

## Puhutun ja kirjoitetun kielen välisistä syntaktisista eroista<sup>1</sup>

1. Suomen kirjakieli syntyi 1500-luvulla. On selvää, että se ennen kaikkea rakentui silloisen puhutun kielen varaan. Keskeisin osuus oli Turun seudun murteella, mutta alusta lähtien kirjasuomeen tuli muidenkin murteiden aineksia. Myöhemmin kirjakielen perusta on laajentunut sillä tavoin, että voidaan sanoa sen nykyään tukeutuvan suhteellisen tasapuolisesti kaikkiin päämurteihin.

Kirjakieli ei Suomessa sen paremmin kuin muuallakaan voinut syntyä yksinomaan oman, puhutun kielen varassa. Kirjoitus jo sinänsä asettaa kielelliselle ilmaisulle aina toisenlaiset vaatimukset kuin puhe. Lisäksi on tärkeää pitää mielessä, että Suomessa niin kuin yleensä muuallakin ensimmäiset painetut omakieliset tekstit olivat suurimmaksi osaksi muista kielistä käännettyjä. Olisi mahdotonta ajatella, ettei tällaisissa oloissa lähdekielten vaikutus olisi näkynyt myös käännöksissä. Näin kirjasuomi alkuvaiheistaan lähtien liittyi yhteiseen eurooppalaiseen kirjalliseen traditioon. Esikuvina olivat meillä lähinnä ruotsi, saksa, latina ja kreikka. Tyylin, tekstilingvistiikan ja tietyissä rajoissa syntaksinkin kannalta tarkasteltaessa on kaikissa eurooppalaisissa kirjakielissä paljon yhteistä.

Kun syntaktisesti ja tekstilingvistisesti verrataan toisiinsa puhuttua ja kirjoitettua suomea, niin eroja on paljon. Osa eroista selittyy suoraan viestintävälineiden erilaisuudesta<sup>2</sup>, osa taas siitä, että kirjoitettu kieli mainituissa suhteissa noudattaa vieraiden kielten malleja enemmän kuin puhuttu. Suomen kieli tuskin on tässä kohden missään erityis-asemassa.

Tässä esityksessäni tarkastelen toisaalta suomen kansanmurteiden, toisaalta suomen kirjakielen tai täsmällisemmin sanoen kirjoitetun

<sup>1</sup> Täydennetty versio esitelmästä, jonka kirjoittaja piti VI kansainvälisessä fennougistikongressissa Syktyvkarissa 26. 7. 1985. Saksankielisen kongressiesitelmän otsikko oli "Syntaktische Unterschiede zwischen gesprochener und geschriebener Sprache".

<sup>2</sup> Vrt. Ikola 1972: 18—19.

yleiskielen<sup>3</sup> välisiä syntaktisia ja osaksi myös tekstilingvistisiä eroja käyttäen etupäässä tilastollisia menetelmiä. Tosin on olemassa muitakin puhutun kielen lajeja kuin kansanmurteet ja muitakin kirjoitetun kielen lajeja kuin kirjoitettu yleiskieli, mutta käsitykseni mukaan näiden kahden kielimuodon vertailu paljastaa tyypillisimmät puhutun ja kirjoitetun kielen väliset erot. Nämä erot saadaan myös selkeämmin näkyviin, kun jätetään sivuun moninaiset eri kielimuotojen väliasteet.

Sellaiseen tilastolliseen tarkasteluun, joka tässä esityksessä on keskeisellä sijalla, on entistä paremmat mahdollisuudet nyt, kun Turun yliopistossa sijaitseva Lauseopin arkisto on toimintavalmis. Sen avulla voidaan tietokonetta käyttäen tulostaa toisaalta täydellisiä esiintymäluetteloita konteksteineen tai ilman, toisaalta tilastotietoja moninaisista syntaksin ja tekstilingvistiikan sekä muidenkin kielen osa-alueiden ilmiöistä, joko taulukkoina tai muussa muodossa. Arkistossa on kaikkia suomen murteistoja ja yleiskieltä edustava aineisto, joka käsittää yhteensä toista miljoonaa tekstisanaa, varustettuna koodimerkinnöillä ja muokattuna mm. tietokoneajoja varten. Toimivat haku- ja analysointi-ohjelmat ovat käytettävissä.<sup>4</sup>

Tässä lyhyessä esityksessä kiinnitän päähuomion sellaisiin seikkoihin, jotka koskevat tekstin ja virkkeiden rakennetta: tekstin jakautumista virkkeisiin, virkkeiden ja lauseiden pituutta, lauseiden välisiä alustus- ja rinnastussuhteita sekä eräitä sellaisia piirteitä, jotka ovat yhteydessä lauserakenteen tiiviyn ja löyhyyden kanssa.

<sup>3</sup> Terminologiasta ks. Ikola 1972: 16—.

