

Kirja-arvio

Anna Rantasila

Yhteiskunnallisen teorian ja monimenetelmällisyyden puolustus

Lindgren, Simon. 2020. *Data Theory: Interpretive Sociology and Computational Methods*. Cambridge, UK: Polity Press, 198 s.

Kuten ympäröivä yhteiskunta myös tiede ja tutkimus dataistuvat kiihtyvällä tahdilla. Suurten aineistojen ja niiden tehokkaaseen käsittelyyn soveltuvien laskennallisten menetelmien kehittyessä joillain tahoilla on jo ehditty julistaa teoria tarpeettomaksi (esim. Spinney 2022; Anderson 2008). Näiden näkemysten mukaan yhä suurempi määrä empiiristä materiaalia ja kehittyneet laskennalliset menetelmät kertovat kaiken tarvittavan ihmisistä ja yhteiskunnasta.

Myös median ja viestinnän tutkimuksessa suuret aineistot ja laskennalliset menetelmät ovat olleet kasvava trendi vähintään viimeisen kymmenen vuoden ajan. Laskennalliset menetelmät ovat olleet eturintamassa erityisesti internetin ja sosiaalisen median kotimaisessa ja kansainvälisessä tutkimuksessa (esim. Haapoja, Laaksonen ja Lampinen 2020; Laaksonen ym. 2020, Toivanen ym. 2020). Suurissa aineistoissa ja laskennallisissa menetelmissä on paljon potentiaalia, mutta molempiin on suhtauduttu myös varauksella yhteiskunnan ja kulttuurin tutkimuksen laajoilla kentillä. Tähän paikoin varsin kiihkeänä vellovaan keskusteluun tarttuu Uumajan yliopiston sosiologian professori Simon Lindgren kirjassaan *Data Theory* (2020). Hän tarjoaa näkökulman siihen, kuinka yhteiskuntatieteellinen teoreettinen ajattelu, suuret aineistot ja laskennalliset menetelmät voitaisiin hedelmällisesti tuoda yhteen. Lindgrenin pääväite on, että huhut teorian kuolemasta ovat voimakkaasti liioiteltuja. Monimutkaistunut maailma kuitenkin vaatii hänen mukaansa klassikkoteorioiden ja metodologisen otteen melko reipasta raikastamista.

Data Theory koostuu seitsemästä varsinaisesta luvusta sekä erillisestä johdannosta ja päätösosasta, ja piirun alle 200 sivun pituudessaan teos on varsin helposti lähestyttävä ja tiivis esitys. Ensimmäisessä luvussa Lindgren esittelee kirjansa päätavoitteen.

Hän haluaa houkutellessa sosiologeja ja muita yhteiskuntatieteilijöitä astumaan ulos menetelmällisistä poteroistaan ja katsomaan rohkeammin monimenetelmällisen tutkimuksen ja laskennallisten menetelmien suuntaan. Lindgren (s. 13) esittää, että dataistuneen maailman ilmiöt ovat niin monimutkaisia, että ne harvoin taipuvat yhden menetelmän luupin alle, vaan edellyttävät eri menetelmien ja tutkimusotteiden yhdistämistä. Toisessa luvussa Lindgren käy läpi kirjan empiiristä ainesta yleisemmällä tasolla ja esittelee sosiaalisessa mediassa tapahtuvan poliittisen toiminnan peruskäsitteitä. Kirjan empiirisen kontekstoinnin ohella hän esittelee napakasti kolme sosiologian klassikkoa, Max Weberin, Émile Durkheimin ja Georg Simmelin, ja tarjoaa esimerkkejä siitä, kuinka heidän teorioitaan voi soveltaa 2020-luvun jatkuvassa käymistilassa olevaan sosiaalisen median ympäristöön ja ilmiöihin. Kolmas luku jatkaa tätä käsittelyä Robert Mertonin ja Elinor Barberin (1976) sosiologisen monimielisyyden (*sociological ambivalence*) käsitteen avulla.

