

Kvantitatiivista kirjallisuudentutkimusta: miten ja miksi?

Hannu Launonen

Kvantitatiiviset menetelmät tekevät tuloaan henkítieteiden puolelle, jopa kirjallisuudentutkimukseen. Tästä vakuuttuakseen ei tarvitse kuin silmäillä tuoreimpia The Timesin kirjallisen liitteen vuosikertoja¹ tai muutamia artikkelikokoelmia, joiden nimet usein kalskahtavat manifestilta (The Computer & Literary Style, Computers in Humanistic Research, Litteraturvetenskap, Mathematik und Dichtung jne.).

Väittely menetelmien sopivuudesta on kesken. Kirjallisuudentutkijat ovat esittäneet perinnäisiä humanistien epäilyjä, jotka usein pohjautuvat väärinkäsityksiin («matematiikalla ja runoudella ei ole mitään tekemistä toistensa kanssa»). On pantu merkille myös harkitumpia muistutuksia. Kuten Anders Ternström on huomauttanut (Dagens Nyheterissä v. 1966), saattaa humanisti täydellä syyllä harata vastaan, jos varsinainen teos uhkaa hukkua numeroihin ja käyriin — jos Shakespearen soneteista saadaan analyysin jälkeen logaritmitaulun näköisiä, niin kuin on käynyt eräissä Roman Jakobsonin tutkielmissa.

Etenkin Suomen oloissa voidaan väittää, että kirjailijan »kielen ja sille luonteenomaisten yksilöllisten poikkeamien numeraalinen tarkkailu lienee mielekäästä vasta, kun normaaliarvoista vallitsee jonkinlainen yksimielinen tietämys» (Annamari Sarajas, Uusi Suomi 1968).² Vaikka emme toistaiseksi pystykään viittaamaan yleiskieltä koskeviin normaaliarvoihin, puute voidaan korvata: täytyy turvautua kyllin laajaan aineistoon ja vertailla useita kirjailijoita keskenään.

On myös epäilty, että statistiikan soveltajat ilman muuta luisuisivat paha-mainaiseen positivismiin ja saisivat aikaan »tarkkuutta pelkän tarkkuuden vuoksi». Jan Thavenius, yksi Litteraturvetenskap-teoksen laatijoista, näyt-

¹ Esim. v. 1971 julkaisi The Times Literary Supplement laajan, 13-osaisen artikkelisarjan, jossa käsiteltiin kvantitatiivisten menetelmien tunkeutumista useille aloille; humanistisista tieteistä olivat edusteilla mm. kielitiede ja kirjallisuudentutkimus (ks. erityisesti D. R. Tallentiren kirjoitusta The mathematics of style, joka ilmestyi 13. 8. 1971).

² Toistaiseksi ei ole tiedossa, missä määrin prof. Pauli Saukkosen johdolla Oulussa valmistuva (nykysuomen) kielistatistiikka tulee auttamaan kirjallisuudentutkijoita.

täisi antavan asheet vastustajan käteen: hän arvelee, että statistiikka parhaiten täydentäisi (nykyään kritiikin kohteeksi joutuneita) uskriittisiä menetelmiä. Näin varmasti onkin, jos ajatellaan vain ns. makrostilistiikkaa (runojen tai teosten keskinäistä vertailua). Mutta kvantifiointi tulee käsittää myös laajemmin: teoskohtainen ja yhteiskunnallinen tarkastelu voitaisiin kokeeksi yhdistää ja tutkia esim. kysymystä siitä, miten kirjailijan tausta, lähtökohdat yms. korreloivat tuotannon esteettisten ominaisuuksien kanssa.

Statistiikka on apuväline. Se saadaan toimimaan vain muiden menetelmien kanssa. Koko ajan on kysymys paitsi siitä, mitä tilastoidaan, myös siitä, miten tilastoja tulkitaan. Menetelmää on myös sovellettava varovasti: sillä on heikkoutensa.

Tärkeimmät kvantitatiivista tutkimustapaa vastaan esitetyt huomautukset voidaan ehkä tiivistää seuraavaksi. Ensiksi saatetaan sanoa, että menetelmä on matemaattisesta pohjastaan huolimatta liian karkea taiteellisen tyylin tutkimiseen. Toiseksi: liiallinen tarkkuus koituu vaaraksi, jos aineisto on kovin laaja ja monitahoinen. Kolmanneksi ehkä lisätään, ettei menetelmä tee oikeutta kontekstin vaikutukselle. Neljänneksi: tutkija tulee herkästi vieneeksi samaan kategoriaan aineksia, jotka vain pintapuolisesti muistuttavat toisiaan. Ja viidenneksi kenties painavin muistutus: kvantitatiivinen tutkimus vaatii paljon vaivaa (ja usein myös varoja), vaikka suurin piirtein samoihin tuloksiin päättäisiin myös muita teitä. (Ks. esim. Ullman, *Language and Style* s. 118—119; Thavenius, *Litteraturvetenskap* s. 44—46; Särkilähti, *Oppikoululehti* 28. 5. 1968.)

Monet seikat puhuvat kuitenkin kvantitatiivisen metodin puolesta. Se on usein korvaamaton tutkijan preliminääritoimissa. Kvantifiointin avulla voidaan tunnistaa anonyymeja tai kiistanalaisia tekstejä («the question of disputed authorship»), selvittää tuotannon kronologiaa ja arvioida teoksen eri osien yhtenäisyyttä; tulokset pystytään esittämään täsmällisessä muodossa, ja epävarmuuden aste voidaan yleensä mitata. Edelleen: niin kuin esim. Jan Thavenius on tähdentänyt (*Litteraturvetenskap* s. 39), kirjallisuuden materiaali, kieli, soveltuu hyvin tilastolliseen tarkasteluun. Kieli koostuu helposti tunnistettavista (ja laskettavista) elementeistä, joiden frekvenssit pysyvät riittävän suurissa aineistoissa likimain muuttumattomina. Lainalaista ei ole ainoastaan se, mitä kielen elementtiä (foneettista, leksikaalista tai syntaktista) käytetään, vaan myös se, miten usein sitä käytetään. Jo pelkät frekvenssitiedot voivat antaa tutkijalle tärkeitä johtolankoja.

Tähänastiset tulokset eivät ehkä tyydytä kirjallisuudentutkijoita. Suurin syy lienee se, että kvantitatiivista tyyliintutkimusta ovat toistaiseksi harrastaneet lähinnä lingvistit. Lääke olisi siis yksinkertainen: myös varsinaiset kirjallisuudentutkijat täytyy saada mukaan.

Seuraava katsaus ei missään suhteessa pyri täydellisyYTEEN. Koetan muuta-

mien summittaisten esimerkkien avulla antaa kuvan kvantitatiivisen kirjallisuudentutkimuksen päälinoista; esittelen joitakin tutkimuksia myös siksi, että ne ovat metodisesti tai muuten tuntuneet erityisen kiinnostavilta.

1.

Anonyymin tekijän tunnistamista ja kiistanalaisten teosten kronologiaa koskeissa kysymyksissä pääsee kvantitatiivinen menetelmä oikeuksiinsa: sen avulla on ratkaistu monia ongelmia, joista tutkijat eivät ole tavanomaisten historiaan ja tyyliin perustuvien päätelmien nojalla saaneet selvyttä. Klassinen esimerkki on *De Imitatione Christi* -teoksen vieminen Tuomas Kempiläisen nimiin; pioneerityötä mm. tämän ongelman kimpussa on tehnyt intialainen G. U. Yule (tutkimuksessaan *The Statistical Study of Literary Vocabulary* 1944; ks. myös Thavenius, *Litteraturvetenskap* s. 38).³ Identifiointimetodia on menestyksekkäästi sovellettu englantilaisella kielialueella, tiettävästi ei vielä kuitenkaan Shakespearen-tutkimuksessa.⁴

Identifiointikysymyksissä lähdetään siitä oletuksesta, että tietyt kirjailijan tyylipiirteet säilyvät läpi tuotannon ennallaan (tai muuttuvat ennustettavissa olevalla tavalla). Analyysia varten valitaan riittävästi muuttujia, jotka sitten toimivat diskriminaattoreina: otaksutaan, että niiden avulla tekijä voidaan erottaa muiden aikalaiskirjailijain joukosta. Tekstinäytteet täytyy valita kyllin laajat, niin että tyylin yksilölliset ominaisuudet pääsevät varmasti esiin. Gustav Herdan arvostelee Yulen *Imitatio*-tutkimusta siitä, että analyysiin on kelpuutettu liian värittömiä sanoja (»function words»), jotka eivät välttämättä ole relevantteja kirjailijan tyylin kannalta; hänen mielestään tarvittaisiin myös muunlaista, keskeisiin sanoihin (»content words») kohdistuvaa tutkimusta (*The Advanced Theory of Language as Choice and Chance* s. 172).

Alvar Ellegård puolestaan päätyi hieman erilaisiin tuloksiin kuin Herdan. Kun hän tutki tilastollisesti 1700-luvun englantilaisia poliittisia kiistakirjoituksia (»Junius Letters»), hän ei hyväksynyt diskriminaattoreiksi voimakkaita retorisia kuvioita. Kävi ilmi, että sellaiset tyylikeinot kuin parallellismi tai

³ G. U. Yule on ottanut tarkasteltavakseen 88 sanaa, jotka esiintyvät vähintään 20 kertaa *Imitatio*-teoksessa. Hän on selvittänyt niiden frekvenssin paitsi Kempiläisen muissa teoksissa myös (*Imitatio* oletetun toisen tekijän) Gersonin tuotannossa. Tulos: 2/3 tutkituista sanoista esiintyy vähintään 20 kertaa Kempiläisen muussa tuotannossa, mutta 1/2 niistä tavataan Gersonin tekstissä. (Ks. myös Herdan, *Language as Choice and Chance* s. 59—61 ja *The Advanced Theory of Language as Choice and Chance* s. 169 ja 172.)