<sup>4</sup> Lähemmin ks. Lauseopin arkiston opas. — LA:n koko aineisto sisältää 1 076 452 tekstisanaa, joista murreaineistossa 885 535 ja yleiskielen aineistossa 190 917 sanaa. Tässä tutkimuksessa olen käyttänyt otosta, joka sisältää noin tunnin pituista äänitettä vastaavat tekstinäytteet 60 pitäjältä (yht. 401 047 tekstisanaa) sekä kaikki LA:n kirjoitetun yleiskielen näytteet, joita on 25 pituudeltaan pitäjänaineistoa vastaavaa aineistoyksikköä (yht. 161 287 tekstisanaa, kun puhutun yleiskielen näytteet on jätetty pois). Käytetty murreotos sisältää aineistoyksiköt seuraavista pitäjistä: Luvia *Kalanti Rymättylä* Karjala Tl. Karuna *Perniö Paimio Merikarvia* Kokemäki *Pöytyä* Somero *Vihti Kihniö* Karkku *Vesilahti* Juupajoki Luopioinen Sääksmäki *Hattula Asikkala* Askola *Iitti Vehkalahti Kurikka* Laihia *Ylihärmä* Nurmo *Toholampi* Ylivieska *Pyhäjoki* Haapajarvi Hailuoto *Ylikiuminki Salla Sodankylä* Muonio Pihlajavesi *Soini Jämsä* Laukaa *Konginkangas Mikkeli mlk.* Punkaharju *Rantasalmi* Leppävirta *Nilsä* Vieremä *Sotkamo Suomussalmi Kitee* Kiihtelysvaara *Liperi* Ruokolahti Luumäki Antrea *Muolaa Koivisto* Seiskari Rautu *Lumivaara*. Kirjoitukseni kohdissa 3 ja 7 on käytetty puolta pienempää murreotosta, joka sisältää kursiviilla merkityt 30 pitäjää. Kirjoitetun yleiskielen aineisto jakautuu seuraaviin ryhmiin: hallinnon kieli ja virkakieli (3 aineistoyksikköä), sanomalehtien kieli (7), aikakauslehtien kieli (4), tietokirjallisuuden kieli (9), kaunokirjallisuuden kieli (2). Kirjoitukseni kohdassa 3 on käytetty 13 aineistoyksikön otosta, joka edustaa edellä mainituista kaikkia muita paitsi kaunokirjallisuuden kieltä.

2. Kirjoitetussa yleiskielessä teksti jakautuu selvärajaisiin virkkeisiin, jotka erotetaan toisistaan välimerkeillä, tavallisesti pisteellä tai kysymys- tai huutomerkillä. Jos taas tarkkaamme sellaista puhuttua kieltä, joka on vapaata kirjoitetun kielen suoranaisestä vaikutuksesta, niin tekstin rakenne on hyvin toisenlainen. On kyllä sellaisia kohtia, joita voidaan pitää selvinä virkerajoina, mutta toisaalta on pitkiäkin puhejaksoja, joista on vaikea löytää kiistattomia jakopaikkoja. Kuten olen aikaisemmin osoittanut, voidaan puhuttu tekstikin jakaa sellaisiin jaksoihin, joita voidaan sanoa virkkeiksi. Tärkeimpänä jakoperusteena ovat kieliopilliset seikat. Suuntaa antavana voidaan pitää Bloomfieldin ja Hockettin määritelmää, jonka mukaan virke (*sentence*) on sellainen kieliopillinen rakenne, joka ei ole minkään laajemman kieliopillisen rakenteen osa. Määritelmä ei ole aukoton, ja kieliopillisten seikkojen lisäksi on tarpeen ottaa huomioon myös tauot, intonaatio ja sisällölliset seikat.<sup>5</sup>

Lopputulokseksi jää, että puheessakin on selviä virkerajoja mutta että rajat kuitenkin monessa tapauksessa joudutaan asettamaan harkinnanvaraisesti. Tämä seikka kieltämättä vähentää virkkeiden pituutta koskevien tutkimustulosten luotettavuutta. Tämä varaus koskee ennen kaikkea kirjoitukseni kohtaa 3. Toisaalta mielivaltaisten ja mahdollisesti virheellisten rajanvetojen vaikutus vähenee, kun käytetään suurta ainesmäärää. Niissä tämän esityksen kohdissa, joissa ei ole kysymys varsinaisesti virkkeiden pituudesta, ei virkejaon ongelmallisuudella enää ole kovin suurta merkitystä.

3. Virkkeen pituuden mittana voidaan käyttää sanamäärää. Virkkeiden sanamäärien aritmeettinen keskiarvo on kirjakielen aineistossa n. 12½ sanaa, murreaineistossa n. 15½ sanaa. Murreaineistossa siis virkkeet ovat keskimäärin sanaluvultaan pitempiä kuin kirjakielen aineistossa.

Pelkkä aritmeettinen keskiarvo ei kuitenkaan yksin sano kaikkea. Asiaa voidaan tarkastella myös niin, että jaetaan virkkeet kahteen yhtä suureen ryhmään sanaluvun mukaan. Käyttämässäni kirjakielen otoksessa on jakauma sellainen, että lyhyempiä virkkeitä edustavaan puoliskoon tulevat kaikki 10-sanaiset ja lyhyemmät, toiseen puoliskoon taas kaikki 12-sanaiset ja sitä pitemmät. Raja jakaa siis 11-sanaiset virkkeet kahtia (11-sanaisista 76 % joutuu lyhyempien ja 24 % pitempien virkkeiden puoliskoon). Murreotoksessa raja asettuu 12-sanaisten kohdalle (12-sanaisista 41 % kuuluu lyhyempien ja 59 % pitempien puolelle). Näinkin tarkastellen siis murteet suosivat pitempiä virkkeitä kuin kirjakieli, mutta ero on pieni.

