

## Eri mittarien soveltuvuus toisen kielen oppijan sanaston rikkauten mittaamiseen<sup>1</sup>

Olen käynnistänyt v. 1990 Åbo Akademin laitoksessa Institutionen för lärarutbildning tutkimusprojektin ruotsinkie-

listen koululaisten suomen kielen taitojen karttumisesta. Tutkimuksen tavoitteena on 370 oppilaan ainekirjoitusaineiston valossa selvittää, millaisen kieli-aidon oppilaat suomen kielessä saavuttavat peruskoulun aikana ns. perinteisessä opetuksessa. Tutkimus valaisee suomea kohdekielenä oppivan välikieltä sekä kvantitatiivisesti että kvalitatiivisesti. Oppilaan strategioista sekä produktiivista että reduktiivista strategiaa voidaan tutkia korpuksen avulla. Aineistona on kaksi eri ainekirjoitussarjaa, argumentoituva »Tarvitaanko televisiota» ja kertova »Retkellä». Argumentoitavasta sarjasta on analysoitu mm. kaikki oppilaiden käyttämät verbisaneet ja verbilekseemit. Esittelen tässä yhteydessä tuloksia, jotka koskevat verbisanaston rikastumista suhteessa syntyperäisten vertailuryhmään.

Sanastolaskelmien lähtökohdaksi voidaan ottaa joko saneet (kaikki sanaesiintymät, r. *löpard*) tai eri sanat eli lekseemit tai *lemmat* (r. *lexikonord*). Korpuksen sanastoa ja sen runsastumista voidaan tutkia esim. absoluuttisilla keskiarvoluvuilla tai käyttämällä erilaisia mittareita. Sanaston rikkautta voidaan mitata mm. laskemalla eri sanaluokkien tai muiden kategorioiden tiheyttä suhteessa koko sanemäärään. Suomalaisissa tutkimuksissa on yleensä analysoitu sanaston rikkautta ja vaihtelua muilla kertoimilla, varsinkin saneiden ja eri sanojen määrien suhteilla (tyyppi/esiintymä-suhteilla).

Rikkautta voidaan mitata myös sanafrekvenssien avulla siten, että tutkitaan koko otoksen sanojen taajuutta. Mitä suurempi osa koko tekstistä muodostuu kielen suurtaajuisista sanoista, sitä köyhempää kieli on ja päinvastoin. Rikkauten mittareina on monissa yhteyksissä käytetty myös tekstissä vain kerran esiintyviä ns. *hapaks legomenon* -sanoja

<sup>1</sup> Kielitieteen päivillä Tampereella 8.5.1992 pidetty esitelmä hieman täydennettynä.

## Havaintoja

ja niiden suhteellisia osuuksia (ks. tark. Grönholm 1992).

Eri leksikkosanojen eli lemموjen määriä (ks. termistä Voionmaa 1990: 230) on varsin yleisesti pidetty hyvänä mittarina mitattaessa leksikaalisia taitoja (Hultman et al. 1977: 57). Hultmanin et al. tutkimuksessa eri sanojen määrän ja arvosanan korrelaatio on korkein mitä he ovat löytäneet. Samansuuntaisia tuloksia eri sanojen määrän merkitsevyydestä on saatu muissakin tutkimuksissa (Sjödoff 1989: 81). Voidaan päätellä, että eri sanojen vaihtelun kasvu ainekirjoi-

tuksessa viittaa samalla kielenkehityksen kasvuun. Äidinkielestä tehdyissä tutkimuksissa on yleinen havainto se, että tytöillä on enemmän lemموja kuin pojilla. Saarelan (1991: 137) mukaan tytöillä on sanastossaan 4. luokalla keskimäärin 15,1 eri verbiä ja pojilla vain 5,3.<sup>2</sup> Näin tyttöjen efektiiviset verbimäärät ovat noin kolminkertaisia poikiin nähden. Oman aineistoni tulokset eivät toisen kielen oppijalla ole aivan näin suuria ja dramaattisia. Esitän taulukossa 1 luokitain tyttöjen ja poikien keskimääräiset eri verbilekseemien määrät ainetta kohti.

TAULUKKO 1. Eri verbilekseemien määrät tytöillä ja pojilla.

Tytöt	Porvoo	Vantaa	Parainen	KA	Salo
4	6.79	7.00	5.33	6.37	12.25
6	13.18	6.38	7.00	8.85	12.77
8	16.69	15.30	7.77	13.25	26.60

Pojat	Porvoo	Vantaa	Parainen	KA	Salo
4	8.00	4.29	2.40	4.90	8.11
6	14.6	6.20	5.64	8.81	7.00
8	15.90	9.25	7.73	10.96	17.0

Verbilekseemien määristä käy ilmi pieni ero tyttöjen eduksi. He käyttävät enemmän sanoja ja enemmän eri sanoja; he siis käyttävät sanoja monipuolisemmin kuin pojat aivan kuin äidinkielessäkin, mutta erot eivät ole niin jyrkät. Itse asiassa eroa on vain 4. ja 8. luokalla, 4. luokalla vain puolitoista verbilekseemiä ja 8. luokalla hieman enemmän, runsaat kaksi lekseemiä. 6. luokalla ei ole mitään eroa eri sukupuolten välillä. Jos näitä lukuja verrataan syntyperäisten

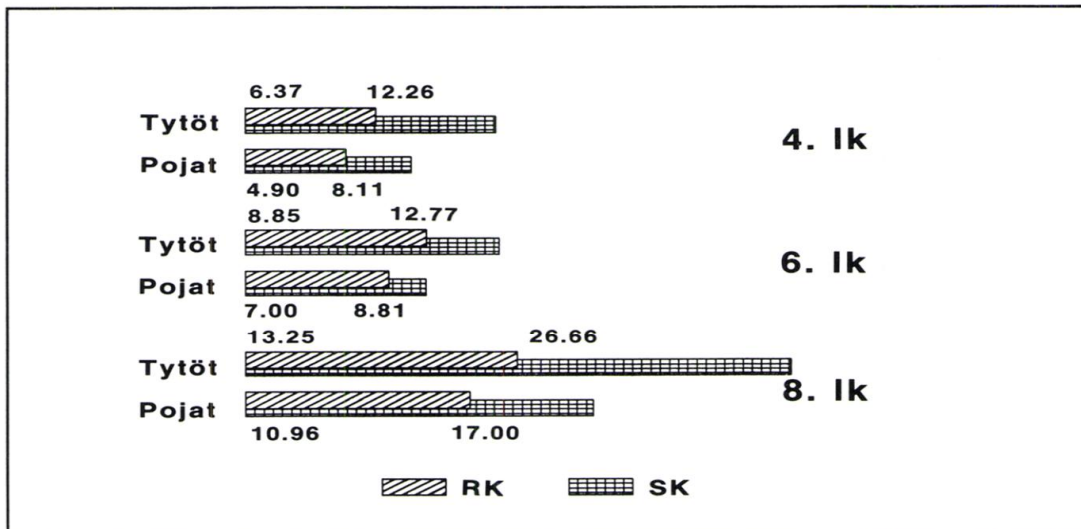
ryhmään, voidaan havaita tyttöjen ja poikien välillä hyvin jyrkkä ero kuten Saarelan aineistossakin. Ero jopa vain kärjistyy yläasteelle tullessa, jolloin suomenkieliset<sup>3</sup> tytöt kirjoittavat noin yhdeksän ja puoli lekseemiä enemmän kuin pojat, eli SK-pojat saavuttavat vain 64,9 % SK-tyttöjen lekseemimäärästä (ruotsinkieliset pojat 8. luokalla 82,7 %). Mitä sitten keskimääräiset eri verbien kappalemäärät kertovat? Ne kertovat tyttöjen suuremmasta sanavarastosta kaiken kaikkiaan ja kyvystä käyttää kieltä vaihtelevammin. Nimenomaan toi-

<sup>2</sup> Tekstin keskellä on käytetty yleiskielen mukaisesti desimaalipilkkuja, mutta taulukoissa ja korrelaatiokertoimissa on tilastollisten esitysten tapaan käytetty desimaalipilkkua.

