

PERTTI RAUTIO
PERTTI SUHONEN

Mihin viiteanalyysi kelpaa?

Silmäys viimeaikaiseen tutkimukseen ja keskusteluun

Rautio, Pertti, ja Suhonen, Pertti, Mihin viiteanalyysi kelpaa? Silmäys viimeaikaiseen tutkimukseen ja keskusteluun [On the usefulness of citation analysis; a review of recent research and discussion]. Kirjasto-tiede ja informatiikka, 1 (1): 8—16. 1981.

On the basis of their recently completed study, titled (in translation) »Flow of information in the social sciences and the Finnish researcher», together with a critical review of recent papers on citation analysis, the authors stress the importance of studies in the citation behaviour adhered to in different branches of science. A deeper understanding of such patterns would add to the knowledge of the reliability and sources of error of citation indexes. This would render the indexes a better usability as tools in the study of scientific communication and the evaluation of scientists and their publications.

Address: Research Institute of Social Sciences, University of Tampere, PB 607, SF-33101 Tampere 10.

Vähitellen laajenevassa informatiikan tutkimuksessa on ollut huomattava sija yhdellä bibliometrisellä menetelmällä, jota voidaan kutsua viiteanalyysiksi (citation analysis). Sitä ovat tutkimuksissaan soveltaneet Suomessa mm. Haapala 1972, Ginman 1975, Sullström 1976, Kiiski 1977 sekä Suutarinen 1975 ja 1978. Myös oma tutkimuksemme tieteellisestä kommunikaatiosta yhteiskuntatieteissä perustui pääosaltaan viiteanalyysin kehittämään (Rautio ja Suhonen 1978, 1980 ja 1981). Viiteanalyysi ei ole kuitenkaan erityisen suomalainen metodi, vaan pikemminkin kuvastaa tutkimusideoiden varsin nopeaa omaksuamista meillä.

Laajojen tietokantojen ja tietojenkäsittelyjärjestelmien kehittyminen (esimerkiksi Science Citation Index, SCI) ovat lisänneet viiteanalyysin sovellutusmahdollisuuksia ja käyttöä. Informatiikan tutkijoiden ohella myös tieteentutkijat ovat keksineet uusille menetelmille runsaasti aiheellista ja aiheellontakin käyttöä. Viime vuosina kansainväliset ja kansalliset tiedepolitiikan elimet ovat kiinnostuneet viiteanalyysin tarjoamista

mahdollisuuksista erityisesti tiedeindikaattorien kehittämisessä (kts. Garfield, Malin and Small 1978, 179—207).

Viiteanalyysin käytön laajentuminen on herättänyt myös keskustelua aineiston ja metodin luotettavuudesta ja käyttömahdollisuuksien rajoista. Näyttää siltä, että varauksettomimmin ja innokkaimmin ovat viiteanalyysiin suhtautuneet luonnontieteilijät ja luonnontieteiden alueilla työskentelevät informatiikan ja tieteen tutkijat. Eniten epäilijöitä löytyy yhteiskuntatieteilijöiden piiristä (esim. Wiener 1978).

Tässä artikkelissa selvitetään aluksi mistä viiteanalyysissa on kysymys. Tämän jälkeen tarkastellaan sen käyttöä erilaisissa tutkimuksen ja käytännön yhteyksissä. Alan tutkimustoiminta on jo niin laajaa, että kaikkien tärkeidenkään tutkimusten esille ottaminen on mahdotonta. Sen sijaan pyrimme esittelemään tyypillisiä esimerkkejä viiteanalyysin erityyppisistä sovellutuksista ja samalla arvioimaan niiden tarkoituksenmukaisuutta ja luotettavuutta. Painopiste on yhteiskuntatieteiden alueen tutkimuksissa ja oma näkökul-

mamme on niinkään yhteiskuntatieteellisesti painottunut. Esille tulevat näkökohdat tosin pätevät usein yhteiskuntatieteiden alueen ulkopuolellakin.

Käsitteistä

Tarkasteltavalla alueella on vallinnut sekä suomen- että englanninkielinen käsitesekeanus. Englanninkielisessä kirjallisuudessa on termejä 'reference' ja 'citation' käytetty osaksi toistensa vastineina ja osaksi niillä on ollut oleellinen merkitysero. Suomenkielisessä puhutaan viitteistä, viittauksista ja lähteistä tekstistä riippuen yhteisin tai eri merkityksin. Vastaavasti puhutaan viite-, viittaus- ja lähdeanalyysistä ilman johdonmukaisuutta.

Varsin selkeällä ja käyttökelpoisella tavalla viiteanalyysin peruskäsitteet on määritelty brittiläisessä DISISS-projektissa, joka on viime vuosien merkittävimpiä tieteellisen kommunikaation tutkimuksia yhteiskuntatieteiden alueella. Projektin viiteanalyytisessä osatutkimuksessa (DISISS 1979, 1) määritellään peruskäsitteet seuraavasti:

'Reference' tarkoittaa julkaisua, johon kohdistunutta viittausta tarkastellaan **viittaavan** julkaisun kannalta. Kysymys on siis **lähteistä**.

'Citations' tarkoittaa johonkin teokseen kohdistuneita viittauksia kun niitä tarkastellaan lähinnä **viitatus** teoksen kannalta. »Thus article A refers to article B; this is a reference. Article B is cited by article A; this is a citation.»

Tätä erottelua voidaan soveltaa myös suomenkieliseen käsitteistöön. Tutkija saa **viittauksia** julkaisuihinsa ja käyttää **lähteitä**. Vastaavasti voidaan näkökulmasta riippuen puhua **viittausanalyysistä** ja **lähdeanalyysistä**. Tämä käsite-erottelu ei ilmeisestikään ole ristiriidassa tähänastisen terminologisen käytännön kanssa. Esimerkiksi Suutarinen (1978) käytti termiä viittausanalyysi tutkimuksessaan, jossa ensisijaisesti oltiin kiinnostuneita suomalaisten lääketieteellisten tutkimusjulkaisujen kansainvälisestä käytöstä. Hän siis tutki suomalaisten tutkijoiden saamia viittauksia. Itse käytimme termiä lähdeanalyysi tutkiessamme suomalaisten sosiologian ja politologian väitöskirjojen käyttämää lähdekirjallisuutta, sen rakennetta ja alkuperää.

