

# Suomen yleiskielen ja murteiden sanastojen suhteista

RAIMO JUSSILA

1. Suomen yleiskielen ja murteiden yhtäläisyyksiä ja eroja ei ole systemaattisesti tutkittu. Välistä kuulee kuitenkin esitettävän, että jonkin seudun puheenparsi on erityisen murteellista tai että paikkakunnan kieli on tavallista lähempänä yleiskieltä. Tavallinen, ilmeisesti E. N. Setälältä periytyvä käsitys on, että Keski-Suomessa Keuruun tienoilla puhuttavat murteet olisivat »puhtainta» suomen kieltä. Mitä puhtaudella tässä yhteydessä tarkoitetaan, on hieman epäselvää. Ilmeisesti ensinnäkin sitä, että näissä murteissa tuntuu käytettävän tavallista vähemmän nuoria ruotsin- ja venäjänperäisiä lainoja, toiseksi sitä, että äännepiirteiltään ne vaikuttavat olevan lähimpänä yleiskieltä.<sup>1</sup>

Onko yleiskielen ja murteiden erojen ja yhtäläisyyksien selvittely lainkaan tutkimuksen arvoinen aihe? Varmasti on. Ensinnäkin se pakottaa keksimään uusia keinoja ja mittareita, joilla aineistoja voidaan luotettavasti analysoida. Toiseksi se saattaa avata uusia näkökulmia suomen yleiskielen kehitykseen: mitkä murteet näyttävät vaikuttaneen eniten yleiskielen muotoutumiseen? Kolmanneksi sen avulla voidaan ehkä päätellä jotakin murteiden muuttumisesta: Millä alueilla yleiskielistyminen on edennyt pisimmälle? Mistä syistä? Neljänneksi tämäntyyppinen tutkimus valaisee koko leksikon ja tekstissä esiintyvän rajallisen sanaston suhteita.

Yleiskielen ja murteen erojen eksaktissa vertailussa on pulmana ensinnäkin sopivien ja vertailukelpoisten tutkimusaineistojen valinta ja käsittely. Tätä vaikeampi ongelma on tutkimusaiheen rajaus. Kielessä on lähes rajattomasti tutkittavia ilmiöitä. Mitkä ilmiöt olisi valittava vertailukohteiksi, jotta voitaisiin väittää pätevästi jotakin eri kielimuotojen

---

<sup>1</sup> Esimerkkinä on usein esitetty Keski-Suomen murteissa Jyväskylän pohjoispuolella esiintyvä *-ea*-yhtymä (*korkea*, *pimmeä*). Samuus on kuitenkin kirjoitusasusta johtuvaa harhaa. Näissä murteissa *-ea*-yhtymä on diftongi, mistä jo geminoituminen on todisteena. Normaalisissa yleiskielessä *ea* on vokaaliyhtymä.

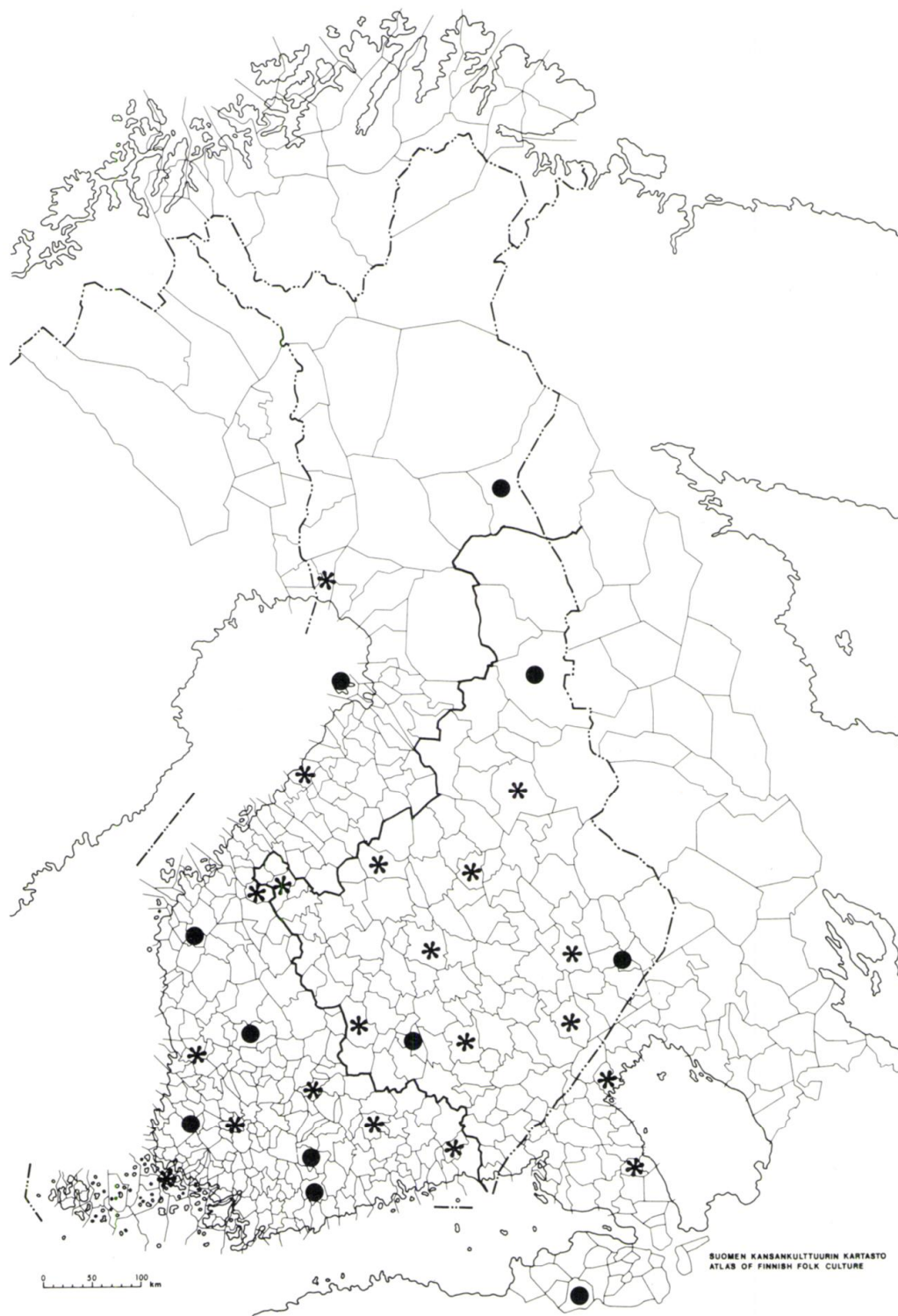
eroista ja yhtäläisyyksistä? Jotta eroja voitaisiin mitata, metodin on oltava kvantitatiivinen ja aineiston riittävän laaja. Jotta yleiskieltä ja murteita – nimenomaan niiden sanastojen eroja ja yhtäläisyyksiä – ylipäättään voitaisiin vertailla, metodin on oltava samalla kvalitatiivinen.