<sup>5</sup> Ikola 1976: 237—. Vrt. Saukkonen 1970: 4—5.

Jos vielä tarkastellaan kutakin sanalukua edustavien virkkeiden määriä yksitellen, niin kirjakielen aineistossa ovat 7—10-sanaisten virkkeiden ryhmät suurimmat (yhteensä nämä ryhmät ovat 28 % kaikista virkkeistä), murreotoksessa taas 4—6-sanaisten ryhmät (yht. 16 %).<sup>6</sup> Suurin on kirjakielessä 8-sanaisten ryhmä (7,6 % kaikista), murteissa vastaavasti 4-sanaisten ryhmä (5,4 %).

Murreaineistossa siis virkkeet ovat keskimäärin pitempiä, mutta toisaalta yleisimpien virkepituuksien kasauma sattuu lyhyihin virkkeisiin, lyhyempiin kuin kirjakielen aineistossa. Tämä näennäinen ristiriita selittyy siten, että murteen puhuja tavallisesti suosii lyhyitä virkkeitä mutta että murrepuhe toisaalta sallii erityisen pitkätkin virkkeet paremmin kuin kirjakieli. Murrepuheessa siis virkepituuksien hajonta on suurempi. Murrepuheen ja -tekstien tutkija huomaa helposti, että puhuja usein liittää pitkän jonon lauseita peräkkäin käyttäen *ja*- tai muuta konjunktioita, usein myös asyndeettisesti. Pysyteltäessä saman asian piirissä puhe voi jatkua pitkäänkin ilman mitään virkerajaksi tulkittavaa. Yleiskielen kirjallinen tyyli sen sijaan vaatii karttamaan kovin pitkiä virkeitä.<sup>7</sup>

4. Virkkeiden pituutta voidaan tarkastella myös käyttäen mittana virkkeen sisältämien lauseiden (pää- ja sivulauseiden) lukumäärää.

Käyttämässäni LA:n aineistossa sisälsi kirjakielen otoksen pisin virke 18 lausetta, murreotoksen pisin 22 lausetta. Kummassakin otoksessa on yksinkertaisten eli yksilauseisten virkkeiden ryhmä suurin. Kirjakielen otoksessa on yksilauseisia n. 51 % kaikista virkkeistä, kaksilauseisia n. 32 %, kolmilauseisia n. 12 % ja neli- tai useampilauseisia n. 5 %. Murreotoksen vastaavat luvut ovat 33 %, 21 %, 15 % ja 31 %.

Jos virkkeet jaetaan niiden sisältämien lauseiden määrän mukaan kahteen yhtä suureen puoliskoon, huomataan, että kirjakielen otoksessa yksilauseiset virkkeet yksinään muodostavat toisen ryhmän — niitähän oli 51 % kaikista. Murreotoksessa taas ensimmäiseen puoliskoon mahtuu yksilauseisten lisäksi suurin osa kaksilauseisista virkkeistä. Yksi- ja kaksilauseisia oli yhteensä 54 %.

Hieman karkeistaen voidaan sanoa, että kirjakielessä puolet kaikista virkkeistä on yksilauseisia, kun taas murteissa niitä on vain kolman-

<sup>6</sup> Murreotoksesta olen ottanut tähän vain kolme ryhmää, koska neljäntenä olisi yksilauseisten ryhmä, joka on erikoisasemassa (yksisanaiset vastaukset, kysymykset ym., joita ei kirjoitetussa kielessä juuri ole).

<sup>7</sup> Hakulinen, Karlsson ja Vilkuva ovat hieman pienemmän kirjakielisen aineiston perusteella (66 851 sanaa, 5 016 virkettä) saaneet virkkeen keskimääräiseksi pituudeksi 13,3 sanaa; suurimmat ovat heidän tilastossaan 8—11-sanaisten virkkeiden ryhmät (Hakulinen — Karlsson — Vilkuva 1980: 137—138). Ero johtunee lähinnä otoksen koostumuksesta.

nes kaikista. Selvä ero näkyy myös niiden virkkeiden määrässä, joissa lauseita on neljä tai enemmän: kirjakielessä tällaisia on vain 5 % kaikista virkkeistä, murrepuheessa taas 31 %.

Kirjakielen otos sisältää pääasiallisesti 1970- ja 80-lukujen kieltä. On mahdollista, että kaksi- ja useampilauseisten virkkeiden osuus vanhemmassa kirjakielessä olisi keskimäärin suurempi. Asia kaipaasi tutkimista.<sup>8</sup>

5. Tässä kohdassa jätämme sivuun yksinkertaiset virkkeet ja tarkastelemme yhdysvirkeitä eli kaksi- tai useampilauseisia virkeitä, joita edellisen kohdan mukaan oli kirjakielen otoksessa n. 49 % ja murreotoksessa n. 67 % kaikista virkkeistä. Nämä voidaan jakaa kahteen ryhmään seuraavasti: (1) jäsensidoksiset virkkeet, joissa on vain yksi päälause ja lisäksi vähintään yksi sivulause; (2) rinnastussidoksiset ja löyhäliitoksiset virkkeet, joissa on vähintään kaksi keskenään rinnasteista tai toisiinsa löyhemmin liittyvää päälauseetta, joihin vuorostaan voi liittyä sivulauseita.<sup>9</sup> Käsitelen rinnastussidoksisia ja löyhäliitoksisia virkeitä yhtenä ryhmänä, koska niiden erottaminen toisistaan olisi hankalaa.