Usein käytetyn viiteanalyysin käsitteen varasimme koko tutkimusalueen ja menetel-

män yleisnimeksi. Viittausanalyysin ja lähdeanalyysin ohella tähän viiteanalyysin kenttään voitaisiin sisällyttää tutkimukset, joissa kohteena ovat tutkijoiden viittauskäytännöt, lähteiden käyttötarkoitukset tms., joille Okko on esittänyt viiteanalyysin nimeä (Rautio ja Suhonen 1978, 34).

Mihin viiteanalyysijä käytetään?

Viiteanalyytiset tutkimukset voidaan monitahoisista ongelmanasetteluistaan huolimatta palauttaa kahteen päätyyppiin, jotka perustuvat erilaisiin viiteanalyysin käytön funktioihin.

Ensinnäkin viiteanalyysit usein tuovat esiin joitakin puolia tieteellisestä kommunikaatiosta. Ne kertovat jotakin siitä miten tieteelliset ideat, teoriat, käsitteet tai empiiriset tutkimustulokset siirtyvät tutkijalta toiselle, julkaisusta toiseen, tiedeyhteisöstä toiseen tai tieteenalalta tai tutkimusalueelta toiselle. Useimpia tehtyjä viiteanalyysiä soveltaneita tutkimuksia voidaan tarkastella tältä kannalta silloinkin kun tutkijoiden omat intressit ovat olleet toisaalla. Tätä funktiota voidaan kutsua kommunikaation tutkimisen aspektiksi.

Toinen funktio liittyy tutkijoiden, tutkimusten ja tieteellisten yhteisöjen hierarkisen aseman tutkimiseen tulosten tieteellisen merkityksen, niiden leviämisen ja sosiaalisen arvostuksen suhteen. Joihinkin harvoin teksteihin, tuloksiin tai julkaisuihin viitataan usein ja useimpiin tuskin koskaan. Vaikka vaihtelun syyt ovat moninaisia, viiteanalyysin avulla on kuitenkin mahdollista osoittaa tiettyjä hierarkisia järjestyksiä. Nimitämme tätä arvostusaspektiksi ajatellen, että sillä voidaan tavoittaa tietoa tutkittujen kohteiden todellisesta, oletetusta tai kuvitellusta tieteellisestä merkityksestä.

Tyypillinen esimerkki tutkimuksesta, jossa painotetaan kommunikaatioaspektia, on oma tutkimuksemme suomalaisten valtio-opin ja sosiologian väitöskirjojen lähteistä II maailmansodan jälkeen (Rautio ja Suhonen 1981). Siinä pyrittiin selvittämään mistä suomalaisten yhteiskuntatieteilijöiden teoreettinen, metodinen ja empiirinen tieto oli peräisin ja millaisia muutoksia tässä suhteessa oli tarkasteltavana aikana tapahtunut. Kun itse olimme kiinnostuneita siitä 'mistä tieto tulee', niin Suutarinen (1978) oli puolestaan

kiinnostunut siitä, 'mihin tieto menee', ts. missä määrin suomalaisia tutkimuksia käytettiin lähteenä muiden kuin suomalaisten lääketieteen tutkijoiden töissä. Tämä näkökulma onkin viiteanalyyseissä ollut tavallisempi ja siinä on jo selvästi mukana myös viiteanalyysin toinen funktio. Jonkin tutkijan saamat monet viittaukset voidaan käsittää osoitukseksi hänen tutkimustensa todellisesta tai oletetusta tieteellisestä merkityksestä. Viiteanalyytissä tutkimuksissa pääpaino onkin useimmiten arviointifunktiossa. Lähdeanalyysit sen sijaan liittyvät pääasiassa kommunikaation tutkimiseen tai olemassa olevan tieteellisen kirjallisuuden rakenteellisten ja 'bibliografisten' ominaisuuksien tutkimiseen.

Uuden tutkimusotteen tai uuteen ilmiöalueeseen kohdistuvan tutkimuksen laajentumisessa voidaan yhdeksi merkkipaaluksi ajatella vaihetta, jolloin ryhdytään kokoamaan sitä koskevia bibliografioita tai kokoavia katsauksia.

Yhteenvedoja yhteiskuntatieteitä koskevista viiteanalyyseistä löytyy Brittainin (1970), Broadusin (1971) ja Fitzgibbonsin (1980) katsausartikkeleista. Sekä tutkimusten perustavoitteet että käytetty aineisto vaihtelevat huomattavasti tutkimuksesta toiseen. Fitzgibbonsin (1980) jokseenkin tuoreessa luettelossa on kasvatustiede mukaanluettuna yhteensä 110 yhteiskuntatieteitä koskevaa viiteanalyysiä. Artikkelin liitteenä julkaistusta luettelosta, jossa on luonnehdittu tutkimuksen päätavoitetta, on laskettavissa, että noin kolmasosassa tutkimuksista oli pyritty selvittämään käytetyn kirjallisuuden peruspiirteitä tieteenalan, kielen, julkaisutyyppin tai muiden julkaisujen bibliografisista tiedoista ilmenevien seikkojen suhteen. Muista tavoitteista yleisin oli alan ydinkirjallisuuden etsiminen tai 'parhaimpien' tai 'eniten vaikuttaneiden' julkaisujen tunnistaminen. Noin kymmenesosassa tutkimuksen keskeinen tavoite liittyi jonkin tieteellisen kommunikaation aspektin tutkimiseen. Joissakin harvoissa tutkimuksissa tavoitteet olivat hyvinkin spesifejä, kuten 'Lotkan lain testaaminen'.