Luvuissa neljä, viisi ja kuusi päästään tarkemmin käsiksi laskennallisten menetelmien, suurten aineistojen ja sosiologisen ajattelun yhteen sovittamiseen. Neljännessä luvussa Lindgren yhdistää toimijaverkkoteoriaa (*actor-network theory*, mm. Latour 2005) laskennallisiin kielen tutkimuksen menetelmiin analysoidessaan ilmastonmuutokseen liittyvää yli miljoonan twiitin aineistoa. Viidennessä luvussa hän puolestaan tuo yhteen vuoden 2018 aikana kaikista Reddit-palvelussa lähetetyistä viesteistä koostuvan aineiston Ernest Laclauin ja Chantal Mouffin (1985) ja Gilles Deleuzen ja Felix Guattarin (1987) käsitteellisten pohdintojen kanssa. Luvun laskennallisena menetelmänä esitellään luonnollisen kielen käsittelyn (*natural language process*, NLP) menetelmiin kuuluvaa Word2vec-tekstianalyysimenetelmää. Kuudes luku keskittyy analysoimaan Ruotsin vuoden 2018 valtiopäivävaaleihin liittyvää twiittimassaa Pierre Bourdieun (1977, 1986) sosiaalisten käytäntöjen ja symbolisen pääoman käsitteiden kautta. Luvussa nämä kaksi käsitettä muunnetaan Twitter-aineistosta mitattaviksi arvoiksi. Seitsemännessä luvussa Lindgren tarjoaa käytännönläheisen tutkimusprosessin mallin, jonka avulla sosiologisen teoreettisen ajattelun ja laskennallisten menetelmien yhdistelmää voi käyttää monenlaisissa empiirisissä asetelmissä.

Lindgren pohjustaa jokaisen luvun teoreettiset näkökulmat ja empiiriset esimerkit huolellisesti. Kirja toimii hyvänä kertauksena tai ensikosketuksena sosiologian klassikoteorioihin ja käsitteisiin sekä laskennallisiin menetelmiin. Mikäli jokin käsitellyistä teorioista tai menetelmistä on lukijalle entuudestaan tuttu, voi yksityiskohtainen selostus kuitenkin olla hieman puuduttavaa. Kirja on selkeä ja ilmaisultaan napakka, joten se on varsin helposti lähestyttävä monen tasoisille lukijoille. Kirjassa on useita kaavioita ja havainnekuvia, mutta ne ovat välillä hieman vaikeasti tulkittavia joko pienen kokonsa, mustavalkopainatuksen tai molempien takia.

Lindgren suhtautuu avoimen innostuneesti laskennallisiin menetelmiin, mutta kirja ei ole big data -hypeopus, vaan aihetta käsitellään sangen kriittisesti ja analyttisesti. Lindgren muun muassa huomauttaa varsin terävästi luvussa yksi, että kaikki aineisto, tuli se millaisesta lähteestä hyvänsä, on aina inhimillisen toiminnan suodattamaa. Niin sanottua raakaa dataa ei toisin sanoen ole olemassa (s. 11). Lindgren ei myöskään pidä laskennallisten menetelmien kautta saatuja tuloksia valmiina vastauksina tutkimuskysymyksiin, ja hän havainnollistaa tätä kirjassa esiteltyjen tapaustutkimusten avulla. Laskennalliset menetelmät ovat hänen mukaansa analyysin ensimmäinen vaihe, jonka jälkeen teoreettinen sosiologinen analyysi astuu mukaan kuvaan, jolloin laskennalliset tu-

lokset asettuvat yhteiskunnalliseen ja kulttuuriseen asiayhteyteensä (s. 33). Tällaiset näkemykset ovat kenties itsestään selviä monelle kriittistä median ja viestinnän tutkimusta tekeväälle lukijalle, mutta ne tuntuvat virkistäviltä ja tarpeellisilta huomioilta vasten edellä mainittuja teorian tarpeettomuutta julistavia puheenvuoroja.

