

Suomen kielen sanaston suomalais-ugrilaiset juuret

Nykysuomen sanasto koostuu eri-ikäisistä aineksista. Pieni osa siitä voidaan palauttaa vanhimpiin tunnettuihin kantakielimuotoihin, uralilaiseen tai suomalais-ugrilaiseen kantakieleen, osa sanoista on aivan äskettäin syntyneitä uudismuodosteita tai lainoja, ja näiden kahden äärimmäisyyden välille mahtuu suuri joukko eri ikäisiä ja erityyppisiä sanakerrostumia.

Minkään luonnollisen kielen sanaston laajuutta ei voi määritellä täsmällisesti, sillä sanasto on avoin systeemi: sanoja syntyy ja kuolee jatkuvasti, ja myös eri kielenkäyttäjien sanastot poikkeavat toisistaan mm. murre-erojen vuoksi. Yleiskielikään ei tältä osin ole selvärajainen. Tämä on helppo havaita esim. selailemalla Nykysuomen sanakirjaa, jonka laajasta materiaalista suuri osa on murteellista, vanhentunutta tai joidenkin kirjailijoiden erityissanastoon kuuluvaa ainesta. Nykysuomen sanakirjan yli 200 000 hakusanasta kuuluu siis vain osa tavallisen suomen kielen puhujan sanavarastoon. Lisäksi näistä sanoista useimmat ovat toissijaisia siinä mielessä, että ne on saatu joko johtamalla tai yhdistämällä muista sanoista. Esim. *käsi*-sanasta on voitu johtaa sanat *kätellä*, *käsittää*, *käsitellä*, *käsine*, *kättilö* jne., ja yhdyssanan osana se esiintyy mm. sanoissa *käsityö*, *käsikirja*, *käsivarsi*, *täyskäsi* jne. Sanat voidaan siis jakaa rakenteensa perusteella kolmeen pääryhmään: yksinkertaisiin perussanoihin eli kantoihin, johdoksiin ja yhdyssanoihin.

Vanhimpien sanojen historiaa tutkittaessa on yleensä tapana kiinnittää huomiota vain kantoihin, joten kantakielten sanarekonstruktiot kuvaavat enimmäkseen johtamattomia perussanoja. Vanhoja johdoksiakin on jonkin verran löydetty, ja varmana pidetään, että sekä johtaminen että yhdistäminen olivat myös etäisimpien kantakielten sananmuodostuskeinoja. Historiallinen johto-oppi on kuitenkin vielä niin puutteellisesti tutkittu, että se sanahistoriassa on pakostakin jäänyt

syryään. Tämä on tärkeätä muistaa silloin, kun aletaan puhua vanhojen omaperäisten sanojen määrästä.

Suomen kielen sanoista palautuu kantaaurialiin tai suomalais-ugrilaiseen kantakieleen varmasti vain kolmisensataa sanaa. Se saattaa kuulostaa mitättömältä määrältä, jos vertaamme sitä Nykysuomen sanakirjan aineistoon. Edellä on kuitenkin käynyt ilmi, ettei tällaista vertailua kannata suorittaa, koska nämä sanastot ovat rakenteeltaan niin erityyppisiä. Kukaan ei toistaiseksi ole laskenut, kuinka monta johtamatonta perussanaa nykysuomessa tällä hetkellä on, joten parasta mahdollista vertailukohtaa ei löydykään. Nykyvirossa on kuitenkin arvioitu olevan noin 6 000 johtamatonta perussanaa, kun aivan nuoret lainat jätetään huomiotta (Andrus Saareste, *Kaunis emakeel* 1, Lund 1952, 63; ks. myös Huno Rätsep, *Eesti keele sõnavara probleeme* V. Tartu riikliku ülikooli toimetised 460, 5). Erkki Itkonen on arvioinut suomen murteiden perussanojen yhteismääräksi yli 10 000 (Kieli ja sen tutkimus, 387). Joka tapauksessa voimme pitää varmana, että määrällisesti arvioiden vanhimpien sanojen osuus nykysuomen sanastosta on melko pieni.

Sanat syntyvät ja ne otetaan käyttöön sitä varten, että niitä tarvitaan. Tarpeellisuusasteessa voi kuitenkin olla huomattavia eroja. Joitakin sanoja me jokainen käytämme joka päivä useita kertoja, joitakin tarvitsemme äärimmäisen harvoin. Sanahistorian tutkijat ovat todenneet, että kielen keskeisimmät sanat säilyvät parhaiten (vrt. esim. Witold Mańczak, *Deux lois de l'évolution phonétique*. Proceedings of the Sixth International Congress of Phonetic Sciences 1970, 605—607). Tämän perusteella voimme päätellä, että vanhimmista kantakielistä nykysuomeen asti periytyneet sanat kuuluvat tähän keskeisimpään sanastoon.