Yksittäisistä viiteanalyysiä soveltaneista tutkimuksista lienee yhteiskuntatieteiden alueella huomionarvoisin Englannissa vuodesta 1971 alkaen toiminut 'Design of Information Systems in the Social Sciences' (DISISS) -projekti (1979), jossa on kiinnitetty huomiota sekä artikkeli- että yhteiskuntatieteissä tärkeään monografiakirjallisuuteen.

Viiteanalyysi tutkimusten laadun tai arvostuksen tutkimisessa

Tutkimusongelma on viiteanalyyseissä usein esitetty niin, että on haluttu tutkia mihin yksittäisiin artikkeleihin, lehtiin tai kirjoihin, keiden tutkijoiden tai laitosten töihin on viitattu eniten. Tämän tapaisten tutkimusten yleistymiseen vaikutti huomattavasti yksityisen Institute for Scientific Information'in tuottamien Science Citation Indexin, Social Sciences Citation Indexin ja Arts & Humanities Citation Indexin aineistojen tarjoamat mahdollisuudet suorittaa helposti laajoja lähdeluetteloihin perustuvia viiteanalyysieja. Tutkimusala- tai tieteenalakohtaisia eniten siteerattujen tutkijoiden, laitosten ja julkaisujen lyhyesti kommentoituja luetteloita on julkaissut erityisesti Eugene Garfield. Merkittävimmät näistä artikkeleista on koottuna teokseen *Essays of an Information Scientist, Volume 1—3 (1977—1980)*. Sama julkaisu sisältää myös huomattavan määrän Garfieldin ja muiden kirjoittajien viiteanalyysin mahdollisuuksia laajemmin käsitteleviä artikkeleita.

Viiteanalyysien avulla on pyritty mitaamaan joko suoraan tutkimusten laatua tai tuloksellisuutta (Cole and Cole 1973), selvittämään laitosten, lehtien ja yksittäisten tutkijoiden tiedeyhteisössä nautimaa arvostusta (Roche and Smith 1978) tai sen avulla on pyritty paljastamaan ketkä tai mitkä tutkijat ja tutkimukset ovat vaikuttaneet eniten josakin tiedeyhteisössä tietynä aikana (Rautio ja Suhonen 1981, 128—143). Monissa näissä tutkimuksista on taustalla funktionaalisen sosiologian mukainen käsitys tiedeyhteisön kerrostuneisuudesta ja tiedeyhteisön jäsenten saavutuksiin perustuvasta eriarvoisuudesta. Viiteanalyysin tulosten perusteella on ehdotettu jaettavaksi Nobelin palkinnot, sitä on ehdotettu puolueettomaksi arviointivälineeksi tieteellisten virkojen täytössä ja sen avulla on uskottu voitavan rationaalistaa kirjastojen hankintoja ja poistoja niiden kokoelmista (kts. esim. Broadus 1977).

Viimeksi mainitut ehdotukset liittyvät läheisesti yrityksiin käyttää viiteanalyysiä hyväksi tiedeindikaattorien kehittämisessä. Tieteen käytännöllisen merkityksen lisääntyessä on kiinnostus tieteen tavoitteelliseen ohjaukseen ja tiedepolitiikkaan lisääntynyt. Tämä on synnyttänyt tarpeen tuottaa sekä laadullisia että määrällisiä menetelmiä ja indikaattoreita tieteellisen toiminnan tulosten arvioi-

miseksi. Tiedeindikaattoreista alettiin puhua aluksi Yhdysvalloissa, mistä tiedeindikaattoreiden kehittäminen on levinnyt nopeasti niin itään kuin länteenkin. Luukkonen-Gronow (1981) luonnehtii tiedeindikaattoreita sanoen niiden kuvaavan tai mittaavan jonkin maan tieteenharjoituksen tilaa ja siinä tapahtuvia muutoksia tai tiede- ja teknologiajärjestelmän tilaa ja toimintaa. Aivan viime aikoina on erityisesti sosialistisissa maissa kiinnitetty viiteanalyysin tarjoamiin mahdollisuuksiin tiedeindikaattoreiden kehittämisessä suuria toiveita (kts. *Scientometrics*-lehden numeroa 5—6/1980). Hyvän yleiskäsityksen keskustelusta, joka liittyy viiteanalyysin käyttämiseen tiedeindikaattoreiden kehittämisessä, saa Garfieldin, Malin ja Small'in artikkelista 'Citation Data as Science Indicators' (1978, 179—207), ja OECD:n Pariisissa 1980 järjestämän tiede- ja teknologiakonferenssin eräistä seminaaripapereista (esim. Martin ja Irvine 1980, tai de Mey 1980).

Kun saatujen viittausten määrää käytetään jonkin kohteen tieteellisen merkityksen tai tutkijayhteisön sisäisen arvostuksen mittana, on taustalla joukko idealisoivia oletuksia. Pyrkimättä tyhjentävään olettamusten luettelointiin, voidaan todeta, että ainakin seuraavat seikat oletetaan päteviksi:

1. Viittaavan julkaisun ja lähteen välillä on sisällöllinen yhteys.
2. Viittaava julkaisu käyttää lähdeä tieteellisenä auktoriteettina ja viittaus on ilmaus positiivisesta, arvostavan sävyisestä suhtautumisesta lähteeseen.
3. Tutkijat ovat viittaneet kaikkiin todella käyttämiinsä lähteisiin eikä lähteeksi ole merkitty sellaisia julkaisuja, joita ei tosiasiassa ole käytetty.
4. Viittaavan julkaisun kaikki viittaukset lähteisiin ovat merkitykseltään samanlaisia, ts. vallitsee jonkinlainen »lähde ja ääni»-periaate.
5. Viittauskertojen määrä on suorassa suhteessa lähteen objektiiviseen tieteelliseen merkitykseen tai lähteen tutkijayhteisössä nauttimaan tieteelliseen arvostukseen.
6. Viiteanalyysissä käytetty aineisto on edustava kohteena olevan tutkimus- tai sisältöalueen suhteen.