Kirjakielen otoksen yhdysvirkkeistä on jäsensidoksisia n. 56 %, rinnastussidoksisia tai löyhäliitoksisia n. 44 %. Murreotoksen vastaavat luvut ovat n. 20 % ja n. 80 %. Tässä kohden siis puhuttu ja kirjoitettu kieli eroavat erittäin jyrkästi toisistaan: jäsensidoksisten virkkeiden osuus on kirjakielessä hyvin paljon suurempi kuin murrepuheessa, joka taas suosii rinnastussidoksisia ja löyhäliitoksisia, so. yleensä rakenteeltaan löyhempiä virkeitä. Tulos ei ehkä sinänsä ole yllättävä, mutta se voidaan todentaa vain tällaisen tilastollisen tutkimuksen avulla. Tulos

<sup>8</sup> Käyttämässäni otoksessa oli kirjakielen näytteissä 13 343 ja murrenäytteissä 25 048 virkettä. Pajunen ja Palomäki ovat valaisseet virkkeiden lausemääriä hieman toisella tavalla. Heidän tutkimuksensa perustuu myös LA:n aineistoon, mutta huomattavasti pienempään otokseen kuin yllä esittämäni tiedot, koska heillä ei vielä ollut käytettävissään yhtä laajaa aineistoa. Mitään ristiriitaa heidän tulostensa ja yllä esittämäni tulosten välillä ei ole. (Pajunen — Palomäki 1984: 15—16; 1985: 42—44.) Sen sijaan Hakulinen, Karlsson ja Vilku esittävät kirjakielen virkkeiden lausemääristä seuraavat prosenttiluvut, jotka poikkeavat olennaisesti minun esittämästäni: yksilauseisia 38 %, kaksilauseisia 35 %, kolmilauseisia 17 % (Hakulinen — Karlsson — Vilku 1980: 136). Tämä ero johtunee ainakin pääasiallisesti siitä, että tekijät määrittelevät lauseen toisin: he lukevat lauseisiin myös tärkeimmät lauseenvastikkeet (mts. 10, 136). Tällaisilla eroilla ei ole merkitystä arvioitaessa tutkimustulosteni luotettavuutta. Murreaineistosta ja kirjakielisestä aineistosta saadut tulokset ovat vertailukelpoisia vain, jos molemmat aineistot on analysoitu samoja periaatteita noudattaen. Näin on LA:ssa menetelty.

<sup>9</sup> Termeistä ks. Tuomikoski 1969: 67. Tuomikoski kuitenkin jakaa tekstin lauseisiin eräissä tapauksissa hiukan toisin kuin LA, jonka jakoa tässä noudatan (ks. Lauseopin artiston opas Liite 2: 69—75).

on senlaatuinen, ettei sitä murrepuheen virkerajojen osittainen harkinnanvaraisuuskaan horjuta.<sup>10</sup>

6. Lauseopin arkiston analysointiohjelman avulla olen tutkinut virkkeensisäisten alistusasteiden määrää: sitä, kuinka paljon on päälauseita, kuinka paljon päälauseelle suoraan alisteisia sivulauseita (1. asteen alistus), tällaiselle sivulauseelle alisteisia sivulauseita (2. asteen alistus) jne. Aineistoistamme saadaan tulokseksi seuraava taulukko:

|                    | Kirjakieli |        | Murteet |        |
|--------------------|------------|--------|---------|--------|
| Päälauseita        | 16 845     | 72,3 % | 53 529  | 70,7 % |
| Parenteesilauseita | 81         | 0,3 %  | 1 086   | 1,4 %  |
| 1. asteen alistus  | 5 577      | 24,0 % | 17 640  | 23,3 % |
| 2. asteen alistus  | 690        | 3,0 %  | 3 002   | 4,0 %  |
| 3. asteen alistus  | 74         | 0,3 %  | 438     | 0,6 %  |
| 4. asteen alistus  | 11         | 0,0 %  | 46      | 0,1 %  |
| 5. asteen alistus  | 7          | 0,0 %  | 5       | 0,0 %  |
|                    | <hr/>      |        | <hr/>   |        |
|                    | 23 281     | 100 %  | 75 746  | 100 %  |

Tulos on sikäli ehkä odotuksenvastainen, ettei kirjakielen ja murrepuheen välillä tässä suhteessa ole mitään merkitsevää eroa.<sup>11</sup> Yhtäpitävyys ulottuu siihenkin yksityiskohtaan, että ”syvin” alistusaste kummassakin kielimuodossa on tämän aineiston mukaan viides alistusaste.

Murrepuheen alistusasteiden tulkinta on toisinaan vaikeaa, sillä lauseiden väliset suhteet ovat usein epäselvät. Havainnollistaakseni murrepuheen moniasteisten alistusten luonnetta esitän muutaman esimerkin (sivulauseen edellä oleva numero osoittaa alistusasteen):

(Karjala Tl.) Kyl se väliste sanos- (1) et se o sukkel (2) ko (3) vaik pistä ovenkim puami (2) ni siit päästä vaa nii (3) ete tiärk (4) kosk kottit tule ja. l (Haapajarvi) Ja niin kait se on jääny sittä, (1) että ei sitä nyt ennää, ennää tota, kuulekkaa (2) että niitä, sitä tapana pi. . . pijettäsi (3) että, että tuota vejetää(n) sittä, (4) ku riijuula ollaa. l (Rautu) Sem, poikain tyttö ol' just kymmenän kuukaue vanha (1) ko mie sano Amalialle (2) jot soat uskoo (3) et auto on niin täys (4) et työ että mahu.

