

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

HEIKKI LESKINEN

1. Tuomo Tuomen Suomen kielen käänteissanakirja on monella tavoin virkistänyt kielentutkimustamme. Eniten se on tietenkin tarjonnut apua johtaja- ja taivutusopin selvittelijöille, mutta myös muiden alojen edustajat ovat siitä hyötäneet. Itse asiassa Käänteissanakirjan ATK-tietokanta on luonut aikaisempaa huomattavasti paremmat edellytykset sanarakenteemme yksityiskohtaiselle kuvaukselle ja varsinkin sen statistiselle analyysille.

Yksi sellainen tutkimuskohde, johon käänteissanakirjan materiaalia ei ole vielä käytetty, on kirjakielen grafeemisto. Siitä on kyllä viime aikoina julkaistu useitakin esityksiä, mutta ne kaikki on laadittu juoksevan tekstin pohjalta. On kuitenkin itsestään selvää, että grafeemien tekstifrekvenssit eivät lankea yhteen niiden »sanafrekvenssien» kanssa, ts. kirjainten ja tietysti myös äänteiden lukumääräsuhteet ovat tekstiaineistossa kokonaan toiset kuin sanavarastossa. Jonkinlaiseksi näytteeksi tästä esittelen artikkelissani Käänteissanakirjan ATK-tietokantaan perustuvia havaintoja sanastomme 1. ja 2. position grafeemien sekä niiden muodostamien sekvenssien esiintymistiheydestä. Periaatteet, joiden mukaisesti aineisto on Nykysuomen sanakirjasta (NS) koostettu, käyvät ilmi Käänteissanakirjan (KS) sivulta XVI (kohdalla 2.3.). Eri grafeemien ja grafeemikombinaatioiden absoluuttiset ja prosenttiset frekvenssit olen saanut valmiiksi laskettuina prof. Tuomelta. Siitä lausun hänelle parhaat kiitokset.

Vertailukohteiksi sopivista, juoksevaa tekstiä edustavista tutkimuksista on laajin Matti Pääkkösen laatima nykyisen yleiskielen tilasto (YK; 1973: 318–322), jonka aineistona on peräti 3 130 382 grafeemiesiintymää. Sen julkaistussa versiossa on kuitenkin noteerattu vain kunkin konsonantin ja vokaalin kokonaismäärät ottamatta huomioon eri asemia. Tässä suhteessa ovat käyttökelpoisempia Vilho Setälän (1972) ja Kaisa Häkkisen (1983: 39–56) laskelmat, koska niissä on käsitelty erikseen sananalkuisia grafeemeja. Toisaalta taas niiden yhteisenä heikkoutena on materiaalipohjan yksipuolisuus: Setälän lähteenä on Uuden testamentin suomennos (UT) ja Häkkisen kansansatumukaelma (KSM).

2. Käänteissanakirjan hakusanojen alkugrafeemien absoluuttiset ja prosenttiset frekvenssit (f ja % I) käyvät ilmi taulukosta 1. Aineiston kokonaismäärä on 71 454. Siihen sisältyy 62 570 konsonanttia ja 8 884 vokaalia. Konsonanteista on »kotoperäisiä» 60 474 ja vieraslähtöisiä 2 096. Ainoa epäsuomalainen vokaali on yhden kerran esiintyvä å. Alkugrafeemien yleisyysjärjestys on k p s t h l v r m a n j e i o u (d) y (f) ä (b) (g) ö (c) (z) (w) (q) (å) (19/28). Luettelon kärkipäätä hallitsevat konsonantit; vokaaleista vain a mahtuu kahden oman konsonantin (n:n ja j:n) edelle. Tekstifrekvensseihin pohjautuvat aineistot poikkeavat selvästi sanakirjamateriaalista: grafeemien yleisyysjärjestys on KSM:ssa k s j p t m v n h o l e a i r y u ä ö (19) ja

TAULUKKO 1: KS:n hakusanojen alkugrafeemien absoluuttiset ja prosenttiset frekvenssit verrattuina kansansatumu-kaelman (% II) ja Uuden testamentin suomennoksen (% III) sananalkuisten grafeemien prosenttisiin tekstifrekvensseihin

G _a	f	% I	% II	% III
a-	2 697	3,77	3,0	2,62
b-	426	0,60	—	0,00
c-	58	0,08	—	—
d-	680	0,95	—	0,01
e-	1 451	2,03	3,6	5,01
f-	552	0,77	—	0,08
g-	330	0,46	—	0,01
h-	5 906	8,27	5,7	9,49
i-	1 403	1,96	2,4	1,93
j-	1 822	2,55	9,4	12,95
k-	10 645	14,90	15,1	9,91
l-	5 302	7,42	4,0	2,70
m-	3 907	5,47	7,8	8,13
n-	2 673	3,74	5,7	4,41
o-	1 287	1,80	4,6	7,87
p-	7 464	10,45	8,9	5,28
q-	6	0,01	—	—
r-	4 650	6,51	2,3	1,55
s-	6 662	9,32	11,3	10,86
t-	6 256	8,76	8,9	9,26
u-	876	1,23	0,7	0,88
v-	5 187	7,26	6,0	5,43
w-	16	0,02	—	—
y-	578	1,81	1,0	1,10
z-	28	0,04	—	—
å-	1	0,00	—	—
ä-	518	0,72	0,4	0,47
ö-	73	0,10	—	0,03
(28)	71 454	100,00	100,0	100,00

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

UT:ssa j s k h t m o v p e n l a i r y u ä (f) ö (d) (g) (b) (19/23). Konsonanttien rivin katkaisee nyt ensin o, sitten e, ja a on vasta kolmas vokaali; lisäksi i sijoittuu r:n, kaikkein pientaajuisimman konsonantin, edelle. Vierasperäiset konsonantit (yht. 2 097 eli 2,93 %), joihin tässä on luettu myös sananalkuinen d, jäävät odotuksenmukaisesti frekvenssilistojen loppuun.

Alkukonsonanttien ja -vokaalien yleisyysuhde on KS:ssa 88 : 12, KSM:ssa Häkkisen mukaan 84 : 16 (1983: 43) ja UT:ssa 80 : 20. Luvut ovat kyllä varsin lähellä toisiaan, mutta vokaalialkuisten sanojen (oik. saneiden) esiintymistiheys näyttää juoksevassa tekstissä olevan aina jonkin verran suurempi kuin leksikossa.