⁴ Bernard O'Donnell on viitannut siihen kiusalliseen tulevaisuudennäkymään, joka saattaa odottaa kaavoihin piintyneitä Shakespearen-tutkijoita: tietokone voi analyysin jälkeen hyvinkin tavata Shakespearen nimen seuraavasti: M-A-R-L-O-W-E (*The Computer & Literary Style* s. 107). — Shakespearen näytelmien kronologiaan kohdistuvista statistisista tutkimuksista ks. esim. Herdan, *The Calculus of Linguistic Observations* s. 79.

antiteesi kuuluivat aikakauden kuosiin; ne eivät toimineet erottavina tekijöinä. Tuntui paremminkin siltä, että diskriminaattoreiden täytyi olla jollakin tavoin huomaamattomia, kenties alitajuisia kielellisiä tottumuksia («linguistic habits»). Näitä olisivat esim. lauseiden pituus (oik. lauseiden pituuden vaihtelu) ja kirjailijan sanaston ns. keskittyneisyyden aste, joita G. U. Yule on tarjonnut kriteereiksi (ks. Ellegård, *A Statistical Method for Determining Authorship* s. 10). Ellegårdin mielestä nämäkään eivät riitä. Hän nimittää omaa tutkimustapaansa autenttisuustestiksi («authenticity test»), ja hän operoi lähinnä Pierre Guiraudin kehittämällä »tyyppisanojen» metodilla. Ellegård on nähty työssään paljon vaivaa: tutkittavaa tekstiä (150 000 sanaa) on sekä sanaston että fraseologian ja kieliopillisten rakenteiden puolesta verrattu oletetun tekijän Sir Philip Francisin ja 59 muun aikalaiskirjailijan teksteihin. Menetelmän pätevyys tulee hyvin todistetuksi. (Ellegårdin metodia pohtii Gustav Herdan teoksessaan *Quantitative Linguistics* s. 149—150.)

Pitkälle meneviä tilastollisia menetelmiä (mm. faktorianalyysia) on sovellettu tutkimuksissa, jotka ovat kohdistuneet esim. Paavalin kirjeisiin (H. H. Somers, *The Computer & Literary Style* s. 128—140), eräisiin Uuden testamentin teksteihin (John W. Ellison, *Computers in Humanistic Research* s. 160—169) tai Herodotoksen, Platonin ja Thukydeiden teoksiin (Andrew Q. Morton ja Michael Levison, *The Computer & Literary Style* s. 141—179).

Bernard O'Donnell on tarkastellut Stephen Cranen kiistanalaisen postuumin teoksen (*The O'Ruddy*) tekijyyttä 18 variaabelin avulla (mm. verbien, adjektiivien ja adverbien määrä, verbien ja adjektiivien määrän suhde, partiippien ja infinitiivien osuus, metaforat) ja päätynyt dramaattiseen tulokseen: romaanin 24 ensimmäistä lukua ovat kirjailijan omaa työtä, 25. luku edustaa siirtymävaihetta, mutta lukujen 26—33 täytyy olla teoksen toimittajien käsialaa (*The Computer & Literary Style* s. 107—115). Ivor S. Francis näyttää turvautuneen normaaliin sanastodistributioon tutkimuksessaan, joka sivuaa amerikkalaista ns. *Federalist*-kiistaa (mts. 38—78).

Esimerkkinä siitä, millaisia pikku tarkennuksia tilastollinen identifikaatiotesti saattaa tuoda Suomen kirjallisuuden historian itsestäänselvyyksiin, voidaan mainita Pentti Leinon tutkielma *Kieli ja kaava* (Vir. 1969 s. 247—264). Leino on käyttänyt mm. alkusointufrekvenssejä apuna selvitellessään Winasta *Walitus Wirsi* -runon tekijyyttä. Tilastollinen analyysi osoittaa, että (toisin kuin näihin asti on luultu) runoa ei ole kirjoittanut Henrik Achrenius; kysymyksessä on »tuntemattoman» 1700-luvun runoilijan teksti.

2.

Sanastoa koskevia tilastollisia tyyllintutkimuksia on tehty runsaasti, ja monet perustuvat laajaan aineistoon. Esim. Josephine Miles lähtee englannin-

kielistä runoutta käsittelevässä teoksessaan (Eras & Modes in English Poetry) siitä, että kvantitatiivinen tarkastelu sopii erityisen hyvin kirjallisiin ryhmiin ja aikakausiin, ylipäänsä laajoihin vertailuihin.

Milesin otokseen sisältyy 150 englantilaista ja 50 amerikkalaista runoilijaa Chaucerista nykyaikaan (1400-luvun jälkipuoliskolta 1900-luvun alkuun). Runoilijat on valittu tasaisesti siten, että kultakin vuosisadalta on tutkimukseen kelpuutettu noin 30 tekijää. Jokaiselta tekijältä on tutkimuksen pohjaksi poimittu tuotannon 1 000 ensimmäistä riviä. Miles huomauttaa itekin tekstivalinnan ontuvuudesta: joskus 1 000 riviä kattaa kirjailijan koko tuotannon, joskus se on siitä vain murto-osa (mts. 249). Kultakin runoilijalta tuli tutkimuksen kohteeksi 7 000—8 000 sanaa: suurin piirtein puolet substantiiveja, adjektiiveja ja verbejä, puolet artikkeleita, pronomineja ynnä adverbereja.

Milesin päätulokset kiinnostanevat muitakin kuin Englannin kirjallisuuden tutkijoita. Tilastollisessa analyysissä seuloutui esiin kolme runouden rakenneytymätyyppiä. Ensimmäinen (»phrasal type») suosii nomineja, etenkin adjektiiveja, verbien partisiippimuotoja sekä lauseenvastikkeita; tämä on Milesin terminologiassa kumulatiivinen esitystapa (»cumulative way of speaking»). Toinen rakennetyyppi (»clausal type») välttää lauseenvastikkeita ja suosii alisteisia sivulauseita, käyttää verbejä partisiippimuotojen asemesta; kysymyksessä on loogisesti etenevä, diskursiivinen esitystapa (»discursive way of speaking»). Kolmanneksi Miles erottaa jonkinlaisen välityylin, jota hän nimittää tasapainoiseksi (»balanced»). Englantilaisessa runoudessa voidaan Milesin mukaan nähdä näiden tyylien aaltoliike suurin piirtein seuraavasti:

- 1) 1500-luvun verbityyli, joka vuosisadan lopulla muuttuu tasapainoiseksi;
- 2) 1600-luvun verbityyli (muuttuu tasapainoiseksi edellisen tapaan);
- 3) 1700-luvun adjektiivityyli (muuttuu tasapainoiseksi edellisen tapaan);
- 4) 1800-luvun verbityyli (muuttuu tasapainoiseksi edellisen tapaan).

Kullekin vuosisadalle hahmottuu oma tyylikuvionsa: jokainen alkaa äärimmäistytyllä ja päättyy tasapainoiseen. Kahta äärimmäistytyliä ei kertaakaan satu peräkkäin. Neljästä kuviosta on kolme täsmälleen samannäköistä; vain 1700-luku muodostaa poikkeuksen, sillä se alkaa adjektiivityylyllä.

Miles väittää, että tätä englantilaisen runouden miltei symmetristä aaltoliikettä vastaavat lisäksi erinäiset rakennepiirteet. Diskursiivityylyiset runot ovat aktiivisia, argumentoivia tai kertovia (ja yleensä säkeistömuodossa); hyvä esimerkki on John Donne. Kumulatiivityylyiset eivät asetu säkeistömuotoon eivätkä välttämättä ole riimillisiä. Nämä runot venähtävät usein pitkiksi (esim. Walt Whitman). 1500-luvun tiukasti säkeistömuotoinen runous muuttui vuosisadan lopulla silosäkeen ja parisäkeen aikakaudeksi. Seuraavalla vuosisadalla kehkeytyi metafyyssikkojen säkeistömuodosta uusklassikkojen parisäe. 1700-luku alkoi toisin: loppusuoinnista luovuttiin ja silosäkeen lisäksi sovellettiin epäsäännöllistä oodimuotoa. Vuosisadan lopulla tämä äärimmäistyli mie-

dontui Grabben ja Goldsmithin parisäkeiden ansiosta. 1800-luvun alussa palattiin verbityyliin, jonka ilmauksia oli mm. balladimuoto.

Verbaalityylin hallitsemiin kausiin liittyy Milesin mukaan käsitteellinen sanasto. Adjektiivityyli taas kytkeytyy onomatopoeettiseen ilmaisuun ja yleeseen («sublime») sanastoon. Vaikka sanaston yleissävy vaihtuu tyylikausien mukana, näyttäisi muutamia keskeisiä sanoja säilyvän runoissa läpi vuosisatojen, metrumin tai lauserakenteen muutoksista huolimatta. Tällaisia yleisinhimillisiä sanoja ovat Milesin aineistossa mm. *god* ('jumala'), *heart* ('sydän'), *life* ('elämä'), *love* ('rakkaus'), *good* ('hyvä'), *great* ('suuri') ja *old* ('vanha').