<sup>10</sup> Yhdysvirkkeitä oli kirjakielen otoksessa 6 573, joista jäsensidoksisia 3 713 ja rinnastussidoksia tai löyhäliitoksia 2 860. Murreotoksessa oli yhdysvirkkeitä 16 706, niistä jäsensidoksia 3 271 sekä rinnastussidoksia tai löyhäliitoksia 13 435.

<sup>11</sup> Pajunen ja Palomäki ovat pienemmän murreaineistonsa nojalla tulleet likipitään samaan tulokseen. Heidän kirjakieltä koskevissa tiedoissaan on sellainen ero, että päälauseiden osuus on suurempi (76,9 %). Ero johtuu siitä, että heillä oli kirjakielestä käytettävissään suppeampi ja yksipuolisempi aineisto. (Ks. Pajunen — Palomäki 1984: 16; 1985: 44.)

Ne harvinaiset tapaukset, jotka on tulkittu viidennen asteen alistuksiksi, ovat sen verran vaikeaselkoisia, etten ota niitä esimerkeiksi.

7. Yleisen käsityksen mukaan lauseet ovat kirjakielessä keskimäärin pitempiä kuin puhutussa kielessä. LA:n aineiston avulla asiasta saadaan seuraava kuva.

Kirjakielen aineistossa lauseen sanamäärä on keskimäärin 6,9, murreaineistossa keskimäärin 5,2. Jos lauseet jaetaan sanaluvun mukaan kahteen yhtä suureen ryhmään, niin kirjakielen aineistossa raja asettuu 6-sanaisten kohdalle, murreaineistossa taas 5-sanaisten kohdalle. Suurimmat yksittäiset ryhmät muodostuvat kirjakielen aineistossa 4—7-sanaisista lauseista (nämä yhteensä 50 % kaikista lauseista), murreaineistossa vastaavasti 3—6-sanaisista lauseista (62,5 % kaikista). Vielä voidaan mainita, että suurin sellainen lauseen pituus, josta on aineistossamme enemmän kuin yksi esiintymä, on kirjakielen aineistossa 42 ja murreaineistossa 25 sanaa.<sup>12</sup>

Nämä kaikki tarkastelutavat osoittavat samansuuntaisesti, että kirjakieli suosii jonkin verran pitempiä lauseita kuin murrepuhe. Voidaan vielä lisätä, että lausepituuksien jakauma murrepuheessa on hyvin säännöllinen ja hajonta pienempi kuin kirjakielessä. Tekstin jakaminen lauseiksi on murrepuheessakin enimmäkseen ongelmatonta. Lauseiden pituudesta puhuttaessa ollaan paljon varmemmalla pohjalla kuin puhuttaessa virkkeiden pituudesta.

8. Vain lyhyesti otan esille syntaktiset epäsäännöllisyydet, jotka ovat murrepuheessa ja muutenkin puhutussa kielessä paljon yleisempiä kuin kirjoitetussa kielessä. Tällaisia ovat esimerkiksi anakoluutit ja erilaiset ellipsit. Nämä eivät ole kirjoitetussa kielessäkään tuntemattomia, mutta kuitenkin monin verroin harvinaisempia kuin puhutussa kielessä. Vapaassa puhutussa kielessä ovat yleisiä myös täytelisäkkeet ja sellaiset sanan tai muun äännöksen toistot, jotka eivät tuo ajatukseen mitään lisää. Tähän kuuluu myös tankkaukseksi sanottu ilmiö. Täytelisäkkeet, merkityksettömät toistot ja tankkaus ovat ilmiöitä, jotka lienevät kirjoitetussa kielessä aivan tuntemattomia — paitsi jos kirjoittaja nimenomaan pyrkii jäljittelemään arkista, vapaata puhetta.<sup>13</sup>

<sup>12</sup> Kirjakielen otokssamme oli sanoja 161 287 ja lauseita 23 275, murreotoksessa vastaavasti 200 136 ja 38 427 (vrt. alav. 4). Pajusen ja Palomäen tutkimustulos ei ole ristiriidassa tässä saadun tuloksen kanssa (Pajunen — Palomäki 1985: 4). Hakulisen, Karlssonin ja Vilkun mukaan lauseiden keskipituus kirjakielessä on 6,6 sanaa (Hakulinen — Karlsson — Vilku 1980: 137). Ero omaan tulokseeni verrattuna on pieni, ja sekin selittyy siitä, että tekijät määrittävät lauseen eri tavalla (vrt. alav. 8).

<sup>13</sup> Vrt. Saukkonen 1970: 7.

Tässä kohdassa mainittuja seikkoja en ole tutkinut tilastollisesti. Kuitenkin on aihetta tässä yhteydessä kosketella termejä *kompetenssi* ja *performanssi*. Kielitieteen uusissa suuntauksissa on kompetenssi ollut keskeisemmässä asemassa kuin ennen. Kansanmurteiden tutkimus kuitenkin nojautuu aina pääasiallisesti performanssiin eli suoritukseen, koska tutkijalla vain poikkeustapauksissa on murteen käyttäjänä riittävä kompetenssi; näissä poikkeustapauksissakaan kompetenssi tuskin voi koskea muuta kuin yhtä ainoata suppeata murretta.