Sananalkuisten konsonanttigrafeemien yleisyysjärjestys on lähteissäni seuraava: KS k p s t h l v r m n j, KSM k s j p t m v h n l r, UT j s k h t m v p n l r (ks. taulukkoa 2). Eniten vaihtelevat konsonanttien j (2,6–9,4–13,0 %), r (6,5–2,3–1,6 %), l (7,4–4,0–2,7 %), m (5,5–7,8–8,1 %) ja k (14,9–15,1–9,9 %) prosenttiset osuudet. Lähimpänä toisiaan ovat tietenkin juoksevaa tekstiä edustavat KSM ja UT, mutta niidenkin välillä on huomattavia eroja. Ne osoittavat, että alkugrafeemien suosio riippuu suuresti määrin tekstityypistä. Vilho Setälä on ilmeisen oikeassa arvellessaan, että j:n yleisyys UT:ssa johtuu paitsi lukuisista ja-konjunktioista myös sanojen *Jeesus*, *Jumala* ja *juutalainen* runsaudesta ja että h:n lukua ovat lisänneet uskonnollisessa kielessä usein toistuvat sanat *henki*, *herra*, *he*, *hän* ja *hyvä* (mts. 10).

TAULUKKO 2: KS:n hakusanojen konsonanttisten alkugrafeemien absoluuttiset ja prosenttiset frekvenssit (% I konsonanteista, % II kaikista) sekä KSM:n (% III) ja UT:n (% IV) sananalkuisten konsonanttigrafeemien prosenttiset tekstifrekvenssit

C _a	f	% I	% II	% III	% IV
k-	10 645	17,0	14,9	15,1	9,9
p-	7 464	11,9	10,5	8,9	5,3
s-	6 662	10,6	9,3	11,3	10,9
t-	6 256	10,0	8,8	8,9	9,3
h-	5 906	9,4	8,3	5,7	9,5
l-	5 302	8,5	7,4	4,0	2,7
v-	5 187	8,3	7,3	6,0	5,4
r-	4 650	7,4	6,5	2,3	1,6
m-	3 907	6,2	5,5	7,8	8,1
n-	2 673	4,3	3,7	5,7	4,4
j-	1 822	2,9	2,6	9,4	13,0
muut ¹	2 096	3,3	2,9	—	0,1
Yht.	62 570	100,0	87,6	85,1	80,1

¹ Grafeemit b-, c-, d-, f-, g-, q-, w- ja z-

Eri lähteiden sananalkuinen (omaperäinen) konsonantisto jakaantuu ään-
töpaikan mukaan seuraavasti:

Ryhmä	KS		KSM	UT
	f	%	%	%
Dentaalit	25 543	42,2	37,8	36,0
Labiaalit	16 558	27,4	26,7	23,5
Palataalit	12 467	20,6	28,8	28,6
Laryngaalit	5 906	9,8	6,7	11,9
Yhteensä	60 474	100,0	100,0	100,0

Aineistot poikkeavat toisistaan yllättävän vähän. Dentaalit ovat niissä kaikissa suurtaajuisin ryhmä. Niiden asema ei tosin ole sanan alussa yhtä hallitseva kuin koko grafeemistossamme, jossa peräti kolme konsonanttia neljästä on dentaaleja (Karlsson 1983: 77). Alkukonsonanttien frekvenssi-jakaumassa herättää huomiota oikeastaan vain se, että sekä KSM:ssä että UT:ssä on palataaleja enemmän kuin labiaaleja, kun taas KS:n sanamateriaalissa tilanne on päinvastainen. UT:n laryngaalien (= h:n) paljouden syy on todettu edellä.

Alkukonsonanttien ryhmittely ääntötavan perusteella antaa puolestaan seuraavanlaisen tuloksen:

Ryhmä	KS		KSM	UT
	f	%	%	%
Klusiilit	24 365	40,0	38,7	30,6
Frikatiivit	12 568	20,8	20,0	25,5
Likvidat	9 952	16,5	7,4	5,4
Nasaalit	6 580	10,9	15,9	15,6
Puolivokaalit	7 009	11,6	18,1	23,0
Yhteensä	60 474	100,0	100,0	100,0

Taulukosta pistää silmään obstruenttien eli klusiilien ja frikatiivien taajuus-ylivoima: niitä on leksikkoaineistossa runsaat 60 % ja juoksevaa tekstiä edustavissa lähteissäkin lähes saman verran. Luvut ovat jopa hieman isommat kuin koko grafeemiaineistosta lasketut (klusiileja 34,2 %, frikatiiveja 18,7 % eli obstruentteja yhteensä 52,9 %; ks. Karlsson 1983: 76). Toinen merkille pantava seikka on se, että KS:ssä on likvidojen osuus huomattavasti suurempi mutta nasaalien ja puolivokaalien osuus paljon pienempi kuin KSM:ssä ja UT:ssä. Likvidalla alkavia sanoja käytetään siis ainakin näissä teksteissä suhteellisesti harvemmin kuin nasaalilla ja puolivokaalilla alkavia.

Alkukonsonanttien määrää lisää KS:n aineistossa sananalkuinen kaksi-konsonanttisuus, joka on erityisen yleistä lainasanoissa. Sen vaikutus ei kui-

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

tenkaan ole kovin merkittävä, sillä obstruenttialkuisten yhtymien kokonaisluku on vain 1 971 (kC- 423, tC- 350, pC- 658, sC- 531, hC- 9; ks. taulukkoa 8). Vaikka kaksikonsonanttiset tapaukset (yht. 2 365) jätettäisiin laskelmista sivuun, nousee obstruenttien osuus 59,8 %:iin (klusiilit 39,2 %, frikatiivit 20,6 %).

TAULUKKO 3: KS:n hakusanojen vokaalisten alkugrafeemien absoluuttiset ja prosenttiset frekvenssit (% I konsonanteista, % II kaikista) sekä KSM:n (% III) ja UT:n (% IV) sananalkuisten vokaaligrafeemien prosenttiset tekstifrekvenssit

V _a	f	% I	% II	% III	% IV
a-	2 697	30,4	3,8	3,0	2,6
e-	1 451	16,3	2,0	3,6	5,0
i-	1 403	15,8	2,0	2,4	1,9
o-	1 287	14,5	1,8	4,6	7,9
u-	876	9,9	1,2	0,7	0,9
y-	578	6,5	0,8	1,0	1,1
ä-	518	5,8	0,7	0,4	0,5
ö-	73	0,8	0,1	—	0,0
å-	1	0,0	0,0	—	—
Yht.	8 884	100,0	12,4	15,7	19,9

Alkuvokaalien yleisyysjärjestys on eri aineistoissa seuraava: KS a e i o u y ä ö (å), KSM o e a i y u ä (ö), UT o e a i y ä ö. Tässä näkyy sana- ja tekstiaineistojen välinen ero harvinaisen selvänä: molemmissa tekstiaineistoissa o on — ilmeisesti usein toistuvan *olla*-verbin ansiosta — taajuuslistan kärjessä, e toisena ja a vasta kolmantena. Itse asiassa KS:n alkuvokaalien ja yleiskielen (YK) kaikkien vokaalin (a i e o u ä y ö å) yleisyysjärjestys ovat melko lähellä toisiaan (ks. Pääkkönen 1973: 320). Taajuussuhteet eivät kylläkään ole samat. Esimerkiksi a:n osuus vokaaleista on KS:ssa 30,4 %, mutta YK:ssa 24,6 %, i:n osuus edellisessä 15,8 %, jälkimmäisessä 22,3 % jne.