Kaikkiaan Milesin tutkimus osoittaa, että tuloksiin voidaan päästä jo veraten yksinkertaisen kvantifioinnin avulla. Hienosyiset analyysit saattaisivat tietenkin antaa enemmän kuin pelkät frekvenssilaskelmat. Ankarinta arvostelua voidaan kohdistaa Milesin otosmenettelyyn. Vaikka päätulokset pitänevät paikkansa, löytyy yksityiskohdista lieviä vääristymiä.⁵

Milesin tutkimukseen rinnastuu ainakin yhdessä suhteessa Peter Hallbergin teos *Natursymboler i svensk lyrik från nyromantiken till Karlfeldt I—II*. Sekään ei hätkähdytä metodin hienouksilla, mutta tuloksia kertyy runsaasti. Paljotöiset taulukot ja huolelliset vertailut herättävät kunnioitusta. Hallberg on pyrkinyt selvittämään, miten ruotsalaiset runoilijat ovat eri aikoina käytelleet väri-, ääni-, tuoksu- ja tunnevaikutelmia. Kysymys on eräänlaisesta suosikkisanojen (kuten suosikkivärien) noteeraamisesta; tyylintutkimus teoksesta tulee sikäli, että kirjailijoita vertaillaan keskenään ja ilmiöt kytketään aikakauden yleisiin kirjallisiin virtauksiin.⁶

Metodisesti kiintoisa on muuan Gustav Herdanin koe. Hän on logaritminen menetelmällä tarkastellut vierasperäisten sanojen ja tekstin pituuden suhdetta Puškinin runo- ja proosanäytteissä (*Mathematik und Dichtung* s. 85—106). Josephine Miles on yhdessä Hanan C. Selvinin kanssa kehittänyt edelleen sitä menetelmää, jota hän sovelsi teoksessaan *Eras & Modes in English Poetry*, ja tutkinut faktorianalyysin avulla 30 englantilaisen 1600-luvun runoilijan sanaston ominaisuuksia (*The Computer & Literary Style* s. 116—127).

Kun Pierre Guiraud tutki Ranskan symbolistien sanavaroja (teoksessa *Les caractères statistique du vocabulaire*), hän pyrki suurempaan yleispätevyteen kuin useimmat muut kirjallisuuden sanaston tutkijat. Guiraudin päätulokset

⁵ Esim. Ralph Maud on kritisoinut otantamenettelyä, jota Miles on käyttänyt Dylan Thomasin teosten tutkimisessa (*Entrances to Dylan Thomas' Poetry* s. 164). 1 000 ensimmäistä riviä eivät anna oikeaa kuvaa runoilijan koko tuotannosta. Maud on tehnyt omia laskelmia ja havainnut, että tilasto muuttuu heti, kun otetaan huomioon Thomasin myöhäistuotanto.

⁶ Hieman samaan tapaan kuin Hallberg on Annikki Yrjänäinen menetellyt opintotutkimuksessaan, joka tosin koskee vain yhtä suomalaista runoilijaa (L. Onervan lyriikan kuva-ainesten tarkastelua). — Pekka Tarkka on soveltanut mm. Hallbergin ja Jan Thaveniuksen esittämiä malleja Lehtos-tutkimukseen (Joel Lehtosen kaksi Matalleena — hyödyllinen mikrotilistiikan kohde. — Suomenkieliset Tieto-Sanommat 1967 s. 21—23).

alkavat jo olla yleisesti tunnettuja.⁷ Tuoreimpia kvantitatiivisia tyyliintutkimuksia ranskalaiselta kielialueelta on Jean Cohenin teos *Structure du langage poétique*. Se keskittyy lähinnä Ranskan runouden kolmeen tyyliisuuntaan, klassismiin, romantiikkaan ja symbolismiin. Teosta sävyttää strukturalistinen ajattelu nimeä myöten, ja se sisältää eräänlaisen runouden teorian ohella runsaasti tilastotietoja mm. epiteeteistä ja riimisisanoista.

3.

Kvantitatiivisten menetelmien avulla on toistaiseksi tutkittu vain harvoja suomalaisia kirjailijoita. Muutamat opintotutkielmat sisältävät kiinnostavia kokeiluja; ne tosin keskittyvät yleensä vain yhteen kirjailijaan kerrallaan.

Jalo Heikkinen on ottanut tilastollisen tutkielman aiheeksi Veijo Meren novellit ja romaanit, joita sitten vertailee Heinrich Böllin teoksiin (Veijo Meri ja Heinrich Böll — rinnastusta sekä sotanovellien ja -romaanien vertailua. Eräitten uusien metodien kokeilua). Hänen tähtäimessään ovat lähinnä sanaluokkien väliset suhteet, osittain myös kielikuvien käyttö. Sanastovertailuilta putoaa osittain pohja pois, kysymys kun on kahdesta erikielisestä kirjailijasta (esim. rinnastettaessa Meren ja Böllin partikkeleiden tai pronominiin käyttöä). Huomiota herättää joka tapauksessa Meren runsaalta tuntuva kielikuvien osuus. Tuloksia on pyritty testaamaan khin neliö -testin avulla. — Timo Kukkola esittää pro gradu -tutkielmassaan tilastotietoja F. E. Sillanpään tuotannosta. Hän rajoittuu tekstin yhteen piirteeseen, tunnetta ilmaiseviin sanoihin (substantiiveihin ja adjektiiveihin). Tutkielman nimi on lupaa-va (Eräs tilastollinen metodi ja sen käyttömahdollisuudet kirjallisuudentutkimuksessa). Käytännössä metodi tarkoittaa frekvenssilaskelmia ja tulosten vertaamista niihin tietoihin, joita on esitetty Sillanpäästä koskevassa kirjallisuudessa. Tutkielmaan liittyy ekskurssi Ahon, Järnefeltin, Mannerkorven ja Hyryn tekstinäytteisiin. Sen päätulos: Hyryllä on tunnetta ilmaisevia sanoja selvästi vähiten, Aholla ja Järnefeltillä eniten. — Juha Vakkuri on laatinut yksityiskohtaisen esityksen Pertti Niemisen tuotannon sanastollisista ominaisuuksista (Sisällön analyysi Pertti Niemisen kokoelmista *Kivikausi*, *Uurnat*, *Päivät* kuin *nuolet* ja *Silmissä maailman maisemat*). Tutkielma antaa pohjan Niemiseen kohdistuvalle tekstikritiikille.

Itse olen tarkastellut attribuuttien käyttöä eräiden suomalaisten runoilijain teoksissa. Pyrin ennen kaikkea vertailemaan vuosina 1925 ja 1965 ilmes-

⁷ Osa Guiraudin tuloksista (v:lta 1954) oli itse asiassa entisten tietojen varmentamista (esim. se havainto, jonka G. K. Zipf oli tehnyt v. 1949: sanojen pituuden ja niiden esiintymisuseuden välillä vallitsee käänteinen suhde; ks. myös Osgood, *Style in Language* s. 298). Paljon siteerattu on mm. se Guiraudin väite, että hyvin pieni määrä sanoja kattaa pääosan kaikista teksteistä: 100 yleisintä sanaa 60 % ja 4 000 yleisintä sanaa peräti 97,5 % (esim. Thavenius, *Litteraturvetenskap* s. 49; Särkilähti, *Juhlakirja Kauko Kyyrön täyttävässä 60 vuotta* 24. 11. 1967 s. 259—260).

tyneitä runokokoelmia.⁸ Työni lähempi tarkastelu antaa kuvan yhdestä kvantitatiivisesta menetelmästä.

Runoilija	1.		2.		3.	
	Sanamäärä	Adj.-attr.		Adjektiivit		
		Abs.	%	Abs.	%	
1. Vaara (A)	780	73	9,4	70	9,0	
2. Vala	363	69	19,3	66	18,0	
3. Jylhä	528	34	6,6	12	2,3	
4. Lauri	693	115	16,6	77	11,0	
5. Pimiä	825	121	14,9	90	11,0	
6. Viljanen	690	112	16,3	105	15,0	
7. Hämeenkoski	1 980	98	5,0	55	2,8	
8. Kailas	4 130	337	8,2	400	9,7	
9. Kivimaa	2 690	369	13,8	288	11,0	
10. O. Manninen	7 850	690	8,8	656	8,4	
11. Mustapää	5 310	571	10,8	372	7,0	
12. Onerva	5 670	480	8,5	540	9,5	
13. Savulinna	406	33	8,6	42	10,0	
14. Törmänen	2 360	215	9,2	219	9,3	
15. Uotinen	11 000	749	6,9	700	6,3	
16. Vaara (B)	2 760	315	11,5	480	17,0	
17. Vuorela	4 240	389	9,2	339	8,0	
Absoluuttisten arvojen summa ja prosenttilukujen keskiarvo	52 275	4 770	9,2	4 511	9,7	

Taulukko I. Sanamäärä sekä adjektiivien ja adjektiiviattribuuttien absoluuttinen ja (sanamäärästä laskettu) prosenttinen osuus vuoden 1925 runokokoelmissa.

⁸ Seuraavat tiedot ovat suomen kielen pro gradu -tutkielmastani (Attribuutin ja pääsananjärjestys vuosien 1925 ja 1965 suomenkielisissä runokokoelmissa ja sen lingvivististä taustaa) ja yleisen kirjallisuustieteen laudaturtyöstäni (Lyriikan attribuutti-tyylistä. Esimerkkeinä 38 suomalaista runoilijaa). Keskityn tässä kirjallisuudentutkimuksen kannalta kiinnostavimpiin kohtiin ja jätän kokonaan pois työn lingvistisen pääosan, attribuutin ja pääsananjärjestystä koskevat tilastot.

Vuodelta 1925 sisältyy tarkasteluun 16 runoilijaa ja 17 tekstinäytettä (Elina Vaaralta on mukana kaksi näytettä; suluissa kunkin teoksen nimi, kuuden ensimmäisen kohdalla Hurmioituneet kasvat -antologiaan kuuluvan runosikermän nimi): 1. Elina Vaara (A) (Palmurannat), 2. Katri Vala (Oleanderisatu), 3. Yrjö Jylhä (Painajaisuunta), 4. Olavi Lauri [Olavi Paavolainen] (Rypäleterttu), 5. Ilmari Pimiä (Näkinenkä), 6. Lauri Viljanen (Tähtikeinu), 7. Toivo Hämeenkoski (Baabelin virroilla), 8. Uuno Kailas (Purjehtijat), 9. Arvi Kivimaa (Palava laiva), 10. Otto Manninen (Virrantyven), 11. P. Mustapää (Laulu ihanista silmistä), 12. L. Onerva (Maan tomuurna), 13. Savulinna [Bernhard Husgafvel] (Iltavarjot), 14. V. E. Törmänen (Tunturilauluja), 15. Mikko Uotinen (Valkoisten ristiä maa...), 16. Elina Vaara (B) (Satu sydäimestä ja auringosta), 17. Einari Vuorela (Varjoleikki).

