidaan joko täyttää tai jättää

KUVA 1.



täyttämättä: ensimmäinen ja neljäs positio voidaan täyttää vain konsonantilla, kolmas positio voidaan täyttää joko konsonantilla tai vokaalilla. Suomen 10 tavutyypin tuottamiseen liittyy siis kolme binaarista ja yksi ternaarinen valinta: (1) tavun ydinvokaalin edelle voidaan sijoittaa konsonantti tai ei mitään; (2) ydinvokaalin jälkeen voidaan sijoittaa konsonantti, vokaali tai ei mitään; (3) ydinvokaalia seuraavan konsonantin jälkeen voidaan sijoittaa toinen konsonantti tai ei mitään; ja (4) ydinvokaalia seuraavan vokaalin jälkeen voidaan sijoittaa konsonantti tai ei mitään. Nämä neljä valintaa ovat lisäksi sellaiset, että kaksi niistä, valinnat (3) ja (4), ovat toisensa pois sulkevia: ne eivät milloinkaan tule molemmat kyseeseen yhtä ja samaa tavutyyppeä tuottaessa. Maksimimäärä valintoja, jotka tulevat kyseeseen yhtä tavutyyppeä tuottaessa, on siis kolme.

Käsittelen seuraavaksi suomen kymmentä tavutyyppeä kiinnittäen huomion niihin rajoituksiin, jotka koskevat tavun alkukonsonantismia, vokalismia ja loppukonsonantismia. Rajoituksia saadaan ensin kaikkiaan kuusi:

- (1) Tavun alussa voi olla korkeintaan yksi konsonantti.
- (2) Tavussa on vähintään yksi vokaali.
- (3) Tavussa on enintään kaksi vokaalia.
- (4) Tavussa mahdollisesti esiintyvät kaksi vokaalia ovat peräkkäisiä.
- (5) Tavun lopussa voi olla korkeintaan kaksi konsonanttia.
- (6) Tavunloppuista kahden konsonantin jonoa voi edeltää korkeintaan yksi vokaali.

Nämä kuusi tavun rakennetta koskevaa rajoitusta voidaan esittää myös kahtena, siten että toinen koskee vokaaleja, toinen tavun pituutta.

- (a) Vokaalivaatimukset: tavussa on aina 1—2 peräkkäistä vokaalia.
- (b) Pituusrajoitukset: tavun ensimmäistä vokaalia voi edeltää yksi ja seuraava kaksi segmenttiä.

Vokaalivaatimukset ovat yhdistelmä rajoituksista (2), (3) ja (4), pituusrajoitukset rajoituksista (1), (5) ja (6). Vokaalivaatimukset ovat selvästi primaarisempia kuin pituusvaatimukset. Niinpä uusissa lainasanoissa ei sal-

lita rikkomuksia vokaalivaatimuksia vastaan, mutta pituusrajoitusten vastaiset tavut (esim. *kravatti*, *strategia*, *ekstra*) kyllä hyväksytään. Jos käytetään binaarista kysymys—vastaus -järjestelmää (vain kyllä—ei -kysymyksiä), voidaan ajatella meneteltävän seuraavasti suomen kymmenen tavutyyppin johtamiseksi: tuotetaan ensin tavujen »universaalinen joukko», ts. kaikki mahdolliset CV-jonot (esim. CCCCCVVC, VVVVVCCC, CC, VCVVVVVC jne.) ja suodatetaan sitten tästä joukosta pois ne, jotka eivät täytä vokaalivaatimuksia ja pituusrajoituksia. Suodattaminen voi tapahtua esim. siten, että kysytään seuraavat neljä kyllä—ei -kysymystä:

- (1a) Onko CV-jonossa 1—2 vokaalia?
- (1b) Ovatko CV-jonon vokaalit peräkkäin?
- (2a) Edeltääkö ensimmäistä vokaalia 0—1 segmenttiä?
- (2b) Seuraako ensimmäistä vokaalia 0—2 segmenttiä?

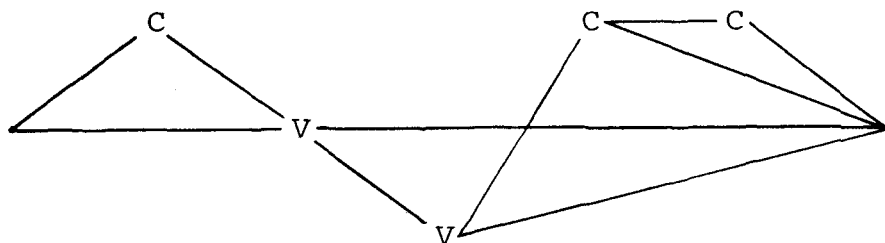
Ne CV-jonot ja vain ne CV-jonot, joiden suhteen vastaus on jokaiseen neljään kysymykseen myöntävä, ovat suomen kielessä hyväksyttäviä tavutyypppejä. Ne CV-johdot, jotka aiheuttavat yhteenkin neljästä kysymyksestä kielteisen vastauksen, ovat »epäsuomalaisia» tavutyypppejä.