Takavokaalit näyttävät olevan sananalkuisina kaikissa aineistoissa yleisempiä kuin etuvokaalit:

Ryhmä	KS		KSM	UT
	f	%	%	%
Etuvok.	4 023	45,3	47,1	42,9
Takavok.	4 860	54,7	52,9	57,1
Yhteensä	8 883	100,0	100,0	100,0

Tilanne on päinvastainen kuin koko grafeemistossa, missä takavokaalien osuus on 646,1 % ja etuvokaalien 53,9 % (Karlsson 1983: 75). — Takavokaa-

lien yleisemmyys korostuu, kun otetaan huomioon, että kielessämme on viisi etuvokaalia mutta vain kolme takavokaalia. KS:ssa on yhden sananalkuisen takavokaalin keskimääräinen frekvenssi 1620, yhden etuvokaalin vain 804.

Vokaalien väljyysasteen mukainen jakauma on varsin kiintoisa. Koko grafeemistoa edustavassa tekstiaineistossa (YK) ovat puolisupeat vokaalit (29,1 %) selvästi harvinaisempia kuin suppeat (36,5 %) ja väljät (34,4 %); ks. Karlsson 1983: 75. Sanan alussa tämä ei pidä paikkaansa:

Ryhmä	KS		KSM	UT
	f	%	%	%
Suppeat	2 857	32,2	26,1	19,6
Puolisupeat	2 811	31,6	52,2	64,9
Väljät	3 215	36,2	21,7	15,5
Yhteensä	8 883	100,0	100,0	100,0

KS:n sana-aineistossa on väljien vokaalien määrä suurin; suppeiden ja puolisupeiden määrät ovat melkein tasoissa. Juoksevassa tekstissä (KSM ja UT) yleisyysjärjestys on suorastaan nurinkurinen: puolisupeat, suppeat ja väljät. Lisäksi puolisupeiden osuus on reippaasti isompi kuin kahden muun ryhmän yhteensä.

Laveat vokaalit ovat koko grafeemistossa lähes kolme kertaa yleisempiä kuin pyöreät (Karlsson *ibid.*). Sanan alussa suhteet vaihtelevat aineistotyypin mukaan:

Ryhmä	KS		KSM	UT
	f	%	%	%
Laveat	6 069	68,3	59,9	50,4
Pyöreät	2 814	31,7	40,1	49,6
Yhteensä	8 883	100,0	100,0	100,0

KSM:ssa ja UT:ssa on pyöreiden alkuvokaalien tekstifrekvenssi erittäin korkea, UT:ssa jopa melkein sama kuin laveiden. Muistettakoon jälleen o-alkuisten sanojen, erityisesti *olla*-verbin, suurtaajuus. KS:n sana-aineistossa laveiden ja pyöreiden alkuvokaalien suhde on sitä vastoin likimain 3:7.

3. Taulukossa 4 on esitelty KS:n hakusanojen 2. position grafeemien eli ns. kakkosgrafeemien absoluuttiset ja prosenttiset frekvenssit. Vertailukohteiksi on nyt otettu UT:n kakkosgrafeemien ja yleiskielen (YK) kaikkien grafeemien prosenttiset tekstifrekvenssit. KSM:n kakkosgrafeemeista ei Häkkisen artikkelissa ole tietoja. — KS:n kakkosgrafeemien kokonaismäärä on 71 453

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

eli yhtä vähemmän kuin ykkösgrafeemien; syynä on yksivokaalinen interjektio *o*. Konsonantti- ja vokaalijakauma on melkein peilikuva ykkösgrafeemien jakaumasta: konsonantteja on 9 921 (13,9 %) ja vokaaleja 61 532 (86,1 %). Konsonanteista on vierasperäisiä 222 (0,31 %), d mukaan luettuna 449 (0,63 %).

Kakkosgrafeemien yleisyysjärjestys on eri lähteissä seuraava: NS *a i u e o ä y r l ö n s h k t m p d v j* (b) (g) (f) (c) (x) (z) (w) (20/27), UT *a i e u ä o l n y t s h r v p m k d ö j* (20), YK *a i t n e s l k o u ä m v r j h y p d ö* (g) (b) (f) (c) (w) (å) (q) (20/27). KS:n aineistossa sijoittuvat nyt vokaalit – tosin hieman oudosti vaihdellen – luettelon kärkeen; konsonanteista vain likvi-

TAULUKKO 4: KS:n hakusanojen kakkosgrafeemien absoluuttiset ja prosenttiset (% I) frekvenssit sekä UT:n kakkosgrafeemien (% II) ja yleiskielen kaikkien grafeemien (% III) prosenttiset tekstifrekvenssit

G ₂	f	% I	% II	% III
a	14 570	20,39	22,74	11,80
b	85	0,12	0,00	0,06
c	9	0,01	0,00	0,03
d	227	0,32	0,31	0,83
e	8 044	11,26	12,59	8,22
f	52	0,07	0,00	0,06
g	65	0,09	0,00	0,12
h	697	0,98	0,88	1,82
i	10 646	14,90	14,60	10,68
j	159	0,22	0,14	1,92
k	667	0,93	0,31	5,32
l	1 724	2,41	5,33	5,73
m	535	0,75	0,53	3,41
n	1 207	1,69	4,05	8,68
o	7 688	10,76	7,72	5,27
p	508	0,71	0,57	1,72
q	0	0,00	0,00	0,00
r	2 239	3,13	0,72	2,25
s	890	1,25	1,72	7,84
t	631	0,88	2,38	9,77
u	10 134	14,18	12,11	5,02
v	215	0,30	0,62	2,47
w	2	0,00	0,00	0,01
x	5	0,01	0,00	—
y	4 016	5,62	2,74	1,79
z	4	0,01	0,00	—
å	0	0,00	0,00	0,00
ä	4 763	6,67	9,67	4,69
ö	1 671	2,34	0,25	0,47
(29)	71 453	100,00	100,00	100,00

dat r ja l tunkeutuvat pientaajuisimman vokaalin eli ö:n edelle. UT:n kakkosgrafeemien tekstifrekvenssiin pohjautuva järjestys on muuten samantapainen, mutta siinä pääsevät konsonantit l ja n paitsi ö:n myös y:n ohitse. Huomiota herättää sitä paitsi ö:n jääminen toiseksi viimeiseksi. Koko grafeemiston yleisyysjärjestyksessä (YK) vokaalit ja konsonantit vuorottelevat. Yhteistä on sentään se, että vokaalit a ja i ovat listan alussa ja vierasperäiset konsonantit sen lopussa.