Tuskin voidaan ajatella niin, että esimerkiksi täytelisäkkeet ja sisälön kannalta merkityksettömät toistot olisivat murrepuheessa aina poikkeamia kompetenssista. Ne ovat päinvastoin puheen normaaleja aineksia, joita puhuja tarvitsee ilmaistessaan ajatuksiaan. Kielitieteen viimeaikaisissa suuntauksissa on, päinvastoin kuin aikaisemmin, kirja- ja yleiskieli suuressa määrin ollut etualalla. Kompetenssin ja performanssin käsitteiden soveltaminen murrepuheeseen on usein ongelmallista, ja sitä kannattaisi tutkia.

9. Olen toisaalla julkaissut tilastotietoja sanaluokkien yleisyydestä suomen murteissa ja kirjakielessä.<sup>14</sup> Kun Lauseopin arkisto silloin vielä oli varsin keskeneräinen, saatoinkin käyttää vain aineistoa, joka edusti eräitä suomen länsimurteita sekä toisaalta sanomalehtien pääkirjoituksia. Viittasin myös eräisiin muiden tutkijoiden laskelmiin, jotka koskivat kirjakieltä.

Näiden aineistojen perusteella sain sellaisen tuloksen, että partikkelit olivat murreaineistossa likimain kaksi kertaa niin yleisiä kuin kirjakielisessä aineistossa. Kaikkein suurin oli ero demonstratiivisten pronominaaliadverbien yleisyydessä; niistä taas oli murreaineistossa sanoilla *niin* ja *sitten* erityisen suuri frekvenssi. Näiden kahden adverbien frekvenssi oli murreaineistossani yhteensä 7 % kaikista tekstisanoista (sanomalehtien pääkirjoituksissa vain 0,5 %). Usein nämä sanat ovatkin puheessa aivan tai melkein merkityksettömiä, monesti täytelisäkkeisiin verrattavia. Myös merkityksetön toisto, josta oli puhetta edellisessä kohdassa, koskee erityisen usein juuri näitä sanoja.

Toinen piirre, joka tuli mainitussa tutkimuksessa selvästi esille, oli demonstratiivipronominien *se* ja *ne* silmiinpistävä yleisyys murrepuheessa. Näiden pronominiin eri muotojen frekvenssi oli n. 9 % kaikista sanoista (sanomalehtien pääkirjoituksissa n. 2 %). Osaksi tämä selittyy siitä, että näitä pronomineja puheessa yleisesti käytetään painottomina, lähes merkityksettöminä, toisinaan myös muiden kielten määräisen ar-

<sup>14</sup> Ikola 1979.



tikkelin käyttöä läheten. Toinen puheessa yleinen tapa on se, että lauseen teema (usein subjekti) ensin ilmaistaan pelkällä pronomiinilla ja vasta lauseen lopussa annetaan täsmällisempi tieto (esim. *Se oli siellä se poika*). Nämä pronominit ovat usein myös merkityksettömän toiston kohteena.

Huomiota ansaitseva havainto oli, että sanojen *niin*, *sitten* ja *se*, *ne* osuus kaikista tekstisanoista murrepuheessa oli kokonaista 16 % (kirjakielisessä aineistossa vain 3 %).<sup>15</sup> Vaikka käytettävissä oleva murreaineisto tätä asiaa tutkiessani olikin yksipuolinen, ei ole todennäköistä, että kuva kovin mullistavasti muuttuisi laajemmankaan aineiston valossa.<sup>16</sup>

10. Suomen kielen tyyppillisiin piirteisiin kuuluvana pidetään ns. lauseenvastikkeiden runsasta käyttöä. Yleisen käsityksen mukaan lauseenvastikkeet ovat luonteenomaisia nimenomaan kirjalliselle tyyllille ja harvinaisempia vapaassa puheessa. Yleisluonteisena arviona tämä käsitys epäilemättä pitääkin paikkansa, vaikkei asiaa ole perusteellisesti tutkittu ja vaikka on esitetty joissain kohdin virheellisiäkin arvioita. Olen eräässä v. 1985 ilmestyvässä tutkielmassa selvittänyt referatiivisen lauseenvastikkeen ja sen eri lajien esiintymistiheyttä eri murrealueilla ja kirjakielessä.<sup>17</sup> LA:n laajaan aineistoon perustuvan vertailun keskeinen tulos oli seuraava.

Murreaineistossa oli keskimäärin yksi referatiivirakenne kutakin n. 3000 tekstisanaa kohti, kirjakielen aineistossa taas yksi jokaista n. 300 sanaa kohti. Näin laskien siis referatiivirakenteen esiintymistiheys oli kirjakielessä kymmenkertainen verrattuna murreaineistoon. Tulos tukee selvästi edellä mainittua yleistä käsitystä. Toisaalta referatiivirakenteen eri tyyppien välillä on tässä suhteessa huomattavia eroja, joita olen selvittänyt mainitussa kirjoituksessani.