<sup>3</sup> Tästä lähtien usein suomenkielinen = SK, ruotsinkielinen = RK.

sen kielen oppimisessa opittujen uusien sanojen lukumäärä ja niiden aktiivinen käyttö paljastaa tärkeitä seikkoja kielel-

lisestä kapasiteetista ja on havainnollisempi tason kuvain kuin erilaiset suhdeluvut.



Kuvio 1. Eri verbilekseemien kasvu ruotsinkielisillä (RK) tytöillä ja pojilla verrattuna suomenkielisiin (SK).

Kuvio 1 havainnollistaa eri luokkatasojen tulokset tytöillä ja pojilla eri kieli-ryhmissä. Vaakapylväät ovat sillä tavalla päällekkäisiä, että niistä käy ilmi, miten pitkälle toisen kielen oppijan taidot ylittävät verrattuna suomenkielisen vertailuryhmän taitoihin. Pylväistä näkyy, että erot ovat pienimmillään 6. luokan poikien ryhmässä. Tällöin SK-vertailuryhmän tulokset ovat niin vaatimattomia, että kielenoppijan välimatka syntyperäiseen jää vähäiseksi.

Sanaston rikkautta voidaan mitata kvantitatiivisesti useilla sen vaihtelevuutta kuvaavilla kertoimilla. Tällaista tutkimusta on harrastettu erityisesti äidinkielisistä korpuksista kirjailijoiden kielen tutkimuksessa. Kvantitatiivinen tutkimus voi yhtä hyvin antaa eräänlaisen makrokuvan myös toisen kielen oppijan saavuttamasta sanaston rikkaudesta ja vaihtelevuudesta. Yleensä kvantitatiivisissa mittauksissa käytetään lähtökohtana kaikkien saneiden (N) ja eri leksikkosa-

nojen (V) määrien suhteita (Broeder 1989: 90). Tähän saakka eniten käytetty mittari on ollut TTR eli sanojen tyyppi/esiintymä -suhde (type-token-ratio).

Omassa tutkimuksessani olen käyttänyt tällaisina kertoimina TTR-arvoa, logaritmiarvoa, Guiraud'n indeksiä, M-kerrointa ja OVIX-kerrointa. Hyvää mittaria on vaikea valita ennakkoon, koska ei ole vertailukelpoisia tuloksia juuri toisen kielen oppijasta. Broeder (1989: 91) on esittänyt kriteeriksi mittarin valinnalle, että validin ja reliaabelin mittarin tulee kielenoppijalla signaloida sanaston kasvua. Pyrin saamaan selville, mitkä käyttämästäni kertoimista näyttäisivät erityisen hyvin soveltuvan juuri toisen kielen oppijan sanaston vaihtelevuuden ja rikkauden mittaamiseen.

Ilman aikaisemmin tehtyjen tutkimustulosten tukea voisi luulla, että pitempään tekstiin, siis myös pitempään ainekirjoitukseen, mahtuu enemmän eri sanoja. Tällainen hypoteesi on kuitenkin jo aiemmin osoitettu virheelliseksi TTR-kerrointa käyttämällä (ks. Muikku 1982: 127). Myös muissa tutkimuksissa on osoitettu selvä korrelaatio tekstin pite-

## Havaintoja

nemisen ja TTR-arvon alenemisen välillä (ks. esim. Crystal 1988: 87–88). Herdandin kehittämä TTR-kerroin ottaa huomioon sanojen ja saneiden suhteen, mutta se on varsin altis muutoksille. Lyhyissä teksteissä kerran esiintyvien sanojen määrä lisääntyy aluksi nopeammin kuin saneiden määrä; vasta pitemmissä teksteissä uusien sanojen ilmaantumismuutos hidastuu (ks. esim. Särkilahti 1977: 66). Lyhyet tekstit saavat siis tällä kertoimella liian eksoottisia ja persoonallisväritteisiä tuloksia, mikä on otettava huomioon omassa otoksessani eripituisten aineiden analyysissä.

TTR-ratiota on ainakin 50-luvulta lähtien käytetty myös lasten kielen sanavaraston vaihtelua tutkittaessa. Alkuaikoina luultiin, että korkeampi arvo indikoi vähäisempää sanaston toistumista ja että vanhemmilla lapsilla arvo aina nousisi (ks. Crystal et al. 1988: 87). Tällaisen suoraviivaisen päätelmän on myöhemmin havaittu aiheuttavan monia teoreettisia ja metodologisia ongelmia mittausten yhteydessä. Crystal (1988: 88) suosittelee, ettei tätä mittaria käytettäisi alle 400 saneen otoksissa, koska se reagoi niin herkästi tekstin pituuteen

pienemmissä otoksissa.

Olen laskenut verbien TTR-arvot, jotka osoittavat kaikkien käytettyjen verbisaneiden ja kaikkien eri verbien (lekseemien) suhteen. Tuloksien analyysissä on kuitenkin otettava huomioon se aiemmissä tutkimuksissa havaittu seikka, että kerroin on varsin altis muutoksille. Kaikkein luotettavimpia tulokset ovat, jos vertailaan hyvin samankokoisia tekstejä (Niemikorpi 1991: 62). Omassa aineistossani ainekirjoitukset ovat kuitenkin hyvin erimittaisia. Niemikorven (1991: 255) mielestä TTR-kerroin on käyttökelpoinen myös, kun pyritään erottelemaan laajempia tekstilajeja tai rekistereitä toisistaan. Myös Broederin (1989: 93) mukaan TTR soveltuu joidenkin asioiden analysointiin. Hän on saanut tällä kertoimella positiivisia korrelaatioita esim. erilaisiin keskustelutyyppeihin. Samoin sillä voidaan erotella toisistaan erilaisia tekstityyppejä.

Koska on mahdollista olettaa, että TTR-arvoa voi käyttää paljastamaan eroja tekstilajeissa, on tutkimuksen koko ruotsinkielisten korpusta ja suomenkielisten korpusta verrattu keskenään (taulukko 2).

TAULUKKO 2. RK- ja SK-oppilaiden koko korpuksen TTR-arvot.