Niin kuin edellä on todettu, konsonanttisten kakkosgrafeemien osuus on leksikkoaineistossa pieni, vain 13,9 % grafeemien kokonaismäärästä. Kovin paljon suurempi se ei näytä olevan juoksevassa tekstissäkään (UT:ssa 17,6 %). Tämä on tietenkin odotuksenmukaista, koska sanojemme valtaenemmistö alkaa yksinäiskonsonantilla. Hämmästyttä saattaa sen sijaan herättää konsonanttisten kakkosgrafeemien yleisyysjärjestys; ks. taulukkoa 5.

TAULUKKO 5: KS:n hakusanojen konsonanttisten kakkosgrafeemien absoluuttiset ja prosenttiset (% I ja II) frekvenssit sekä UT:n konsonanttisten kakkosgrafeemien (% III) ja yleiskielen kaikkien konsonanttigrafeemien (% IV) prosenttiset tekstifrekvenssit

C	f	% I	% II	% III	% IV
r	2 239	22,57	3,13	0,72	2,25
l	1 724	17,38	2,41	5,33	5,73
n	1 207	12,17	1,69	4,05	8,68
s	890	8,97	1,25	1,72	7,84
h	697	7,03	0,98	0,88	1,82
k	667	6,72	0,93	0,31	5,32
t	631	6,36	0,88	2,38	9,77
m	535	5,39	0,75	0,53	3,41
p	508	5,12	0,71	0,57	1,72
d	227	2,29	0,32	0,31	0,83
v	215	2,17	0,30	0,62	2,47
j	159	1,60	0,22	0,14	1,92
muut ¹	222	2,24	0,31	0,00	0,28
Yht.	9 921	100,00	13,88	17,56	52,06

¹ Grafeemit b, c, f, g, q, w, x ja z

Konsonanttien useusjärjestys on KS:ssa r l n s h k t m p d v j, UT:ssa l n t s h r v p m d k j, YK:ssa t n s l k m v r j h p d. Leksikkoaineiston suurtaajuisimmat konsonantit ovat siis r, l ja n. Niiden osuus 2. position konsonanttien kokonaismäärästä on peräti 52 % (sanan alussa vain 20,2 %). UT:n juoksevassa tekstissä asetelma on samantapainen: kärkipaikkoja pitävät hallussaan l ja n; r tosin sijoittuu vasta kuudenneksi, mutta näiden kolmen

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

konsonantin osuus on silti 57,5 % (sanan alussa 10,9 %). — Yleiskielen kaikkien konsonanttigrfeemien tekstifrekvenssit (% IV) poikkeavat jyrkästi edellisistä.

Seuraavassa taulukossa on esitetty KS:n ja UT:n konsonanttisten kakkosgrfeemien ääntöpaikan mukainen jakauma sekä vastaavat tiedot yleiskielen konsonanttigrfeemistosta.

Ryhmä	KS		UT	YK
	f	%	%	%
Dentaalit	6 981	69,7	82,6	67,4
Labiaalit	1 397	14,1	9,8	14,8
Palataalit	891	9,0	2,6	14,1
Laryngaalit	697	7,0	5,0	3,5
Yhteensä	9 901	100,0	100,0	100,0

Toisen position konsonantistoa hallitsevat ylivoimaisesti dentaalit. Juoksevassa tekstissä niiden osuus yhä vain kasvaa kaikkien muiden ryhmien kustannuksella. Vertailu s:n 404 taulukkoon osoittaa, että sanan alussa dentaaileja käytetään tuntuvasti vähemmän (KS 42,2 %, UT 36,0 %) kuin kakkosasemassa.

Ääntötavan mukainen ryhmittely ilmaisee kakkosaseaman konsonanttien erikoislaatuisen jakauman varsin havainnollisesti:

Ryhmä	KS		UT	YK
	f	%	%	%
Klusiilit	2 183	22,0	20,3	34,2
Frikatiivit	1 639	16,5	14,8	18,7
Likvidat	3 963	39,9	34,5	15,3
Nasaalit	1 742	17,6	26,1	23,2
Puolivokaalit	374	3,8	4,3	8,5
Yhteensä	9 901	100,0	100,0	100,0

Likvidojen ja nasaalien osuus on KS:n sana-aineistossa 57,5 %, mutta obstruenttien vain 38,5 %. UT:n tekstimateriaalin vastaavat luvut ovat 60,6 % ja 35,1 %. Sanan alussa (ks. s. 404) asetelma on aivan päinvastainen: KS 27,4 % ja 60,8 %, UT 21,0 % ja 56,1 %. Samoin on koko konsonantiston (YK) laita.

Edellä todetusta ei kuitenkaan vielä selviä, miksi r on KS:n hakusanojen 2. position yleisin konsonantti (22,6 %), vaikka se sijoittuu UT:n tekstimateriaalissa vasta kuudennelle sijalle (4,1 %). Asiantila käy ymmärrettäväksi, kun otetaan huomioon r:n suosio sananalkuisten konsonanttiyhtymien jälkimmäisenä jäsenenä (ks. taulukkoa 8): Cr-tapauksia on KS:ssa kaikkiaan

1 219 eli 51,5 % CC-alkuisten kokonaismäärästä. Aiheellista lieneekin tarkastella erikseen 2. position konsonanttien esiintymistä a) konsonanttiyhtymissä ja b) vokaali + konsonantti -kombinaatioissa:

a)			b)		
CC	f	%	VC	f	%
Cr	1219	51,5	Vl	1340	17,7
Cl	384	16,2	Vn	1169	15,5
Ct	225	9,5	Vr	1020	13,5
Cp	120	5,1	Vs	772	10,2
Cs	118	5,0	Vh	657	8,7
Ck	99	4,2	Vk	568	7,5
Cv	73	3,1	Vm	517	6,8
Ch	40	1,7	Vt	406	5,4
Cn	38	1,6	Vp	388	5,1
Cm	18	0,8	Vd	227	3,0
Cj	10	0,4	Vj	149	2,0
Muu	21	0,9	Vv	142	1,9
Yht.	2365	100,0	Muu	201	2,7
			Yht.	7556	100,0

Ääntämispaikan ja -tavan mukaan nämä aineistot jakaantuvat seuraavasti:

a)			b)		
Ryhmä	(C)C %	(V)C %	Ryhmä	(V)C %	(V)C %
Dentaalit	83,9	65,3	Klusiilit	18,8	23,0
Labiaalit	9,3	15,6	Frikatiivit	7,1	19,5
Palataalit	4,6	10,3	Likvidat	67,8	31,2
Laryngaalit	1,7	8,7	Nasaalit	2,4	22,3
Yht.	100,0	100,0	Puolivok.	3,5	3,9
			Yht.	100,0	100,0

Esiintymistajuuksien erot ovat jyrkimmät konsonanteissa r (a: 51,5 %— b: 13,5 %), n (1,6 %— 15,5 %), m (0,8 %— 6,8 %), s (5,0 %— 10,2 %) ja h (1,7 %— 8,7 %). Merkille pantavaa on, että ainoastaan r:n prosenttinen osuus on a-aineistossa olennaisesti (peräti 38 %-yksikköä) suurempi kuin b-aineistossa. Muista tapauksista (10) seitsemässä on tilanne päinvastainen, yhdessä (p) tasan ja vain kahdessa (t ja v) kääntynyt lievästi a:n eduksi.