(Vrt. Subramanyam 1980, 356; Griffith, Drott and Small 1977).

On selvää, että olettamusten mukainen idealisoitu tilanne ei ole koskaan voimassa kaikilta osin. Tieteellistä merkitystä tai arvos-

tusta mittaavien viiteanalyysien tulosten arvioinnin kannalta olisi tärkeätä tietää miten suuria poikkeamat ovat ideaalitapauksesta ja mitä tekijöitä tulisi taustaolettamusten lisäksi ottaa huomioon vertailtaessa esim. eri tieteitä tai tutkimusaloja keskenään. Seuraavassa pohdimme viiteanalyysiin liittyviä ongelmia edellä esitettyjen seikkojen suhteen. Vaikka tarkastelu kohdistuuakin lähinnä arviointiaspektiin ovat ongelmat usein samoja myös kommunikaatioaspektia tutkittaessa.

Viittauskäytännöstä

Viiteanalyysien validiteetin kannalta oleellisia kysymyksiä ovat, missä tarkoituksessa aikaisempiin julkaisuihin viitataan, mitä funktioita viittauskäytäntöön liittyy ja mitä viite tai aikaisemman julkaisun esiintyminen lähdeluettelossa oikeastaan paljastaa viittaavan ja viitatuksen välisestä suhteesta. Ongelmaa ovat pohtineet eri puolilta mm. Kaplan (1965), Weinstock (1971), Suhonen ja Rautio (1981, 22—26), ja empiirisen tutkimuksen avulla ovat viittauskäytännön luonnetta selvittäneet mm. Moravcsik ja Murugesan (1975, 1979), Shubin ja Moitra (1975) ja myös omaan tutkimukseemme (Suhonen ja Rautio 1981) sisältyy tällaista aines-

Selvitellessämme suomalaisten sosiologien ja politologien väitöskirjojen viittauskäytäntöä kiinnitimme huomiota siihen, että lähdeluetteloihin oli merkitty varsin runsaasti sellaisia lähteitä, joihin ei oltu viitattu tekstissä lainkaan, tai viittaus oli triviaali tai ei välittänyt varsinaista tieteellistä tietoa viittaavaan julkaisuun. Esim. tutkija ilmoitti käsiteltävää asiaa jollakin tarkentamattomalla tavalla sivuavan julkaisun. Osa lähdeluetteloista oli laadittu selvästi tutkimusaihetta käsitteleväksi aihebibliografiaksi. Runsaasti työn kannalta epäoleellisia lähteitä sisältävät lähdeluettelot tietenkin johtavat viiteanalyysin tekijää harhaan ja alentavat tulosten luotettavuutta. Omassa aineistossamme tieteellisistä lähteistä oli 32 % sellaisia, jotka eivät varsinaisesti välittäneet tieteellistä informaatiota viittaavaan väitöskirjaan. (Käytetystä lähteiden luokitus- ja karsintamenettelystä sekä sen pohjana olleesta väitöskirjojen sisällönanalyysistä kts. Rautio ja Suhonen 1981, 27—47.)

Viittauskäytännössä esiintyy siis liioittelevia ja jopa irrationaalisia piirteitä esimerkiksi silloin, kun siteeraaminen toimii jonkinlaise-

na tieteellisen statuskampailun välineenä. Väitöskirjoissa tätä piirrettä selittää osittain vaatimus, että tutkijan on osoitettava perehtyneensä tutkimusalueen kirjallisuuteen. Turhaa tai liioittelevaa viittaamista esiintyy myös silloin, kun sinänsä yhdentekevä teksti pyritään koristelemaan tieteelliseksi, ts. viittauksia tieteellisen tekstin ulkoisena tuntomerkkinä käytetään lähinnä kosmeettisessa tarkoituksessa. Viittausten avulla voidaan myös osoittaa oma pyrkimys kuulua tiettyyn koulukuntaan, niiden avulla voidaan palkita ystäviä ja samalla tavalla ajattelevia, ja luopumalla viittauksesta voidaan tieteellisiä kiistakumppaneita »rangaista». Perusteellista tietoa näistä viittauskäytännön sosiaalisista sivuilmioista ei kuitenkaan ole olemassa. Kuitenkin myös ne ovat vaikuttamassa viittausanalyysien tuloksiin.

Kirjallisuudessa on kiinnitetty varsin runsaasti huomiota tutkijan tai yhteisön itseviittausten ja negatiivisten viittausten aiheuttamiin ongelmiin (Tagliacozzo 1977; Chubin and Moitra 1975). Useimmissa itseviittauksissa tutkijoiden on yleensä havaittu osoittavan julkaisunsa yhteyksiä aikaisempiin julkaisuihinsa eivätkä ne poikkea sinänsä muusta viittauskäytännöstä. Luonnontieteellisessä artikkelikirjallisuudessa itseviittausten tavallinen määrä on 10—20 % (Tagliacozzo 1977, 252). Eräillä kapeilla erikoistuneilla tutkimusalueilla itseviittausten määrä voi olla huomattavasti suurempikin. Kovin isoa ongelmaa itseviittaukset eivät viittausanalyysien kannalta muodosta, ja ne voidaan aineiston automaattisessa käsittelyssä jopa karsia pois. Negatiivisten viittausten on toisinaan epäilty huomattavasti vähentävän viittausanalyysien luotettavuutta. Onhan yleisesti tiedossa, että heikko tai virheellisen tiedon sisältävä julkaisu voi joutua runsaankin negatiivisen huomion kohteeksi. Chubin ja Moitra (1975) tutkivat 43 fysiikan alalle sijoittuvaa artikkelia, ja totesivat, että viittauksista ei yksikään ollut täysin kielteinen ja osittain kielteisiäkin oli vain 5 %. Myös suomalaisia väitöskirjoja koskevassa tutkimuksessaamme (Rautio ja Suhonen 1981) saatoimme todeta, että kielteisiä viittauksia on erittäin vähän ja etteivät ne vaikuta tuloksiin merkittäväällä tavalla.