Suomen tavujen rakennetta voidaan kuvata myös seuraavasti (ks. kuvaa 1 s. 266). Tavussa on neljä mahdollista positiota, kohtaa: alku, ydin, ytimenjatke ja loppu. Näistä muut kuin ydin ovat valinnaisia: ne joko esiintyvät tai ovat esiintymättä. Neljä positiota voidaan kukin täyttää vain yhdellä segmentillä, siten että alku ja loppu ovat aina konsonantteja, ydin on aina vokaali ja ytimenjatke on joko konsonantti tai vokaali. Suomen tavujen rakenne on selvästi sen sonorisuusperiaatteen mukainen, jonka mukaan vähäsonoriset äänteet (konsonantit) sijaitsevat tavun laidoilla, runsassonoriset äänteet (vokaalit) tavun keskuksessa. Kuten jäljempää käy ilmi, myös ytimenjatke sopii tähän kaavaan: ytimenjatkeena voivat esiintyä vain kaikkein sonorimmat konsonantit (likvidat ja nasaalit) tai vähäsonoriset vokaalit (muut kuin väljät vokaalit).

Kiintoisa seikka, joka käy ilmi edellä olevasta kaaviosta ja käsittelemistäni rajoituksista, on tavun kokonaispituus. Supisuomalainen tavu on korkeintaan neljän segmentin (äänteen/kirjaimen) mittainen. Jos tavu päättyy kahteen konsonanttiin, tavussa ei voi olla kahta vokaalia (esim. **kiers-ki*, **taunt-ti*), eli jos tavussa on kaksi vokaalia, tavu ei voi päättyä kahteen konsonanttiin. Juuri tämä rajoitus tekee mahdolliseksi sen, että esim. tavujen *kaut* ja *kant* kolmannet segmentit saadaan rinnasteisiksi: nämä segmentit esiintyvät tavun ytimen ja lopun välissä. Näin oikeastaan ytimenjatkeen käsite perustuu juuri kyseiseen tavun maksimipituuteen. Kaisa Häkkinen on kiinnittänyt huomioni siihen, että esim. sukunimessä *Suorttanen* (joka tajuttaneen supisuomalaiseksi, huolimatta siitä että se alkujaan tulee ruotsin sanasta *svart*) neljän segmentin

rajoitus rikkoutuu. Saattaa hyvin olla, että tämä rajoitus ei enää ole suomessa puhtaasti äänneopillisena rajoituksena voimassa. (Se ei ehkä enää ole produktiivinen.) Jos *suort*-tyyppiset tavut hyväksytään supisuomalaisiksi tavuiksi, edellä oleva tavarakennetta kuvaava kaavio on muutettava seuraavanlaiseksi:

KUVA 2.



Tämän mukaisessa viisipositioisessa tavarakenteessa eivät tavun ensimmäistä vokaalia seuraava toinen vokaali (esim. *u* tavussa *kaut*) ja ensimmäistä vokaalia seuraava sonoranttinen konsonantti (esim. *n* tavussa *kant*) enää muodostakaan yhtä kategoriaa, ytimenjatketta. Nyt tavu jäsenyy pikemmin niin, että tavun koko vokaaliaines (tavun positiot 2 ja 3) muodostavat tavun ytimen ja tavun lopun koko konsonanttiaines (positiot 4 ja 5) muodostavat tavun lopun. Sekä tavun ydin että loppu voivat tämän tulkinnan mukaan käsittää siis joko yhden tai kaksi segmenttiä.

Huolimatta *suort*-tyyppisistä viisisegmenttisistä tavuista, joita esiintyy erisnimissä ja murteellisissa sanoissa, tulkitsen suomen tavujen maksimipituuden olevan neljä segmenttiä. Tämä ratkaisu on sopusoinnussa sen periaatteen kanssa, jonka mukaan esitykseni perustuu hyvin suppeaksi ymmärrettyyn suomen yleiskieleen.

Edellä mainitut kymmenen tavutyyppiä saadaan viideksi, jos jätetään huomiotta se, esiintyykö tavun alussa konsonantti vai ei. Nämä tavutyyppit ovat seuraavat:

- (1) (C)V
- (2) (C)VC
- (3) (C)VV
- (4) (C)VVC
- (5) (C)VCC

Tavutyyppien käsittelyn jälkeen voidaan nyt ryhtyä tarkastelemaan yksityisiä tavuja eli sitä, mitkä yksityiset konsonantit ja vokaalit voivat esiintyä C:nä ja V:nä kyseisissä tavutyypeissä. Tehtävä on kaksijakoinen: ensin on esitettävä konsonanttien ja vokaalien inventaari ja sitten ne rajoitukset, jotka koskevat yksityisiä konsonantteja ja vokaaleja eri tavuasemissa.