Vuodelta 1965 kuuluvat tutkielmaan seuraavat 22 runoilijaa: 1. Raimo Hartzell (Joku soittaa pianoa), 2. Helvi Hämäläinen (Poltetut enkelit), 3. Pirkko Jaakola (Älä välitä Arnold), 4. Harri Kaasalainen (Nootti), 5. Väinö Kirstinä (Luonnollinen tanssi), 6. Laura Latvala (Toisillemme; yhteiskokoelma Olavi Siippaisen kanssa), 7. Mauno Manninen (Kaks silmää vain), 8. Aila Meriluoto (Tuoddaris), 9. Pekka Piirto (Niin monta kertaa Mozart), 10. Mirkka Rekola (Ilo ja epäsymmetria), 11. Matti Rossi (Näytelmä henkilöitä), 12. Yrjö Saarenpää (Kirvesvarrenveistäjän omakuva), 13. Pentti Saarikoski (Kuljen missä kuljen), 14. Pentti Saaritsa (Pakenevat merkit), 15. Tyyne Saastamoinen (Olen lähtenyt kauas), 16. Kalevi Seilonen (Tosiasioita minusta), 17. Olavi Siippainen (Toisillemme; yhteiskokoelma Laura Latvalan kanssa), 18. Lassi Sinkkonen (Harhaileva aamupäivä), 19. Pekka Suhonen (Kootut runot), 20. Lauri Tarkiainen (Tahdosta, viisaudesta, uuvuttavasta tunteesta), 21. Elina Vaara (C) (Lokikirja), 22. Pekka Virtanen (Aurinko silmissämme).

Runoilija	1.		2.		3.	
	Sanamäärä	Adj.-attr.		Adjektiivit		
		Abs.	%	Abs.	%	
1. Hartzell	2 070	174	8,5	124	6,0	
2. Hämäläinen	4 350	475	10,9	348	8,0	
3. Jaakola	2 140	195	9,1	102	5,2	
4. Kaasalainen	1 120	109	9,7	102	9,1	
5. Kirstinä	7 020	788	11,3	594	8,5	
6. Latvala	340	31	9,1	35	10,0	
7. M. Manninen	4 030	305	7,6	336	8,3	
8. Meriluoto	1 800	154	8,6	200	11,0	
9. Piirto	2 010	191	9,5	106	5,2	
10. Rekola	1 600	98	6,2	94	5,9	
11. Rossi	3 770	404	10,9	644	17,0	
12. Saarenpää	3 320	305	9,2	234	7,0	
13. Saarikoski	3 950	251	6,4	250	6,3	
14. Saaritsa	2 680	263	9,9	240	9,0	
15. Saastamoinen	1 520	194	13,1	138	9,1	
16. Seilonen	2 330	144	6,4	132	5,7	
17. Siippainen	3 010	203	7,2	350	12,0	
18. Sinkkonen	3 300	295	8,9	370	9,1	
19. Suhonen	4 070	192	4,7	259	6,4	
20. Tarkiainen	3 560	95	2,7	160	4,5	
21. Vaara (C)	7 500	950	12,7	1 008	13,0	
22. Virtanen	460	44	9,8	54	12,0	
Absoluuttisten arvojen summa ja prosenttilukujen keskiarvo	65 950	5 860	8,9	5 880	8,9	

Taulukko II. Sanamäärä sekä adjektiivien ja adjektiiviattribuuttien absoluuttinen ja (sanamäärästä laskettu) prosenttinen osuus vuoden 1965 runokokoelmissa.

Jotta attribuuttien vertailulle saataisiin taustaa, olen ensin arvioinut kunkin runoilijan käyttämien eri sanojen koko määrän; adjektiiviattribuuttien määrän olen laskenut tarkalleen. Jokaiselle runoilijalle olen erikseen mitannut adjektiiviattribuuttien prosenttisen osuuden koko sanamäärästä. Vertailun vuoksi olen arvioinut myös runoilijain käyttämien adjektiivien määrän. (Näiden prosenttinen osuus on laskettu kunkin runoilijan sanamäärästä.) Tulokset selviävät taulukoista I ja II.

Taulukoista näkyy, että runoilijain käyttämät sanamäärät vaihtelevat verrattain suuresti. Vuoden 1925 joukossa (taulukko I) sisältyy eniten sanoja Uotisen kokoelmaan (11 000 sanaa), vähiten Valan (363) ja Savulinnan (406); eroa on siis yli 10 000 sanaa. On selvää, että myös attribuuttien määrät tällöin ovat erilaiset. Kaikkien Hurmioituneet kasvot -antologiaan osallistuneiden sanamäärä, Vaaran, Valan, Jylhän, Laurin, Pimiän ja Viljasen, on alle 1 000.

Sanojen määrä on suurin piirtein verrannollinen adjektiiviattribuuttien absoluuttiseen määrään. Eniten sanoja käyttänyt Uotinen on viljellyt eniten myös adjektiiviattribuutteja (749 tapausta). Savulinna, Jylhä ja Vala, joiden sanamäärät jäivät pienimmiksi, käyttävät vastaavasti vähiten adjektiiviattri-

buutteja (33, 34 ja 69 tapausta). Aivan kaikkien suhteen ei tällainen verranto kuitenkaan päde. Hämeenkoski on käyttänyt miltei kolme kertaa enemmän sanoja kuin Viljanen ja yli kaksi kertaa enemmän kuin Pimiä (vastaavat arvot 1 910, 690 ja 825), mutta hänen kokoelmaansa sisältyy tuntuvasti vähemmän adjektiiviattribuutteja (98 tapausta) kuin Viljasen (112) ja Pimiän (121) teksteihin. Hämeenkosken adjektiiviattribuuttien suhteellinen osuus (5,0 %) onkin vuoden 1925 joukon pienin. Jylhän ja Valan välillä vallitsee samantapainen suhde kuin Hämeenkosken ja Pimiän. Jylhän sanamäärä (528) on selvästi suurempi kuin Valan (363), mutta hän on silti käyttänyt vähemmän adjektiiviattribuutteja (vastaavat arvot 34 ja 69).

Taulukosta II havaitaan, että vuoden 1965 runoilijoista ovat selvästi eniten sanoja käyttäneet Vaara (C) ja Kirstinä (7 500 ja 7 020 sanaa); heidän kokoelmiinsa sisältyy myös eniten adjektiiviattribuutteja (950 ja 788). Sanamäärältään pienin (340 sanaa) on Latvalan kokoelma, joka itse asiassa onkin vain osa Toisillemme-teosta. Latvala on niin ikään viljellyt absoluuttisesti vähiten adjektiiviattribuutteja (31 tapausta). Miltei yhtä suppea on Virtasen kokoelma, johon sisältyy vain 460 sanaa ja 44 adjektiiviattribuuttia.

Vuoden 1965 runoilijoista ovat Latvala ja Virtanen ainoat, joiden sanamäärä on alle 1 000. Vuoden 1925 joukossa esiintyy sen sijaan seitsemän yhtä vähän sanoja käyttänyttä: Savulinna lisäksi ne kuusi runoilijaa, jotka ovat osallistuneet Hurmioituneet kasvot -antologiaan. Tarkiaisen asema on vuoden 1965 joukossa jossakin määrin poikkeuksellinen. Hänen sanamääränsä (3 560) kohoo miltei yhtä suureksi kuin Saarenpään (3 320). Siitä huolimatta sisältyy Tarkiaisen kokoelmaan vain 95 adjektiiviattribuuttia, vajaa neljäsosa Rossin (404) ja vajaa kolmasosa Saarenpään (305) vastaavista arvoista. Miltei yhtä lailla herättävät huomiota Rekolan absoluuttiset arvot. Hänen sanamääränsä (1 600) on suurempi kuin Saastamoisen (1 520) ja lähes yhtä suuri kuin Meriluodon (1 800), mutta adjektiiviattribuutteja hän on käyttänyt vain 98 (Saastamoisen ja Meriluodon vastaavat arvot 194 ja 155).

Vuoden 1925 runoilijoille olen laskenut adjektiiviattribuuttien keskimääräistä käyttöä osoittavaksi prosenttiluvuksi 8,9. Yksityisistä runoilijoista Vala viljelee adjektiiviattribuutteja eniten; hänelle mitattu luku (19,3 %) on joukon suurin. Miltei yhtä korkea on kuitenkin myös Laurin arvo (16,6 %), seuraavina Viljasen (16,3 %) ja Pimiän (14,9 %) arvot. Yli 10 prosentin kohoavat lisäksi Kivimaan, Vaara (B):n ja Mustapään vastaavat luvut (13,8, 11,5 ja 10,8 %). Tulenkantajien antologiaan osallistuneet runoilijat eivät suinkaan käynteille adjektiiviattribuutteja yhdenmukaisesti: Valan arvo osoittautui joukon suurimmaksi, Jylhän prosenttiluku 6,6 taas edustaa miltei toista äärimmäisyyttä. Hämeenkoski käyntelee attribuutteja vieläkin niukemmin: hänelle laskettu arvo on 5,0 %.