Onkin ehkä niin, että negatiivinen arviointi sijoittuu aivan tiettyihin julkaisutyyppeihin, kuten kirjallisuusarvosteluihin, katsausartikkeleihin ja poleemiisiin keskustelupuheenvuoroihin. Uutta tieteellistä tietoa tuottavissa julkaisuissa negatiivisia viittauksia on

vain satunnaisesti. Koko asian tekee tietenkin vielä mutkikkaammaksi se, että tieteen historiasta voidaan poimia lukuisia esimerkkejä julkaisuista, jotka ovat joutuneet voimakkaan ja laajan kritiikin kohteeksi, mutta joilla on osoittautunut olevan jopa uutta paradigmaa luova merkitys. Samoin voidaan poimia esimerkkejä »hedelmällisistä virheistä». Negatiivinenkin viittaaminen voi siis olla oire julkaisun tieteellisestä merkityksestä.

Lähteet eivät ole »tasa-arvoisia»

Edellä käsitellyjä seikkoja huomattavasti vakavampi ongelma liittyy olettamukseen, jonka mukaan kaikki lähteet olisivat merkitykseltään samanlaisia viittaavan julkaisun kannalta. Jos oletamus ei ole tosi, niin i myöskään voida väittää, että julkaisun saamisen viittauskertojen määrä on suorassa suhteessa sen objektiiviseen tieteelliseen merkitykseen tai sen tutkijayhteisössä nauttimaan tieteelliseen arvostukseen. Sen, että lähteet eivät ole tutkimuksen ja niistä kertovan julkaisujen kannalta saman arvoisia tiennee jokainen hiemankin tutkimustyötä tehnyt. Jonhonkin tutkimukseen keskeisiä teoreettisia ideoita tai oleellisen metodisen tiedon välittänyt julkaisu on tärkeämpi kuin lähde, josta on poimittu irrallinen empiirinen numerotieto vertailuaineistoksi. Vastaavasti voidaan tärkeämpänä pitää teoksia, joista on sata kertaa poimittu teoreettisia vaikutteita tai metodisia innovaatioita kuin tutkimuksia, joihin on viitattu tuhat kertaa vähemmän tärkeiden empiiristen tietojen välittäjänä.

Pelkkiin lähdeluetteloihin perustuvan viittausanalyysin kannalta lähteiden merkityksen vaihtelu on ongelmallista siksi, että ne eivät kerro mitään siitä, millaista tietoa viittaavaan julkaisuun ja sitä ennen sen kirjoittajalle on välittynyt. Ongelma on usein kierretty olettamalla ainakin implisiittisesti, että jonhonkin julkaisuun kohdistuneiden viittausten määrä korreloi sen tieteelliseen merkitykseen. Tällainen yleistys voidaan esittää, mutta on merkittäviä poikkeuksia. Tästä taas seuraa, että viittausanalyysiin, jotka perustuvat pelkästään lähdeluetteloiden erittelyyn, on syytä suhtautua jossakin määrin varovasti. Lähdeluettelossa olevien julkaisujen eriarvoisuus niiden tieteellisen merkityksen suhteen viittaavan julkaisun kannalta paljastuu vasta kun viittaavaa ja viitattua julkaisua tarkastellaan suhteessa toisiinsa välittyneiden

ideoiden ja informaation kannalta. Eräs hedelmällinen suunta viittausanalyysien kehittämiseksi voisikin olla tutkimus, jossa perinteisiä kvantitatiivisia viittausanalyysijä systemaattisesti täydennettäisiin sisällöllisiä suhteita erittelevillä laadullisilla analyyseillä. Näin voitaisiin saada entistä tarkempi käsitys viittausanalyysien todellisesta validiteetista.

Millaiset seikat sitten vaikuttavat siihen viitataanko johonkin lähteeseen ehkä aiheetomankin runsaasti vai jääkö se ehkä viittausanalyysissä merkitystään vastaamattomalla tavalla varjoon? — Vakiintuneen tavan mukaan tutkijat pyrkivät käyttämään lähteenä alkuperäisiä julkaisuja. Ei liene kovinkaan harvinaista, että silloinkin kun tieto on poimittu toisen käden lähteistä, viittaus kuitenkin osoitetaan alkuperäislähteeseen. Tästä käytännöstä seuraa, että monet julkaisutyypit, joilla tieteellisessä tiedonvälityksessä on varsin tärkeä asema, jäävät viittausanalyysissä näkymättömiin. Tyypillisiä tällaisia julkaisuja ovat oppikirjat, sanakirjat, käsikirjat ja muut hakuteokset, samoin lehdet, jotka julkaisevat pääasiassa uutisia, keskustelupuheenvuoroja ja kirja-arvosteluja, mutta eivät varsinaisia tieteellisiä artikkeleita. Satariano (1978) on osoittanut, että viittausanalyysit jättävät vaille huomiota joukon lehtiä, jotka välittävät tutkijoille oleellista tietoa, mutta joihin ei kuitenkaan viitata.