Tulkitsen suomen foneemijärjestelmän käsittävän 13 konsonanttia ja 8 vokaalia. Nämä ovat seuraavat (esitän luettelot karkeasti siten, että vaakakseli kuvaa ääntymäpaikkaa ja pystyakseli ääntymätapaa, sonorisuutta):

| | | |
|---|---|---|
| p | t | k |
| | s | |
| | d | h |
| m | n | ŋ |
| | r | |
| | l | |
| v | j | |

| | | |
|---|---|---|
| i | y | u |
| e | ö | o |
| | ä | a |

Jos jokainen 13 konsonantista ja 8 vokaalista voisi esiintyä missä tahansa tavuasemassa, yllä mainitut 10 tavutyyppejä olisivat edustuneina suomen kielessä kaikkiaan 33 042 erilaisena tavuna. Eri tavutyyppejä edustaisivat seuraavat määrät erilaisia tavuja:

| tavutyyppi: | erilaisten tavutyyppeiden teoreettinen lukumäärä: |
|-------------|--|
| (1) V | = 8 |
| (2) CV | $13 \times 8 = 104$ |
| (3) VC | $8 \times 13 = 104$ |
| (4) CVC | $13 \times 8 \times 13 = 1\,352$ |
| (5) VV | $8 \times 8 = 64$ |
| (6) CVV | $13 \times 8 \times 8 = 832$ |
| (7) VVC | $8 \times 8 \times 13 = 832$ |
| (8) CVVC | $13 \times 8 \times 8 \times 13 = 10\,816$ |
| (9) VCC | $8 \times 13 \times 13 = 1\,352$ |
| (10) CVCC | $13 \times 8 \times 13 \times 13 = 17\,576$ |
| | <hr/> yhteensä 33 042 |

Konsonantteja ja vokaaleja koskee kuitenkin joukko rajoituksia, joiden takia erilaisten tavujen määrä on todellisuudessa huomattavasti pienempi kuin

33 042. Käsittelen seuraavaksi niitä rajoituksia, jotka koskevat yksityisten konsonanttien ja vokaalien esiintymistä eri tavuissa.

(1) Tavun alun konsonanttina voi esiintyä mikä tahansa 13 konsonantista, mutta sananalkuisissa tavuissa alkukonsonanttina eivät tule kyseeseen *d* ja *ŋ*. Nämäkin konsonantit esiintyvät tavunalkuisina sanan sisässä (esim. *kah-den* ja *lan-ŋat*). Koska en tässä artikkelissa erota sananalkuisuutta ja sanansisäisyyttä, *d* ja *ŋ* eivät mitenkään poikkeaa muista konsonanteista.

(2) Jos tavussa on vain yksi vokaali, tämä voi olla mikä tahansa 8 vokaalista. Jos vokaaleja on kaksi, vokaalijonoa koskee kaksi huomattavaa rajoitusta.

(a) Vokaalisoinnun takia toisena vokaalina ei voi olla takavokaali *u*, *o* tai *a* silloin, kun ensimmäinen vokaali on *y*, *ö* tai *ä*. Teoreettisesti mahdollisia kahden vokaalin jonoja on 8×8 eli 64. Vokaalisoinnun takia kiellettyjä vokaalijonoja (esim. *yu*, *öu*, *äu*, *yo*) on kaikkiaan 3×3 eli 9. Näin siis VV-jonon lukumäärä pienenee 55:ksi.

(b) Jos vokaalisointu olisi ainoa diftongeja (samaan tavuun kuuluvia kahden eri vokaalin jonoja) koskeva rajoitus, odottaisi suomessa olevan kaikkiaan 38 diftongia (*i* ja *e* voisivat olla 7 eri vokaalin edellä ja muut vokaalit voisivat kukin olla 4 eri vokaalin edellä: $7 + 7 + 4 + 4 + 4 + 4 + 4 = 38$). Todellisuudessa suomessa esiintyy kuitenkin vain 18 diftongia. Luku on pienempi kuin 38 siitä erityisestä syystä, että (a) väljä vokaali *a/ä* ei voi esiintyä diftongin jälkimmäisenä jäsenenä (esim. **ia*, **iä*, **ea*, **eä*, **ua*, **yä*) ja (b) välivokaali *e/ö/o* voi esiintyä diftongin jälkimmäisenä jäsenenä ainoastaan, jos edellinen jäsen on samapaikkainen suppea vokaali (vain diftongit *ie*, *yö* ja *uo* esiintyvät, esim. **ye*, **äe*, **iö*, **äö*, **eo*, **ao* eivät). Väljiä vokaaleja koskeva rajoitus pienentää teoreettista lukua 38 kahdeksalla, välivokaaleja koskeva rajoitus pienentää sitä kahdelatoista; näin päädytään teoreettisesta luvusta 38 diftongien lukuun 18 ($38 - 8 - 12 = 18$).