	Verbisaneet	Eri verbit	TTR
Ruotsinkieliset	3240	1664	51.4
Suomenkieliset	1863	963	51.7

Koko Tarvitaanko televisiota -ainesarjan aineistosta laskettuna ovat tyyppi/esiintymä -suhteet aivan samanlaiset toisen kielen oppijalla ja syntyperäisellä. Eri verbien määrä on noin puolet kaikkien verbisaneiden määrästä eli juuri sen verran kuin on tilastollisestikin arvioitu, että yli tuhannen saneen otoksissa on (ks. taulukko 2). Verbeistä lasketut arvot eivät suurissa otoksissa näytä mitenkään poik-

keavan tuloksista, joita on saatu laskeamalla kaikki sanaluokat. Esim. Broederin (1989: 92) mittauksissa toisen kielen oppijan TTR-arvot olivat toistokertomuksissa 48,0–53,1 ja vapaassa keskustelussa 51,6–63,0.

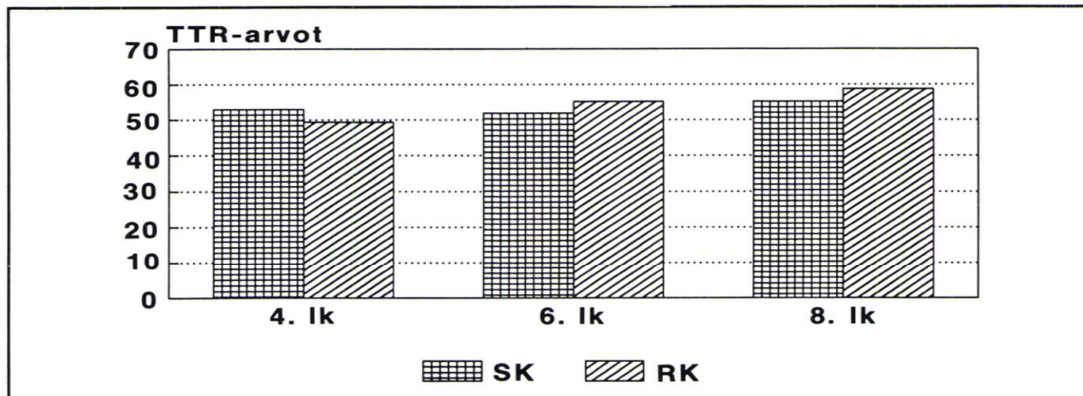
Taulukossa 3 ovat TTR-arvot prosenttilukuina. Luvuista voi verrata eri paikkakuntien ja luokkatasojen eroja.

TAULUKKO 3. Prosentitiset TTR-arvot luokittain ja paikkakunnittain.

	Porvoo	Vantaa	Parainen	KA	Salo
4	42.9	47.3	57.6	49.3	53.0
6	35.3	68.5	61.5	55.1	51.8
8	50.9	52.9	72.3	58.7	55.2

Eri luokkien väliset vaihtelut eivät ole yleensä kovin suuria ja arvot vaihtelevat, kuten Hultmaninkin tuloksissa, 50:n kummallakin puolella. Tuloksista joka tapauksessa näkyy, että toisen kielen oppijan lyhyissä teksteissä arvot voivat

nousta liian korkeiksi syntyperäisiin verrattuna. Näin on esim. 6. ja 8. luokalla. Kuviosta 2 voidaan nähdä, kuinka L2-oppijoiden pylväät ylittävät mainituilla luokilla syntyperäisten pylväät.



Kuvio 2. Verbien TTR-arvot toisessa kielessä verrattuna ensikieleen.

Yksilötasolla vaihtelut ovat suuria, ja käytännössä joidenkin oppilaiden TTR-arvot ovat täydet 100 prosenttiakin. Aine-esimerkeissä 1–2 on korkeat TTR-arvot.

(1) Vantaa 4. lk., tyttö. Minun mielestä televisiosta voi olla haittaa sillä sieltä tulee aika paljon ohjelmia missä murhataan ihmisiä. Semmosia elokuvia ei saisi lähettää sillä silloin lapset alkavat leikkivät että he murhaavat toisensa ja silloin voi käydä huonosti. TTR 81,8.

(2) Parainen 6. lk., poika. Minun mielestä televisiosta on hyötyä, koska sinä voi katsoa urheilutuloksia. Ja koska sinä näkee uutiset ja urheilu ohjelmat. TTR 100,0.

Toisaalta toisen kielen oppijan lyhyessäkin aineessa saattaa olla erittäin alhaiset TTR-arvot, mikäli hallittujen verbilekseemien määrä on pieni ja oppilas yleistää *olla-* ja *tulla-*verbiä. Esim. aineessa 3 oppilas on osannut aktiivisesti käyttää vain yhtä verbiä (*olla*):

(3) Parainen 4. lk., poika. Televisiosta on enemmän hyötyä, koska televisioohjelmat ovat niin jännittäviä ja hauskoja. Mutta siinä on aika paljon väkivaltaa, mutta kyllä sekin jännä. TTR 25,0.

Samojen verbien yleistäminen ei ole vain kaikkein lyhyimpien aineiden ongelma. Myös selvästi pitemmissä aineissa voi olla alhaiset TTR-arvot.

## Havaintoja

(4) Porvoo 4. lk., tyttö. Televisiosta on kyllä enemmän haittaa mutta se on kuitenkin aika hauska. Television parhain ohjelma on tähtien sota ja Dallas. Se olisi hauskaa jos televisiossa olisi enemmän hauskia ohjelmia. Minä en tiedä miksi Dallas on hauska, mutta se on hauska. Tähtien sodassa on yksi robotti tai ihminen tai mitä se nyt on ja sen nimion Darth Vader. Televisiossa on vain kaksi hyvää ohjelmaa ja dallas loppui ja Tähtien sota oli filmi. Nyt ei tule yhtään hauskia ohjelmia. TTR 26,6.

Saarela (1991: 13) on saanut hyvin samankaltaisia ja korkeita TTR-arvoja neljäsluokkalaisten natiivikirjoittajien ainekirjoituksista. Leskisen et al. (1974: 102) tutkimuksessa on saatu merkittävästi alempia verbien arvoja (6. lk. 26,8, 8. lk. 30,1 ja lukio 24,9), mutta on otettava huomioon, että myös tuotetut tekstit olivat vastaavasti pitempiä. Samoin on muistettava, että nykysivistysten mukaan arvot madaltuvat lapsen iän kasvaessa (Crystal et al. 1988: 88) – mikä tietenkin johtuu siitä, että tekstit säännönmukaisesti pitenevät. Pienet erot sukupuolten kesken johtunevat suunnilleen samansuuruisesta erosta tuotettujen tekstien pituuksissa, vaikka Saarela (1991: 13) erehtyykin nimittämään poikien saamaa arvoa »paremmaksi» kuin tyttöjen. Mm. Hultman on pitänyt tyttöjen ja poikien välistä yhden prosentin eroa (tyttöjen hyväksi) aivan mitättömänä käytännössä.