Satariano (1978) teki amerikkalaisia sosiologeja edustavalle näytteelle kyselyn, jossa hän tiedusteli 53:n nimeltä mainitun lehden säännöllistä seuraamista. Vastaajilla oli mahdollisuus nimetä myös muita seuraamia lehtiä. Vastausten perusteella muodostettiin lista, jossa lehdet olivat järjestyksessä sen mukaan miten säännöllisesti niitä seurattiin. Kun listaa verrattiin Baughman'in (1974) viittausanalyysin perusteella laatimaan sosiologien eniten viittaamien lehtien listaan, voitiin todeta, että 55 % lehdistä oli samoja. Listoissa oli tiettyjä systemaattisia eroja. Viittausanalyysien perusteella listalle tuli useammin mukaan monitieteisiä ja muiden tieteiden lehtiä, kuin mitä oli säännöllisesti seurattujen lehtien listalla. Säännöllisellä lukulistalla olivat siis selvemmin oman alan lehdet. Satariano päätteleekin, että viittausanalyysien seurauksena helposti tullaan vähätelleeksi yleistajuisten ja alueellisten (regional) lehtien merkitystä, samoin tutkijoiden ammattikäytännöstä annetaan todellista monitieteisempi kuva.

American Sociological Review oli ensimmäisenä molemmilla listoilla, mutta eniten luettujen lehtien listan toiseksi sijoittui »American Sociologist», joka ei mahtunut 20:n ensimmäisen joukkoon viiteanalyysilistalla. Lehti käsittelee lähinnä sosiologien ammatillisia kysymyksiä ja välittää tutkimustointaa koskevaa tietoa esim. tekeillä olevista tutkimuksista ja keskustelua, mutta ei julkaise varsinaisesti merkittäviä tieteellisiä artikkeleita. Vaikka lehteä luetaan runsaasti, siihen ei silti useinkaan viitata.

Viittausanalyysin tietopohjalta on periaatteessa edellytetty edustavuutta tutkittavan alan suhteen, mutta monissa tutkimuksissa on aineistoa valitaessa painavammaksi seikaksi muodostunut käytettävissä olevan aineiston helppo saatavuus ja käsittelyn helppous. Useimmat tutkimukset ovat perustuneet Science Citation Indexin tapaisiin tiedostoihin, jotka ovat painottuneet englanninkieliseen materiaaliin ja amerikkalaisperäisiin julkaisuihin. Tätä ei silti ole pidetty erityisenä ongelmana tulosten yleistettävyyden kannalta. Kuitenkin erityisesti yhteiskuntatieteellisillä ja humanistisilla aloilla, jossa tutkimuksella on usein selvästi kansallisesti rajoittunut luonne, tarkasteltavan aineiston julkaisukiellä ja julkaisupaikoilla on oleellinen vaikutus siihen millaisia tuloksia saadaan.

Kun viittausanalyysijä tehdään arvostusfunktioon liittyen, tulisi vertailuja suoritettaessa ottaa huomioon, että tutkimusalueen laajuus, siis alalla toimivien tutkijoiden, laitosten ja ilmestyvien julkaisujen määrä sekä samojen kysymysten parissa työskentelevien tutkijoiden määrä vaikuttavat johonkin julkaisuun kohdistuvien viittausten määriin (Cole 1971). Laajan tutkimusalueen yleisluontoisia ongelmia käsittelevä standardiesitys saa todennäköisesti huomattavasti viittauksia, kun taas jonkin kapean erityisalueen spesifejä kysymyksiä käsittelevä tieteellisesti hyvinkin ansiokas teksti voi jäädä vähälle huomiolle vain siksi, että ei ole olemassa muita tutkijoita, jotka olisivat kiinnostuneita samoista kysymyksistä.

Viittausanalyysi arvostaa siis kohteitaan »tieteen parlamentin voimasuhteiden mukaisesti» hyvinkin monessa mielessä. Jos tuloksia käytetään tiedepoliittisessa päätöksenteossa tai kirjallisuuden valinnassa kirjastoihin ja yliopistojen kurssivaatimuksiin ottamatta tätä huomioon, päädytään todennäköisesti laajojen tieteenalojen standarditutkimuksen suosimiseen uutta etsivän tutkimuksen ase-

masta. Viittausanalyysi on potentiaalisesti konservatiivinen, keskiarvoihin sidottu tapa tarkastella tieteellisen toiminnan tuloksellisuutta.

Viiteanalyysi tieteellisen kommunikaation tutkimuksen välineenä

Tutkijoiden ja julkaisujen välisiin suhteisiin tavalla tai toisella huomiota kiinnittävässä menetelmänä viiteanalyysi lähes aina tuottaa tietoja myös tieteellisestä kommunikaatiosta. Viittaus aikaisempaan julkaisuun paljastaa asianmukaisella tavalla tehtynä sen, että tutkimuksen kokonaisprosessissa informaatiota on siirtynyt tutkijalta toiselle, ja samalla julkaisun lukija saa vihjeen tiettyyn asiaan liittyvän lähteen olemassaolosta. Tekstiyhteydessä olevat viittaukset ja niiden pohjalta laaditut lähdeluettelot sisältävät sen perusaineiston, jonka avulla tieteellisten ideoiden leviämisen perusvirrat ovat jälkeensä osoitettavissa. Vaikka tutkijoiden välisillä epävirallisilla yhteyksillä onkin tärkeä merkitys tieteellisen tiedon leviämässä, näistä yhteyksistä jää tietoa vain satunnaisesti ja epäsystemaattisesti. Lähdeviitteet sen sijaan ovat lähes rajaton »fossiilikokoelma» sille, joka etsii jälkiä tieteellisen tiedon ja ideoiden leviämisestä.