Kun jokainen kahdeksasta vokaalista voi esiintyä yksinään tai toisen identtisen vokaalin vieressä (esim. *ii*, *ee*, *yy*) ja kun diftongeja on kaikkiaan 18, saadaan tavun vokaaliston kokonaismääräksi $8 + 8 + 18$ eli 34. Kyseeseen tulevat vokaalit ja vokaalijonot ovat seuraavat: *i*, *y*, *u*, *e*, *ö*, *o*, *ä*, *a*; *ii*, *yy*, *uu*, *ee*, *öö*, *oo*, *ää*, *aa*; *yi*, *ui*, *ei*, *öi*, *oi*, *äi*, *ai*; *iy*, *ey*, *öy*, *äy*, *iu*, *eu*, *ou*, *au*; *ie*, *yö*, *uo*.

(3) Tavun lopun konsonantteja koskee kaksi rajoitusta. Toinen on yksinäiskonsonantteja, toinen kahden konsonantin yhtymiä koskeva.

(a) Jos tavunloppuisia konsonantteja on vain yksi, tämä voi olla mikä tahansa muu konsonantti kuin *d* tai puolivokaali *j/v*. Mahdollisia tavunloppuisia yksinään esiintyviä konsonantteja on siis 10 (esim. *tup-su*, *kat-se*, *lak-ki*, *kas-te*, *lah-ti*, *ram-pa*, *ran-ta*, *ran-ko*, *var-si* ja *kal-ja*).

(b) Jos tavun loppuun kuuluu kaksi konsonanttia (esim. *lant-ti*, *kars-ki*), edellinen konsonantti on aina nasaali *m/n/ŋ* tai likvida *r/l* ja jälkimmäi-

nen soinniton obstruentti $p/t/k/s$. Näiden ääntymätapaa koskevien rajoitusten lisäksi näitä kahta konsonanttia koskee rajoitus, jonka mukaan nasaalia voi seurata ainoastaan samapaikkainen obstruentti (ts. konsonanttijonot $*mt$, $*mk$, $*ms$, $*np$, $*nk$, $*np$, $*nt$ ja $*ns$ ovat kielenvastaisia). Jos konsonanttijonon edellisena konsonanttina esiintyvää nasaalia merkitään arkkifoneemin symbolilla N ja ajatellaan tämän arkkifoneemin ääntymäpaikan määräytyvän aina seuraavan konsonantin mukaiseksi, konsonanttijonon järjestelmä voidaan esittää seuraavasti:

$$\left\{ \begin{array}{c} \text{N} \\ \text{r} \\ \text{l} \end{array} \right\} \quad \left\{ \begin{array}{c} \text{P} \\ \text{t} \\ \text{k} \\ \text{s} \end{array} \right\}$$

Edellisena konsonanttina voi esiintyä yksi kolmesta mahdollisesta ja jälkimmäisenä yksi neljästä. Näin kahdesta konsonantista koostuvien tavunloppujen määräksi saadaan 3×4 eli 12. Esiintyvät CC-tyyppiset tavunloput ovat: mp , nt , nk , ns ; rp , rt , rk , rs ; lp , lt , lk ja ls (esim. *tamp-pi*, *kant-ta*, *ranh-ka*, *kans-sa*; *karp-pa*, *kurt-tu*, *tark-ka*, *kars-ta*; *vilpp-i*, *helt-ta*, *tilk-ka* ja *vals-si*). Luku 12 on varsin pieni: teoreettisesti mahdollisia kahden konsonantin jonoja olisi kaikkiaan 13×13 eli 169.

Kun tavun lopussa voi olla 10 erilaista yksinäiskonsonanttia ja 12 erilaista kahden konsonantin yhtymää, tavunloppuisten konsonanttien määräksi saadaan $10 + 12$ eli 22.

Olen edellä päätenyt siihen, että (1) tavun alussa voi olla 13 erilaista konsonanttia, (2) tavussa voi olla 8 yksinäisvokaalia, 8 kaksoisvokaalia ja 18 diftongia eli kaikkiaan 34 V/VV-jonoa ja (3) tavun lopussa voi olla 10 yksinäiskonsonanttia ja 12 kahden konsonantin jonoa eli kaikkiaan 22 C/CC-jonoa.

Esittämiäni yleisten tavun alkukonsonantteja, vokaalistoja ja loppukonsonantteja koskevien rajoitusten lisäksi suomessa on eräitä yksityisiä segmenttejä ja tavutyyppisiä koskevia erityisrajoituksia. Näistä huomattavimmat ovat seuraavat neljä.

- (1) Koska d ja η eivät voi esiintyä sananalkuisina ja koska tavutyyppi (C)VCC on mahdollinen ainoastaan sananalkuisena tavuna, tavutyyppin CVCC alkukonsonanttina voi tulla kyseeseen vain 11 (ei 13) konsonanttia.
- (2) Koska diftongit iy ja ey eivät esiinny sanan ensimmäisessä tavussa (tulkiten morfeemin *lepyhkä* sellaiseksi, että se ei kuulu käsittelemääni kieli-muotoon) ja koska nämä kaksi diftongia voivat esiintyä ensimmäistä tavua edempänä ainoastaan avotavussa (esim. *leiriytyä*, *kiteytyä* voivat sisältää iy - ja ey -diftongin, mutta sanoissa *leiriyttää* ja *kiteyttää* on aina tavunraja y :n edellä), nämä kaksi diftongia eivät tule kyseeseen (C)VVC-

tyyppisissä tavuissa. Tässä tavutyypissä ei ole siis 18, vaan ainoastaan 16 mahdollista diftongia (ei 26, vaan ainoastaan 24 kahden vokaalin jonoa). Tämä rajoitus vähentää teoreettisesti mahdollisten tavujen määrää 260:llä.