2. Tavoitteenani on selvittää, kuinka lähellä suomen yleiskieli ja murteet ovat toisiaan sanastoltaan. Ongelmaa voi lähestyä hedelmällisesti kahdesta näkökulmasta. Voidaan vertailla keskenään yleiskielen ja murteiden sanavarastoja, leksikkoja. Silloin tutkimusaineistoina olisivat lähinnä Nykysuomen sanakirjan esittelemä ja Suomen murteiden sanakirjan esittelemä leksikko tai jonkin yksittäisen murteen leksikko. Toinen lähestymistapa on se, että vertaillaan jonkin murteisen puhunnoksen eli äänitteestä litteroidun tekstin sanastoa yleiskielen leksikkoon. Tässä tutkimuksessa murreaineistoina ovat Suomen kielen näytteitä -sarjan sanastot ja yleiskielisenä vertailuaineistona Nykysuomen sanakirjan sisältämä leksikko.

Kielen, yksittäisen murteen tai yksittäisen murteenpuhujan koko sanastosta käytän nimitystä *leksikko*. Määrätekstin sanastosta käytän nimitystä *tekstileksikko*. Vertailtavina ovat tässä siis ryhmä murteen tekstileksikkoja ja yleiskielen koko leksikko. Tulokset kuvaavat tavallaan sitä, missä määrin yleiskieltä täydellisesti hallitseva ja murteen äänteellismuoto-opillisen systeemin taitava suomenpuhuja ymmärtää murretekstin sanastoa. Tulokset olisivat toisentyypisiä, jos kohteena olisi myös murteesta koko leksikko, puhumattakaan siitä, että verrattaisiin murteen ja yleiskielen tekstileksikkoja.

3. Perusaineistona ovat ne Suomen kielen näytteitä -sarjan 16 näytettä, joista on tehty täydelliset sanahakemistot. Aineiston laajuus on 208 056 sanetta ja 25 861 analysoitua sanaa (yhteensä n. 1 170 painosivua; yhdessä näytevihkossa keskimäärin 13 000 sanetta ja 1 600 sanaa). Yhdessä vihkossa on kahden kielenoppaan tunnin pituinen näyte. Perusaineisto sisältää 32 tuntia litteroitua puhetta 32 puhujalta, joista 18 on miehiä ja 14 naisia. Näytteistä kahdeksan on länsimurteiden ja kahdeksan itämurteiden puolelta. Tämän perusaineiston lisäksi olen käyttänyt täydennysaineistona Suomen kielen näytteitä -sarjan numeroita 2–16, joissa on arviolta 200 000 sanetta ja 24 000 sanaa. Tarkkoja laskelmia niiden perusteella ei voi kuitenkaan tehdä, koska näiden osien sanastot eivät ole täydellisiä. Koko aineiston laajuus on runsaat 400 000 sanetta ja 50 000 analysoitua sanaa. Aineiston jakautuminen suomen kielen alueelle näkyy oheisesta kartasta. Taulukossa 1 on yksityiskohtaiset tiedot perusaineiston sana- ja sanemääristä. Myös näytteiden yksilölliset kokovaihtelut on eritelty (1. näyte, 2. näyte).

## Suomen yleiskielen ja murteiden sanastojen suhteista



- = keskimääräistä murteellisempi pitäjänmurteen tekstileksikko.
- \* = keskimääräistä yleiskielisempi pitäjänmurteen tekstileksikko.
- = itä- ja länsimurteiden raja.

TAULUKKO 1. Murreaineiston tekstileksikot.