TTR-arvon käyttöä kohtaan on esitetty vakavaa kritiikkiä äidinkielellä kirjoitettujenkin tekstien analyysissä (ks. esim. Särkilahti 1977: 80–81; Särkkä 1987: 136). Vielä varovaisempi on oltava, jos käyttää tätä mittaria kielenoppijan kohdekielen rikkauden tutkimiseen, koska korkeat arvot harvoin viittaavat todelliseen rikkautta kuvastavaan sanaston variaatioon. Samoin muihin, varsinkin hyvin analyttisistä kielistä saatuihin tuloksiin, on suhtauduttava varoen. Mm. Zobl on saanut verbien TTR-arvoksi vain 22,8 korkeakoulutasoisilla toisen kielen (L2 englanti) oppijoilla. Tällaisiin lukuihin

englannin osalta vaikuttavat artikkelien ja tavallisimpien prepositioiden toistuminen.

Kyky vaihdella sanavarastoa on joka tapauksessa tärkeä mittari verrattaessa oppilaan kielenkehitystä. On selvää, että useammilla eri sanoilla voi kuvailla tarkemmin sisältöä ja ajatuksiaan kuin harvoilla sanoilla. Hultman et al. (1977: 56–) ovat ruotsinkielisten lukiolaisten sanaston variaatiota mitatessaan käyttäneet kehittämänsä ns. OVIX-arvoa (= *ordvariationsindex*), jonka kaava on  $V = N(2 - Nk)$ .  $N$  on saneiden määrä ja  $V$  on eri sanojen eli lekseemien määrä ja  $k$  on vakio ( $OVIX = 1/k$ ) (Hultman et al. 1977: 264). Heidän mukaansa OVIX on käyttökelpoinen erimittaisiinkin teksteihin eli se on melko riippumaton tekstin pituudesta. Arvon korrelaatio tekstin pituuteen on laskettu useista erityyppistä teksteistä. Mm. lukiolaisaineissa korrelaatio on ollut negatiivinen (–0,008) ja sanomalehtiteksteissä positiivinen (0,31).

Myös Josephson et al. (1990: 45) ovat käyttäneet OVIX-arvoa eri-ikäisten koululaisten (peruskoululaisen ja lukiolaisten) sanaston variaation mittaamiseen. He ovat suuresta ruotsiksi (= L1) kirjoitetusta ainekirjoitusaineistosta voineet päätellä, että OVIX-arvot kasvavat oppilaan iän mukana, päinvastoin kuin TTR-arvot. Heidän mukaansa oppilaiden äidinkielellään kirjoittamissa teksteissä arvot ovat 4. luokalla 0,50, 6. luokalla 0,55, 9. luokalla 0,60 ja lukiossa 0,67.

OVIX ottaa siis huomioon koko sane määrän ja eri sanojen määrän. OVIX voidaan laskea oppilaan käyttämistä kaikista sanoista tai vain jostakin sanaluokasta, kuten itse olen tehnyt. Hultmanin et al. mukaan OVIXilla laskettaessa teksti, jossa on suhteellisesti enemmän eri sanoja, saa korkeamman arvon kuin teksti, jossa on suhteellisesti vähemmän eri sanoja. Mikäli OVIX-arvolla todella voidaan tehdä päätelmiä vaihtelevammasta sanastosta, on se käyttökelpoisempi mittari kuin em. muut mittarit myös toisen kielen oppijan sanastoa arvioitaessa. OVIX-ar-

volla mitattaessa saadaan omasta aineistoni taulukon 4 osoittamat tulokset luokittain.

TAULUKKO 4. OVIX-arvot luokittain ja paikkakunnittain.

	Porvoo	Vantaa	Parainen	KA	Salo
4	51.5	50.6	66.3	56.1	35.6
6	42.1	50.3	52.5	48.3	35.8
8	33.8	35.5	26.9	32.0	27.3
KA	42.5	45.5	48.6	45.5	32.9

OVIX-arvot on laskettu jokaiselta oppilaalta erikseen, ja näin otos on voitu jakaa erikseen tyttöjen ja poikien ryhmään ja kieli-indeksiin perusteella hyvin ruotsin-

kielisten ja kaksikielisten ryhmään. Tyttöjen ja poikien luokittaiset keskiarvotulokset ovat taulukossa 5.

TAULUKKO 5. Tyttöjen ja poikien OVIX-arvot eri paikkakunnilla.

Tytöt	Porvoo	Vantaa	Parainen	KA	Salo
4	0.107	0.062	0.147	0.105	0.062
6	0.068	0.074	0.087	0.076	0.062
8	0.060	0.054	0.048	0.054	0.039
Pojat	Porvoo	Vantaa	Parainen	KA	Salo
4	0.088	0.126	0.374	0.196	0.101
6	0.056	0.047	0.089	0.064	0.082
8	0.056	0.079	0.110	0.081	0.055

Toisen kielen oppijan tuloksia OVIXilla analysoitaessa ei voi tehdä suoraviivaisia johtopäätöksiä korkeamman arvon ja vaihtelevamman sanaston välillä. Arvot näyttävät yleisesti ottaen laskevan iän mukana kuten TTR-arvotkin. On otettava huomioon, että Hultmanin et al. aineiston lukiolaiset eivät kirjoittaneet tavanomaisia aineita, vaan laajoja, keskimäärin 600 sanan kirjoitelmia, joissa sai käyttää tietoteoksia lähdemateriaalina. Oman aineistoni L2-kirjoittajat tuottavat vielä peruskoulun ala-asteella niin lyhyitä tekstejä, ettei OVIXillakaan pystytä täy-

sin eliminoimaan tekstin pituutta. Se näkyy aivan selvästi alempien luokkien liian korkeina arvoina, jotka monesti lähentelevät sataa tai ylittävätkin sen. Samaten toisen kielen oppijalle tyypillinen suuri vaihtelu on tälläkin arvolla nähtävissä.

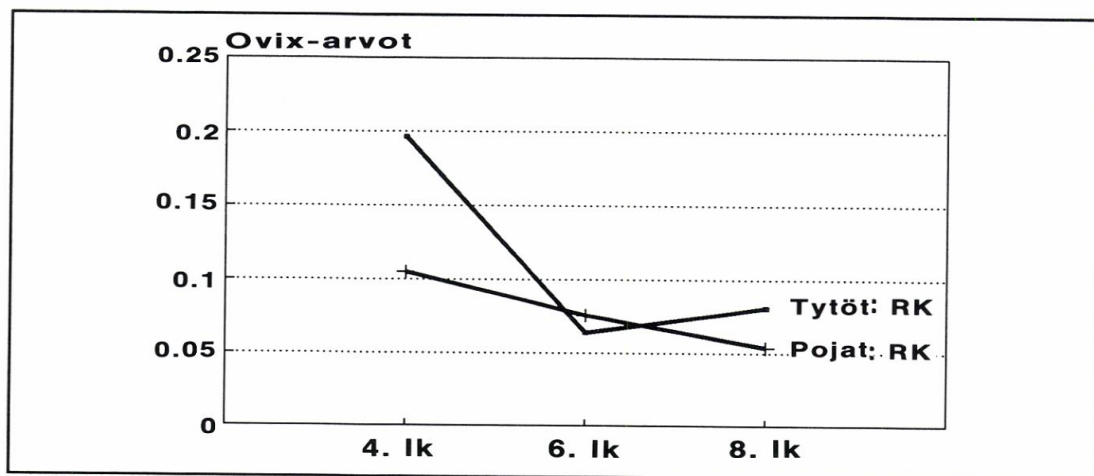
Myös Sjödoiff (1989: 81–82) on todennut, että OVIX ei sovellu aineisiin, jotka ovat 150:tä sanaa lyhyempiä. OVIX-arvon pituusherkyys tulee kärjistyneenä esille juuri L2-oppijan lyhyitä tekstejä mitattaessa, ja arvoista tulee epärealistisen korkeita (Sjödoiff 1989: 83). Tämän vuoksi Sjödoiff on käyttänyt kaavaa vain 7.-