- (3) Suomen η on distribuutioltaan sillä tavoin suppea, että se ei voi esiintyä kahden samaan tavuun kuuluvan vokaalin jäljessä, esim. **kaaŋka*, **kienko*, **eunŋki*). (C)VVC-tyyppisissä tavuissa loppukonsonanttina ei siis tule kyseeseen 10 vaan ainoastaan 9 konsonanttia. Tämän rajoituksen takia teoreettisesti mahdollisten tavujen määrä pienenee edelleen 312:lla. On kuitenkin huomattava, että tämä η -äännettä koskeva rajoitus ei päde sellaisissa tapauksissa, joissa η on morfinrajalla tapahtuvan valinnaisen ääntymäpaikka-assimilaation tulos, esim. *saanŋko*, *tuonŋkin*. Näissä assimilaatiotapauksissa en ole tulkinnut η -äännettä aivan tavalliseksi η -äänneeksi, vaan olen pitänyt sitä n :nä.
- (4) r :n esiintyminen on siinä mielessä »epäsymmetristä», että tämä konsonantti ei voi esiintyä molemmilla puolilla vokaalia. Tämän rajoituksen takia suomessa ei voi esiintyä esim. tavuja *rir*, *ryr*, *rur*; *riir*, *rair*, *ryör*; *riir*, *rurp*, *rurs*. Teoreettisesti mahdollisten tavujen lukumäärä vähenee tämän rajoituksen takia edelleen kaikkiaan 66 tavulla.

Kun nyt tiedetään, mitä tavutyyppejä suomessa esiintyy ja mitä konsonantteja ja vokaaleja voi esiintyä tavujen eri asemissa, voidaan päästä selville myös siitä, montako erilaista teoreettisesti mahdollista tavua suomessa on. Eri tavutyyppien teoreettiset lukumäärät selviävät seuraavasta taulukosta (taulukon laskutoimituksia osoittavassa osassa olevat vähennyslaskut tarkoittavat r -konsonantin erityisrajoituksesta johtuvia tapauksia):

| tavutyyppi: | erilaisten tavutyyppien teoreettinen lukumäärä, kun mainitut rajoitukset otetaan huomioon: |
|-------------|--|
| (1) V | = 8 |
| (2) CV | $13 \times 8 = 104$ |
| (3) VC | $8 \times 10 = 80$ |
| (4) CVC | $13 \times 8 \times 10 - 8 = 1\ 032$ |
| (5) VV | = 24 |
| (6) CVV | $13 \times 24 = 312$ |
| (7) VVC | $24 \times 9 = 216$ |
| (8) CVVC | $13 \times 24 \times 9 - 26 = 2\ 782$ |
| (9) VCC | $8 \times 12 = 96$ |
| (10) CVCC | $11 \times 8 \times 12 - 32 = 1\ 024$ |
| | <hr/> |
| | yhteensä 5 678 |

Sivulla 269 olevasta taulukosta käy ilmi, että erilaisten tavujen lukumäärä olisi 33 042, jos kaikki konsonantit ja vokaalit voisivat esiintyä rajoituksitta kaikissa tavuasemissa. Nyt kun rajoitukset voidaan ottaa huomioon eli kun voidaan vähentää teoreettisesti mahdollisista tavuista systemaattiset aukot, lukumäärä pienenee 5 696:ksi. Tämä luku on vain n. 17 % luvusta 33 042. Saatu tavujen luku 5 696 ei kuitenkaan vielä edusta todellisten suomen kielessä esiintyvien tavujen lukumäärää. On selvää, että kieli ei käytä hyväkseen kaikkia mahdollisuuksiaan; kielessä on tässäkin suhteessa satunnaisia aukkoja. Seuraava tehtävä on pyrkiä luetteloimaan ja laskemaan kaikki suomen kielessä todella esiintyvät tavut.