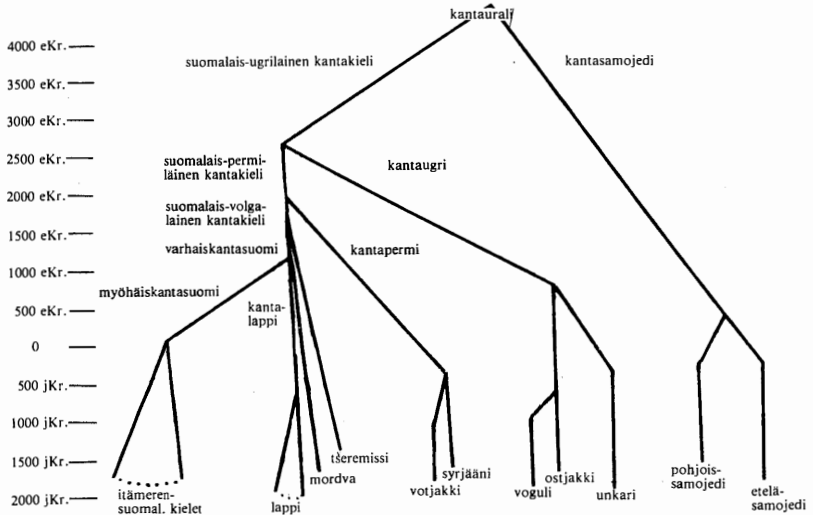
Sanojen ikä voidaan määritellä vain tutkimalla, onko suomen sanoille löydettävissä etymologisia eli samaa historiallista alkuperää olevia vastineita sukukielistä ja mistä sukukielistä vastineita löytyy. Mistään kantakielestä ei ole säilynyt kirjallisia muistomerkkejä, joten kaikki kantakieliä koskeva tieto on itse asiassa nykykielten yhtäläisyyksien pohjalta asetettua hypoteesia. Kaikkia yhteisiltä näyttäviä ominaisuuksia ei tosin voi palauttaa kantakieliin, sillä yhtäläisyydet voivat johtua myös esim. siitä, että kielet ovat erikseen kehittyneet samaan suuntaan (= konvergenssi), yhteiset ainekset on lainattu kielestä toiseen, joko oman kielikunnan sisällä tai kielikunnan ulkopuolelta, tai sitten yhtäläisyys voi johtua sattumastakin. Yhteisten ainesten tunnistaminenkaan ei ole aina helppoa, sillä kaikki kielet ovat muuttuneet enemmän tai vähemmän erilliskehityksensä aikana.

Vanhoja omaperäisiä sanoja etsittäessä on käytettävä kriteerinä se-

kä äänneasua että merkitystä. Äänteellinen samankaltaisuus on perusedellytys rinnastamiselle, mutta yksinään se ei ole ratkaiseva. Usein tämä samankaltaisuuskin voidaan todeta vain tiettyjen äännehistoriallisten ennakkoehtojen avulla. Ensi silmäyksellä on ehkä vaikea uskoa, että suomen *pii* ja unkarin *fog* 'hammas' ovat sama sana, niin kuin ovat myös *kota* ja *ház* 'talo' sekä *ydin* ja *velð* 'id.'. Näitä rinnastuksia pidetään varmoina, vaikka sanojen nykymuodoissa ei ole yhtäkään identtistä äännettä. Tärkeätä on tietää, mitkä äänteet voivat eri kielissä vastata toisiaan. Yleensä vastineita on useampia kuin yksi, joten vastaa- vuussuhteiden verkostot ovat monimutkaisia. Sitä paitsi suuri osa rinnastuksista on katsottava tavalla tai toisella poikkeukselliseksi. Tästä huolimatta rinnastusta voidaan pitää varmana. Esim. lukusanat ja pronominat ovat äännehistorian kannalta todellisia kiusankappaleita, mutta merkityksen perusteella yhdistäminen on kuitenkin monissa tapauksissa mahdollista suorittaa varmuuden veroisella todennäköisyydellä. Tämä kaikki on kuitenkin aivan odotuksenmukaista sen valossa, mitä kielen keskeisimpien sanojen rakenteesta yleensä tiedetään. Näiden sanojen on havaittu olevan keskimäärin lyhyempiä kuin muut saman kielen sanat, ja lisäksi niiden on todettu muuttuvan epäsäännöllisesti (vrt. Mańczak ed.). Tämä on helppo uskoa, kun tarkastelee vaikkapa persoonapronominien asuja suomen eri kielimuodoissa (*mä, mää, mie, miä, mnää* jne.). Äännehistorian tutkimuksen kannalta tämä on joka tapauksessa ikävää sikäli, että äänne- muutoksia koskevat säännönmukaisuudet on määriteltävä sellaisten tapausten pohjalta, jotka keskimääräistä suuremmalla todennäköisyydellä ovat poikkeuksia näistä säännöistä. (Vrt. esim. Häkkinen, *Finnisch-ugrische Mitteilungen* 5/1981, 185—188.)

Omaperäisten sanojen iän määrittelyn kannalta on ratkaisevaa, missä sukukielissä sanalla on vastineita. Uralilaisen kielikunnan oletetaan hajaantuneen tiettyssä järjestyksessä, jota on tapana kuvata ns. kielten sukupuun avulla (kuva 1). Kuviosta on helppo havaita, että esim. suomi ja samojedikielet ovat yhteydessä toisiinsa vain vanhimman tunnetun kantakielen, kantauralin kautta. Jos siis näissä kielissä on yhteisiä ominaisuuksia, ne voidaan hypoteettisesti palauttaa kantauraliin. Muita suomen sanoja onkin teknisesti mahdotonta todistaa yhtä vanhoiksi. Käytännössä tilanne saattaa monien sanojen kohdalla olla sellainen, että ne palautuvat kantauraliin, ovat säilyneet suomessa mutta hävinneet samojedikielistä. Tällöin ei kuitenkaan ole mitään keinoa osoittaa niiden uralilaista alkuperää. Jos siis halutaan saada selville suomen kielen vanhimmat sanat, niille on osoitettava vastine samojedikielistä (= uralilaiset sanat) tai ugrilaiskielistä (= suomalais-ugrilaiset sanat). (Ks. myös Häkkinen, *Suomalais-*

ugrialaisten kielten etymologisen tutkimuksen asemasta ja ongelmista, teoksessa *Sanomia*, juhla-kirja Eeva Kangasmaa-Minnin 60-vuotispäiväksi, s. 196—199. Esimerkkejä sanojen iän määrittämisestä ks. Häkkinen, *Finnisch-ugrische Mitteilungen* 5/1981, s. 180, 183.)