Periaatteessa kommunikaatiotutkija voi olla yhtä hyvin kiinnostunut ketkä ja missä viittaavat tiettyyn artikkeliin, kirjaan tai laajempaan julkaisujen joukkoon, kuin siitäkin mistä tietyn julkaisun tieto on peräisin. Kommunikaatiotutkimuksessa voidaan käyttää siis yhtä hyvin viittaus- ja lähdeanalyysiä. Löytyypä sellaisiakin tutkimuksia, joissa on kiinnitetty saman aikaisesti molempiin suuntiin huomiota ja päädytty jonkinlaiseen input-output-systeemiin (Inhaber ja Alvo 1978). Silloin kun viiteanalyysissä ollaan kiinnostuneita tieteellisestä kommunikaatiosta tavallisimmin kiinnitetään huomiota informaatiolähteesen vastaanottajanäkökulmasta. Tällöin lähdeanalyysin ydinongelma koskee sitä, mistä tutkija on vaikutteensa saanut ja miten tutkija ja tutkimus kytkeytyy tieteenalan kommunikaatioverkkoon vastaanottamansa informaation suhteen.

Lähdeanalyysin avulla voidaan tarkastella jossakin kiinnostuksen kohteena olevassa julkaisujoukossa käytettyjen lähteiden tiettyjä ominaisuuksia analysoimalla pelkästään lähteiden bibliografisia tietoja. Fitzgibbons (1980)

on esittänyt tieteenaloittain yhteenvedoja lähdekirjallisuuden julkaisutyypeistä, kielestä ja iästä. Joissakin tutkimuksissa on kiinnitetty huomiota myös julkaisumaahan ja lähteiden tieteenalaan.

Myös tieteellistä kommunikaatiota tutkitessa viiteanalyysi tutkimusaineistona on altis niille virhemahdollisuuksille ja epätarkkuuksille, joita edellä todettiin arviointifunktiota tarkasteltaessa. Tutkijoiden viittauskäytännön epäsystemaattisuus ja horjuvuus vaikeuttavat myös tieteellisten ideoiden välittymisen jäljittämistä.

Jos tieteellinen kommunikaatio ajatelleen yksinkertaisesti vain informaation siirtymiseksi formaalissa muodossa lähettäjältä — siis tutkijalta tai julkaisijalta — vastaanottavalle tutkijalle tai hänen ulottuvilleen, tarjoaa lähdeluetteloihin perustuva analyysi kohtalaisen validia materiaalia. Kertovathan lähdeluettelot niistä julkaisuista, joita tutkija on lukenut tai ainakin pitänyt kädessään sen verran, että on voinut merkitä muistiin tarvittavat bibliografiset tiedot. Mutta jos tieteellinen kommunikaatio ajatellaan informaation vastaanottamiseksi siinä määrin, että siitä näkyy jälkiä uudessa tutkimuksessa, lähdeluetteloihin perustuva viittausanalyysi osoittautuu hyvin epätarkaksi, jopa virheelliseksi. Asiaa valaisevat taulukon 1. esittämät numerot omasta tutkimuksestamme. Siinä väitöskirjojen lähteet sisällönanalyysin perusteella jaettiin tärkeätä ja vähemmän tärkeätä teoreettista, metodista ja empiiristä informaatiota välittäneisiin.

Jos pelkästään sosiologian väitöskirjojen lähdeluetteloiden perusteella olisi tehty päätelmiä tieteellisten vaikutteiden kulusta, olisi päädytty kotimaisen lähdekirjallisuuden keskeiseen asemaan. Tarkempi erittely kuitenkin osoitti, että merkittävin informaatio — tärkeä teoreettinen ja metodinen tieto — omaksuttiin varsin harvoin kotimaisista lähteistä ja useimmiten USA:ssa julkaistusta kirjallisuudesta.

Tämäntapaiset esimerkiksi sisällönerittelyyn perustuvat aineistot ovat varsin suuri-koisia eikä niillä voi korvata Science Citation Indexin tapaisia laajoja atk-pohjaisia tiedostoja. Oman työmme perusteella arvioisimme kuitenkin tarpeelliseksi sellaisen tutkimuksen lisäämisen, joka tutkijoiden viittauskäytäntöjä selvittelemällä ja lähdekirjallisuuden välittämää informaatiota erittelemällä täydentäisi perinteisen viiteanalyysin tarjoamaa tietoa. Tällainen reaalin tietä viiteindek-

Taulukko 1. Suomalaisten sosiologian väitöskirjojen teoreettisten, metodisten ja empiiristen lähteiden alkuperä (%)

Julkaisumaa	Teoreettiset lähteet		Metodiset lähteet		Empiiriset lähteet		Kaikki
	Tärkeät	Muut	Tärkeät	Muut	Tärkeät	Muut	
Suomi	13	21	14	31	49	51	39
Pohjoismaat	9	8	10	7	10	9	8
Englanti	13	13	8	7	8	8	9
Saksan Ltv (Saksa)	4	4	4	1	4	2	2
Ranska	1	2	0	1	1	1	1
Muu Länsi-Eurooppa	2	2	1	1	2	1	1
Neuvostoliitto	0	0	0	0	0	0	1
DDR	2	2	—	—	—	0	—
Muut sosialistiset maat	1	1	—	—	0	0	0
USA, Kanada	55	48	65	51	22	27	36
Muut	—	1	0	1	3	2	1
Yhteensä	100 %	100 %	100 %	100 %	100 %	100 %	98
N =	253	1762	52	635	182	3224	5497

(Aineistosta tarkemmin: Rautio ja Suhonen 1981, 5—15.)

sien luotettavuudesta, virhelähteistä ja niiden käyttömahdollisuuksien rajoista itse asiassa lisäisi niiden käytettävyyttä tieteellisen kommunikaation sekä tutkijoiden ja tulosten arvioinnin välineenä.