Vuoden 1925 runoilijain attribuuttien käyttöä osoittava keskimääräinen prosenttiluku (8,9) käy yksiin sen luvun kanssa, joka ilmoittaa samojen runoilijain keskimääräistä adjektiivien käyttöä. Yhtäpitävyyttä voidaan havaita useimmilla yksityisillä runoilijoilla. Siten Vala, joka käyttää eniten adjektiiviattribuutteja, saa myös suurimman adjektiivien määrää osoittavan arvon (18,0 %); Jylhän ja Hämeenkosken prosenttiluvut (2,3 ja 2,8) jäävät tässäkin pienimmiksi.

Eniten herättää huomiota Vaara (B). Hän käyntelee adjektiiveja selvästi enemmän kuin adjektiiviattribuutteja (vastaavat luvut 17,0 ja 11,5 %). Vaara (B):n adjektiiviattribuuttien määrä tosin jää lähelle keskiarvoa, mutta

17,0 % on miltei joukon ääriarvo. Verraten suuri osa Vaara (B):n adjektiivista on siis lauseasemaltaan muita kuin attribuutteja. (On otettava huomioon, että runoilijain adjektiiviattribuutit eivät kaikki suinkaan ole varsinaisia adjektiiveja tai adjektiivisesti käytettyjä partisiippeja; osa on adjektiivisia lukusanoja ja pronomineja.)

Mustapään prosenttilukujen tendenssi on päinvastainen kuin Vaaran: attribuuttien prosenttinen määrä nousee suuremmaksi kuin adjektiivien (vastaavat arvot 10,8 ja 7,0 %). Samaa suuntaa osoittavat Jylhän (6,5 ja 2,3 %) ja Hämeenkosken (5,0 ja 2,8 %) selvästi pienemmät arvot.

Vuoden 1965 runoilijat käyttävät attribuutteja keskimäärin 8,9 %, siis likipitään saman verran kuin vuoden 1925 runoilijat. Adjektiivipitoisuus sen sijaan on vuoden 1925 kokoelmassa keskimäärin suurempi kuin vuoden 1965.

Yksityisten runoilijoiden attribuuttien määrissä ilmenee selviä vuosien välistä eroja. Vuoden 1965 runoilijoiden suurimmat prosenttiluvut eivät nouse yhtä korkeiksi ja pienimmätkin jäävät hieman pienemmiksi kuin 40 vuotta aiemmin julkaistuiden. Suurimmat luvut, Saastamoisen (13,1 %) ja Vaara (C):n (12,7 %), ovat huomattavasti pienemmät kuin vuoden 1925 korkeimmat arvot, Laurin (16,6 %) ja Viljasen (16,3 %). Saastamoisen ja Vaara (C):n arvojen lisäksi vain viidelle runoilijalle mitatut luvut nousevat yli 10 prosenttiin.

Tarkiaisen (2,7 %) ja Suhosen (4,7 %) arvot ovat vuoden 1965 joukossa hieman irrallisia ja jäävät selvästi pienimmiksi. Rossin suhteen kiintyy huomio siihen, että hän käyttää samaan tapaan kuin vuoden 1925 runoilijoista Vaara (B) attribuutteja prosenttisesti vähemmän kuin adjektiiveja (vastaavat arvot 10,9 ja 17,0 %); adjektiiveja esiintyy siis Rossin kokoelmassa verraten runsaasti muissa asemissa kuin attribuutteina. Sama tendenssi, joskaan ei yhtä selvänä, näkyy Meriluodon (8,6 ja 11,0 %), Siippaisen (7,2 ja 12,0 %), Suhosen (4,7 ja 6,4 %), Tarkiaisen (2,7 ja 4,5 %) ja Virtasen (9,8 ja 12,0 %) arvoista.

Täytyykö runoilijain käyttämiä adjektiivimääriä pitää yleisesti ottaen suurina vai pieninä? Kysymys on hankala, sillä edustavaa tilastoa ei meillä ole laadittu edes yleiskielestä (saati runoudesta).⁹ Jonkinlaisen vertailukohdan tarjoavat esim. Sirkka-Liisa Särkilahden laskelmat (Juhlakirja Kauko Kyyrön täyttäessä 60 vuotta 24. 11. 1967 s. 255). Ne tosin koskevat suomalaisia proosakirjailijoita. Aineistoon ovat kuuluneet mm. Eeva Joenpellon romaani Ralli, Marja-Liisa Vartion romaani Se on sitten kevät ja F. E. Sillanpään muis-

⁹ Suomalaisen kielistatistiikan miltei täydellistä olemattomuutta ilmentää esim. se, että niinkin huolellisesti dokumentoidussa teoksessa kuin Lauri Hakulisen Suomen kielen rakenne ja kehitys on täytynyt esittää tilastollisesti riittämättömään aineistoon perustuvaa tietoa. Hakulisen teoksen sisältyy mm. suomen omaperäisten ja lainasanojen prosenttijakauma, joka on laskettu vain 2 000 sanaa käsittävästä otoksesta (mts. 388); useimmista kielen ilmiöistä, esim. sijoista ja sanaluokista, ei ole esittää minkäänlaista tilastoa. — Saksalaisella ja englantilaisella kielialueella tilanne on toinen. Kun Hans Eggers työryhmineen tutki tilastollisesti lauseiden pituutta kirjallisuudessa, hän sovelsi 50 000 sanan otantaa (Meier, Deutsche Sprachstatistik s. 185 ja 199). Josephine Miles on käyttänyt tyyliintutkimuksensa pohjana noin 1 500 000 sanaa käsittävää otosta (ks. Eras & Modes in English Poetry s. 249).

telmateos Päivä korkeimmillaan; kaikista on poimittu 2 000 — 2 500 sanaa käsittävä näyte. Tutkimus osoitti, että kirjailijat ovat käyttelleet adjektiiveja seuraavasti:

	Absoluuttinen arvo	Prosenttiarvo
Vartio	102	6,0
Joenpelto	130	10,1
Sillanpää	249	16,8

Tulos vastaa suurin piirtein odotuksia. Vartio on yleensä pyrkinyt kirjoittamaan puhekielen mukaisesti, ja hänen tekstiinsä sisältyy prosenttisesti vähiten adjektiiveja. Impressionistisena pidetty Sillanpää puolestaan viljelee adjektiiveja eniten. Kun Särkilähti on lisäksi havainnut, että Vartio käyttää eniten verbejä (näytteessä 50,8 %), hän nimittää Vartion romaaniyyliä verbaaliseksi. Sillanpään tyyliä hän ei luonnehdi lähemmin, mutta Vartioon verrattuna sitä kai voisi kutsua adjektiiviseksi (verbejä vain 32,7 %).¹⁰

Vuoden 1925 runoilijoista rinnastuvat Sillanpään näytteeseen Vala ja Vaara (B), joiden vastaavat adjektiivien prosenttiarvot ovat vain vähän suuremmat (18,0 ja 17,0 %). Verraten lähelle tulee myös Viljaselle mitattu 15,0 %. Vuoden 1925 runoilijoille laskettu keskiarvo 9,7 % sijoittuu lähimmäs Joenpellon prosenttilukua (10,1). Kristillis-isänmaallinen runoilija Mikko Uotinen on kenties hieman yllättävästi käyttellyt adjektiiveja likimain samassa suhteessa kuin Vartio. Lähelle Vartion arvoa sijoittuu myös Mustapäälle mitattu 7,0 %. Jylhän ja Hämeenkosken arvot (2,3 ja 2,8 %) sen sijaan jäävät Vartionkin prosenttilukua selvästi pienemmiksi.

Vuoden 1965 runoilijoista vain Rossi käyttää adjektiiveja suurin piirtein samassa suhteessa kuin Sillanpää (Rossille mitattu arvo: 17,0 %). Useimpien tämän vuoden runoilijain arvot nousevat lähelle Joenpellon prosenttilukua: esim. Latvalan (10,0 %), Meriluodon (11,0 %) ja Sinkkosen (9,1 %); ei liioin keskiarvo (8,9 %) jää siitä kauas. Hartzellin (6,0 %), Saarikosken (6,3 %) ja Suhosen (6,4 %) kokoelmassa tavataan adjektiiveja likipitään samassa suhteessa kuin Vartion näytteessä.

Pikainen vertailu osoitti kaikkiaan, että molempien vuosien keskimääräiset prosenttiluvut ovat lähimpänä Joenpellon arvoa. (Joenpellon Ralli-teos ilmestyi v. 1959, ja hänen tähänastisen tuotantonsa pääosa sijoittuu 1950- ja 1960-luvulle.) — Sillanpäälle mitattua prosenttilukua muistuttavat useammat vuoden 1925 runoilijain arvot kuin vuoden 1965. Tämä tuntuu luonnolliselta. Vaikka Sillanpään muistelmateos ilmestyi vuonna 1956, hän on kirjailijana lähinnä 1925-vuoden runoilijain »aikalainen». — Marja-Liisa Vartio julkaisi romaaninsa (Se on sitten kevät) vuonna 1957; lähimmäksi hänen arvoaan

¹⁰ Särkilahden tutkimuksen luotettavuutta tosin vähentää toisaalta otoksen pienuus ja toisaalta se, että Sillanpäältä on valittu tuotannon periferiaan kuuluva muistelmateos.

jäävät odotuksenmukaisesti monet vuoden 1965 runoilijain prosenttiluvut. Vuodelta 1925 sen sijaan vain yksi runoilija käyntelee adjektiiveja samassa suhteessa kuin Vartio.

Sillanpäälle oli Särkilahden tutkimuksessa laskettu suurin adjektiivien määrää osoittava prosenttiluku. Vuoden 1925 runoilijoista sen ylittää kaksi, vuodelta 1965 vain yksi. Särkilahden mittauksessa sai Vartio puolestaan pienimmän arvon. Sen alittaa vuoden 1925 runoilijoista vain kaksi mutta vuoden 1965 runoilijoista viisi; vuoden 1965 joukkoon sisältyy lisäksi useita, joiden arvot vain hienokseltaan ylittävät Vartion prosenttiluvun. Nykyrunoilijoissa näyttäisi tämän mukaisesti olevan Vartion kaltaisia niukasti adjektiiveja käytteviä enemmän kuin vuoden 1925 joukossa.