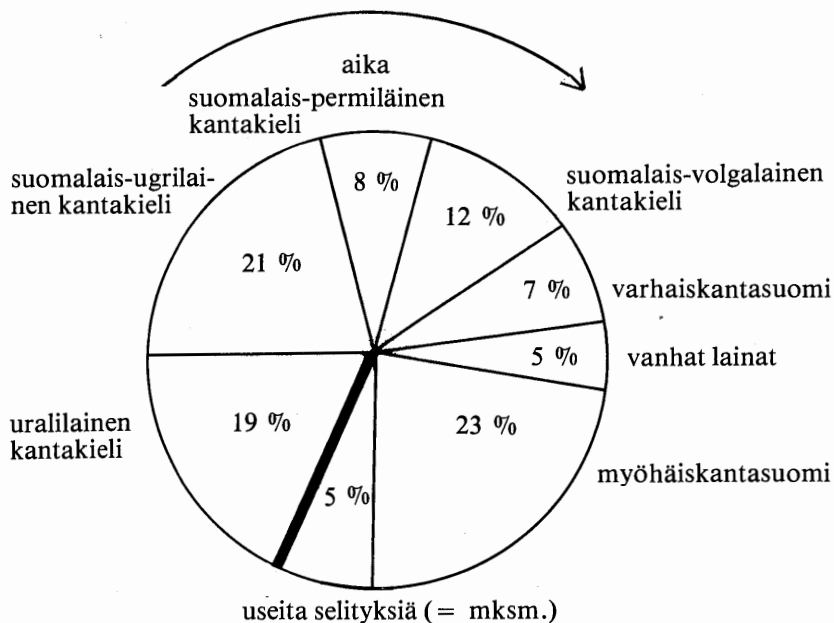


Kuva 1. Uralilaisten kielten sukupuu

Edellä esitetyn perusteella voisimme jo asiaa tarkemmin tutkimatta ennustaa, että nyky-suomen yleisimmät sanat ovat iältään vanhempia kuin sanasto keskimäärin. Suomen kielen sanafrekvenssejä on arvioitu Oulun yliopistossa, jossa Pauli Saukkosen johdolla on laadittu Suomen kielen taajuussanasto (1979). Tämän perusteella suomen kielen 50 yleisintä sanaa ovat:

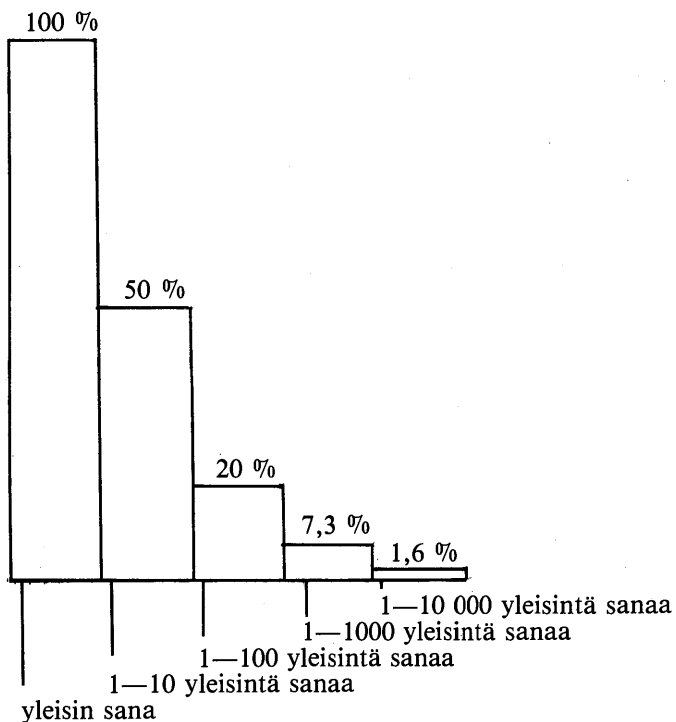
1. olla	sgr.	uralilaisia	14
2. ja	< germ.	suomalais-ugrilaisia	12
3. se	ur.	suomalais-permiläisiä	2
4. ei (e-/neg.)	ur.	suomalais-volgalaisia	5
5. joka (jo-)	svolg.	varhaiskantasuomalaisia	3
6. että (e-/pron.)	ur.	balttilaisia lainoja	2
7. tämä (tä-)	ur.	germaanisista lainoja	2
8. hän	sgr.	itämerensuomalaisia	10
9. voida	ims.		
10. saada	sperm.	Huom! Kaikki ajoitukset ovat	
11. mutta (muu-)	sgr.	SKES:n mukaisia. Vain var-	
12. niin (ne-)	ims.	moiksi arvioidut sukukielen	
13. kuin (ku-)	ur.	vastineet on otettu huomioon.	
14. ne	ims.		
15. kun (ku-)	ur.		
16. tulla	svolg.		
17. myös (myö-)	sgr.		
18. tai (tahto-)	vksm.		
19. kaikki	< baltt.		
20. aika	ims.		
21. me	ur.		
22. suuri	ims.		
23. vuosi	sgr.		
24. mikä (mi-)	ur.		
25. toinen (tuo-)	ur.		
26. muu	sgr.		
27. minä	ur.		
28. antaa	sgr.		
29. pitää	sgr.		
30. tehdä	sgr.		
31. jo	< germ.		
32. he	sgr.		
33. vain	ims. (us. mahd. kantoja)		
34. nyt (nyky-)	svolg.		
35. sekä (se-)	ur.		
36. sanoa	ims.		
37. kuitenkin (ku-)	ur.		
38. nämä (nä-)	sperm.		
39. käyttää (käy-)	ims.		
40. jos (jo-)	svolg.		
41. asia	sgr.		
42. maa	ur.		
43. sitten (se-)	ur.		
44. ottaa	ims.		
45. mukaan	ims. (us. mahd. kantoja)		
46. ihminen	svolg.		
47. vielä	< baltt.		
48. osa	vksm.		
49. oma	vksm.		
50. uusi	sgr.		

Kunkin sanan perään on merkitty viittaus alkuperään, yleensä siihen kantakieleen, johon sana todennäköisimmin voidaan palauttaa Suomen kielen etymologisen sanakirjan mukaan. Sadan yleisimmän sanan ikäjakauma ei olennaisesti poikkea edellisestä (ks. graafista esitystä kuva 2).