*Itämurteet*

SKN nro	Pitäjä	Sanoja			Saneita		
		1. näyte	2. näyte	Yht.	1. näyte	2. näyte	Yht.
1	Suomussalmi	1 243	971	1 863	7 122	8 529	15 651
17	Liperi	910	1 069	1 629	7 145	5 942	13 087
18	Rautalampi	1 046	1 142	1 822	6 968	7 173	14 141
24	Kurkijoki	928	925	1 541	6 403	6 269	12 672
26	Moloskovitsa	894	1 070	1 568	8 068	8 417	16 485
28	Sotkamo	1 172	942	1 761	8 045	6 444	14 489
29	Kerimäki	905	1 056	1 585	5 908	6 844	12 752
31	Joutsa	687	1 112	1 730	5 378	7 950	13 328
Yhteensä				13 499			112 605

*Länsimurteet*

SKN nro	Pitäjä	Sanoja			Saneita		
		1. näyte	2. näyte	Yht.	1. näyte	2. näyte	Yht.
19	Kalajoki	1 225	1 148	1 935	7 356	5 641	12 997
20	Ikaalinen	996	830	1 495	7 349	5 077	12 426
21	Loppi	1 120	919	1 683	6 554	5 824	12 378
22	Alatornio	854	714	1 265	5 105	3 499	8 604
23	Hollola	894	836	1 421	5 843	4 796	10 639
25	Alastaro	750	867	1 296	5 335	5 298	10 633
27	Velkua	1 096	836	1 608	7 714	5 500	13 214
30	Hinnerjoki	923	1 137	1 659	6 399	8 161	14 560
Yhteensä				12 362			95 451
Kaikki murteet yhteensä				25 861			208 056

Näytteiden sanastot on hakusanoitettu pääpiirtein Suomen murteiden sanakirjan karkeistusohjeita noudattaen (ks. tarkemmin Suomen murteiden sanakirjan toimitusohjeet). Tällöin on karsittu äänneopilliset murteellisuudet ja suurin osa morfologisista murteellisuuksista. Johdostyyppinä on erotettu enemmän kuin SMS:ssa (esim. omina hakusanoinaan ovat kaikki -MINEN, -STI ja -JA -johdokset). Metakielenä on siis nykyinen kirjakieli, joten sanojen vertailu yleiskielen ja murreaineiston välillä on teknisesti mahdollista.

Tyypillisiä säännönmukaisia äänteellisiä murteellisuuksia, jotka on hakusanoitettaessa suodatettu aineistosta, ovat esim. pitkän vokaalin diftongiutumisen, diftongien redusoituminen, yleis- ja erikoisgeminaatio, EA-yhtymän edustus, TS:n edustus ja liudennus. Epäsäännöllisistä äänneilmiöistä on metakielistettäessä tasoitettu mm. ensi tavun pitkän ja lyhyen vokaalin vaihtelu (*tyytyä* ~ *tytyä* → TYYTYÄ) ja eräät labiaalistumistapaukset (*pölätä* → PELÄ-

TÄ). Epäsäännölliset äänneilmiöt voitaisiin tulkita myös leksikkotason ilmiöiksi. Niiden esiintymistajuus on joka tapauksessa niin pieni, etteivät ne juuri vaikuta kokonaistuloksiin.

Myös muoto-opillisia piirteitä on metakielistettäessä tasoiteltu. Esimerkiksi yhdyssanoissa monikon genetiivin murteellisia variantteja ei ole säilytetty. Samoin on menetelty myös adverbien sijojen suhteen. Sen sijaan johto-opilliset (sekä johdinten että vartalovokaalien) murteellisuudet on noteerattu: omina hakusanoinaan ovat mm. HOPIO, JÄNES, SEKAANTUA, SEKAUTUA, SEKAINUTUA, SEKAUDA, SEKAAPUA, VANHATA, ALES).

Semantiikkaa on otettu huomioon vain karkeasti. Esimerkiksi SILTA-sanan merkitys 'lattia' ja HAKO-sanan merkitys 'lieko' on tulkittu murteelliseksi. Jos sanalla on semanttisesti sekä yleiskielisiä että murteellisia esiintymiä, sana on luokiteltu yleiskieliseksi.

4. Aineiston olen analysoinut vertaamalla SKN-sarjan sanastoja sana sanalta Nykysuomen sanakirjan sanastoon. Tässä analyysissä kahtiajako toisaalta nykysuomen sanoihin ja toisaalta Nykysuomen sanakirjan murteelliseksi luonnehtimiin sanoihin tai sanakirjasta puuttuviin sanoihin on liian karkea. Tästä syystä analyysissä on käytetty seuraavaa viisijakoa:

A) *Murteelliset sanat* = sanat, jotka puuttuvat Nykysuomen sanakirjasta (ei kuitenkaan b- tai d-kohtiin kuuluvia sanoja) tai jotka NS luonnehtii kansanmurteiksi (luonnehdinnat »kansanomainen», »kansanrunoissa», »murteellinen»; joskus myös »kansatieteessä», »mytologiassa»). Esim. HEHVO, KELA/PAATTI,<sup>2</sup> KISA/MUJEEN/VETO, KOTTI/KUORMA, PALTSAMO, SUTSKAHTAA, TAUTA, TRULLI.

B) *Muodostustavaltaan murteelliset sanat* = johdokset ja yhdyssanat, jotka puuttuvat NS:sta tai jotka NS luonnehtii kansanmurteisiksi mutta joiden kantasana tms. on NS:n mukaan normaalia yleiskieltä. Tähän ryhmään on luokiteltu myös sanat, joilla on NS:ssa esiintymätön tai NS:n kansanmurteiseksi kirjaama merkitys. Esim. EDELLIMMÄINEN (vrt. edellä, edellinen), EDUITSI (etu, edessä), ETELÄNÄ (etelässä), HARJAUDA (harjaantua), HYLKI/JÄÄ (hylje), JÄLKIIN (jälkeen), KAUKA/NIITTY (kaunoniitty, kauka-), KAUOIN (kauan, kauka-), MUSTILAINEN (mustalainen), PUITTAA (puu), PYSTÖSSÄ (pystyssä), PÖTKYLÄ (pötky).