Työn lingvistisen osuuden kulku oli lyhyesti seuraava. Perusaineistoon kuului noin 18 000 attribuuttitapausta. Laskin eri tyyppien (joista tärkeimmät olivat normaalitapaus *sininen varjo* ja inversiotyyppi *varjo sininen*) kaikki esiintymät ja taulukoin ne. Tässä vaiheessa ilmenneiden kahdeksan muuttujan perusteella laadittiin Yliopiston laskentakeskuksessa korrelaatiomatriisi. Korrelaatioiden valossa tarkastelin yksityisten runoilijain välisiä eroja ja yhtäläisyyksiä. Attribuuttien käyttelyn mukaan hahmottuvia runoilijaryhmiä sen sijaan seulottiin matriisiin perustuvan klusterianalyysin avulla. Analyysissa saatiin esille kolme selväpiirteistä ryvästä, joihin useimmat runoilijat liittyivät; lisäksi kävi ilmi, että muutamille lasketut arvot poikkesivat niin paljon kaikista muista, että ne jäivät rypäistä erilleen. Kiinnostavaa oli mm. se, että Laurin ja Mustapään arvot eivät liittyneet vuoden 1925 pääklusteriin vaan sijoittuivat nykyrunoilijain suurimpaan ryhmään. Toisaalta myös Latvala ja Meriluoto putosivat pois ilmestymisvuoden edellyttämästä klusterista. Vuoden 1965 klusteri osoittautui kaikkiaan kiinteämmäksi kuin vuoden 1925. Tämä tuntuisi kuvastelevan jonkinlaista nykyrunouden syntaksin yhdenmukaistumista.

Yleiskuvaa täydentääkseni otin lähempään tarkasteluun adjektiiviattribuuttien semanttiset ominaisuudet. Tässä vaiheessa turvauduin 20 prosentin otantaan. Attribuutteja koskevat tiedot lävistettiin reikäkortteihin ja käsiteltiin Yliopiston laskentakeskuksessa. Vuoden 1925 kokoelmista kertyi näytteeseen yhteensä 987 ja vuoden 1965 kokoelmista 1 212 attribuuttia. Reikäkorttitarkasteluun sisältyi siis kaikkiaan 2 199 attribuuttia.

Koekoodauksen jälkeen päädyin seuraavaan semanttiseen ryhmittelyyn (joka tietenkin on vain yksi monista mahdollisista):¹¹

¹¹ Esim. Stephen Ullmann on synestesiaa koskevissa tutkimuksissaan soveltanut kuusijakoa: 1. kosketusta, 2. kuumuutta, 3. makua, 4. hajua, 5. ääntä ja 6. näköä ilmoittavat aistimukset (The Principles of Semantics s. 277). B. Loman on puolestaan adjektiivikombinaatioita käsitellessään viitannut vain kolmeen semanttiseen kategoriaan: kokoa, kansallisuutta ja väriä ilmoittavaan (Adjektivkombinationer s. 80). Nämä jaotellut osoittautuivat yhdistettyinäkin liian ahtaiksi.

1. väriä ja valoilmiötä merkitsevät adjektiivit (myös sellaiset kuin *palava*, *heijastava*; samoin *pimeä*, *kirkas*, usein myös *kultainen*);
2. kuuloaistiin vetoavat (useimmiten partisiippeja, kuten *laulava*, *rämi-sevä*; myös *hiljainen*);
3. hajuaistiin vetoavat (useimmiten partisiippeja, esim. *tuoksuva*);
4. makuaistiin vetoavat (harvinaisia; lähinnä vain *imelä*, *karvas* ja *väkevä* selvästi makua tarkoittavassa yhteydessä);
5. tuntoaistiin vetoavat (*kylmä*, *kuuma*, *märkä* yms.);
6. mittaa, kokoa ja etäisyyttä ilmaisevat (ovat yleensä jollakin tavoin kvantifioitavissa: *tyhjä*, *täysi*; *pieni*, *suuri*; *läheinen*, *kaukainen*);
7. liikettä ilmaisevat (yleensä partisiippeja, esim. *juokseva*; myös *hidas* ja *nopea*);
8. kansallisuutta tai kieltä merkitsevät (myös *itämainen*, *länsimainen* yms.);
9. tunnetta ilmaisevat (kuten *julma*, *kammottava*; myös *ystävällinen* ja *onnellinen*, jotka voisivat yhtä hyvin kuulua 10. ryhmään);
10. arvottavat adjektiivit (*hyvä*, *paha*; *kaunis*, *ruma* yms.; sen sijaan sellaiset superlatiivimuodot kuin *pienin*, *nopein* jne. olen sijoittanut perusmerkityksen mukaan, en superlatiivisuuden, johon jo sinänsä liittyy arvottamista).

Lyriikan kieleen olennaisesti kuuluva moniselitteisyys aiheuttaa sen, että useita attribuutteja ei voi ilman muuta sijoittaa mihinkään ryhmään. Miten suhtautua esim. sellaiseen ilmaisuun kuin Lauri Viljasen *vihmanraskas yö*? On selvää, ettei merkitys siinä määräydy yksinomaisesti perusosan *raskas* mukaan. Lähinnä kai *vihmanraskas* täytyisi koodata tunnetta merkitseviin (joihin olen sijoittanut myös *raskas*-adjektiivin, mikäli sitä on selvästi käytetty metaforisesti). Kun olen sijoitellut synesteettisiä ilmauksia eri ryhmiin, olen joutunut käyttämään useampaa kuin yhtä koodia. Tyypillinen esimerkki on Katri Valan (hajuherneistä) käyttämä ilmaus *kuultava tuoksu*: attribuutille *kuultava* täytyy merkitä sekä näkö- että kuulovaikutelmaa ilmoittava koodi.

Reikäkorttitarkasteluun alistettujen attribuuttien jakaantuminen eri semanttisiin ryhmiin selviää taulukoista III ja IV.

Taulukoista käy ilmi, että vuoden 1925 runoilijat ovat keskimäärin käyttelleet eniten 10. ryhmään kuuluvia (arvottavia) attribuutteja (29,6 %). Vuoden 1965 runoilijoille laskettu vastaava arvo jää jonkin verran pienemmäksi (22,6 %). Nykyrunoilijat näyttäisivät mieltyneen mitan ja koon ilmaisemiseen: vuoden 1965 ryhmälle mitattu keskiarvo on tässä ryhmässä lähes 5 prosenttia korkeampi kuin vuoden 1925 arvo. Kolmas semanttinen ryhmä, jossa lasketut arvot keskimäärin kohoavat huomattavan suuriksi, on väriä tai valoilmiötä merkitsevien kategoria. Vuoden 1965 tekijät ovat myös värin runoilijoita hieman selvemmän kuin vuoden 1925 (vastaavat arvot 27,3 ja 23,9 %).

Muissa ryhmissä jäävät keskiarvot suhteellisen pieniksi: ne eivät kummaankaan vuoden kohdalla nouse yli 10 prosentin. Huomiota herättää kuitenkin se, että vuoden 1925 attribuutit sisältävät enemmän tunnetta, tuntoa ja kuulo-

Runoilija	1. Väri	2. Kuulo	3. Haju	4. Maku	5. Tunto	6. Koko	7. Liike	8. Kans.	9. Tunne	10. Arvott.
1. Vaara (A)	20,0					60,0				20,0
2. Vala	33,3	11,1			22,3	11,1		11,1	11,1	
3. Jylhä		20,0			20,0		20,0		40,0	
4. Lauri	25,0	33,3			16,8	8,3			8,3	8,3
5. Pimiä	26,7	6,6			13,3		26,7			26,7
6. Viljanen	60,0	10,0			20,0				10,0	
7. Hämeen- koski	16,7				8,3	25,0			8,3	41,7
8. Kailas	15,0	5,0			5,0	20,0	5,0		10,0	40,0
9. Kivimaa	24,4	6,7			24,4	26,8	4,4		4,4	8,9
10. O. Manni- nen	10,8	4,0			2,8	27,0	1,4		4,0	50,0
11. Mustapää	38,0				8,0	22,0	2,0		8,0	22,0
12. Onerva	8,2	2,0			8,2	28,5	10,2		8,2	34,7
13. Savulinna		25,0			25,0	50,0				
14. Törmänen	26,2				4,3	17,4	4,3		17,4	30,4
15. Uotinen	22,3				2,2	20,0		1,1	12,2	42,2
16. Vaara (B)	46,2				7,7	12,8			12,8	20,5
17. Vuorela	32,7	7,7		1,9	7,7	23,1			9,6	17,3
Keskiarvo	23,9	4,1		0,1	8,1	21,8	3,2	0,3	8,9	29,6

Taulukko III. Vuoden 1925 runoilijain käyttämien attribuuttien jakaantuminen eri semanttisiin ryhmiin (luvut prosentteja).

aistimuksia kuin nykyrunoilijain attribuutit. Haju- ja makuaisti ovat selvästi niukkimmin edusteilla. Vaikka näihin ryhmiin kuuluvia attribuutteja ei ole näytteeseen sattunut, ei se tietenkään merkitse niiden täydellistä puuttumista perusaineistosta. Kaikkinaiset aistimusvaliteetit ovat yleisiä nimenomaan osassa vuoden 1925 runoutta (tulenkantajilla), joskin ne usein esiintyvät muussa kuin attribuutin funktiossa. Näyttää siltä, että esim. 1920-luvun dekadentiksi leimattu Olavi Lauri (Paavolainen) on erityisen mieltynyt hajuaistiin vetoaviin ilmauksiin. Hänen runosikermästään löytyy helposti sellaisia ilmauksia kuin *raukeitten kukkien lemusta* tai *oman ruumiini tuoksu*. Olisi houkuttelevaa vertailla yksityisiä runoilijoita taulukoiden III ja IV perusteella, mutta otoksen pienuuden takia lienee tosiasiallista merkitystä vain keskiarvoilla.