Kuva 2. Suomen 100 yleisimmän sanan ikäjakauma

Taajuussanakirjan avulla voidaan myös osoittaa, että vanhimpien sanojen suhteellinen osuus vähenee siirryttäessä yleisimmistä sanoista harvinaisempiin. Tämän sanakirjan avulla voidaan tosin tutkia vain hiukan yli 10 000 yleisintä sanaa, mutta jo tämä aineisto riittää kyseisen tendenssin osoittamiseen (ks. kuva 3).



Kuva 3. Vanhojen omaperäisten sanojen osuus suomen yleisimpien sanojen joukossa

Keskeisimpien sanojen laatuun on myös syytä kiinnittää huomiota. Sanathan on tapana jakaa lähinnä taivutuksen perusteella sanaluokkiin, ja sanaluokkien taajuussuhteet poikkeavat tekstissä ja sanastossa selvästi:

	sanasto	teksti	
substantiivit	75,25 %	35,83 %	(1.)
adjektiivit	10,03 %	9,22 %	(3.)
verbit	8,16 %	24,33 %	(2.)
adverbit	1,74 %	7,94 %	(5.)
numeraalit	0,39 %	0,75 %	(7.)
interjektiot	0,32 %	0,14 %	(9.)
pronominit	0,09 %	8,42 %	(4.)
post- ja prepos.	0,05 %	0,31 %	(8.)
konjunktiot	0,04 %	6,86 %	(6.)
sanaluokattomat ja sanaluokkayhdistelmät	3,93 %	6,20 %	

(Suomen kielen taajuussanasto, 11).

Taulukosta käy selville, että suhteellisen suppean osajärjestelmän muodostavat luokat, pronominit ja konjunktiot, ovat tekstissä huomattavasti yleisempiä kuin niiden osuus sanastosta edellyttäisi. Näihin ryhmiin kuuluu siis hyvin suurifrekvenssisia elementtejä, ja niitä onkin runsaasti aivan keskeisimpien sanojen joukossa: suomen 100 yleisimmästä sanasta on pronomineja 17 ja konjunktioita 8. Suomalais-ugrilaisissa kielissä ei ole vanhoja omaperäisiä konjunktioita, vaan nämä on joko lainattu muista kielistä tai johdettu omaperäisistä aineksista, usein pronominivartaloista. Näin ollen pronominien luokan osuus korostuu vielä, jos otamme huomioon varsinaisten pronominien lisäksi myös keskeisessä sanastossa esiintyvät pronominivartalot. Seuraava taulukko luettelee suomen 100 yleisimmän sanan joukosta pronominit tai pronominivartalosta muodostetut sanat. (Huom! Nykykielen kannalta kaikki eivät ole pronomineja; toisaalta eräät nykyiset pronominit palautuvat substantiivivartaloihin, esim. *itse*):

e-: että (6) N	1
erä-: eräs (78), eri (82)	2
he (32)	1
hän (8)	1
itse (53)	1
jo-: joka (5), jos (40), jokin (60) N	3
kaikki (19) < baltt.	1
ko-: koska (93)	1
ku-: kuin (13), kun (15), kuitenkin (37) N	3
me (21) N	1
mi-: mikä (24) N	1
mi-: minä (27) N	1
moni (57) N	1
mu-: mukaan (45)	1
mū-: mutta (11), muu (26)	2
ne: niin (12), ne (14) N	2
nä-: nämä (38), näin (76) N	2
sama (52) < germ.	1
se-: se (3), sekä (35), sitten (43), sellainen (61), siis (94), siellä (95)	6
si-: sinä (63) N	1
tuo-: toinen (25)	1
tä-: tämä (7), tässä (75) N	2
	Σ 36

N = mahdollisesti nostraattinen elementti

Yhteensä 36 pronominista tai pronominivartalosta muodostetusta sanasta kaksi on lainattua (*kaikki*, *sama*), muut ovat omaperäisiä. Näiden joukossa on useita ikivanhoiksi arveltuja elementtejä. *N*-

kirjaimella merkittyjä sanoja arvellaan näet monissa etymologisissa lähdelehdissä aineksiksi, jotka todistavat ns. nostraattihypoteesin puolesta. Tämän hypoteesin mukaan uralilainen, indoeurooppalainen ja altaalainen kielikunta ja eräät muutkin ryhmät (mm. dravidakielet) saattaisivat olla alkusukulaisuussuhteessa keskenään (ks. esim. V. M. Illič-Svityč, *Opyt sravnenija nostratičeskih jazykov*, Moskva 1971). Kaikissa näissä kieliryhmissä esiintyy yllä esitetyn kaltaisia pronomini-variantteja, joiden voidaan olettaa kuuluvan etymologisesti yhteen. Mainittujen pronominiin lisäksi on tietysti olemassa muitakin nostraattiseksi arveltuja sanoja, mutta juuri pronominit ovat kielisukulaisuutta tutkittaessa sikäli keskeisessä asemassa, ettei niiden lainautuminen kielestä toiseen käy päinsä yhtä helposti kuin esim. substantiivien. Mahdotonta lainautuminen ei tietenkään ole, ja luultavaa on, että esim. indefiniittipronominit lainautuvat monin verroin helpommin kuin persoonapronominit, sillä indefiniittipronominiin heterogeeninen joukko ei muodosta persoonapronominisysteemiin verrattavaa kiinteää järjestelmää. Uralilaisissa kielissä persoonapronominit ovat pohjana myös possessiivisuffiksi- ja persoonapäätysesysteemeille, joten näissä pronomineissa tapahtuvat muutokset koskisivat välttämättä kielen keskeisimpiin kuuluvaa viittaussuhteiden verkostoa. Sen sijaan muutaman substantiivin lainautumisella tai häviämällä ei ole sanastosysteemin kannalta merkitystä.