Materiaalin kahtiajako on vaikuttanut konsonanttien foneettisten ominaisuuksien frekvenssijakaumaan seuraavasti: 1) Osa-aineistossa a eli toisen konsonantin jäljessä ovat dentaalien ja likvidojen osuudet kasvaneet ja kaikkien muiden ryhmien supistuneet. 2) Osa-aineistossa b eli vokaalin jäljessä ovat vastaavasti dentaalien ja likvidojen osuudet jonkin verran laskeutuneet ja muiden nousseet. Muutosten pääsyy on epäilemättä r-loppuisten konsonanttiyhtymien runsaus.

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

TAULUKKO 6: KS:n hakusanojen vokaalisten kakkosgrafeemien absoluuttiset ja prosenttiset (% I ja II) frekvenssit sekä UT:n vokaalisten kakkosgrafeemien (% III) ja yleiskielen kaikkien vokaaligrafeemien (% IV) prosenttiset tekstifrekvenssit

V	f	% I	% II	% III	% IV
a	14 570	23,68	20,39	22,74	11,80
i	10 646	17,30	14,90	14,60	10,68
u	10 134	16,47	14,18	12,11	5,02
e	8 044	13,07	11,26	12,59	8,22
o	7 688	12,49	10,76	7,72	5,27
ä	4 763	7,74	6,67	9,67	4,69
y	4 016	6,53	5,62	2,74	1,79
ö	1 671	2,72	2,34	0,25	0,47
å	0	0,00	0,00	0,00	0,00
Yht.	61 532	100,00	86,12	82,42	47,94

Toisen position vokaaligrafeemien yleisyysjärjestys on KS:n sana-aineistossa a i u e o ä y ö ja UT:n tekstiaineistossa a i e u ä o y ö. Labiaalivokaalit sijoittuvat jälkimmäisessä selvästi heikommin kuin edellisessä; niillä on siis alhainen tekstifrekvenssi. Samaan suuntaan viittaa myös YK:n tekstimateriaali. Siinä järjestys on a i e o u ä y ö.

Alkuaseman ja kakkosaseman vokaalistot poikkeavat toisistaan ennen kaikkea kvantitatiivisesti: sananalkuisia vokaaligrafeemeja on KS:n aineistossa 8 884, toiseksi sijoittuvia 61 532. Vokaalien yleisyysjärjestyskään ei säily samana 1. positiosta (a e i o u y ä ö) 2:seen siirryttäessä, vaan i ja u nousevat nyt e:n edelle sekä ä y:n edelle. Kakkosasemassa a:n, e:n ja o:n suhteelliset osuudet pienenevät (a:n 6,7, e:n 3,2 ja o:n 2,0 %-yksikköä) ja i:n, u:n, ä:n ja ö:n kasvavat (i:n 1,5, u:n 6,6, ä:n 1,9 ja ö:n samoin 1,9 %-yksikköä).

Etu- ja takavokaalien jakauma on sana- ja tekstiaineistossa (KS ja UT) lähes sama. Yleiskielen koko materiaalissa (YK) ryhmien suhteet sen sijaan kääntyvät päinvastaisiksi:

Ryhmä	KS		UT	YK
	f	%	%	%
Etuvokaalit	29 140	47,4	48,3	53,9
Takavokaalit	32 392	52,6	51,7	46,1
Yhteensä	61 532	100,0	100,0	100,0

Sananalkuisista grafeemeista on etuvokaaleja 45,3 % ja takavokaaleja 54,7 %. Edellä todetut yleisyysjärjestyksen muutokset ovat siis vain hieman lisänneet etuvokaalien osuutta.

Väljyysasteen mukaisessa luokituksessa pistää silmään suppeiden vokaalien paljous:

Ryhmä	KS		UT	YK
	f	%	%	%
Suppeat	24 796	40,3	35,7	36,5
Puolisuppeat	17 403	28,3	24,9	29,1
Väljät	19 333	31,4	39,3	34,4
Yhteensä	61 532	100,0	100,0	100,0

KS:n sana-aineistossa on suppeiden vokaalien osuus huomattavasti suurempi ja väljien pienempi kuin juoksevaa tekstiä edustavissa aineistoissa, varsinkin UT:ssa. Tämä osoittaa, että edellisten tekstifrekvenssi on alhainen ja jälkimmäisten korkea. Myös puolisupeita vokaaleja esiintyy UT:n kaltaisen tekstin kakkosgrafeemeina suhteellisesti vähemmän kuin leksikossa. — Suppeita vokaaleja on KS:n 2. positiossa peräti 8,1 %-yksikköä enemmän kuin 1:ssä (32,2 %), puolisupeita ja väljiä vastaavasti vähemmän (puolisuppeita -3,3 ja väljiä -4,8 %-yksikköä).

Laveiden ja pyöreiden vokaalien yleisyysuhde on KS:n sana-aineistossa lähes 3 : 2.

Ryhmä	KS		UT	YK
	f	%	%	%
Laveat	38 023	61,8	72,3	73,8
Pyöreät	23 509	38,2	27,7	26,2
Yhteensä	61 532	100,0	100,0	100,0

Pyöreiden vokaalien osuus on KS:ssa selvästi suurempi kuin juoksevassa tekstissä, missä se jää vajaaseen 30 prosenttiin, YK:n materiaalissa jopa lähelle 25:tä. Sanan alussa pyöreitä vokaaleja on vain 31,7 % eli 6,5 prosenttiyksikköä vähemmän kuin kakkosasemassa.

4. Kahden ensimmäisen grafeemin kombinaatiot jakaantuvat KS:n sana-aineistossa seuraavasti:

	f	%
CV	60 205	84,26
CC	2 365	3,31
VC	7 556	10,57
VV	1 327	1,86
Yht.	71 453	100,00

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

CV-alkuisten sanojen alkugrafeemien kombinoitumista on esitelty yksityiskohtaisesti taulukoissa 7 (C = kotoperäinen konsonantti) ja 7a (C = vierasperäinen konsonantti). Taulukossa 7 kiintyy huomio lähinnä sellaisiin tapauksiin, joissa yhtymän frekvenssi poikkeaa tuntuvasti keskiarvosta. Näitä ovat ennen muita puolivokaalilla alkavat jonot, sillä kuten tunnettua sananalkuinen j ei hevin liity i:hin eikä e:hen ja sananalkuinen v on varsin harvinaisen labiaalivokaalin edellä. Näissä kohdin taulukon poikkeamat ovatkin jyrkimmät: je -11,4, ji -16,2, vo -7,9, vu -13,3 ja vy -5,9 %-yksikköä keskiarvosta. Toisaalta taas esimerkiksi jä:n osuus on 17,8 ja va:n 11,4 %-yksikköä keskimääräistä suurempi. KS:n sana-aineiston muut 5 %-yksikköä ylittävät