Myös tässä analyysissä tuli ilmi eräitä kvantifioinnin valo- ja varjopuolia. Toisaalta tämäntapaisissa töissä päästään lujalle pohjalle, vertailukelpoisiin lukuihin ja täsmälliseen ilmaisuun. Toisaalta juuri kiinnostavimmat tapaukset täytyy käsitellä erikseen, semminkin kun on kysymys runoudesta.

4.

Tilastollinen menettely on osoittautunut mielekkääksi useilla tyylintutkimuksen lohkoilla, joihin ei tässä ole kiinnitetty minkäänlaista huomiota. Tu- loksiin on meilläkin päästy mm. metriikkaa, erityisesti klassisia mittoja, ja

Runoilija	1. Väri	2. Kuulo	3. Haju	4. Maku	5. Tunto	6. Koko	7. Liike	8. Kans.	9. Tunne	10. Arvott.
1. Hartzell	16,7				8,3	16,7	16,7		16,7	24,9
2. Hämäläinen	50,9					26,3	5,3		3,5	14,0
3. Jaakola	33,3					13,3			13,3	40,1
4. Kaasalainen	36,4	9,1								9,1
5. Kirstinä	25,4	3,2			9,1	27,2	9,1			27,0
6. Latvala					6,3	25,4	4,8		7,9	
7. M. Manninen						25,0			75,0	
8. Meriluoto	36,4				3,0	12,1	9,1		6,1	33,3
9. Piirto	33,3				11,1	22,2	11,1			22,3
10. Rekola	27,3	9,0			18,2	27,3				18,2
11. Rossi	30,0				10,0	60,0				
12. Saarenpää	13,5	8,1			2,7	32,4		5,4	13,5	24,4
13. Saarikoski	13,8	3,4				44,8	10,4		6,9	20,7
14. Saaritsa	22,2					38,9	11,1	5,6	11,1	11,1
15. Saastamoinen	27,3			4,5	4,5	41,0	4,5		4,5	13,7
16. Seilonen	58,3				8,3	16,8		8,3		8,3
17. Siippainen	33,3				8,3	16,7				41,7
18. Sinkkonen	17,4	26,1			8,7	4,4		4,4	12,9	26,1
19. Suhonen	26,1		4,3		8,8	26,1			4,3	26,1
20. Tarkiainen	28,6					35,7	7,1			28,6
21. Vaara (C)	66,6								16,7	16,7
22. Virtanen	18,7	1,8	1,8	1,8	4,1	27,6	8,1	4,1	9,0	23,0
	50,0									50,0
Keskiarvo	27,3	2,9	0,5	0,5	4,4	26,6	5,7	1,8	7,7	22,6

Taulukko IV. Vuoden 1965 runoilijain käyttämien attribuuttien jakaantuminen eri semanttisiin ryhmiin (luvut prosentteja).

sointuisuutta koskevista töissä, sellaisissa kuin Matti Sadeniemen teos *Metriikkamme perusteet* tai Pentti Leinon (lähinnä folkloristis-lingvistinen) tilastanalyysi *Strukturaalinen alkusointu suomessa*.

Mihin suuntaan kvantitatiivista kirjallisuudentutkimusta voitaisiin kehittää? Vihjeen kenties antaa John B. Carrollin huomiota herättänyt kirjoitus *Vectors of Prose Style* (*Style in Language* s. 283—292; ks. myös Thavenius, *Litteraturvetenskap* s. 57—58, samoin Kirstinä, *Runo ja lukija* s. 127—143, jossa on testattu Carrollin tulosten sopivuutta kouluopetukseen).

Carroll pyrki tutkimaan proosatyylin perusolottuvuuksia. Hänen aineistoonsa sisältyi 150 näytettä englanninkielisestä kirjallisuudesta, mm. romaaneista, esseistä, sanomalehtien pääkirjoituksista, oppikirjoista ja saarnoista. Poikkeuksellista oli, että tutkimuksessa sovellettiin sekä objektiivisia että subjektiivisia mittapuita.

Objektiiviset mittapuu (yhteensä 39) olivat sellaisia kielellisiä tosiseikkoja kuin virkkeiden ja tavujen määrä, sanojen pituus, persoonapronominien ja prepositioiden osuus, latinalaisperäisten suffiksien määrä jne. Subjektiivisia arvioita varten valittiin kahdeksan koehenkilöä, jotka esittivät mielipiteensä

kaikista 150 näytteestä 29 asteikon avulla. Asteikot koostuivat adjektiiveista, joiden arveltiin kattavan tyylin pääpiirteet, ja ne olivat muotoa »täsmällinen — epämääräinen».

Mittapuita kertyi yhteensä 68 (39 objektiivista ja 29 subjektiivista). Niitä sovellettiin kaikkiin 150 tekstinäytteeseen. Taulukoinnin perusteella laadittiin korrelaatiomatriisi, joka tuli sisältämään 68×150 korrelaatiokerrointa. Matriisi alistettiin faktorianalyysiin, jonka oli määrä seuloa esiin aineiston perusluottuvuudet. Analyysin tuloksena Carroll sai kuusi kaksinapaista faktoria A, B, C, D, E ja F, jotka käyvät ilmi oheisesta kuviosta 1. Kuvioon sisältyy myös tulosten sovellusta: kahden proosanäytteen tyyliprofiili (näytteinä katkelma F. Scott Fitzgeraldin novellista *A Diamond as Big as the Ritz* ja katkelma Mickey Spillanen salapoliisiromaanista *Vengeance is Mine*).

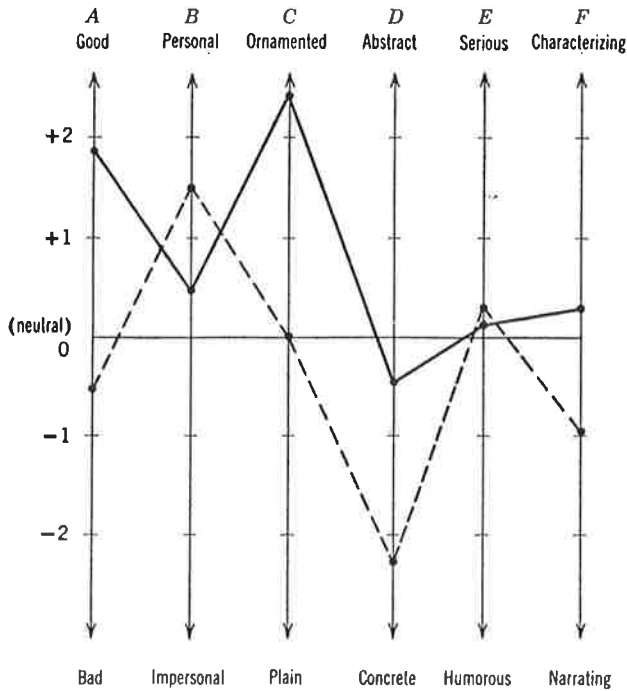
A-faktori merkitsee tyylin yleistä arvottamista (»General Stylistic Evaluation»); se sisältää sellaisia subjektiivisen arvottamisen vastakohtapareja kuin hyvä—huono, kiinnostava—pitkäveteinen, täsmällinen—epämääräinen ja syvällinen—pintapuolinen. Ainoakaan objektiivinen mitta ei osunut tähän faktoriin. Sen sijaan faktoria B (»Personal Affect») luonnehtivat paitsi subjektiiviset arviot (kuten persoonallinen—epäpersoonallinen, intiimi—etäinen, emotionaalinen—järkipерäinen) myös muutamat objektiiviset mittapuut, esim. pronomien ja erityisesti persoonapronominien määrä (mitä enemmän näitä pronomineja, sitä tuttavallisempaa tyyliä).

Kolmatta faktoria kuvaavat parhaiten arviot asteikoilla koristeellinen—yksinkertainen, monisanainen—tiivis, teennäinen—luonteva ja siro—kömpelö. Myös objektiivisilla tekijöillä (esim. lauseiden pituudella ja kuvailevien adjektiivien määrällä) on merkitystä tämän faktorin suhteen. Carrollin yleisluonnehdinta: C-faktori mittaa koristeellisuutta (»Ornamentation»).

Faktoria D Carroll nimittää abstraktisuudeksi (»Abstractness»). Samoin kuin faktorit B (»Personal Affect») ja C (»Ornamentation») on tämäkin riippumaton tyylin arvottamisesta eli faktorista A (»General Stylistic Evaluation»). Tyylin tunnepitoisuus tai abstraktisuus (tai näiden ominaisuuksien puuttuminen) ei toisin sanoen ole yhteydessä siihen, pitävätkö koehenkilöt tekstiä hyvänä vai huonona. D-faktori koostuu toisaalta subjektiivisista arvioista (abstraktinen—konkreettinen, yksinkertainen—mutkikas, omintakeinen—tavanomainen), toisaalta objektiivisista indikaattoreista (esim. numeraalien ja determinatiivipronominien vähyys tekee tekstistä abstraktista).

E-faktorin (»Seriousness») suhteen on eniten merkitystä asteikoilla vakava—sukkela ja vakava—humoristinen (osittain myös sellaisilla kuin maskuliininen—feminiininen ja syvällinen—pintapuolinen). Sitä luonnehtivat myös muutamat objektiiviset tekijät, esim. epämääräisten artikkelien vähyys ja determinatiivisten sanojen runsaus.