## Havaintoja

luokkalaisten tai sitä vanhempien oppilaiden kirjoittamia aineita analysoidessaan, ja näin menetellen hän on pitemmistä aineista saanut 7.–9. luokkatasolta kasvavia lukuja. Esim. suomenkielisten (L1) oppilaiden OVIX-arvot ruotsin kielessä (L2) ovat kasvaneet 7. luokasta 9. luokkaan tasolla 44,6 → 45,6 → 49,4 (luvut ovat prosenttisia). Luvut on laskettu kaikista sanaluokista, niin että niitä ei voi suoraan verrata verbeistä laskettuihin arvoihin. Arvojen voi myös aina odottaa olevan hieman matalampia ruotsin kielestä laskettuna, koska artikkelit ja prepositiot vähentävät sanaston vaihtelua verrattuna suomen kieleen, josta tällaiset muoto sanat puuttuvat. Lyhyemmistä aineista arvot nousevat liian korkeiksi. Näin ollen oman aineistoni toisen kielen oppijoiden hyvin lyhyistä aineista tulee korkeita arvoja, jotka luonnollisesti laskevat pituuden »normalisoituessa» ja alkavat nousta

vasta, kun pituus saavuttaa vähintään 150 sanetta.

OVIX-arvoista piirretyt ruotsinkielisten käyrät (kuvio 3) ja suomenkielisten käyrät (kuvio 4) ovat kummassakin ryhmässä yleensä laskusuuntaisia. Käyristä näkyy myös havainnollisesti, kuinka poikien arvot näyttävät korkeammilta kuin tyttöjen, koska poikien aineet ovat lyhyempiä. Poikien ja tyttöjen väliset erot ovat tälläkin mittarilla mitattaessa suuremmat syntyperäisten ryhmässä, ja erot ovat suurimmillaan 4. luokalla. Huomionarvoista on, että OVIX näyttää saman tuloksen, joka tuli esille jo tyttöjen ja poikien eri lekseemien määrissä. 6. luokan ruotsinkieliset pojat ovat edenneet kielkehityksessään harppauksenomaisesti ja ovat tilapäisesti jopa parempia kuin tytöt (kuten alempi käyrä on tulkittava). Poikien taso normalisoituu tyttöihin nähdessä uudestaan 8. luokalle tultaessa.

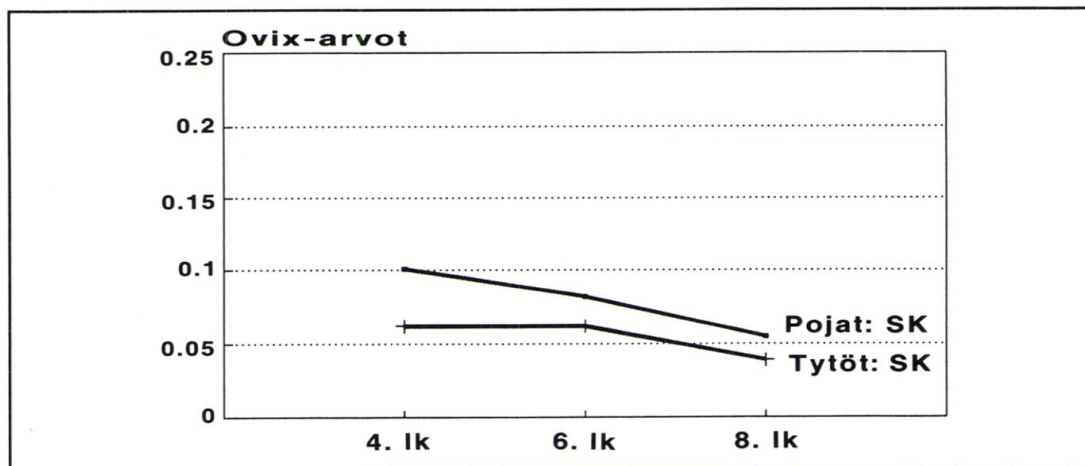


Kuvio 3. Ruotsinkielisten tyttöjen ja poikien OVIX-arvojen kehitys.

RK-tyttöjen ja RK-poikien tuloksia arvioitaessa nähdään, että porvooolaisilla pojilla on joka luokka-asteella alemmat ja siis normaalimmat arvot kuin tytöillä. Tämä tulos ja tällainen tulkinta alemmista arvoista normaalimpina korreloi täysin taulukko 5:n esittämän tiedon kanssa, et-

tä porvooolaisilla pojilla todettiin olevan keskimäärin enemmän eri leksikkosanoja kuin tytöillä. On myös pääteltävissä, että paraislaisten 4.-luokkalaisten poikien korkein arvo on epänormaalein, mikä johtuu ylilyhyistä, jopa yhden virkkeen aineista. Niissä luonnollisesti eri verbien määrä voi olla sama kuin verbisaneiden määrä ja arvo voi kohota jopa lähelle 0.4:ää (0,374). 8.-luokkalaisten pitemmät aineet ovat jo paremmin vertailukelpoisia





Kuvio 4. Suomenkielisten tyttöjen ja poikien OVIX-arvojen kehitys.

suhteessa muiden tutkimusten OVIX-arvoihin. Luvuista käy ilmi mm., että RK-tytöt saavuttavat 8. luokalla saman arvon (0,054) kuin SK-pojat (0,055), mikä on sinänsä arvokas OVIXin avulla saatu tieto. Tyttöjen sanaston vaihtelu ei yllä tässä luokassa poikien tasolle.

Edellä esitettyjen tulosten valossa voidaan odottaa kaksikielisten saavan matalampia ja normaalimpia arvoja.<sup>4</sup> Näin näyttää olevankin joka luokka-asteella (RK 0,206 → 0,072 → 0,069 – KK 0,093 → 0,061 → 0,055). Samaten kaksikielisten arvot ovat 4. ja 8. luokalla hieman korkeampia kuin suomenkielisten. Erot kaksikielisten ja suomenkielisten välillä ovat OVIXilla mitattaessa yllättävän pieniä (KK 0,093 → 0,061 → 0,055 – SK 0,079 → 0,069 → 0,046). Esitän näytteissä 5 ja 6 rinnakkain kaksikielisten ja suomenkielisten aineen, joiden verbeistä lasketut OVIX-arvot ovat samat.