Suomen kielen taajuussanaston ulkopuolelle jää vajaat sata vanhaksi osoitettua sanaa. Osa näistä on intuitiivisesti arvioiden käypää sanastoa (*kupla, kyy, pieksää, punoa*), mutta osa tuntuu selvästi marginaaliselta. Tämän vuoksi olen tutkinut kokeellisesti, kuuluvatko nämä kyseenalaiset sanat nuorten nykysuomalaisten sanavarastoon. Tutkimuksen perusaineistoksi valitsin sellaiset suomen sanat, joilla on varma etymologinen vastine vähintään kahdessa lappia etäisemmässä etasukukielihaarassa. Käytännössä sanat palautuvat siis joko suomalais-permiläiseen, suomalais-ugrilaiseen tai uralilaiseen kantakieleen. Jättämällä pois ne sanat, joilla on vastine vain yhdessä kielihaarassa, halusin varmistua siitä, ettei joukossa ole satunnaiseen äänteelliseen samankaltaisuuteen perustuvia rinnastuksia. Kokeen tarkoituksena oli lähinnä määritellä prosentuaalisesti se osuus, joka vanhasta sanastosta on katsottava nykykielen kannalta marginaaliseksi: yksittäisten sanojen tunnettuuden selvittäminen oli vasta toisella sijalla.

Koesanalistaa laatiessani jätin pois ne sanat, jotka oman intuitioni perusteella kuuluivat tavallisen suomen kielen puhujan aktiiviseen sanavarastoon (vrt. yllä). Vaikka marginaaliseksi epäiltyjen sanojen valintakriteeri kieltämättä oli epämääräinen, en katso sen vaikuttavan haitallisesti lopputulokseen, sillä varsinaisia rajausongelmia ei synty-

nyt. Yhden liuskan pituiseen koekaavakkeeseen mahtui näet runsaasti sellaisiakin sanoja, joiden arvasin jo etukäteen olevan kaikille tuttuja. Otin nämä kuitenkin mukaan, sillä arvelin koehenkilöiden tarpeettomasti masentuvan joutuessaan täyttämään listaa, joka on täynnä pelkästään tuntemattomia sanoja. Vaikeampi ongelma syntyi homonymiasta: suomen kielessä on useissa tapauksissa vanhan omaperäisen sanan rinnalla uudempi, täysin eri alkuperää oleva ja erimerkityksinen sana. Myös vanhassa sanastossa on muutamia homonyymeja (esim. *kuu* 'Mond' ~ *kuu* 'rasva', joista jälkimmäistä epäilin marginaaliseksi). Näitä varten oli koeliuskkaan merkittävä sama sana useita kertoja, minkä lisäksi kokeen suorittamisohjeissa ilmoitettiin kaikilla sanoilla olevan eri merkitys. Ennen koetta kokeen johtaja ilmoitti myös, että kaikki sanat ovat perusmuotoisia eikä joukkoon sisälly erisnimiä. Lopullisessa koeliuskassa oli 77 sanaa, joista varsinaisia tutkittavia sanoja oli 60. Koetilanteessa johtaja luki jokaisen sanan ääneen, jotta huolimaton lukeminen tai jonkin sanan huomaamatta jääminen ei olisi päässyt vaikuttamaan lopputulokseen. Koe suoritettiin Kouvolassa, Salparinteen yläasteella ja Kouvonrinteen lukiassa. Tämä on hyvä panna merkille, sillä sanojen tuntemuksessa saattaa olla murteiden välisiä eroja. Koehenkilöinä toimi 115 12—18-vuotiasta koululaista. He saivat pohtia kunkin sanan merkitystä 20 sekunnin ajan, minkä jälkeen siirryttiin seuraavaan kohtaan. Merkityksen kuvaamisen sai kukin suorittaa haluamallaan tavalla, joko synonyymien tai parafrasien avulla, tärkeitä ominaisuuksia luonnehtimalla tai jopa piirtämällä. Annettuja vastauksia arvioidessani pyrin ottamaan huomioon sen tosiseikan, että todellisuudessa ihmiset eivät juuri joudu arvioimaan sanojen merkitystä irrallaan käyttöyhteydestä ja koetilanne sinänsä oli koehenkilöille outo. Toisaalta kuitenkin oletin, että tuttujen sanojen merkityssisällöstä olisi näissäkin olosuhteissa mahdollista antaa jonkin verran viitteitä. En siis vaatinut täydellisiä semanttisia määritelmiä, vaan pyrin ainoastaan täysin outojen sanojen erottamiseen. Kokeessa löytyi 22 sanaa, joiden merkityksestä kenelläkään ei ollut aavistustakaan. Nämä olivat:

hähnä	'kirjava tikka'
kaha	'kori, aitaus'
kalin	'nuotan osaverkko'
keri	'poistetun tuohen tilalle kasvanut uusi tuohi'
keso	'pasuri (kala)'
kuiru, -i	'pitkä ja kapea vene; lusikka'
kultaa	'haalia t. haroa (vedestä), lippoa (siikaa)'
kuu	'rasva'
kyyny	'kyynysilmä 'sirrisilmä'
käpsä	'kämpä; näpistelijä'

lansi	'matala; alanko'
lietsku	'keinu, kiikku'
liippo	'perhonen'
lämsä	'lasso, nahkahihna, silmukka'
läyli	'tukala, raskas, vaikea'
nysi, l-	'viikatteen varsi'
purku	'lumipyry'
sepä	'reen tai kelkan kaartuva etuosa'
sonsar	'kirppu'
säkiä	'monni (kala)'
tohtaja	'kuikka; vesilintu'
yrkä	'sulhanen'

Aineistosta löytyi lisäksi yhdeksän sellaista sanaa, jotka olivat tuttuja vain 1—5 koehenkilölle. Nämä olivat:

ora	'piikki, naskali'
otava	'lohiverkko; kalteva asento'
ponsi	'kädensija, varren pää'
pura	'taltta; naskali; kivipora'
rita	'loukkupyydyys'
saivar	'tän muna'
salava	'pajunsukuinen puu'
tymä	'liima; tasainen massa; sohjo'