Yksi mahdollisuus olisi käydä läpi tekstejä ja kerätä niistä kaikki erilaiset tavut. Tällä menetelmällä saisi tietenkin loppujen lopuksi selville suomen tavujen lukumäärän, mutta menetelmä olisi erittäin vaivalloinen. Jotta pääsisi edes lähelle kielen todellisten tavujen lukumäärää, tulisi käydä läpi miljoonien ja taas miljoonien sanojen laajuisia tekstejä. Olenkin menetellyt toisin: Olen ensin laatinut luettelon teoreettisesti mahdollisista 5 696 tavusta. Tämän jälkeen kolmihenkinen työryhmä (Kaisa Häkkinen, Jussi Niemi ja minä) ryhtyi etsimään sanoja, joissa eri tavut esiintyvät. Kukin toimi ensin muista erillään pyrkien löytämään asianomaisia sanoja »omasta päästään», Nykysuomen sanakirjasta ja Käänteissanakirjasta. Lopuksi kolmen henkilön itsenäisesti saamat tulokset yhdistettiin toisiinsa. Näin satunnaisten aukkojen määrä saatiin mahdollisimman pieneksi. On luonnollista, että tämä menetelmä tuottaa tuloksia, jotka ovat suurestikin riippuvaisia siitä, kuinka nokkelia tai ei-nokkelia sanojen etsijät sattuvat olemaan. Uskon, että saadut tulokset tästä huolimatta ainakin jossakin määrin kuvastavat todellista suomen kielen tavujen määrää.

Löytämiemme supisuomalaisten ja meille tutuissa sanoissa esiintyvien tavujen kokonaisluku on 2 983 eli pyöreästi 3 000. Tämän lisäksi olemme löytäneet 308 sellaista tavua, jotka esiintyvät ainoastaan erisnimissä, hiukan murteellisissa tai harvinaisissa (meille usein tuntemattomissa) sanoissa. Jos nämäkin tavut lasketaan mukaan, löytämiemme tavujen kokonaisluku on 3 291 eli pyöreästi 3 300. Näiden lukujen mukaan suomen kieli käyttää hyväkseen teoreettisesti mahdollisesta 5 696 tavusta ainoastaan n. 52 % (58 %); satunnaisten aukkojen osuus on siis lähes puolet teoreettisesti mahdollisista tavuista.

Löydetyt tavut voidaan seuraavaksi lajitella tavutyypeiksi. Tällöin voidaan nähdä, että kieli käyttää hyväkseen täydellisemmin toisia tavutyypejä kuin toisia (ts. satunnaisia aukkoja on toisissa tavutyypeissä vähemmän kuin toisissa). Löydettyjen tavujen tyypit ja kunkin tavutyyppin »käyttöprosentti» (paljonko teoreettisesti mahdollisista tavuista on todella käytössä) ilmenevät seuraavasta taulukosta. Sulkeissa olevat luvut perustuvat materiaaliin, jossa

mukana ovat myös erisnimet, hiukan murteelliset sanat, harvinaiset sanat jne.

| tavutyyppi: | | teoreettisesti mahdollisia tavuja: | löydettyjä tavuja: | tavutyypin käyttöprosentti: |
|-------------|------|--|-----------------------|--------------------------------|
| (1) | V | 8 | 8 | 100 |
| (2) | CV | 104 | 104 | 100 |
| (3) | VC | 80 | 75 (76) | 94 (95) |
| (4) | CVC | 1 032 | 912 (933) | 88 (90) |
| (5) | VV | 24 | 22 (23) | 92 (96) |
| (6) | CVV | 312 | 280 (281) | 90 (90) |
| (7) | VVC | 216 | 75 (83) | 35 (38) |
| (8) | CVVC | 2 782 | 1 233 (1 293) | 44 (46) |
| (9) | VCC | 96 | 12 (33) | 13 (34) |
| (10) | CVCC | 1 024 | 264 (459) | 26 (45) |

Taulukosta nähdään mm., että suomen tavuinventaarin kaksi ylivoimaisesti yleisintä tavutyyppeä ovat CVVC ja CVC. (Tämä ei tietenkään tarkoita sitä, että nämä kaksi tavutyyppeä olisivat yleisimmät myös puhutussa tai kirjoitetussa tekstissä.) Tavutyypin käyttöprosentin nähdään olevan (a) lyhimmissä tavutyypeissä 100 %, (b) keskipitkissä tavuissa (niissä, joissa ydinvokaalia seuraa vain yksi segmentti) noin 90 % ja (c) pisimmissä tavuissa alle 50 %. Tendenssi on varsin selvä: mitä pitempi tavu on ydinvokaalista tavun loppuun, sitä pienempi on tavun käyttöprosentti (eli sitä enemmän tavutyypissä on satunnaisia aukkoja). Tämä tendenssi on mitä ilmeisimmin universaalinen, ei ainoastaan suomen kieltä koskeva. Kielikohtainen sen sijaan saattaa olla se havainto, joka on taulukosta tehtävissä tavunalkuisten konsonanttien suhteen. Pisimmissä tavuissa (tavutyypit 7—10) käyttöprosentti on pienempi silloin, kun tavun alussa ei ole vokaalia (ts. kun tavu tässä mielessä on oikeastaan lyhyempi). Tätä suomen kielen tavujen tendenssiä voidaan nimittää tasapainon pyrkimykseksi: jos tavun ydinvokaalia seuraa maksimimäärä segmenttejä eli kaksi, niin kernaasti myös tavun alussa tällöin esiintyy tasapainottava konsonantti.