C) *Yleiskielen tyyliväritteiset sanat* = sanat, jotka NS:n mukaan eivät kuulu neutraaliin yleiskieleen (lukuun ottamatta kansanmurteisiksi luokiteltuja a- ja b-kohdan sanoja). Tähän luokkaan kuuluviksi on analysoitu sanat, jotka NS on luonnehtinut seuraavin tyyli- tai käyttömerkinnöin: alatyylissä, arkityylissä, halveksiva(sti), harvinainen, harvoin, hengellisessä kielessä, his-

<sup>2</sup> Vinoviivaa olen käyttänyt yhdyssanojen osien rajan merkkinä.

toriassa, koululaiskielessä, lastenkielessä, leikillisesti, puhekielessä, raamatun kielessä, runokielessä, vanhentunut, vanhastavassa tyyliässä, ylioppilaskielessä, ylätyylissä. Tähän on sisällytetty myös sanat, joiden tilalle NS suosittaa jotakin toista sanaa (merkinnät »paremmin» ja »tavallisemmin» sekä nuoli →, joka viittaa hylättävästä muodosta suositeltavaan; esim. »kauaa adv. kielt. yhteyksissä; par. kauan», »ohkainen a. tav:mmin ohut», »käsinkivi s. → käsikivi»). C-ryhmään on luokiteltu lisäksi sellaisia sanoja, joita NS luonnehtii epäformaalein tavoin (esim. NS: »kissi vars. hyväilevästi: kissa»; »varrota harv. asiatyylissä: odottaa»; »sauhu ei juuri puhtaassa asiatyylissä»).

D) *Yleiskielisiksi tulkittavissa olevat sanat* = johdokset ja yhdyssanat, joista NS:ssa ei ole artikkelia mutta jotka tai joiden yhdysosat ovat normaalia yleiskieltä. Esim. HAMMAS/VIKAINEN, HERATON, IMELÄ/JAUHO/JUOMA, KÄÄRMEINEN, LOIS/MUMMO, LUTERILAIS/KANSA, OLKI/HUONE, OSTO/KENKÄ, PAULOITTAJA, PIENIMMÄINEN, RUOSTE/NAULA/LAATIKKO, SETÄ/UKKO, TORKKUJA, TUMMANLAINEN. Tähän luokkaan on sijoitettu muutamassa näytteessä esiintyvät raakalainat (HEER, MAKLER, MIRNAA). Rajatapaukset on pyritty luokittelemaan kuitenkin ryhmään B (esim. EDELLIMMÄINEN, PÖTKYLÄ).

E) *Yleiskieliset sanat* = sanat, jotka esiintyvät NS:ssa ja ovat sen mukaan neutraalia yleiskieltä.

Yhdyssanat on analysoitu jonkin yhdysosan perusteella niin, että ne kuuluvat vain yhteen »murteellisuusluokkaan». Ryhmä on valittu murteellisimman vaihtoehdon mukaan (esim. A-ryhmään kuuluvat AMPUMA/HOLLI ja RII/VIKKO/AIKA, B-ryhmään KUUSEN/KARAKKA ja TUKIN/AJU/PAIKKA, C-ryhmään METSÄ/TYÖ/KAMPEET ja HELLUNTAI/PÄIVÄ/EHTOO).

Analyysin yhteydessä olen kirjannut myös Nykysuomen sanakirjan kansanmurteisuutta ja tyyliä koskevat merkinnät (A-, B- ja C-ryhmät). Näin voidaan selvittää sitä, missä määrin NS pystyy tulkitsemaan murteellista sanastoa (NS:n murre-sanastosta ks. tarkemmin Tuomi 1989b).

Aineiston olen käsitellyt siten, että olen tallentanut A-, B-, C- ja D-luokkiin kuuluviksi arvioimani sanat murrenäytteittäin pitäen eri kielenoppaiden sanastot erillään. E-luokan sanojen ja saneiden määrät olen laskenut näytteen sanojen ja saneiden kokonaismäärien ja A–D-luokkiin kuuluvien sanojen ja saneiden määrien erotuksina. Taulukossa 2 on luokitusnäytteenä otos Sotkamon murteen sanastosta (Räisänen 1987).

Suomen yleiskielen ja murteiden sanastojen suhteista

TAULUKKO 2. Analyysinäyte.

Selityksiä: luokitusperuste on lihavoitu; \$ = merkitys on murteellinen.

Luokitus	Hakusana	Saneita kpl		Yht.	Luokitteluperuste
		1. näyte	2. näyte		
B	AJU <sup>2</sup>	0	2	2	murteellinen johdostyyppi, NS: murt.
B	AJU/MATKA	1	0	1	murt. johdostyyppi
C	AKKA	0	2	2	NS: us. halv.
D	ARKKU/MESTARI	1	0	1	yhdysosat yleiskieltä
B	AYRSHIRELAINEN	1	0	1	murt. johdos
C	ELUKKA	2	0	2	NS: us. ark.
B	ENNEN/VANHAINEN	2	0	2	murt. yhdyssanatyyppi
C	EVAKKO	3	0	3	NS: vars. puhek.
A	<b>HAAHKA</b> <sup>2</sup>	0	1	1	ei NS:ssa
C	HAALATA	0	2	2	NS: ark.
D	HALME/RUIS/PELTO	1	0	1	yhdysosat yleiskieltä
C	HALVATTU	1	0	1	NS: ark.
B	HALVIO	3	0	3	murt. johdostyyppi
A	<b>HARSOA</b>	0	1	1	ei NS:ssa
A	HAVUN/ <b>RANKI</b>	0	1	1	<i>ranki</i> -sanaa ei NS:ssa
B	HELKUTELLA	1	0	1	takavokaalisuus murt.; NS: helkytellä
C	HELLA	1	0	1	NS: vars. ark.
C	HOMMA	0	2	2	NS: us. ark.
B	HUONE/MIES \$	2	0	2	NS: kans. (merkitys luokitteluperusteena)
B	TUNKIO \$	0	1	1	murt. merkitys