(5) Vantaa 6. lk., tyttö, KI 3,09 (kaksikielinen). Minun mielestä televisiosta on hyötyä. Sieltä tulee uutisia niin että tietää mitä muissa maissa tapahtuu.

<sup>4</sup> Oppilaiden kieli-indeksi perustuu 7 eri variaabeliin, jotka on saatu kyselyn tuloksena. Kaksikielisten kieli-indeksin rajat ovat 1–3,5 ja ruotsinkielisten 3,6–5.

Usein tulee opettavia ohjelmia. OVIX 0,054.

(6) Salo 6. lk., tyttö (suomenkielinen). Minun mielestäni televisiota tarvitaan, onhan sitä mukava katsella kun ulkona on huono ilma tai, kun ei ole muuta tekemistä, mutta ei televisiota pitäisi aina katsella pitäisihän ihmisellä olla muitakin harratuksia kuin television katselu. OVIX 0,054.

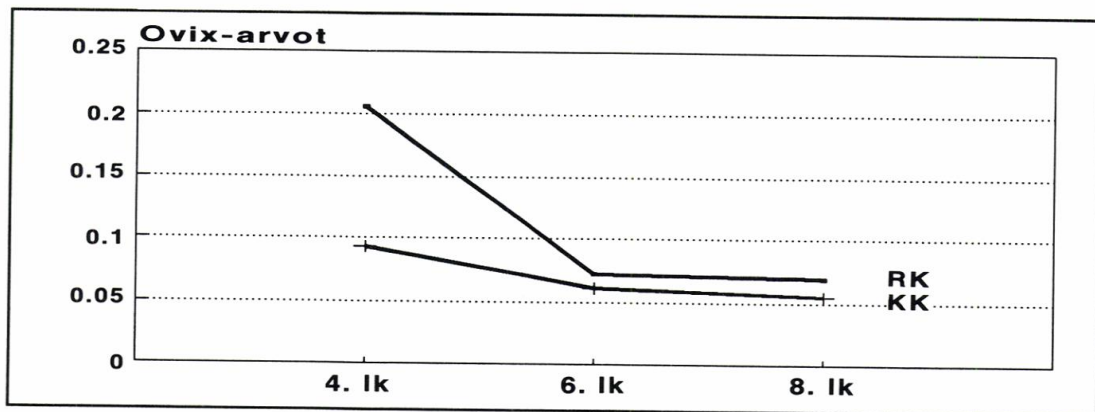
Ruotsinkielisten ja kaksikielisten erot eivät ole suuria. Suurimmillaan ne ovat 4. luokalla, jolloin erot kärjistyvät varsinkin paikkakuntaakohtaisesti. Porvoon 4.-luokkalaisilla ruotsinkielisillä oppilailla on sen sijaan jo suunnilleen sama arvo (0,110) kuin kaksikielisten neljäsluokkalaisten kaikkien paikkakuntien keskiarvoluvulla (0,093). Kaksikielisillä on luonnollisesti keskimäärin vaihtelevampi sanasto kuin vasta oppimisensa alussa olevilla ruotsinkielisillä L2-oppijoilla. 4. luokan ruotsinkielisten OVIX-arvot ovat kautta linjan joka tutkimuspaikkakunnalla poikkeuksellisen korkeita, erityisen korkeita Paraisilla. Tämän ilmiön on katsottava johtuvan lyhyistä teksteistä, kuten TTR-arvojenkin yhteydessä. Jo 6. ja 8. luokalla arvot normalisoituvat alle 0,0:n tasolle, kuten suomenkielisessä vertailuryhmässäkin. Esitän pari RK-oppilaan ainetta, joissa on korkea OVIX.

## Havainnointia

(7) Parainen 4. lk., tyttö KI 4,91. Hyvä. Se on ahuska ja jännittävä. Ja se on hödölisea. Hono: Hej opita miton sistä OVIX 0,286. 3 verbisanetta (N), 2 eri verbiä (V).

(8) Parainen 6. lk., tyttö KI 4,91. Minä en pidä televisiosta. Se on liian kallis ja huonoa ohjelmia. Televisio ottaa niin paljon tilaa. Minä kuulen mielummin radiota. Uutiset ja säät ovat ainoa ohjelmia mitä minä katson. Ihmisten silmät ovat huonoa kun he katsovat televisio. OVIX 0,130. 7 verbisanetta (N), 4 eri verbiä (V).

Kaksikielisten arvot ovat jo 4. luokalla suunnilleen samalla tasolla kuin ruotsinkielisillä oppilaille vasta 6. ja 8. luokalla. Ainoana poikkeuksena on Parainen, jonka 4. luokan kaksikielisillä on poikkeuksellisen korkeat arvot ja selvästi lyhyemmät ja heikommat aineet kuin muiden paikkakuntien kaksikielisillä. Kaksikielisten arvot ovat keskimäärin kaikilla paikkakunnilla ja luokkatasolla matalampia ja normaalimpia kuin ruotsinkielisten, ja heidän arvonsa lähentyvät enemmän suomenkielistä vertailuryhmää.



Kuvio 5. Ruotsinkielisten ja kaksikielisten oppilaiden OVIX-arvojen kehitys.

OVIX-arvokaan ei – kuten ei TTR:kään – sellaisenaan ole käyttökelpoinen L2-kirjoittajien sanaston vaihtelun mittarina. Yksilötasolla OVIX-arvo voi alemmilla luokilla olla aivan liian korkea todelliseen kirjoittamisen tasoon ja sanaston hallintaan nähden. Sillä saadaan kuitenkin esille monia mielenkiintoisia seikkoja, kuten esim. L2-oppijoiden suuri vaihtelu varsinkin alemmilla luokilla. Sen avulla voidaan myös tehdä päätelmiä siitä, missä vaiheessa L2-oppijan sanaston käyttö alkaa olla verrattavissa syntyperäisen käyttöön.

Leksikon rikkautta on tutkittu myös Guiraud'n indeksillä, ja sen käytöstä on saatu hyviä kokemuksia, varsinkin kun