Sanat *jänkä* 'suo, räme', *kaarne* 'korppi', *kajava* 'lokki', *pii* 'saka-ra, piikki' ja *uuhi* 'emälammas' olivat tuttuja 6—20 koehenkilölle. Sanat *kierä* 'kierteinen; lumeton, sileä; pyörä, kehrä', *kolja* 'jättiläinen; suuri, kolkko', *kuusama* 'eräs puutarhakasvi (*Lonicera*)', *lyly* 'käyrän puun kova selkäpuoli; em. puusta tehty lyhyt suksi' ja *niini* 'lehmuksen kuori, josta pinta on poistettu' olivat tuttuja 21—40 koehenkilölle, ja 41—57 koehenkilöä tunki sanat *koi* 'päivän sarastus', *sotka* 'eräs vesilintu', *säyne* 'säynävä (kala)' ja *sääksi* 'kalasääski (kotkansukuinen lintu)'. Kokeen perusteella löytyi kaikkiaan 45 sanaa, jotka olivat tuttuja vähemmälle kuin puolelle koehenkilöistä. Nämä sanat muodostavat yhteensä n. 16 % kaikista otokseen kuuluvista sanoista (275). Täysin tuntemattomien sanojen osuus on 8 %. Tutkittujen sanojen merkitykset antavat selvän vihjeen siitä, miksi kyseiset sanat ovat jääneet pois kielenkäyttäjien sanavarastosta. Monet niistä ovat vanhanaikaisiksi käyneiden tarve-esineiden nimityksiä. Joukossa on myös eläinten nimiä, ja tällöin unohtumisen syy on ollut joko se, että vanha nimitys on korvattu uudella (*kaarne* → *korppi*; *kajava* → *lokki*), tai sitten kyseinen eläinkin on nuorille nykysuomalaisille tuntematon (*saivar*, *täi*, *ke-so*). Lähes kaikki marginaaliset sanat ovat substantiiveja (vrt. ed.).

Suomea pidetään yleisesti konservatiivisena kielenä (vrt. Mikko

Korhonen, Virittäjä 3/1981). Tällä tarkoitetaan sitä, että suomen kieli on poikkeuksellisen hyvin säilyttänyt uralilaisen kielikunnan vanhoja tunnusomaisia piirteitä. Tätä käsitystä voidaankin perustella mm. viittaamalla siihen, että vanhimmille kantakielille tyypilliset, kaksitavuiset ja vokaaliloppuiset sanahahmot muodostavat vieläkin suomen sanaston kanoniset muodot, mikä heijastuu mm. tavutyyppeiden käyttöfrekvensseissä: suosituimmat tavutyypit ovat CV (40,4 %) ja CVC (27,5 %) (ks. tarkemmin eri tavutyyppeiden frekvensseistä eri sana-asemissa: Häkkinen, Statistische Angaben über die Lautstruktur der finnischen Sprache, Finnisch-Ugrische Mitteilungen, tulossa). Toisaalta äänne- ja sanatasolta on osoitettavissa lukuisia tapauksia, joissa suomea ei voi pitää konservatiivisena. Esim. äännejärjestelmä on muuttunut huomattavasti sekä vokaalien että konsonanttien osalta: konsonanttisysteemi on kutistunut miltei puoleen, ja vokaalijärjestelmä on vastaavasti laajentunut (vrt. esim. Pekka Sammallahti, FUF 43; Juha Janhunen, SUSA 77). Näin ollen myös sanojen äänneasujen on täytynyt monissa tapauksissa muuttua melkoisesti. Jos sanoissa kuitenkin on esiintynyt vain sellaisia ääniteitä, joiden arvellaan suomessa säilyneen muuttumattomina, saattavat vanhimmatkin rekonstruktiot vastata tarkasti suomen kielen nykyisiä asuja (esim. kantaauralin **pesä* > sm. *pesä* ja **maksa* > sm. *maksa*). Näissä tapauksissa ei tietenkään voida varmasti tietää, millä tavalla nykyisten sukukielten toisistaan poikkeavat variantit olisi suhteutettava toisiinsa, mutta äännevastaavuuksien systemointi voidaan usein suorittaa valitsemalla suomen nykymuotoja muistuttava kantamuoto. Toisentyypeistä tukea tarjoavat vanhimmat lainat, joiden kohdalla kantamuotojen rekonstruointi ei tapahdukaan suomen sukukieliä vertailemalla. Mm. vanhimpien indoeurooppalaisien ja arjalaisten lainojen arvellaan säilyttäneen alkuperäisen asunsa poikkeuksellisen hyvin juuri suomessa, ja sama koskee myös balttilaisia ja germaanisista lainoja, joskaan näiden avulla ei päästä yhtä kauas ajassa taaksepäin.

Sanaston kohdalla konservatiivisuus voi tarkoittaa sitä, että vanhat sanat ovat säilyttäneet äänneasunsa suhteellisen muuttumattomina. Toisaalta se voi tarkoittaa myös, että kieli on säilyttänyt huomattavan suuren määrän vanhoja sanoja. Jo edellä totesimme, että suomen kielestä löytyy kolmisensataa suomalais-ugrilaista sanavartaloa. Tätä lukua on tietysti verrattava muissa sukukielissä säilyneiden sanojen määrään, mikäli haluamme päästä selville siitä, osoittaako se erityistä konservatiivisuutta. Täysin vertailukelpoisten tietojen hankkiminen eri sukukielistä on kuitenkin mahdotonta, sillä useimmista näistä ei ole olemassa luotettavaa etymologista sanakirjaa. Suomi ja unkari ovat sanastotutkimuksen alalla huomattavasti muita kieliä edellä, ja ainoas-