5. Luokittelutulokset olen laskenut yhteen kielenoppaittäin ja pitäjänmurteittain sekä taulukoinut. Taulukossa 3 on esitetty sanamäärien mukaan lasketut pitäjänmurteiden tulokset. Näytteiden sanamäärät eivät vaihtele yhtä paljon kuin sanamäärät (vrt. alemmaa): vähiten sanoja on Alatornion näytteessä eli 1 265 ja eniten Kalajoen näytteessä eli 1 935.

Sanastollisesti murteellisia ovat ryhmiin A ja B kuuluvat sanat. Niiden osuus on yhteensä 6,9 %. Yleiskielisiin ryhmiin C, D ja E kuuluu 93,1 % perusaineiston tekstileksikkojen sanoista. Siis vain noin joka kymmenes tekstileksikon (ei juoksevan tekstin) sanoista on tässä tutkimuksessa asetuin kriteerein murteellinen. Täydennysaineiksesta tekemäni arviolaskelma antoi yleiskielisyysarvoksi 93,6 % (ks. tarkemmin taulukosta 5).

Taulukossa 4 on esitetty sanamäärien mukaan lasketut pitäjänmurteiden tulokset. Tuloksia arvioitaessa on muistettava, että näytteiden sanamäärät saattavat vaihdella varsin suuresti (Alatornion näytteen 8 604 saneesta Moloskovitsan 16 485 saneeseen).

RAIMO JUSSILA

TAULUKKO 3. Murretekstileksikon sanojen jakautuminen »murteellisuusluokkiin» A–E

(Lyhenteet: A = murteelliset sanat, B = muodostustavaltaan murteelliset sanat, C = yleiskielen tyyliväritteiset sanat, D = yleiskielisiksi tulkittavissa olevat sanat, E = yleiskieliset sanat.)

	A		B		C		D		E		Yht.
	f	%	f	%	f	%	f	%	f	%	f
Alastaro	15	1,2	61	4,7	50	3,9	41	3,2	1 129	87,1	1 296
Alatornio	35	2,8	46	3,6	62	4,9	17	1,3	1 105	87,4	1 265
Hinnerjoki	60	4,2	70	4,2	85	5,1	33	2,0	1 411	85,1	1 659
Hollola	30	2,1	50	3,5	36	2,5	19	1,3	1 286	90,5	1 421
Ikaalinen	44	3,0	60	4,0	70	4,7	37	2,5	1 284	85,9	1 495
Joutsa	45	2,6	80	4,6	69	4,0	71	4,2	1 465	84,6	1 730
Kalajoki	50	2,6	59	3,1	64	3,3	68	3,5	1 694	87,6	1 935
Kerimäki	42	2,7	63	4,0	54	3,4	37	2,3	1 389	87,6	1 585
Kurkijoki	25	1,6	47	3,1	43	2,8	38	2,5	1 388	90,1	1 541
Liperi	30	1,8	42	2,6	22	1,4	36	2,2	1 499	92,0	1 629
Loppi	60	3,6	81	4,8	88	5,2	72	4,3	1 382	82,1	1 683
Moloskovitsa	125	8,0	99	6,3	27	1,7	57	3,6	1 260	88,0	1 568
Rautalampi	36	2,0	38	2,1	49	2,7	95	5,2	1 604	88,0	1 822
Sotkamo	45	2,6	50	2,8	68	3,9	34	1,9	1 564	88,8	1 761
Suomussalmi	89	4,8	100	5,4	68	3,7	69	3,7	1 537	82,5	1 863
Velkua	33	2,1	66	4,1	69	4,3	49	3,1	1 391	86,5	1 608
Yhteensä	764	3,0	1 012	3,9	924	3,6	773	3,0	22 388	86,6	25 861

TAULUKKO 4. Murretekstien saneiden jakautuminen »murteellisuusluokkiin» A–E.

(Lyhenteet kuten taulukossa 3.)