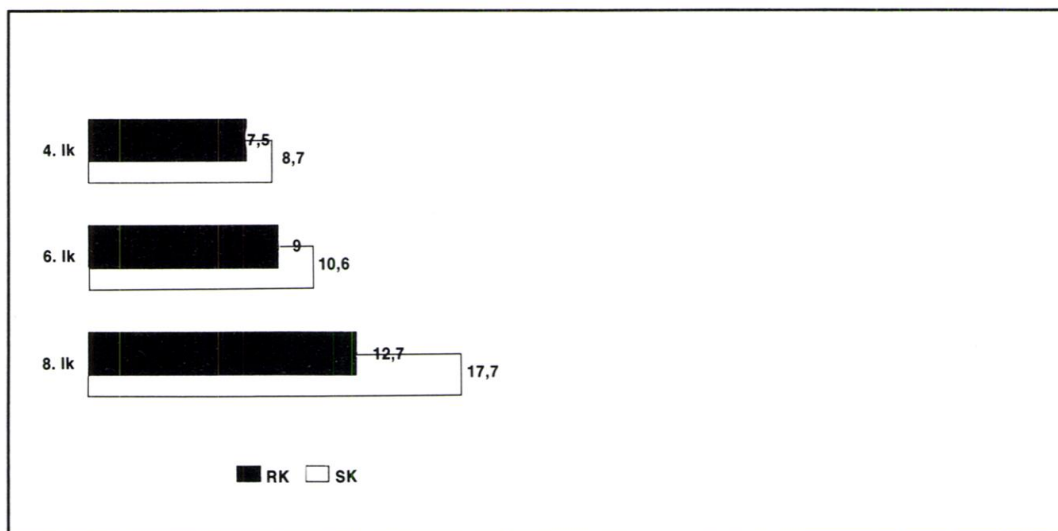
on seurattu sanaston kehittymistä ajallisesti (Voionmaa 1990: 230–231). Guiraud'n indeksi lasketaan kaavasta  $V: \sqrt{N}$  eli eri verbien määrä jaetaan eri verbisaneiden määrän neliöjuurella. Guiraud'n indeksiä käyttämällä saneiden kokonaismäärän vaikutus saatavaan arvoon heikkenee. Tämä on tärkeää, koska lyhyissä ainekirjoituksissa on suhteellisesti enemmän leksikkosanoja kuin pitkissä. Leksikkosanojen määrä ei koskaan kasva tekstin pidentyessä enää samassa suhteessa kuin saneiden määrä. Tätä kaavaa on menestyksekkäästi käytetty varsinkin Ruotsissa kaksikielisten oppilaiden sanavaraston tiheyttä tutkittaessa. Omasta aineistostani saadut luvut ovat taulukossa 6.

TAULUKKO 6. Sanaston rikastuminen Guiraud'n indeksillä mitattuna.

Guiraud'n indeksi	Porvoo	Vantaa	Parainen	KA	Salo
4	8.047	8.328	6.259	7.544	8.737
6	8.776	8.797	9.503	9.025	10.553
8	13.840	12.596	11.562	12.666	17.734

Kun saneiden määrästä otetaan neliöjuuri suhteessa eri sanoihin, saadaan jyrkästi nousevia arvoja, joita ei saatu edellä käsitellyillä mittareilla. Erot eri luokkatasojen välillä tulevat selvemmin esille, ja kielen rikastuminen on tällä mittarilla selvästi havaittavissa ja voimakkaampaa syntyperäisten ryhmässä, jossa arvot kas-

vavat peräti yhdeksän yksikön verran (toisen kielen oppijalla vain viiden yksikön verran). Kuvioista 6 näkyy, että toisen kielen oppijan sanaston rikkaus ei ala-asteella jää huomattavan paljon jälkeen syntyperäisten arvoista; vasta 8. luokalla erot ovat silmiinpistävämpiä.



Kuvio 6. Verbisanan vaihtelun kasvu RK- ja SK-oppilailla Guiraud'n indeksillä laskettuna.

Yläasteelle tullessa ero kasvaa ja syntyperäiset pystyvät kasvattamaan sanastoaan suhteellisesti enemmän. Tosin luokkakohtaisissa tuloksissa on muistettava, että vertailuryhmässä on eniten tyttöjä ja tyttöillä taas on huomattavasti enemmän leksikkosanoja kuin pojilla. Jos vertailuryhmässäkin olisi yhtä paljon poikia kuin tyttöjä, eivät erot toisen kie-

len oppijaan nähden olisi näin näkyviä. Sanaston rikastuminen on toisen kielen oppijalla omassa aineistossani parhaimmillaan suunnilleen samalla tasolla kuin Voionmaan (1990: 230–231) tutkimuksessa, jossa (tosin) aikuisten kielenoppijoiden paras kehitys oli noin neljä ja puoli yksikköä (6,04 → 9,48). Ruotsinkielisten koululaisten paljon parempi kielenkehi-

## Havaintoja

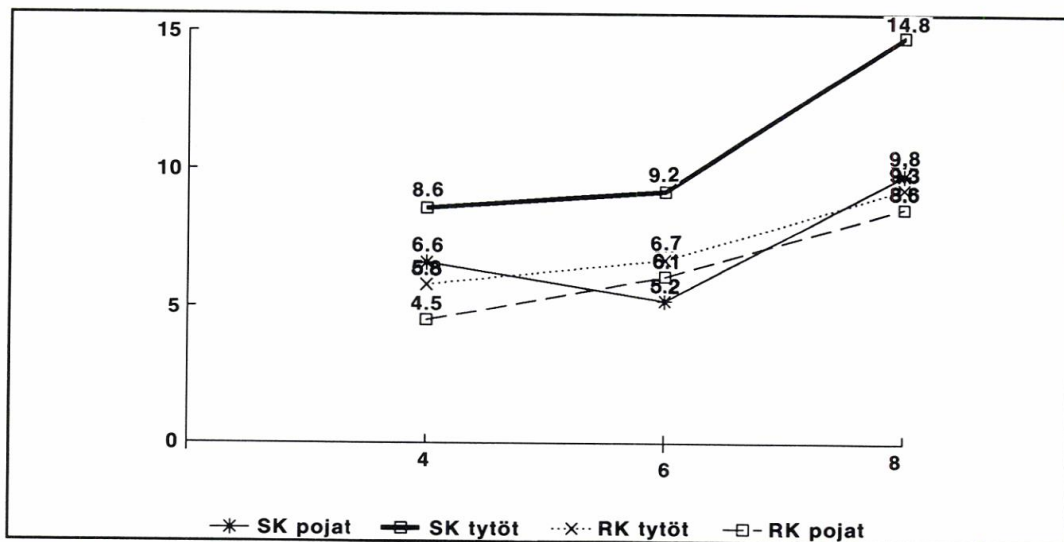
tyksen lopputulos 8. luokalla, 12,66, on odotuksenmukaistakin, koska kielenoppiminen on jo jatkunut vuosia. Voionmaan tutkimuksessa seurattiin toisen kielen oppijan sanaston rikastumista vain muutaman kuukauden. Kuviosta 7 käy ilmi, miten Guiraud'n indeksi eli sanan keskitiheyden neliöjuuri kasvaa otoksen

koon kasvaessa. Ruotsinkielisillä sanaston rikastuminen on hyvin tasaista koko peruskoulun aikana, mutta syntyperäisillä kasvu on jyrkempää otoksen kasvaessa.

Taulukossa 7 ovat erikseen tyttöjen ja poikien arvot Guiraud'n indeksillä laskettuna.

TAULUKKO 7. Tyttöjen ja poikien sanaston rikastuminen Guiraud'n indeksillä mitattuna.