TAULUKKO 7: KS:n CV-alkuisten hakusanojen kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit

Ca	V ₁	a	e	i	o	u	y	ä	ö	Yht.	%
h-	f	1437	1037	816	569	872	449	431	291	5897	10,1
	%	24,4	17,5	13,8	9,6	14,8	7,6	7,2	4,9	100	
i-	f	291	26	9	329	494	181	469	23	1822	3,1
	%	16,0	1,4	0,5	18,1	27,1	9,9	25,7	1,3	100	
k-	f	2488	1048	1537	1997	1861	538	535	218	10222	17,5
	%	24,3	10,3	15,0	19,5	18,2	5,3	5,2	2,1	100	
l-	f	1245	695	1078	634	727	224	540	159	5302	9,1
	%	23,3	13,1	20,3	12,0	13,7	4,2	10,2	3,0	100	
m-	f	1049	569	394	484	756	317	204	130	3903	6,7
	%	30,7	16,7	11,5	14,2	22,1	9,3	6,0	3,8	100	
n-	f	562	247	413	372	463	279	304	32	2672	4,6
	%	21,0	9,2	15,5	13,9	17,3	10,4	11,4	1,2	100	
p-	f	1636	920	1180	921	1111	404	360	272	6806	11,6
	%	24,2	13,6	17,5	13,6	16,5	6,0	4,5	4,1	100	
r-	f	1062	787	780	443	593	379	431	170	4645	7,9
	%	22,9	16,9	16,8	9,5	12,8	8,2	9,3	3,7	100	
s-	f	1072	776	1266	715	1221	597	467	17	6131	10,5
	%	17,5	12,7	20,6	11,7	19,9	9,7	7,6	0,3	100	
t-	f	1358	702	666	678	1367	488	340	309	5908	10,1
	%	23,0	11,9	11,3	11,5	23,1	8,3	5,8	5,2	100	
v-	f	1835	707	1600	245	163	43	579	15	5187	8,9
	%	35,4	13,6	30,8	4,7	3,1	0,8	11,2	0,3	100	
Yht.		14035	7509	9739	7387	9628	3899	4606	1638	58495	100
%		24,0	12,8	16,7	12,6	16,4	6,7	7,9	2,8	100	—

poikkeamat ovat ko (+6,9), ma (+6,7), mi (-5,2), mu (-5,7), sa (-6,5) ja tu (+6,7). Tällainen vaihtelu saattaa johtua pelkästään leksikkomateriaalin epätasaisesta valikoitumisesta.

Taulukkoon 7 on merkitty paitsi eri CV-kombinaatioiden myös alkukonsonanttien ja kakkosvokaalien absoluuttiset ja prosenttiset frekvenssit. Niistä käy ilmi, että kaikkien grafeemien yleisyysjärjestys säilyy samana kuin koko aineistossa (vrt. taulukoihin 2 ja 6), vaikka laskelmassa ovat nyt mukana vain kotoisella konsonantilla alkavat CV-tapaukset.

Vierasperäisten konsonanttien jälkeisten vokaalien yleisyysjärjestys on outo (ks. taulukkoa 7a): e i a o u y ö ä; yhtymät Cö ja Cä ovat kyllä ani harvinaisia. Kolmen suurtaajuisimman vokaalin yhteinen osuus on peräti 73,8 %.

TAULUKKO 7a: KS:n vierasperäisellä konsonantilla alkavien CV-hakusanojen kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit

Ca	V ₁	a	e	i	o	u	y	ä	ö	Yht.	%
b-		136	44	62	60	39	12	0	3	356	14,9
c-		10	7	7	11	3	1	0	0	39	2,3
d-		36	262	227	62	28	14	0	0	629	36,8
f-		117	58	99	94	33	24	1	3	429	25,1
g-		65	93	15	17	16	4	0	0	210	12,3
q-		0	0	0	0	6	0	0	0	6	0,4
w-		5	2	7	0	0	0	0	0	14	0,8
z-		4	4	2	14	1	2	0	0	27	1,6
Yht.		373	470	419	258	126	57	1	6	1710	100
%		21,8	27,5	24,5	15,1	7,4	3,3	0,1	0,4	100	—

Taulukossa 8 esitettyjen CC-tapausten osuus konsonanttialkuisten yhtymien kokonaismäärästä (62 570) on vain 3,78 %. Niillä ei näin ollen ole olennais-ta vaikutusta alkukonsonanttien frekvenssijakaumaan (ks. s. 403). Toisin on 2. position konsonantiston laita: siitä (9 921) CC-tapaukset ovat 23,84 %. Siksi CC- ja VC-yhtymät onkin edellä (s. 410) käsitelty erikseen. Kahden konsonantin kombinaatioita hallitsee ylivoimaisesti Cr (51,5 %). Viiden seuraavaksi yleisimmän yhtymän osuus on 40,0 % (Cl 16,2 + Ct 9,5 + Cp 5,1 + Cs 5,0 + Ck 4,2 %). Alkukonsonanteista ovat yleisimpiä obstruentit: p 658, s 524, k 423, t 348 eli yhteensä 1 953 (= 82,58 % kokonaismäärästä). Suurtaajuisimmat vierasperäiset alkukonsonantit ovat f (123), g (120), b (70) ja d (51).

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

TAULUKKO 8: KS:n CC-alkuisten hakusanojen kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit (% I kaikista, % II CC-tapauksista)

CC	f	% I	% II	CC	f	% I	% II
bj	1	0,00	0,04	ph	2	0,00	0,08
bl	17	0,02	0,72	pj	1	0,00	0,04
br	52	0,07	2,20	pl	142	0,20	6,00
ch	12	0,02	0,51	pn	13	0,02	0,55
cl	2	0,00	0,08	pr	430	0,60	18,18
cr	4	0,01	0,17	ps	64	0,09	2,71
cs	1	0,00	0,04	pt	6	0,01	0,25
dj	1	0,00	0,04	rh	5	0,01	0,21
dr	48	0,07	2,03	sc	9	0,01	0,38
ds	1	0,00	0,04	sf	9	0,01	0,38
dz	1	0,00	0,04	sh	8	0,01	0,34
fl	48	0,07	2,03	sk	99	0,14	4,19
fr	75	0,10	3,17	sl	24	0,03	1,01
gh	2	0,00	0,08	sm	13	0,02	0,55
gl	39	0,05	1,65	sn	6	0,01	0,25
gn	6	0,01	0,25	sp	117	0,16	4,95
gr	73	0,10	3,09	sr	2	0,00	0,08
hj	2	0,00	0,08	ss	2	0,00	0,08
hm	4	0,01	0,17	st	219	0,31	9,26
hs	3	0,00	0,13	sv	22	0,03	0,93
kh	2	0,00	0,08	sw	1	0,00	0,04
kl	111	0,16	4,69	th	6	0,01	0,25
kn	12	0,02	0,51	tj	3	0,00	0,13
kr	238	0,33	10,06	tp	3	0,00	0,13
ks	10	0,01	0,42	tr	297	0,42	12,56
50	0,07	2,11	ts	37	0,05	1,56	
mh	1	0,00	0,04	tv	1	0,00	0,04
mj	1	0,00	0,04	tw	1	0,00	0,04
mm	1	0,00	0,04	wh	2	0,00	0,08
mn	1	0,00	0,04	zl	1	0,00	0,04
nj	1	0,00	0,04	Yht.	2 365	3,31	100