	A		B		C		D		E		Yht.
	f	%	f	%	f	%	f	%	f	%	f
Alastaro	42	0,4	221	2,1	152	1,4	45	0,4	10 173	95,7	10 633
Alatornio	57	0,7	96	1,1	192	2,2	22	0,3	8 237	95,7	8 604
Hinnerjoki	96	0,7	193	1,3	359	2,5	37	0,2	13 875	95,3	14 560
Hollola	71	0,7	87	0,8	59	0,6	20	0,2	10 402	97,8	10 639
Ikaalinen	103	0,8	137	1,1	167	1,3	37	0,3	11 982	96,4	12 426
Joutsa	73	0,5	131	1,0	143	1,1	121	0,9	12 860	96,5	13 328
Kalajoki	73	0,6	160	1,2	213	1,6	69	0,5	12 482	96,0	12 997
Kerimäki	63	0,5	107	0,8	150	1,2	43	0,3	12 389	97,2	12 752
Kurkijoki	52	0,4	187	1,5	57	0,5	46	0,4	12 330	97,3	12 672
Liperi	41	0,3	70	0,5	38	0,3	57	0,4	12 881	98,4	13 087
Loppi	119	1,0	185	1,5	182	1,5	95	0,8	11 797	95,3	12 378
Moloskovitsa	447	2,7	363	2,2	131	0,8	78	0,5	15 466	93,8	16 485
Rautalampi	58	0,4	55	0,4	95	0,7	108	0,8	13 825	97,8	14 141
Sotkamo	62	0,4	100	0,7	172	1,2	49	0,3	14 106	97,4	14 489
Suomussalmi	173	1,1	208	1,3	162	1,0	89	0,6	15 019	96,0	15 651
Velkua	64	0,5	121	0,9	212	1,6	58	0,4	12 759	96,6	13 214
Yhteensä	1594	0,8	2421	1,2	2484	1,2	974	0,5	200 583	96,4	207 861



Saneista murteellisia on 2,0 % ja yleiskielisiä siis 98,0 %. Toisin sanoen SKN-sarjan juoksevassa tekstissä on keskimäärin vain joka 50. sane murteellinen. Täydennysaineistosta ei sanemäärien mukaista arviota voi tehdä.

Sekä sanoina että saneina mitaten murreteksti on kvantitatiivisesti varsin lähellä Nykysuomen sanakirjan yleissuomea. Sanastollisesti suomen murteet ja yleiskieli ovat tämän aineiston valossa yllättävän yhdenmukaisia. Tekstileksikon murteelliset sanat ovat enimmäkseen jonkin erikoisalan – maanviljelyn, kalastuksen, metsästyksen jne. – erikoistermejä. Niiden suhde murteen yleissanastoon on sama kuin yleiskielen ammattitermien suhde yleissanastoon. Murteelliset termit ovat siis murreyhteisöissäkin usein vain pienen asiantuntijajoukon, esim. ammattimiesten, sanastoa, eivät koko yhteisön yhteistä kieltä. Tämäkin näkökulma olisi syytä muistaa, kun murteellisen sanaston painoarvoa arvioidaan koko murreyhteisön kannalta.

Suomen yleiskielen ja murteiden sanastollinen läheisyys on ymmärrettävää, kun ajattelemme kirjakielemme ikää ja syntyä. Kirjakieli ja vielä enemmän yleispuhekieli ovat nuoria. Ja ne ovat syntyneet nimenomaan murteiden pohjalta. Vielä 1800-luvulla kirjakieli oli näiden kielimuotojen symbioosissa selvästi saavana osapuolena. 1900-luvun kuluessa kirja- ja yleispuhekieli ovat nousseet hallitsevaan asemaan. Ne ovat vahvasti vaikuttamassa murteiden sanaston muuttumiseen. Sen sijaan yleiskielen vaikutusta ei ole se, että vanhat murteelliset ammattitermit häviävät. Ne häviävät yhteiskunnallisista syistä, siksi että ammatit ja työtekniikat muuttuvat. Vanhoja sanoja ei enää tarvita, ja uusiin tarpeisiin saadaan uudet termit yleiskielestä.

Suomen yleiskielen ja murteiden leksikkojen likeisyys näkyy myös siinä, että Nykysuomen sanakirjaan on kirjattu varsin paljon murteellisuuksia ja kansanomaisuuksia. Perusaineiston luokkiin A ja B kuuluvista murteellisista tapauksista peräti 31 % esiintyy Nykysuomen sanakirjassa. Toisin sanoen sellaista murretekstileksikon sanastoa, joka ei lainkaan esiinny NS:ssa, on vain 4,6 %. Tämäkin kuvastaa kirjakielemme nuoruutta ja sitä, että murteet ovat olleet vielä tällä vuosisadalla kirjailijoihimme tärkeitä sanalähteitä.

Kirjakielen ja murteiden leksikkoja tutkinut Tuomo Tuomi toteaa, etteivät ne »muodosta keskenään kovin jyrkkää kontrastia». Yleiskielen 70 000 perussanasta ja johdoksesta noin puolet eli 35 000 esiintyy murteissa. Lisäksi hän toteaa kirjakielen saaneen runsaita sanastollisia itämurteista ns. murteiden taistelun aikana (Tuomi 1989a: 12 ja 1989b: 28). SKN-sarjan perusteella tekemäni havainnot tukevat näitä käsityksiä.

6. Tuloksista ei voi tehdä kovin jyrkkiä murremaantieteellisiä päätelmiä. Aineisto on alueelliseen tarkasteluun liian pieni. Tässä yhteydessä ei ole ollut mahdollista tutkia murteenpuhujien yksilöllisten ominaisuuksien, puheen-

aiheiden ja haastattelijan osuuden vaikutusta sanaston yleiskielisyyteen ja murteellisuuteen. Varsinainen tilastollinen käsittely on tässä myös jätetty syrjään.

Taulukkoon 5 on koottu tiedot sanojen yleiskielisyysasteesta sekä perusetä täydennysaineiston osalta. Täydennysaineiston lukuihin tulee suhtautua varauksellisesti, koska siihen kuuluvien näytteiden kokonaissanamäärä on arvioitu sivumäärän perusteella. Niiden murreosanasto on kuitenkin havaintojeni mukaan poimittu SKN-sarjan sanahakemistoihin tyydyttävän tarkkaan. Suurimmat aukot ovat D-luokkaan kuuluvien sanojen poiminnassa. Karttaan (s. 311) on kuvattu keskimääräistä murteellisemmat ja keskimääräistä yleiskielisemmät pitäjänkaksileksikot.