Lähteet

- Brittain, J. M., The Systemic Approach: Studies of Communication Artifacts. In *Information and Its Users*. Wiley-Interscience, New York 1970, 125—144.
- Broadus, R. N., The Literature of the Social Sciences: a Survey of Citation Studies. *International Social Science Journal*, 23 (2). 1971.
- Broadus, R. N., The Applications of Citation Analysis to Library Collection Building. *Advances of Librarianship*, 7: 299—335. 1977.
- Chubin, D. E. and Moitra, S. D., Content Analysis of References: Adjunct or Alternative to Citation Counting? *Social Studies of Science*, 1975 (5): 423—441.
- Cole, J., Measuring the Quality of Sociological Research: Problems in the Use of the Science Citation Index. *The American Sociologist*, 6. (February): 23—29. 1971.
- Cole, J. and Cole, S., *Social Stratification in Science*. The University of Chicago Press, Chicago and London, 1973.
- DISISS: The Structure of Social Science Literature as Shown by Citations. Design of Information Systems in the Social Sciences. Research Reports Series A. No 3. Bath University 1979.
- Fitzgibbons, S. A., Citation Analysis in the Social Sciences. In Stuart R. D. and Miller, G. B. (Jr.) (edit.): *Collection Development in Libraries*, A Treatise. Jai Press Inc., Greenwich, Connecticut, 1980.
- Garfield, E., *Essays of an Information Scientist*. ISI Press, Philadelphia, Vol. I—II 1977, Vol. III 1980.
- Garfield, E., Malin, M. V. and Small, H., Citation Data as Science Indicators. In Elkana, Y. et al. (edit.): *Toward a metric of Science: The Advent of Science Indicators*. John Wiley & Sons, New York 1978, 179—207.
- Ginman, M., Eräitä teoreettisia näkökohtia kirjallisuuden leviämistä ja valinnasta. *Lääketieteellisen keskuskirjaston julkaisuja No 2*, Helsinki 1975.
- Griffith, B. C., Drott, M. C. and Small, G., On the Use of Citations in Studying Scientific Achievements and Communication. *Society for Social Studies of Science Newsletter*, 1977 (2): 9—13.
- Haapala, A., Biologian tutkimustulosten välityminen kansainvälisiin sekundäärijulkaisuihin. *Luonnontutkija* 1972 (2): 45—48.
- Inhaber, H. and Alvo, M., World Science as an Input-Output System. *Scientometrics*, 1 (1): 43—64. 1978.
- Irvine, J. and Martin, B. R., Assessing Basic Research: Some Partial Indicators of Scientific Progress in Radio Astronomy. Paper Presented at the Science & Technology Conference, OECD, Paris 1980.
- Kaplan, N., The Norms of Citation Behavior: Prolegomena to the Footnote. *American Documentation*, 16 (3). 1965.
- Kiiski, K., Viiteanalyysin perusteet ja sovellutuksia sekä empiirinen analyysi suomalaisen psykologian, sosiologian ja sosiaalipsykologian käyttämästä kirjallisuudesta. Kirjastotieteen ja informatiikan pro gradu-tutkielma, Tampereen yliopisto 1977.
- Luukkonen-Gronow, T., Tiedeindikaattorit tutkimuksen arvioinnin välineenä. *Sosiologia* 1981 (3): 224—230.
- de Mey, M., Note on Citation Tracing of Key Publications as a Bibliometric Indicator of the Cognitive Dynamics of a Scientific Speciality. Paper presented at the Science & Technology Conference, OECD, Paris 1980.
- Moravcsik, M. J. and Murugesan, P., Some Results on the Function and Quality of Citations. *Social Studies of Science*, 1975 (5).

- Moravcsik, M. J. and Murugesan, P., Citation Patterns in Scientific Revolutions. *Scientometrics*, 1 (2). 1979.
- Rautio, P. ja Suhonen, P., Viiteanalyysi tieteellisen kommunikaation tutkimuksessa. Julkaisussa Suhonen, P. ja Rautio, P. (toim.) *Tieteellinen kommunikaatio ja tieteen kehitys, kolme artikkelia. Yhteiskuntatieteiden tutkimuslaitos, Kat-sauksia ja keskustelua*, No E 7. 1978.
- Rautio, P. ja Suhonen, P., The Use of Citation Analysis in the Study of Scientific Communication in the Social Sciences. Paper presented at the Science & Technology Conference, OECD, Paris 1980.
- Rautio, P. ja Suhonen, P., Yhteiskuntatieteiden tietovirrat ja suomalainen tutkija. Raportti sosiologian ja valtio-opin väitöskirjojen lähteistöstä. Suomen Akatemian julkaisuja 2/1981, Helsinki 1981.
- Roche, T. and Smith, D., Frequency of Citations as Criterion for the Ranking of Departments, Journals, and Individuals. *Sociological Inquiry*, 48 (1): 49—57. 1978.
- Satariano, W. A., Journal Use in Sociology: Citation Analysis versus Readership Patterns. *Library Quarterly*, 48 (3): 293—300. 1978.
- Subramanyam, K., Citation Studies in Science and Technology. In Stueart, R. D. and Miller, G. B. (Jr.) (edit.): *Collection Development in Libraries, A Treatise*. Jai Press Inc., Greenwich, Connecticut, 1980, 345—372.
- Sullström, R., Suomessa julkaistujen, etupäässä kansantaloustieteellisten väitöskirjojen lähteiden käytöstä. Helsingin yliopiston kansantaloustieteen laitoksen tutkimuksia, No 28, 1976.
- Suutarinen, P., Duodecim-seuran julkaiseman *Annales-sarjan* käytöstä ulkomailla. *Lääketieteellisen keskuskirjaston julkaisuja* No 1, Helsinki 1975.
- Suutarinen, P., Viittausanalyysi suomalaisten lääketieteellisten julkaisujen kansainvälisestä käytöstä. *Kansanterveystieteen julkaisuja* M 39/78, Tampere 1978.
- Tagliacozzo, R., Self-Citations in Scientific Literature. *Journal of Documentation*, 33 (4): 251—265. 1977.
- Weinstock, N., Citation Index. In Kent, A. (edit.): *Encyclopedia of Library and Information Sciences*, New York 1971.
- Wiener, J., Fotnotfetischen. Häften för kritiska studier, No 1/1978, 72—77.