Taulukoissa 9 ja 9a esitelty VC-alku poikkeaa kombinaatioista CV ja CC olennaisesti siinä, että sen jäsenet eivät läheskään aina kuulu samaan tavuun, vrt. esim. *a-ho*, *a-loittaa* ja *ah-taa*, *al-ku* jne. Tämän vuoksi VC-jonon frekvenssitiedoilla on lähinnä vain tilastollista kiinnostavuutta. Sitä vähentää vielä jossain määrin se, että mistään ei paljastu, milloin d esiintyy kotoisessa sanassa t:n heikkoasteisena vastineena (esim. *odottaa*, *ydin*), milloin taas vierasperäisen sanan normaalina kakkosgrafeemina. Tässä Vd on sijoitettu taulukkoon 9a eli vierasperäisten konsonanttien joukkoon, mutta kakkosgrafeemien keskinäisessä vertailussa se on kyllä otettu erikseen huomioon; ks. s. 408.

HEIKKI LESKINEN

TAULUKKO 9: KS:n VC-alkuisten hakusanojen kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit

V _a	C ₁	h	j	k	l	m	n	p	r	s	t	v	Yht.	%
a-	f	111	70	128	309	129	337	157	345	259	69	78	1 992	28,0
	%	5,6	3,5	6,4	15,5	6,5	16,9	7,9	17,3	13,0	3,5	3,9	100	
e-	f	96	2	157	203	86	198	65	169	111	136	32	1 255	17,6
	%	7,6	0,2	12,5	16,2	6,9	15,8	5,2	13,5	8,8	10,8	2,5	100	
i-	f	115	1	56	265	132	356	3	113	155	83	7	1 286	18,0
	%	8,9	0,1	4,4	20,6	10,3	27,7	0,2	8,8	12,1	6,5	0,5	100	
o-	f	110	43	79	112	55	121	112	176	127	65	12	1 012	14,2
	%	10,9	4,2	7,8	11,1	5,4	12,0	11,1	17,4	12,5	6,4	1,2	100	
u-	f	78	28	20	153	42	114	38	85	74	29	4	665	9,3
	%	11,7	4,2	3,0	23,0	6,3	17,1	5,7	12,8	11,1	4,4	0,6	100	
y-	f	104	0	47	212	61	17	9	48	30	15	0	543	7,6
	%	19,1	0,0	8,7	39,0	11,2	3,1	1,7	8,8	5,5	2,8	0,0	100	
ä-	f	37	5	75	61	12	23	4	78	16	8	8	327	4,6
	%	11,3	1,5	22,9	18,7	3,7	7,0	1,2	23,9	4,9	2,4	2,4	100	
ö-	f	6	0	6	25	0	2	0	6	0	1	1	47	0,7
	%	12,8	0,0	12,8	53,2	0,0	4,3	0,0	12,8	0,0	2,1	2,1	100	
Yht.		657	149	568	1340	517	1168	388	1020	772	406	142	7 127	100
%		9,2	2,1	8,0	18,8	7,3	16,4	5,4	14,3	10,8	5,7	2,0	100	—

TAULUKKO 9a: Kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit KS:n sellaisissa VC-alkuisissa hakusanoissa, joissa C on ns. vierasperäinen konsonantti

V _a	C ₁	b	d	f	g	x	z	Yht.	%
a-		43	45	18	47	0	3	156	36,4
e-		2	78	8	11	2	0	101	23,6
i-		2	71	0	2	0	0	75	17,5
o-		38	28	17	1	3	0	87	20,3
u-		0	3	0	4	0	0	7	1,6
y-		0	1	0	0	0	0	1	0,2
ä-		0	0	0	0	0	0	0	0,0
ö-		0	1	0	0	0	0	1	0,2
Yht.		85	227	43	65	5	3	428	100
%		19,9	53,0	10,0	15,2	1,2	0,7	100	—

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

Jos tarkastellaan rinnan KS:n koko aineiston ja VC-jonon alkuvokaaleja ja kakkoskonsonanteja, voidaan tehdä ainakin seuraavat havainnot (ks. taulukoita 3 ja 5). Alkuvokaalien yleisyysjärjestys vaihtuu VC-kombinaatioissa siten, että i ohittaa e:n, vaikka myös e:n osuus jonkin verran kasvaa (i: 15,8 % → 18,0 %; e: 16,3 % → 17,6 %). Muut muutokset eivät vaikuta vokaalien sijoitukseen (mm. a: 30,4 % → 28,0 %; y: 6,5 % → 7,6 %; ä: 5,8 % → 4,6 %). Konsonanttien puolella r:n asema heikkenee tuntuvasti: se siirtyy frekvenssilistan kärjestä l:n ja n:n jälkeen kolmanneksi (r: 22,6 % → 13,5 %; l: 17,4 % → 17,7 %; n: 12,2 % → 15,5 %). Muutosten syistä ks. s. 410. Lisäksi m ohittaa t:n (m: 5,4 % → 6,8 %; t: 6,4 % → 5,4 %) ja j v:n (j: 1,6 % → 2,0 %; v: 2,2 % → 1,9 %). Muilta osin tilanne säilyy ennallaan.

Taulukossa 9a pistää silmään neljän suurtaajuisimman alkuvokaalin (yleisyysjärjestyksessä: a e o i) täydellinen ylivoima: niiden yhteinen osuus on peräti 97,8 %. Kakkoskonsonanteista on tietysti yleisin d (53,0 %); sitä seuraavat tasaisin välein b, g ja f. x ja z ovat vain marginaalisia grafeemeja.