TAULUKKO 5. Murretekstileksikon yleiskielisyysaste prosentteina pitäjittäin sanamäärästä laskettuna (yleiskieliset luokat C, D ja E).

(Selitteet: \* = täydennysaineistoon kuuluva näyte, *kursivointi* = itämurteinen näyte.)

Sanat	%		
1. <i>Rautalampi</i>	95,9	17. Kauhava	93,7*
Noormarkku	95,9*	18. Alatornio	93,6
3. <i>Liperi</i>	95,6	19. <i>Kerimäki</i>	93,4
<i>Mikkeli</i>	95,6*	<i>Rautu</i>	93,4*
5. <i>Kurkijoki</i>	95,3	21. Vihti	93,2*
6. <i>Pihtipudas</i>	95,2*	22. Ikaalinen	93,0
7. <i>Jämsä</i>	95,0*	23. <i>Joutsa</i>	92,8
8. Sippola	94,8*	24. Hinnerjoki	92,2
9. <i>Sotkamo</i>	94,6	25. Salla	92,0*
10. Hollola	94,4	26. Loppi	91,6
Kalajoki	94,4	27. Hailuoto	91,2*
12. Pälkäne	94,3*	<i>Kiihtelysvaara</i>	91,2*
13. Alastaro	94,1	29. Jurva	90,1*
<i>Lappajärvi</i>	94,1*	30. <i>Suomussalmi</i>	89,9
15. <i>Lapinlahti</i>	93,8*	31. <i>Moloskovitsa</i>	85,7
Velkua	93,8		

Perusaineiston yleiskielisyysaste on 93,1 %, täydennysaineiston 93,6 % ja koko aineiston 93,3. Taulukosta 5 näkyy lievä tendenssi, että itämurteiset näytteet ovat hieman yleiskielisempiä sanastoltaan kuin länsimurteiset. Periferiassa ja kielirajaa vasten olevat pitäjät ovat selvästi murteellisimpia (*Moloskovitsa*, *Suomussalmi*, *Jurva*). Suomen keskiosissa olevat itämurteet ovat selvästi yleiskielisempiä. Yleiskielisyyden ja murteellisuuden vastakohta sanastossa heijastanee kuitenkin selvimmin vastakohtaa maantieteellinen keskusta — periferia. Rautalammin pitäjä keskipisteenä piirretty ympyrä, jonka säde on parisensataa kilometriä, sulkee sisäänsä tai sivuaa 18:aa SKN-sarjan

pitäjää. Niistä 15 on keskimääräistä eli 93,3:a %:a yleiskielisempiä ja vain kolme (Joutsa ympyrän sisällä, Ikaalinen ja Kiihtelysvaara ympyrän kehällä) keskimääräistä murteellisempia. Ympyrän ulkopuolelle jäävien pitäjien näytteistä viisi on keskimääräistä yleiskielisempiä ja kahdeksan murteellisempia.

7. Lopuksi on syytä ottaa esiin tulosten luotettavuus. Voidaan aiheellisesti kysyä, heijastavatko tulokset luotettavasti murteisen puhunnan keskimääräistä todellisuutta? Tähän voi suoralta kädeltä vastata, että SKN-sarja ei edusta keskimääräistä todellisuutta vaan keskimääräistä murteellisempaa aineistoa. Murteenkeruun menetelmien mukaan jo kentältä on pyritty valitsemaan sellaiset kielenoppaat, joiden puhe on joka suhteessa, myös sanastollisesti, mahdollisimman vanhakantaista ja murteellista. Sama valintakriteeri on toistunut litteroitavia näytteitä valittaessa: vähintäänkin 30 arkistoidusta äänitetunnista on taas pyritty julkaisemaan parhaat eli murteellisimmat. Toisin sanoen tulokseni ovat aineistopohjan takia varmasti keskimääräistä murteellisempia.

Toinen mahdollinen virhelähde sisältyy analyysimenetelmääni. Aina ei ole yksiselitteisen selvää, mihin luokkaan sana tulisi luokitella. Luokittelussa olen pyrkinyt valitsemaan epäselvissä tapauksissa murteellisemmän vaihtoehdon. Tämäkin on kallistamassa lopputulosta murteellisempaan suuntaan.