taan näitä kieliä voi etymologiselta kannalta pitää perin pohjin tutkituina. Mutta kun sukukielten vertailevasta tutkimuksesta on kysymys, ei kahden kielen yksityiskohtainenkaan tuntemus vielä riitä, sillä uusi vertailumateriaali voi vaikuttaa arvaamattomalla tavalla jo tunnettuun aineistoon. Esim. obinugrialaisten kielten tutkimus saattaa paljastaa uusia rinnastusmahdollisuuksia. Ainoa toistaiseksi käytössä oleva yleisfennougristinen etymologinen sanakirja on Björn Collinderin *Fennougric Vocabulary*. Siihen on koottu vanhimmat etymologiat riippumatta siitä, missä kielissä näillä sanoilla on vastine. On huomattava, että Suomen kielen etymologisessa sanakirjassa on vain suomesta lähteviä etymologioita, joten sitä ei voi pitää koko kielikunnan vanhaa omaperäistä sanastoa kattavana. Sama koskee *mutatis mutandis* myös Unkarin kielen historiallis-etymologista sanakirjaa (*A Magyar nyelv történeti-etimológiai szótára I—III*) ja Unkarin sanaston suomalais-ugrilaiset ainekset -teosta (*A magyar szókészlet finnugor elemei I—III*). Käytännössä myös Collinderin sanakirja painottuu pitkälti suomen ja unkarin suuntaan, näitä käytetään mm. ensisijaisina haku-kielinä. Collinderin sanakirjan perusteella voidaan laatia pieni tilasto eri kielissä säilyneiden uralilaisten/suomalais-ugrialaisten sanojen määrästä. Kunkin kielen kohdalla on ilmoitettu kaksi lukua, joista ensimmäinen tarkoittaa varmoiksi katsottujen sanojen määrää ja jälkimmäinen sisältää myös epävarmat etymologiat:

1. suomi	507/587
2. ostjakki	494/600
3. voguli	436/510
4. lappi	401/490
5. syrjääni	365/435
6. unkari	337/406
7. samojedikielet	327/485
8. mordva	317/357
9. votjakki	305/358
10. tšeremissi	302/357

Tämän tilaston mukaan suomi näyttää todellakin säilyttäneen vanhaa sanastoa erityisen hyvin. On huomattava, että Collinderin ja Suomen kielen etymologisen sanakirjan kanta ei läheskään kaikkialla käy yksiin, joten varmoiksi arvioitujen rinnastusten luku on tässä suurempi kuin aikaisemmin mainittu n. 300. Tämän luvun perusteella suomi sijoittuisi tilastossa viimeiseksi, mutta tällöin olisi myös kaikkia muita lukuja tarkistettava, joten jonkinlaisen vertailukelpoisuuden saavuttamiseksi tässä yhteydessä on tyydyttävä pelkästään Collinderin etymologiaan kannanottoihin. Suomen ja unkarin etymologisten sanakirjo-

jen avulla on kuitenkin mahdollista tutkia, kuinka paljon suomen/unkarin sanoilla on vastineita eri sukukielissä ja päästä selville siitä, kumpi näistä on sanastollisesti lähempänä muita sukukieliä. Esim. Suomen kielen etymologisen sanakirjan perusteella suomen ja sen sukukielten yhteisten sanojen määrä on seuraava (vain varmoiksi arvioitujen rinnastukset on otettu mukaan):

suomi	— lappi	649
	— mordva	334
	— tšeremissi	249
	— syrjääni	238
	— ostjakki	214
	— votjakki	212
	— voguli	201
	— unkari	150
	— samojedikielet	115

Näitä lukumääriä arvioitaessa on muistettava sukupuuhypoteesi. Suomesta katsoen näyttää siltä, että mitä aikaisemmin kieli on eronnut suomen kielen sisältävästä kielikunnan haarasta, sitä vähemmän yhteisiä sanoja sillä on suomen kanssa. Ostjakki, voguli ja unkari ovat oletettavasti kuitenkin eronneet kieliyhteydestä samanaikaisesti, joten tältä osin ei huomattava ero yhteisten sanojen määrässä selity sukupuuhypoteesin avulla. Paremminkin voisi kuvitella, että unkarin sanasto on olosuhteiden muuttumisen takia uusiutunut voimakkaammin kuin muiden sukukielten. Unkarilaisetkin ovat melko varhain siirtyneet perinteisen metsästys- ja kalastuskulttuurin piiristä paimentolaisiksi ja joutuneet kosketuksiin muiden kansojen kanssa kiinteämmin kuin lähimmät sukulaisensa. On myös otettava huomioon se mahdollisuus, että osa eri kielten yhteisestä sanastosta on tulosta kielikunnan sisäisistä lainakosketuksista. Näistä eivät unkarilaisetkaan ole jääneet paitsi (vrt. esim. Károly Rédei, NyK LXVI, 253—), mutta maantieteellisistä syistä ovat kontaktimahdollisuudet olleet paremmat muiden sukukielten kohdalla. Lainojen ohella yllä oleva SKES:n tilasto sisältää myös uralilaista/suomalais-ugrilaista kautta nuoremmat uudennokset, ja tämä kasvattaa erityisesti suomen ja lapin yhteisten sanojen määrää.

A magyar szókészlet finnugor elemei -teoksen mukaan unkarista löytyy eri sukukielten kanssa yhteisiä sanoja seuraavasti:

unkari	— voguli	279
	— ostjakki	252
	— syrjääni	176
	— suomi	164
	— votjakki	159

— mordva	132
— tšeremissi	128
— lappi	120
— samojedikielet	99

Unkarin lähimmät sukukielet voguli ja ostjakki ovat odotuksenmukaisesti ensimmäisinä. Näissä kielissä on kaiken todennäköisyyden mukaan vanhan yhteisen sanaston lisäksi yhteisiä innovaatioita. Etymologioiden yhteismäärä on kuitenkin huomattavasti pienempi (1509) kuin suomea ja sen etäsukukieliä koskevien rinnastusten lukumäärä (2362), joten myös tämä tilasto tukee olettamusta unkarin sanaston voimakkaasta uusiutumuksesta ja suomen kielen suhteellisesta konservatiivisuudesta.