TAULUKKO 10: KS:n VV-alkuisten hakusanojen kahden ensimmäisen grafeemin absoluuttiset ja prosenttiset frekvenssit (% I kaikista, % II VV-tapauksista)

VV	f	% I	% II	VV	f	% I	% II
aa	157	0,28	11,83	oo	25	0,03	1,88
ae	17	0,02	1,28	ou	39	0,05	2,94
ai	196	0,27	14,77	ui	73	0,10	5,50
ao	2	0,00	0,15	uo	3	0,00	0,23
au	177	0,25	13,34	uu	128	0,18	9,65
ee	34	0,05	2,56	ya	2	0,00	0,15
ei	23	0,03	1,73	ye	2	0,00	0,15
eo	4	0,01	0,30	yo	2	0,00	0,15
eu	34	0,05	2,56	yy	1	0,00	0,08
ie	5	0,01	0,38	yö	27	0,04	2,03
ii	19	0,03	1,43	äe	7	0,01	0,53
io	7	0,01	0,53	äi	48	0,07	3,62
iu	2	0,00	0,15	äy	43	0,06	3,24
iä	9	0,01	0,68	ää	93	0,13	7,01
oa	3	0,00	0,23	öi	9	0,01	0,68
oi	120	0,17	9,04	öy	16	0,02	1,21
				Yht.	1 327	1,89	100,00

Taulukon 10 sananalkuisista VV-sekvensseistä on kaksoisvokaaleja 457 (34,44 %), diftongeja 815 (61,42 %) ja muita vokaaliyhtymiä 55 (4,14 %). Kaksoisvokaalien frekvenssijakauma on seuraava: aa 157 (34,4 %), uu 128 (28,0 %), ää 93 (20,4 %), ee 34 (7,4 %), oo 25 (5,5 %), ii 19 (4,2 %) ja yy 1 (0,2 %). Kolmen suurtaajuisimman yhteinen osuus on 82,8 %. Kaksoisvokaaleja ee ja oo tavataan pääasiallisesti vain lainoissa.

Diftongien yleisyysjärjestys on ai 196 (24,0 %), au 177 (21,7 %), oi 122 (15,0 %), ui 73 (9,0 %), äi 48 (5,9 %), ou 39 (4,8 %), eu 34 (4,2 %), yö 27 (3,3 %), ei 23 (2,8 %), öi 9 (1,1 %), ie 5 (0,6 %), uo 3 (0,4 %) ja iu 2 (0,2 %). Lukumääräsuhteet poikkeavat melkoisesti juoksevasta tekstistä lasketuista (ks. Karlsson 1983: 90), mutta KS:n materiaali ei pienuutensa vuoksi ole vertailukelpoinen. Tässä esitellyistä diftongeista on suppenevia 780 ja auke-
nevia 35 (ie, uo, yö); i-loppuisia on 469 (61,1 %), u-loppuisia 252 (32,3 %) ja y-loppuisia 59 (7,6 %).

Sananalkuisen VV-sekvenssin alku- ja jälkikomponenttien yleisyysjärjes-
tykset (V_1 : a u ä o e i y ö; V_2 : i u a ä e y o ö) eroavat lähes täydellisesti
toisistaan; yhteistä niillä on vain u:n sijoittuminen toiseksi ja ö:n viimeiseksi.
Suhteiden sekavuus johtuu monilta osin diftongien, suurimman VV-ryhmän,
fonotaktisista rajoituksista.

LÄHTEET

- HÄKKINEN, KAISA 1983: Suomen kielen äännerakenteen ominaispiirteistä. — Nyky-
suomen rakenne ja kehitys. Tietolipas 93: 39—56. Pieksämäki.
KARLSSON, FRED 1983: Suomen kielen äänne- ja muotorakenne. Juva.
PÄÄKKÖNEN, MATTI 1973: Tilastotietoja suomen yleiskielen grafeemeista. — Suoma-
lais-ugrilaisen Seuran aikakauskirja 72: 318—322. Hki.
SETÄLÄ, VILHO 1972: Suomen kielen dynamiikkaa. Suomi 116:3. Hki.
TUOMI, TUOMO 1972: Suomen kielen kääntheisanakirja. Suomalaisen Kirjallisuuden
Seuran toimituksia 274. Hämeenlinna.

Daten zur Allgemeinheit der Grapheme und Graphemkombinationen im Wortanlaut des Finnischen

HEIKKI LESKINEN

Im vorliegenden Aufsatz werden u.a. die Beleghäufigkeiten der im Fi. begegnenden Grapheme der 1. und 2. Position sowie der durch sie gebildeten Graphemsequenzen in einem Wörterbuch (Rückläufiges Wörterbuch des Finnischen) und einer Textquelle (Finnische Übersetzung des Neuen Testaments) miteinander verglichen. Das Material des ersteren enthält 71 454 verschiedene

Wörter, das des letzteren 804 243 Wortbelege. Hier die wichtigsten Beobachtungen:

1. Zwischen den Anfangskonsonanten und -vokalen besteht im Lexikon ein Allgemeinheitsverhältnis von 88 : 12, im Text von 80 : 20. Die entsprechenden Zahlen der 2. Position sind fast entgegengesetzt: 14 : 86 und 18 : 82.

2. Die Häufigkeitsreihenfolge der

Tietoja sananalkuisten grafeemien ja grafeemikombinaatioiden yleisyydestä

Konsonanten im Wortanlaut sieht im Lexikon folgendermaßen aus: k p s t h l v r m n j, im Text j s k h t m v p n l, in der 2. Position für die Konsonanten im Lexikon r l n s h k t m p d r j, im Text l n t s h r v p m d k j. In beiden Materialien besteht die deutliche Mehrheit der ersten Konsonanten aus Obstruenten (60,8 % und 56,1 %). Durch die außergewöhnlich hohe Textfrequenz des Halbvokals j wird das sonst einheitliche Bild allerdings etwas gestört. Bei den zweiten Konsonanten herrschen wiederum die Liquida und Nasale vor (57,5 % und 60,6 %).

3. Die Anfangsvokale nehmen im Lexikon-Material die Reihenfolge a e i o u y ä ö ein, im laufenden Text die Reihenfolge o e a i u y ä ö. Hier kommt der Unterschied zwischen dem Wort- und dem Textmaterial anschaulich zum Ausdruck: für den Text führt o – offenbar dank des häufigen Auftretens des Verbs *olla* 'sein' – die Liste an, gefolgt von e, während a, der häufigste Vokal im Finnischen, erst an dritter Stelle steht. Für die zweite Position zeigt

sich dagegen in beiden Quellen ein sehr ähnliches Ergebnis: a i u e o ä y ö und a i e u ä o y ö.

4. Die Sequenzen der beiden ersten Grapheme verteilen sich im Lexikonmaterial wie folgt: CV 84,3 %, CC 3,3 %, VC 10,6 % und VV 1,9 %. Bei den CV-Kombinationen fällt auf, daß das j im Wortanlaut kaum mit i oder e verbunden wird und daß das v im Wortanlaut nur selten vor einem Labialvokal steht. Nach den Konsonanten fremder Herkunft (= stimmhafte Klusile und f) begegnen am häufigsten die Vokale e i a o; u und y sind in dieser Position selten, ä und ö ganz exzeptionell. Bei den Verbindungen zweier Konsonanten im Wortanlaut herrscht überwiegend die Sequenz Cr (51,5 %). Für die Vokale der VC-Kombination ergibt sich die Häufigkeitsfolge a i e o u y ä ö. Bei Sequenzen zweier Vokale (VV) sind Doppelvokale bzw. lange Vokale mit 34,5 %, Diphthonge mit 61,4 % und andere Vokalverbindungen mit 4,1 % vertreten.