## LÄHTEET

- JUSSILA, RAIMO 1984: Kalajoen murretta. SKN 19.  
 JÄRVIKOSKI, OLLI 1986: Velkuan murretta. SKN 27.  
 KALLIO, JUSSI – LAHTI, MARKUS 1982: Kauhavan murretta. SKN 11.  
 LEHTINEN, RAIJA 1982: Pihtiputaan murretta. SKN 13.  
 LESKINEN, HEIKKI – MIKKULAINEN, RAIJA 1985: Kurkijoen murretta. SKN 24.  
 LYYTIKÄINEN, ERKKI 1978: Jurvan murretta. SKN 2.  
 LÄNSIMÄKI, MAIJA 1985: Lopen murretta. SKN 21.  
 MIELIKÄINEN, AILA 1980: Mikkelin murretta. SKN 10.  
 NS = Nykysuomen sanakirja 1–6. Hki 1951–1961.  
 PALANDER, MARJATTA 1983: Liperin murretta. SKN 17.  
 PALLONEN, JUHANI 1986: Moloskovitsan murretta. SKN 26.  
 PUNTTILA, MATTI 1979: Sippolan murretta. SKN 4.  
 PÄÄKKÖNEN, MATTI 1979: Hailuodon murretta. SKN 6.  
 REKUNEN, JORMA 1985: Alastaron murretta. SKN 25.  
 RÄISÄNEN, ALPO 1978: Suomussalmen murretta. SKN 1.  
 ——— 1987: Sotkamon murretta. SKN 28.  
 SARVAS, LEENA 1980: Lapinlahden murretta. SKN 9.  
 SIVULA, JAAKKO 1985: Hollolan murretta. SKN 23.  
 SKN = Suomen kielen näytteitä. Kotimaisten kielten tutkimuskeskuksen julkaisuja. Hki.  
 SMS = Suomen murteiden sanakirja 1–2. Kotimaisten kielten tutkimuskeskuksen julkaisuja 36. Hki 1985, 1988.  
 SOUTKARI, PENTTI 1985: Ikaalisten murretta. SKN 20.

RAIMO JUSSILA

- Suomen murteiden sanakirjan toimitusohjeet. Kotimaisten kielten tutkimuskeskuksen julkaisuja 54. Hki 1988.
- SÄRKKÄ, PIRJO 1982: Kiihtelysvaaran murretta. SKN 14.
- TOIVAINEN, KIRSTI 1985: Alatornion murretta. SKN 22.
- TOIVAINEN, KARI 1982: Raudun murretta. SKN 15.
- TUOMI, TUOMO 1989a: Suomen murteiden sanakirja. Johdanto. Kotimaisten kielten tutkimuskeskuksen julkaisuja 36. Hki.
- 1989b: Yleiskielemme murrepohjainen sanasto. (Teoksessa:) Nykysuomen sanavarat (toim. Jouko Vesikansa). Juva.
- VEHMASKOSKI, MAILA 1981: Sallan murretta. SKN 21.
- WHITE, LEILA 1980: Lappajärven murretta. SKN 7.
- 1982: Noormarkun murretta. SKN 16.
- VILPPULA, MATTI 1979: Vihdin murretta. SKN 5.
- VUORENPÄÄ, PIRKKO 1988: Hinnerjoen murretta. SKN 30.
- YLI-LUUKKO, EEVA 1980: Jämsän murretta. SKN 8.
- 1983: Rautalammin murretta. SKN 18.
- 1989: Joutsan murretta. SKN 31 (ilmestymässä).
- YLI-LUUKKO, EEVA — YLI-PAAVOLA, JAAKKO 1987: Kerimäen murretta. SKN 29.
- YLI-PAAVOLA, JAAKKO 1979: Pälkäneen murretta. SKN 3.

## Finnish standard and dialectal lexicons compared

RAIMO JUSSILA

The aim of this study is to examine how close the lexicon of Standard Finnish is to that of dialectal Finnish. Data from transcribed dialect tapes are compared with the total lexicon of Standard Finnish. The dialectal material comprises the vocabulary of the Finnish Language Samples collection, while the comparative material comes from the Dictionary of Modern Standard Finnish. The basic dialectal material represents 16 local dialects, and totals 208,056 word-forms and 25,861 lexemes. Supplementary material has been taken from another 15 Finnish Language Samples for which word-frequencies are lacking (approx. 200,000 word-forms and 24,000 lexemes). The dialectal words have been converted to phonologically and morphologically standard forms according to the rough transcription used by the Dictionary of Finnish Dialects.

The dialectal lexicons have been compared word for word with the Dictionary of Modern Standard Finnish (DMSF). The following classification was used: A)

*Dialectal words* = words not included in the DMSF, or which the DMSF labels as dialectal. B) *Words with a dialectal formation* = derivative and compound forms which are lacking in the DMSF or labelled there as dialectal, but having e.g. a base form which the DMSF takes as normal standard language. C) *Stylistically marked standard words* = those that according to the DMSF are not part of the neutral standard, and are labelled e.g. as vulgar, colloquial, poetic, etc. D) *Words which can be interpreted as standard* = derivatives and compounds for which the DMSF has no entry but which are, or the elements of which are, nevertheless normal standard language. E) *Standard language items* = those which appear in the DMSF and are accepted there as neutral standard forms.

Words belonging to groups A and B are defined as dialectal. They account for 6.9 % of the lexikon; 93.1 % are thus standard, i.e. groups C, D and E. In running text, 2.0 % of word-forms

are dialectal and 98.0 % standard. In terms of both word-forms and lexemes the dialectal texts are fairly close to the standard Finnish represented by the DMSF. This close lexical correspondence is due to the relatively late development of the standard language, and also the strong lexical influence of dialects on the standard. A further sign of this close relationship is that as many as 31 % of the words in groups A and B (i.e. of the dialectal words) are given in the DMSF and labelled there as dialectal.

The material is not extensive enough for a dialect-geographical study. Yet some tentative conclusions can be drawn. The eastern dialects seem to be slightly closer to the standard, as

regards their lexicon, than the western ones. However, the contrast between standard and dialect seems more likely to reflect the opposition between geographical centre and periphery. Localities on the periphery, especially those adjacent to a language border, are clearly more dialectal.

The material has not been selected at random. During the original fieldwork stage, the informants chosen were elderly, conservative folk with particularly marked dialectal speech. And in selecting material to be published from the extensive tapes the same criterion was chosen. The resulting texts are therefore more dialectal than average.