Muiden lauseenvastikkeiden yleisyyttä en toistaiseksi ole tilastollisesti tutkinut. Edellä mainittua tulosta voidaan pitää yleisesti suuntaa

<sup>15</sup> Suomen kielen taajuussanaston mukaan on sanojen *niin*, *sitten*, *se* ja *ne* osuus yhteenlaskettuna 2,8225 % eli pyöristettynä sama kuin omassa kirjakielisessä aineistossani (Saukkonen — Haijus — Niemikorpi — Sulkala 1979).

<sup>16</sup> Kun edellä oleva jo oli kirjoitettu, tulostutin kaikkia murrealueita edustavasta otoksesta (tekstisanoja yht. 200 136; ks. alav. 4) vastaavat frekvenssitiedot. Näistä sanojen *niin* ja *sitten* osuus oli yhteensä 7,2 % ja sanojen *se* ja *ne* 11,7 %. Näiden kaikkien sanojen yhteinen osuus oli 19,0 %, siis vielä suurempi kuin aiemman aineistoni valossa. Tulos on huomiota ansaitseva: murrepuheessa lähes joka viides sana on joko *niin*, *sitten* tai jokin *se*, *ne* -pronominin muoto. Tässä olisi kontrastiivinen tutkimustehtävä. Odotuksenmukaista on, että samantapainen ilmiö tunnetaan murrepuheessa muidenkin kielten alueella.

<sup>17</sup> Ikola 1985; vrt. Saukkonen 1970: 5—6; 1972: 50; 1977: 211—212.

antavana, mutta varmasti eri lauseenvastikkeiden välillä on tässä suhteessa huomattavia eroja. Eräät lauseenvastiketyypit voivat olla murrepuheessa suhteellisen yleisiäkin, toiset taas hyvin harvinaisia, jotkin ilmeisesti aivan esiintymättömiä.

11. Kysymys puhutun ja kirjoitetun kielen välisistä syntaktisista eroista on laaja tutkimusaihe. Tässä lyhyessä esityksessä olen poiminut esille vain muutamia eroja, joista useimpia olen selvittänyt tilastollisesti. Esitystä olen pelkistänyt sillä tavoin, että olen vertailtaviksi kielimuodoiksi ottanut toisaalta vain kirjoitetun yleiskielen, toisaalta vanhakantaiset kansanmurteet. Vaikka eri kielet ovat rakenteeltaan erilaisia, uskallan olettaa, että puhutun ja kirjoitetun kielen väliset erot muidenkin kielten piirissä ovat jossakin määrin samansuuntaisia kuin suomessa. Siksi olen otsikoinut esitykseni väljästi.

Toistan vielä tärkeimmät tulokseni tiivistettyinä:

(1) Jos virkkeiden pituuden mittana käytetään sanalukua, niin murrepuheen tyypillisimmät virkkeet ovat lyhyempiä kuin kirjakielen. Toisaalta murrepuheessa on hyvin pitkiä virkkeitä useammin kuin kirjakielessä. Virkepitouden hajonta on murrepuheessa suurempi.

(2) Virkkeen sisältämien lauseiden lukumäärä on murrepuheessa keskimäärin suurempi kuin kirjakielessä. Kaikista virkkeistä on yksilauseisia kirjakielessä suunnilleen puolet, murrepuheessa suunnilleen kolmannes. Murrepuheessa yksi- ja kaksilauseiset virkkeet yhteenlaskettuna edustavat suunnilleen puolta kaikista virkkeistä.

(3) Kirjakieli suosii jäsensidoksia virkkeitä huomattavasti enemmän kuin murrepuhe, jolle taas rinnastussidoksiset ja löyhäliitoksiset virkkeet ovat tunnusomaisia. Kaikista yhdysvirkkeistä on toisaalta jäsensidoksisten, toisaalta rinnastussidoksisten ja löyhäliitoksisten virkkeiden lukumääräsuhde kirjakielessä 56:44 ja murrepuheessa 20:80.

(4) Jos verrataan päälauseiden ja eriaisteisten sivulauseiden jakautumia, niin siinä ei murrepuheen ja kirjakielen välillä ole merkitsevää eroa.

(5) Kirjakieli suosii jonkin verran pitempiä lauseita kuin murrepuhe. Eri laskutapojen perusteella voidaan tätä eroa kuvata sanomalla, että tyypillisimpiä ovat kirjakielessä 6-sanaiset, murrepuheessa 5-sanaiset lauseet.

(6) Murrepuheessa ovat monenlaiset epäsäännöllisyydet, täytelisäkkeet, merkityksettömät toistot ym. paljon yleisempiä kuin kirjakielessä.

(7) Eräät painottomat partikkelit ja pronominit ovat murrepuheessa monin verroin yleisempiä kuin kirjakielessä. Tällaisia ovat suomessa mm. *niin, sitten, se, ne*.

(8) Eräät lauseenvastiketyypit ovat kirjakielessä yleisempiä kuin murrepuheessa, mm. referatiivirakenne.