Tavujen tilastolliseen käsittelyyn kuuluu tietenkin myös tavujen esiintymisfrekvenssi. Tämän selville saamiseksi on ainoa keino edellä mainitsemani vaivalloinen laajojen tekstien läpikäyminen. Tingin kuitenkin tekstin laajuudesta ja kävin läpi verrattain suppean, vain 1 000 tavua käsittävän tekstin, Tuntemattoman sotilaan ensimmäiset 1 000 tavua. En kiinnittänyt huomiota yksityisiin tavuihin vaan ainoastaan kymmeneen tavutyyppeihin. Tavutyypin esiintymisfrekvenssit näkyvät seuraavasta taulukosta.

| tavutyyppi: | tavutyypin määrä: | tavutyypin esiintymistodennäköisyys: |
|-------------|----------------------|---|
| (1) V | 52 | .052 |
| (2) CV | 385 | .385 |
| (3) VC | 51 | .051 |
| (4) CVC | 299 | .299 |
| (5) VV | 13 | .013 |
| (6) CVV | 110 | .110 |
| (7) VVC | 2 | .002 |
| (8) CVVC | 84 | .082 |
| (9) VCC | 0 | .000 |
| (10) CVCC | 4 | .004 |
| yhhteensä | 1 000 | 1.000 |

Taulukosta nähdään mm., että vokaali alkuiset tavut ovat harvinaisempia kuin vastaavat konsonantti alkuiset: V-alkuisten tavujen esiintymistodennäköisyys on .118, C-alkuisten .382. Jos jätetään huomiotta se, esiintyykö tavun ensimmäisen vokaalin edellä konsonantti vai ei, tavutyypit saadaan viideksi ryhmäksi. Näiden esiintymistodennäköisyydet ovat seuraavat:

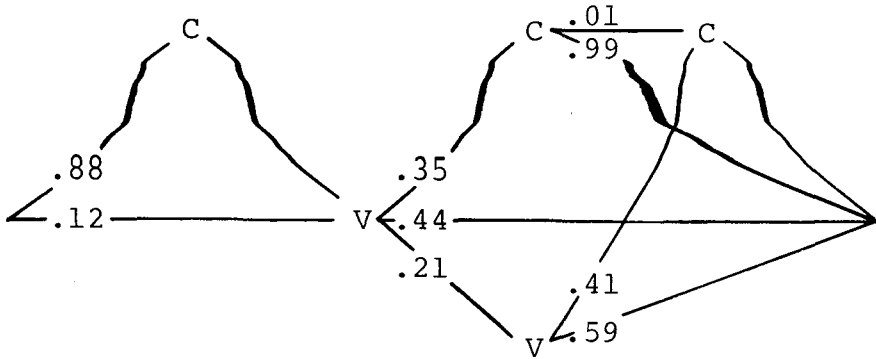
| tavutyyppi: | esiintymistodennäköisyys: |
|-------------|---------------------------|
| (C)V | .437 |
| (C)VC | .350 |
| (C)VV | .123 |
| (C)VVC | .086 |
| (C)VCC | .004 |

Taulukosta on nähtävissä mm. se odotuksen mukainen tendenssi, että pitemmät tavut ovat harvinaisempia kuin lyhyemmät. Tämän lisäksi taulukoista nähdään mm. se, että vokaalia edeltävä konsonantti ja vokaalia seuraava konsonantti vaikuttavat vastakkaisella tavalla tavutyypin esiintymistodennäköisyyteen. Yleisempiä ovat ne tavutyypit, joissa vokaalia edeltää konsonantti, kuin ne, joissa vokaalia ei edellä konsonantti (vokaalia edeltävä konsonantti on siis »luonnollinen», tunnusmerkitön). Toisaalta ne tavutyypit, joissa vokaalia seuraa konsonantti, ovat säännöllisesti harvinaisempia kuin ne, joissa vokaalia ei seuraa konsonantti (vokaalia seuraava konsonantti ei siis ole yhtä »luonnollinen» ja tunnusmerkitön kuin vokaalia edeltävä konsonantti).

Käyttämällä hyväksi saatuja tavutyypin esiintymistodennäköisyyksiä, kuvan 1 tavutyypin »tuottamisautomaatti» voidaan varustaa todennäköi-

syyskertoimilla, jotka osoittavat kaavion kussakin haaraumakohdassa, mikä valintojen todennäköisyys on silloin, kun suomalainen puhuu (kirjoittaa) äidinkieltään.

KUVA 3.