Guiraud'n indeksi					
Tytöt	Porvoo	Vantaa	Parainen	KA	Salo
4	7.523	5.563	4.395	5.827	8.587
6	6.920	5.888	7.202	6.670	9.179
8	10.501	8.969	8.475	9.315	14.814
Pojat					
	Porvoo	Vantaa	Parainen	KA	Salo
4	3.909	6.221	4.465	4.465	6.609
6	5.471	6.535	6.200	6.068	5.223
8	9.016	8.844	7.858	8.572	9.766



Kuvio 7. Verbisanan vaihtelun kasvu tytöillä ja pojilla Guiraud'n indeksillä laskettuna.

Keskiarvotulosten perusteella RK-tyttöillä on vain vähän korkeammat arvot kuin RK-pojilla. Ero on mitätön 6. luokalla,

kuten monien muidenkin mittausten yhteydessä on havaittu. Syntyperäisten ryhmässä sen sijaan SK-tyttöjen ja SK-

poikien arvojen ero on odotuksenmukaisesti suuri. Kuviosta 7 näkyy selvästi kuudesluokkalaisten SK-poikien taantuma, jolloin ruotsinkielisten kasvukäyrät kulkevat tilapäisesti SK-poikien yläpuolella. Toisen kielen oppijoiden tulokset eivät muutoin yleensä yllä syntyperäisten tuloksiin.

Guiraud'n indeksi erottaa toisen kielen oppijan ja syntyperäisen huomattavasti selvemmin kuin muut käytetyt mittarit ja antaa tulokseksi kasvavia arvoja. Tästä voitaisiin päätellä, että neliöjuurinen tyyppi/esiintymä -suhde on tämääntyyppisistä suhdelukummittareista kaikkein käytökelpoisin mitattaessa toisen kielen oppijan sanaston variaatiota suhteessa syntyperäisiin. Huono puoli on kuitenkin se, että arvot nousevat otoksen mukaan varsin jyrkästi, jolloin hyvin erikokoisia otoksia ei voi täysin verrata toisiinsa. Itse asiassa jyrkästi nousevien arvojen kasvulinjat eivät juuri poikkea kasvulinjoista, joita on saatu tulokseksi pelkkää sanemäärien kasvua tai eri lekseemien määrän kasvua mitattaessa. On siis varsin vaikeaa osoittaa sellaista mittaria, joka olennaisesti objektiivisemmin tai tarkemmin kuin absoluuttisista luvuista saatava informaatio paljastaisi leksikaalisen rikkauden ja variaation.

MAIJA GRÖNHOLM

## LÄHTEET

- BROEDER, P. — EXTRA, G. — VAN HOUT, R. 1989: Processes in the developing lexicon of adult immigrant learners. *Nation*, P. — Carter, r. (toim.), *Vocabulary acquisition*. AILA Review-*Revue de L'AILA* 6 (1989), s. 86—109. Free University Press, Amsterdam.
- CHRYSAL, D. — JOHNSON, K. (toim.) 1988. *Analyzing Children Language*. T. J. Press Padstow Ltd., Cornwall.
- GRÖNHOLM, M. 1991: Ruotsinkielisten oppilaiden sanaston kehityksestä suomen kielessä. *Finlance. The Finnish Journal of Language Learning and Language Teaching*. Volume IX, s. 17—32. Korkeakoulujen kielikeskus, Jyväskylän yliopisto, Jyväskylä.
- GRÖNHOLM, M. 1992: Tavalliset ja harvinaiset sanat suomea kohdekielenä oppivan leksikossa. Tella, S. (toim.), *Joustava ja laaja-alainen opettaja*. Ainedidaktiikan symposium Helsingissä 7.2.1992. Helsingin yliopiston opettajankoulutuslaitos. *Tutkimuksia* 100, s. 129—142. Yliopistopaino, Helsinki.
- GRÖNHOLM, M. (työn alla): Yksikielisten ja kaksikielisten oppilaiden suomen kielen taitojen kehittyminen ruotsinkielisessä peruskoulussa. *Käsikirjoitus*.
- HULTMAN, T. — WESTMAN, M. 1977: *Gymnasistsvenska*. Liber Lärmedel, Lund.
- JOSEPHSON, O. — MELIN, L. — OLIV, T. 1990: *Elextext. Analyser av skoluppsatser från åk 1 till åk 9*. Studentlitteratur, Lund.
- KARVONEN, J. — TAKALA, A. — RÖMAN, K. — YLIENTALO, O. 1970: *Opettajan sanastokirja*. Gummerus, Jyväskylä.
- KORKMAN, Ch. 1990: *Tvåspråkiga elever skriver uppsats*. Licentiatavhandling i nordiska språk. Helsingfors universitet.
- LESKINEN, H. — SAVIJÄRVI, I. — SÄRKÄ, T. 1974: *Koululaisten kirjallisen ilmaisutaidon kehityksestä*. Jyväskylän yliopiston suomen kielen laitoksen julkaisuja 8.
- LINNARUD, M. 1987: *Lexikal täthet och betydelsesamband*. Hyltenstam, K. (utg.), *Svenska i invandrarperspektiv*, s. 219—230. Liber Förlag, Lund.
- MUIKKU, P. 1982: *Kritiikin kieli — havainnoja eräiden kirjallisuusarvostelujen tyylistä ja sisällöstä*. Suomen kielen lisensiaatintyö. Jyväskylän yliopisto, Jyväskylä.
- NIEMIKORPI, A. 1991: *Suomen kielen sanaston dynamiikkaa*. *Acta Wasaensia* No 26. *Kielitiede* 2. Vaasan yliopisto, Vaasa.

## Havaintoja

- PÄÄKKÖNEN, I. 1973: Oppikouluun pyrkineiden sanavarat. Lisensiaatintyö. Oulun yliopisto.
- SAARELA, L. 1991: Neljäsluokkalaisten tyttöjen ja poikien sanaston eroista. Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja 33. Oulun yliopisto, Monistus- ja kuvakeskus, Oulu.
- SAUKKONEN, P. — HAIPUS, M. — NIEMIKORPI, A. — SULKALA, H. 1979: Suomen kielen taajuussanasto. WSOY, Porvoo.
- SJÖDOFF, I.-L. 1989: Med svenska som mål. Effekter av två undervisningsprogram på invandrarelevets svenska i skrift. Skrifter utgivna av institutionen för nordiska språk vid Uppsala universitet 2. Graphic Systems, Stockholm.
- SÄRKILAHTE, S.-L. 1977: Tyyliintutkimuksen kvantitatiiviset metodit. Suomen sovelletun kielitieteen yhdistyksen julkaisuja N:o 19. Turun yliopiston offsetpaino, Turku.
- VIBERG, Å. 1989: Vägen till ett nytt språk 1. Andraspråksinlärning i ett utvecklingsperspektiv. Natur och Kultur, Stockholm.
- VIBERG, Å. 1991: Utvärdering av Skolförberelsegrupper i Rinkeby. En longitudinell djupstudie av språkutvecklingen. Rapport 4. Stockholms universitet. Centrum för tvåspråkighetsforskning. International Tryck AB, Uppsala.
- VOIONMAA, K. 1990: Aikuiset sanaston oppijoina. Tommola, J. (toim.), Vieraan kielen ymmärtäminen ja tuottaminen, AFinLan vuosikirja 1990, s. 227–256. Turun yliopiston offsetpaino, Turku.