Kuudenneksi hahmottunut F-faktori koostuu pelkästään objektiivisista



Kuvio 1. Proosatyylin perusulottuvuudet John B. Carrollin mukaan (*Style in Language* s. 291). Lisänä kahta tekstinäytettä kuvaavat tyyliprofiilit (—: F. Scott Fitzgerald, *A Diamond as Big as the Ritz*; - - -: Mickey Spillane, *Vengeance is Mine*).

mittapuista. Jos teksti on karakterisoivaa, siinä esiintyy niukasti transitiivisia mutta runsaasti intransitiivisia verbejä, niukasti propreja mutta runsaasti adjektiiveja: liikutaan akselilla kuvailu—kerronta («Characterization versus Narration»).

Carrollin tutkimus on ollut paljotöinen. Se on vaatinut laajan työryhmän ja runsaasti apurahoja, mutta myös tulokset tuntuvat pätevilä. Carrollin kuusi faktoria antanevat pohjaa ainakin englanninkielisen proosan tutkijoille.

Kiintoisimmalta vaikuttaa subjektiivisten ja objektiivisten mittapuiden yhdistäminen. Se onnistui useimmissa tapauksissa mutta epäonnistui ensimmäisen, kirjallisuudentutkimuksen kannalta keskeisen faktorin (arvottamisen) suhteen. Tämäkin on tulos: kirjallisuuden ominaislaatua voidaan mekaanisesti eritellä varsin pitkälle, mutta arvoarvostelmat täytyy jokaisen tehdä itse.

L Ä H T E I T Ä

- CARROLL, JOHN B. *Vectors of Prose Style. — Style in Language.* Ed. by THOMAS A. SEBEOK. Cambridge, Mass. 1968.
- COHEN, JEAN *Structure du langage poétique.* Paris 1966.
- ELLEGÅRD, ALVAR A *Statistical Method for Determining Authorship.* The Junius Letters, 1769—1772. Gothenburg Studies in English, 13. Göteborg 1962.
- ELLISON, JOHN W. *Computers and the Testaments. — Computers in Humanistic Research.* Ed. by EDMUND A. BOWLES. Englewood Cliffs 1967.
- FRANCIS, IVOR S. *An Exposition of a Statistical Approach to the Federalist Dispute. — The Computer & Literary Style.* Ed. by JACOB LEED. Kent, Ohio 1966.
- HAKULINEN, LAURI *Suomen kielen rakenne ja kehitys. 3., korjattu ja lisätty painos.* Kcuruu 1968.
- HALLBERG, PETER *Natursymboler i svensk lyrik från nyromantiken till Karlfeldt I—II.* Göteborg 1951.
- HEIKKINEN, JALO Veijo Meri ja Heinrich Böll — rinnastusta sekä sotanovellien ja -romaanien vertailua. Eräitten uusien metodien kokeilua. Pro gradu -tutkielma 1967 (Helsingin yliopiston yleisen kirjallisuustieteen ja teatterintutkimuksen laitos).
- HERDAN, GUSTAV *Eine Gesetzmässigkeit der Sprachenmischung. Mit einem Exkurs über Goethes »West-Östlichen Divan«.* — *Mathematik und Dichtung.* 3. durchgesehene Auflage. München 1969.
- *Language as Choice and Chance.* Groningen 1956.
- *Quantitative Linguistics.* London 1964.
- *The Advanced Theory of Language as Choice and Chance.* Kommunikation und Kybernetik in Einzeldarstellungen. Herausgegeben von H. Wolter und W. O. Keidel. Band 4. Berlin — Heidelberg — New York 1966.
- *The Calculus of Linguistic Observations.* *Janua Linguarum* 9. 's-Gravenhage 1962.
- KIRSTINÄ, VÄINÖ *Runo ja lukija.* Tapiola 1971.
- KUKKOLA, TIMO *Eräs tilastollinen metodi ja sen käyttömahdollisuudet kirjallisuuden tutkimuksessa. Pro gradu -tutkielma 1962 (Helsingin yliopiston yleisen kirjallisuustieteen ja teatterintutkimuksen laitos).*
- LAUNONEN, HANNU *Attribuutin ja pääsanana järjestys vuosien 1925 ja 1965 suomenkielisissä runokokoumissa ja sen lingvististä taustaa. Pro gradu -tutkielma 1969 (SS).*
- *Lyriikan attribuuttityylistä. Esimerkkeinä 38 suomalaista runoilijaa. Laudaturtutkielma 1970 (Helsingin yliopiston yleisen kirjallisuustieteen ja teatterintutkimuksen laitos).*
- LEINO, PENTTI *Kieli ja kaava — Henrik Achreniuksen runot. — Vir.* 1969.
- *Strukturaalinen alkusointu suomessa. Folklore pohjainen tilastoanalyysi.* SKST 298. Forssa 1970.
- LOMAN, B. *Adjektivkombinationer. Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning I.* Stockholm 1964 (moniste).
- MAUD, RALPH *Entrances to Dylan Thomas' Poetry.* New York 1963.
- MEIER, HELMUT *Deutsche Sprachstatistik I/II.* Hildesheim 1967.
- MILES, JOSEPHINE *Eras & Modes in English Poetry.* Berkeley — Los Angeles 1964.
- — SELVIN, HANAN C. *A Factor Analysis of the Vocabulary of Poetry in the Seventeenth Century. — The Computer & Literary Style.* Ed. by JACOB LEED. Kent, Ohio 1966.
- MORTON, ANDREW Q. — LEVISON, MICHAEL *Some Indicators of Authorship in Greek Prose. — The Computer & Literary Style.* Ed. by JACOB LEED. Kent, Ohio 1966.
- O'DONNELL, BERNARD *Stephen Crane's The O'Ruddy: A Problem in Authorship Discrimination. — The Computer & Literary Style.* Ed. by JACOB LEED. Kent, Ohio 1966.
- OSGOOD, CHARLES E. *Some Effects of Motivation on Style of Encoding. — Style in Language.* Ed. by THOMAS A. SEBEOK. Cambridge, Mass. 1964.
- SADENIEMI, MATTI *Metriikkamme perusteet ja sovellutusta moderneihin ja antiikin mittoihin.* SKST 236. Helsinki 1949.
- SARAJAS, ANNAMARI *Tehtävien asettelu Suomen kirjallisuuden tutkimuksessa. — Uusi Suomi* 6. 10. 1968.
- SKST = Suomalaisen Kirjallisuuden Seuran toimituksia.

- SOMERS, H. H. *Statistical Methods in Literary Analysis*. — The Computer & Literary Style. Ed. by JACOB LEED. Kent, Ohio 1966.
- SÄRKILAHTI, SIRKKA-LIISA Kirjallisuudentutkimus ja tilastotiede. — Oppikoululehti 28. 5. 1968.
- Voidaanko kirjailijoiden kieltä tutkia kvantitatiivisin metodein? Eeva Joenpellon, Vilho Niitemaan, Marja-Liisa Vartion ja F. E. Sillanpään tekstinäytteiden tilastollinen analyysi. — Juhlakirja Kauko Kyyrön täyttävässä 60 vuotta 24. 11. 1967. Vammala 1967.
- TALLENIRE, D. R. *The Mathematics of Style*. Thinking by Numbers, 2. — The Times Literary Supplement 13. 8. 1971.
- TARKKA, PEKKA Joel Lehtosen kaksi Matalceenaa — hyödyllinen mikrostilistiikan kohde. — Suomenkieliset Tieto-Sanommat 1967.
- TERNSTRÖM, ANDERS Litteraturforskning med och utan statistik. — Dagens Nyheter 20. 8. 1966.
- THAVENIUS, JAN Kvantitativa metoder i stilistiken. — Litteraturvetenskap. Nya mål och metoder. Stockholm 1966.
- ULLMANN, STEPHEN *The Principles of Semantics*. London 1959.
- *Language and Style*. Oxford 1966.
- VAKKURI, JUHA Sisällön analyysi Pertti Niemisen kokoelmista Kivikausi, Uurnat, Päivät kuin nuolet ja Silmissä maailman maisemat. Pro gradu -tutkielma 1969 (Helsingin yliopiston kotimaisen kirjallisuuden laitos).
- YRJÄNÄINEN, ANNIKKI L. Onervan lyriikan kuva-ainesten tarkastelua. Pro gradu -tutkielma 1967 (Helsingin yliopiston kotimaisen kirjallisuuden laitos).

Quantitative Literaturforschung: wie und warum?

Hannu Launonen

Der Aufsatz behandelt das Eindringen der quantitativen Methoden auf den verschiedenen Sektoren der Literaturforschung. Es wird festgestellt, dass die statistische Erfassung verschiedenen Grades am meisten Nutzen bringt in Fragen z.B. der Bestimmung eines vorläufig anonymen Verfassers sowie hinsichtlich der Chronologie umstrittener Werke (Musterbeispiel: Alvar Ellegård, *A Statistical Method for Determining Authorship*), ferner in Stiluntersuchungen, die auf Wortschatzvergleichen fussen, vor allem wenn es um umfangreiches Material geht (wie in der Abhandlung von Josephine Miles: *Eras & Modes in English Poetry*).

In Finnland ist die quantitative Stilforschung noch nicht angelaufen; zum Teil liegt das daran, dass hier vorläufig

die Grundstatistiken über die Gemeinsprache fehlen. Vor allem im Bereich der Erforschung der Lexik hat man jedoch interessante Experimente unternommen, die gern weitergeführt werden sollten.

Die quantitativen Methoden könnten zugunsten der Literaturforschung so ausgebaut werden, dass — wie es John B. Carroll in seiner Arbeit *Vectors of Prose Style* in dem Buch *Style in Language* tut — die messbaren objektiven Faktoren (die grammatischen Eigenschaften des Textes) möglichst vereint werden mit den subjektiven Einschätzungen der Forscher oder Versuchspersonen. Bei der Bearbeitung der Ergebnisse wären dann feingefächerte statistische Methoden angebracht, z.B. die Cluster- oder Faktorenanalyse.