Kunkin tavutyypin esiintymistodennäköisyys saadaan kertomalla keskenään kaavioon merkityt todennäköisyydet. Kymmenen tavutyypin esiintymistodennäköisyyksiksi saadaan automaattista seuraavat luvut, joiden nähdään yleensä varsin hyvin täsmävän edellä mainittujen todellisten esiintymisprosenttien kanssa.

| tavutyyppi: | kaaviosta laskemalla saatavat esiintymis- todennäköisyydet: | todelliset esiintymis- todennäköisyydet: |
|-------------|---|--|
| (1) V | $.12 \times .44 = .053$ | .052 |
| (2) CV | $.88 \times .44 = .352$ | .385 |
| (3) VC | $.12 \times .35 \times .99 = .042$ | .051 |
| (4) CVC | $.88 \times .35 \times .99 = .277$ | .299 |
| (5) VV | $.12 \times .21 \times .59 = .015$ | .013 |
| (6) CVV | $.88 \times .21 \times .59 = .109$ | .110 |
| (7) VVC | $.12 \times .21 \times .41 = .010$ | .002 |
| (8) CVVC | $.88 \times .21 \times .41 = .076$ | .084 |
| (9) VCC | $.12 \times .35 \times .01 = .000$ | .000 |
| (10) CVCC | $.88 \times .35 \times .01 = .003$ | .004 |
| | <hr/> yhteensä | <hr/> 1.000 |

Taulukko osoittaa kuitenkin, että yhden kaaviosta laskemalla saatavan tavutyypin esiintymistodennäköisyys on melkoisen virheellinen: tavutyypin VVC saadaan kaaviosta viisi kertaa yleisemmäksi kuin tämä tavutyypin todellisuudessa on. Kyseessä on selvästi sama pisimpien tavujen tasa-

painon pyrkimys, joka löydettiin edellä tarkastelemalla eri tavutyyppien käyttöprosentteja. Tämän pyrkimyksen takia kaavio ei edusta puhdasta Markovin ketjua, jossa kunkin valinnan todennäköisyys on riippumaton aikaisemmin tehdyistä valinnoista: se, valitaanko tavuun ydinvokaalin ja sitä seuraavan segmentin jälkeen vielä loppukonsonantti vai ei, on jossakin määrin riippuvainen siitä, onko tavun alkuun valittu alkukonsonantti vai ei. (Tässä lienee paikallaan huomauttaa, että puhuessani »valinnoista» en viittaa psykologiseen todellisuuteen. Kyseessä on vain tapa kuvata tavujen rakennetta.)

Olen tässä artikkelissa käsitellyt suomen yleiskielen tavuja monelta eri kannalta. Olen toiminut seuraavasti:

- (1) Ryhmitin tavut kymmeneksi erilaiseksi C- ja V-symbolein ilmaistuksi tavutyypiksi.
- (2) Esitin kuvassa 1 kaavion, eräänlaisen tuottamisautomaatin, joka tuottaa kyseiset kymmenen tavutyyppeä.
- (3) Pohdin, millä eri tavoin suomen tavujen CV-jonoja koskevat rajoitukset voidaan esittää mahdollisimman yksinkertaisina sanallisina väittäminä.
- (4) Esitin suomen kielen konsonanttien ja vokaalien inventaarin ja ryhdyin pohtimaan suomen tavuja (ei enää CV-jonoina vaan) yksityisten konsonanttien ja vokaalien muodostamina jonoina.
- (5) Havaitsin, että esittämiäni suomen kielen tavujen CV-rakenteita ja foneemi-inventaaria käyttäen voitaisiin tuottaa kaikkiaan noin 33 000 erilaista tavua.
- (6) Väitin, että suomen tavujen teoreettinenkaan luku ei voi kuitenkaan olla suurempi kuin noin 5 700, siitä syystä että suomen konsonanteilla ja vokaaleilla on joukko systemaattisia esiintymisrajoituksia.
- (7) Pyrin luetteloimaan kaikki suomen tavut ja päädyin siihen, että suomessa esiintyy todellisuudessa noin 3 000—3 300 erilaista tavua.
- (8) Esitin alustavia (varsin suppeaan materiaaliin perustuvia) havaintoja suomen kymmenen tavutyypin esiintymisfrekvensseistä kirjoitetussa tekstissä; ks. kuvaa 3 s. 276.

Lopuksi minun olisi pitänyt käsitellä yksityisten tavujen (ei vain CV-jonoina ilmaistujen tavutyyppien) esiintymisfrekvenssejä. Tähän tehtävään minulla ei kuitenkaan ole ollut mahdollisuuksia.

On Finnish syllables

Kalevi Wiik

My purpose in this article is to give a concise description of the structure of the Finnish syllable and to figure out the number of different syllables in Finnish. I start by presenting a Markovian process (Fig. 1) which produces the ten existing native Finnish syllable types in terms of vowels and consonants. If the eight vowels and thirteen consonants of Finnish had no distributional restrictions, the total number of syllables would be about 33 000. Because of various kinds of systematic distributional restrictions such

as vowel harmony, and restrictions on consonant clusters, the number of theoretically possible syllables is much lower, about 5 700. Of these only about 3 000, or slightly more than 50 %, actually exist in the language. I conclude with tentative statistics on the frequency of occurrence of the ten syllable types. In Fig. 3 rough estimates (based on a text of only 1 000 syllables) are presented for the probabilities of choice at the various points of the automaton that produces the ten syllable types.

Osoite:

Address:

*Turun yliopisto
Fonetikan laitos
Linnankatu 9—11 A
SF-20100 Turku 10*