## LÄHTEET JA LYHENTEET

- HAKULINEN, AULI — FRED KARLSSON — MARIA VILKUNA 1980: *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Publications of the Department of General Linguistics, University of Helsinki. Helsinki.
- IKOLA, OSMO 1972: Kirjakieli ja puhekieli, yleiskieli ja murre. *Sananjalka* 14 s. 16—26.
- 1976: Virke puheen yksikkönä. *Virittäjä* 80 s. 237—246.
- 1979: Hiukan sanaluokkien yleisyysuhteista suomen murteissa ja kirjakielessä. *Explanationes et tractationes Fenno-Ugricae in honorem Hans Fromm*. Edidit Erhard F. Schiefer. München.
- 1985: Referatiivisen lauseenvastikkeen käyttö suomen murteissa ja kirjakielessä. Paul Aristen kunniaksi julkaistava juhlaKirja. Tallinna.
- LA = Lauseopin arkisto, Turun yliopisto.
- Lauseopin arkiston opas*. Toimittanut Osmo Ikola. Kirjoittajat Osmo Ikola — Olli Järvi-koski — Ulla Palomäki — Pekka Porri. Lauseopin arkiston julkaisuja 1. Turku 1985.
- PAJUNEN, ANNELI — ULLA PALOMÄKI 1984: *Tilastotietoja suomen kielen rakenteesta 1*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 30. Helsinki.
- 1985: *Tilastotietoja suomen kielen rakenteesta 2*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 31. Helsinki.
- SAUKKONEN, PAULI 1970: Puhekielen luonteesta. *Kielikello* 3 s. 3—9.
- 1972: Puhekielen lauseopillisista tunnusmerkeistä. *Puhekieli ja ilmaisu*. Toimittanut Esko Vierikko. Porvoo. S. 49—52.
- 1977: Spoken and written language. *Folia linguistica*. Tomus XI 3/4. The Hague. S. 207—215.
- SAUKKONEN, PAULI — MARJATTA HAIPUS — ANTERO NIEMIKORPI — HELENA SULKALA 1979: *Suomen kielen taajuussanasto*. Porvoo.
- TUOMIKOSKI, R. 1969: Lauseiden ja virkkeiden terminologiasta. *Virittäjä* 73 s. 62—68.

OSMO IKOLA: *Syntaktische Unterschiede zwischen gesprochener und geschriebener Sprache*

Die syntaktischen Verschiedenheiten zwischen gesprochener und geschriebener Sprache bilden ein umfangreiches Forschungsgebiet, aus dem in einem kurzen Aufsatz nur einige vereinzelte Erscheinungen herausgegriffen werden konnten. Es wurde lediglich finnisches Forschungsmaterial benutzt. Wir können aber vermuten, dass die betreffenden Unterschiede in vielen Sprachen ähnlichen Charakter besitzen.

Der vorliegende Aufsatz behandelt syntaktische, mit statistischen Methoden erfasste Unterschiede zwischen der Schriftsprache und der Mundarten im Finnischen. Die Bereitschaft zu derlei Untersuchungen ist heute grösser als früher, weil das Syntaxarchiv in Turku jetzt für Auswertungen zur Verfügung steht. Für die folgende Untersuchung wurde ein Mundartenmaterial von etwa 400 000 Textwörtern und ein schriftsprachliches Material von etwa 160 000 Wörtern benutzt.

Der finnische Terminus *virke* wird hier mit *Ganzsatz* wiedergegeben (vgl. Admoni). Bei gesprochener Rede kann es oft unterschiedliche Meinungen darüber geben, wo ein Ganzsatz endet und der nächste anfängt. Aber auch gesprochene Texte sind in Abschnitte zerlegbar, die mit Recht Ganzsätze genannt werden können.

Die wichtigsten Ergebnisse dieser Untersuchung waren:

(1) Die typischen Ganzsätze in mundartlicher Rede sind kürzer als in der Schriftsprache, andererseits sind aber sehr lange Ganzsätze in mundartlicher Rede häufiger als in der Schriftsprache.

(2) Die Anzahl der Sätze innerhalb eines Ganzsatzes ist in mundartlicher Rede durchschnittlich grösser als in der Schriftsprache, wo ungefähr die Hälfte aller Ganzsätze nur je einen Satz enthält, während in mundartlicher Rede die Ganzsätze mit einem oder zwei Sätzen die Hälfte von allen Ganzsätzen ausmachen.

(3) Von allen zusammengesetzten Sätzen sind die Satzgefüge (hypotaktische Satzverbindungen) in der Schriftsprache sehr viel häufiger als in mundartlicher Rede (Schriftsprache 56 %, Mundarten 20 %). Diese bevorzugt Satzreihen, d. h. Satzverbindungen mit nebengeordneten Hauptsätzen.

(4) Im Rahmen des Computerprogramms im Syntaxarchiv kann auch die Häufigkeit der verschiedenen Subordinationsstufen von Sätzen innerhalb eines Ganzsatzes ermittelt werden. In diesem Fall erwies es sich, dass zwischen geschriebener und gesprochener Sprache kein statistisch signifikanter Unterschied besteht.

(5) Die Schriftsprache bevorzugt etwas längere Sätze als die mundartliche Rede. Wir können sagen, dass die typischen Sätze in der Schriftsprache aus 4—7 Wörtern, in den Mundarten hingegen aus 3—6 Wörtern bestehen.

(6) In mundartlicher Rede werden mancherlei Unregelmässigkeiten wie auch gewisse unbetonte Partikeln und Pronomina weitaus häufiger benutzt als in der Schriftsprache.

(7) Mehrere Typen von Satzentsprechungen sind in der Schriftsprache üblicher als in der mundartlichen Rede. Im Vergleich zu den Mundarten ist in der Schriftsprache die Frequenz der referativen Satzentsprechung (= der Partizipialkonstruktion) zehnmal so hoch.