Jos otamme huomioon vain Collinderin esittämät uralilaiset/suomalais-ugrilaiset sanat, saamme eliminoiduksi innovaatioiksi todetut ainekset:

suomi — ostjakki	225	unkari — voguli	182
— lappi	213	— suomi	172
— voguli	213	— ostjakki	171
— mordva	196	— syrjääni	156
— syrjääni	187	— votjakki	142
— unkari	172	— mordva	139
— tšeremissi	168	— tšeremissi	129
— votjakki	162	— lappi	124
— samojedikielet	140	— samojedikielet	98

Tämä tilasto ei enää noudattele sukupuuhypoteesia sellaisenaan. Sopii kuitenkin pohtia, minkä verran siihen vaikuttaa rekonstruointitekniikka ja eri sukukielten tuntemus: obinugrilaisilla kielillä on keskeinen asema osoitettaessa sanoja suomalais-ugrilaisiksi (vrt. ed. s. 9), ja lappologian alalla suomalaiset fennougristit — myös etymologit — ovat päässeet pitemmälle kuin minkään muun etäsukukielen tutkimuksessa. Nämä tilastot eivät kuitenkaan anna mitään aihetta olettaa, että jokin muu sukukieli olisi sanastollisesti konservatiivisempi kuin suomen kieli.

Sanaston muuttuminen heijastaa kulttuurin, olosuhteiden, asenteiden ja arvostusten muuttumista. Perehtyminen sanaston historiaan on tärkeä keino suomen kielen ja kansan menneisyyden kartoittamisessa. Kielen menneisyys ei kuitenkaan koskaan jää lopullisesti taakse. Jokaisessa luonnollisessa kielessä elää traditio osana produktiivista ja jatkuvasti muuttuvaa systeemiä. Muutoksista ja muotivirtauksista huolimatta suomen kielen sanasto seisoo tukevasti monituhatuotisilla suomalais-ugrilaisilla juurillaan.

KAISA HÄKKINEN: *The Finno-Ugric roots of the Finnish vocabulary*

It is only possible to determine the age of the old, indigenous words in Finnish by reference to other languages of the same family. It is thought that the Uralian group scattered in a certain chronological order (Fig 1), and words are dated by trying to discover which of the related languages possesses etymological equivalents.

According to the etymological dictionary, there are in the Finnish language slightly less than 300 word stems for which an equivalent can with certainty be demonstrated in the most distant related languages, i.e. in the Ugric or Samoyede languages. Although the group of words in question is small in number, it forms the nucleus of the everyday vocabulary: something like half the hundred commonest words in Finnish go back to the oldest known proto-languages and they are therefore at least 4,000—5,000 years old (the distribution of the hundred commonest words according to age is shown in Figure 2).

A large number of the ancient words are among those most frequently used, but the proportion tails off evenly as we reach the less frequent vocabulary (Figure 3).

Many of the central words are pronouns and conjunctions. These parts of speech form a relatively dense network of referential relationships within the language and their frequency in written texts is far greater than one would expect from their quantitative proportion in the lexicon (cf. Table p. 13). Among the hundred commonest words in Finnish there are seventeen pronouns and eight conjunctions. The Uralian languages do not possess ancient indigenous conjunctions: they are either derived from indigenous stems, usually pronouns, or they have been borrowed. The proportion of pronouns in the central vocabulary is even greater if one takes into account both present-day pronouns and words derived from old pronoun stems, which may nowadays belong to some other part of speech (see list on page 14). Of the thirty-six words listed, almost a third are mentioned in etymological literature as providing evidence in support of the so-called "nosstrate" hypothesis, according to which there was an aboriginal relationship between the Uralian, the Altaic and the Indo-European families, as well as several other groups.

Some of the old Finnish words are nowadays marginal and no longer form part of the vocabulary of an ordinary user of the language. In this connection, however, one must remember individual and dialectal differences. In order to get a general picture of the marginal zone I organised a recognition test, in which the subjects were 115 Kouvola school-children aged 12—18 years. The results show that some 8 % of the oldest words must be considered unknown to young Finns of today. A total of sixteen per cent of the words studied were unknown to more than half of the children. Almost all the words that turned out to be marginal were nouns (see lists page 16—17).

Finnish is generally considered to be a conservative language, and as far as vocabulary is concerned this has been explained, among other things, by the fact that the original word form of the Uralian language (disyllabic, terminal vowel) has been preserved as the most popular form of stem. As far as one can tell from the etymological literature available, it seems that Finnish has also preserved the ancient vocabulary better than the related languages. This impression may, however, partly be due to the fact that the etymological study of Finnish has — with the exception of Hungarian — been more extensive than is the case for the related languages. In any case, according to the only dictionary of general Uralian, Björn Collinder's *Fenno-Ugric Vocabulary*, the greatest number of Finno-Ugric or Uralian words are found in Finnish (p. 19; undisputed, as well as possible, parallels are mentioned separately; note the difference between FUV and the information presented earlier by SKES). The fact that Ostyak and Vogul head the list statistically may in part be due to their central importance as testimony of Finno-Ugricism. Other statistics suggest that Finnish has more words in common with the related languages than Hungarian, and it may therefore be said that Hungarian at least cannot be considered conservative in the matter of vocabulary.

The vocabularies of all natural languages are open-ended: they are continuously changing systems. Not all the components, however, are equally prone to change or disappearance. It is known that the most central part of the vocabulary of a language

changes least, and this is confirmed by etymological studies of Finnish and word frequency. The relationship between the history of words and the history of sounds is not simple however: the most central vocabulary is dominated by a powerful tendency towards simplification and one effect of this is that the words are shorter and simpler than in the vocabulary as a whole; often, moreover, the movement towards simplification results in irregular sound changes. As far as the history of sounds is concerned this leads to a paradox: the rules concerning sound changes must be based on material which in all probability deviates from the norm.