Lindgrenin kirja saattaa vaikuttaa laskennallisten menetelmien puolestapuhujien mielestä suorastaan puolustelevalta, mutta hän ei käsittele perinteistä yhteiskunnallista tutkimustakaan silkkihansikkain. Lindgren haluaa rohkaista niin sanotusti perinteisiä yhteiskuntatieteilijöitä tarkastelemaan avoimemmin mielin sitä, millaisilla aineistojen, menetelmien ja teoreettisten käsitteiden tai viitekehysten yhdistelmillä tutkia dataistuneen nyky-yhteiskunnan monimutkaisia ilmiöitä. Tämä kuitenkin edellyttää hänen mukaansa ”ripausta teoreettista anarkismia” (s. 20) ja ”avoimen lähdekoodin lähestymistä teoriaan” (s. 89).

Teoreettisella anarkismilla ja avoimen lähdekoodin lähestymistavalla Lindgren tarkoittaa joustavaa ja luovaa tapaa käsitellä (klassikko)teoriaa suhteessa käsillä olevaan aineistoon ja tutkimuskysymyksiin. Lindgren (s. 88–90) väittää, että yhteiskunnallinen tutkimus suhtautuu teoriaan usein melko jäykästi keskittyen turhan paljon siihen, mikä valitun teorian olemus on tai miten sitä sovelletaan ”oikein”. Lindgrenin mukaan monimenetelmällistä tutkimusta usein kritisoidaan siitä, ettei tutkimus ota riittävästi huomioon hyödynnetyn teoreettisen viitekehysten kaikkia nyansseja. Tähän kritiikkiin vastaukseksi Lindgren (s. 90) tarjoaa Michel Foucaultin ajatteluun nojaavaa anarkistin asennetta: teorit on tehty hyödynnettäväksi, joten niitä tulisi lähestyä kysyen, mitä kyseisellä teorialla on tarjottavana tietulle tutkimukselle, eikä toisin päin.

Lindgren (s. 169) tiedostaa, että edellä kuvatun kaltainen ”brutaali” teorioiden lähestyminen vaatii keskustelua siitä, miten pitkälle teorioita voidaan venyttää ja soveltaa, ilman että niiden selitysvoima laimenisi. Hän kuitenkin vastustaa ajatusta teorioista muuttumattomina monoliitteina. Nojaten sekä Herbert Blumerin ja Pierre Bourdieun kaltaisiin klassikoihin että tuoreempaan sosiologian teorianmuodostusta koskevaan kirjallisuuteen Lindgren (s. 170–172) esittää, että dataistuneen maailman ilmiöitä tutkittaessa teoreettinen ja metodologinen puhdasoppisuus ansaitsee jäädä vähemmälle huomiolle. Sen sijaan Lindgren toivoo yhteiskunnan tutkijoilta avointa ja uskaliaista teorioiden ja menetelmien yhdistelyä, sillä aineiston ja teorian tulisi olla jatkuvassa vuoropuhelussa. Hänen mukaansa vain niin voidaan saada edes jonkinlainen ote jatkuvasti liikkeessä olevasta maailmasta, jossa asiat, ihmiset ja ilmiöt limittyvät toisiinsa yllättävillä tavoilla.

Vetävästi kirjoitettuna ja tiiviinä teoksena *Data Theory* voi suositella kaikille laskennallisista menetelmistä edes varovaisen kiinnostuneille yhteiskunnan, median ja viestinnän tutkijoille ja opiskelijoille. Teos esittelee yhteensä seitsemän erilaista tapaa yhdistellä suuria sosiaalisen median aineistoja ja sosiologian klassikkoteorioita. Näiden melko seikkaperäisten tapaustutkimusten lisäksi kirjaa voi lukea myös yleisempänä monimenetelmällisen tutkimuksen oppaana, sillä se sisältää monia mielenkiintoisia ja melko provokatiivisia huomioita menetelmien, käsitteiden ja tutkimusprosessin suhteesta. Erityisesti seitsemäs luku, *Theoretical I/O*, olisi varsin helposti sovellettavissa oppimateriaaliksi mille tahansa tutkimusmenetelmiä käsittelevälle kurssille. Varsinainen ohjekirja laskennallisen analyysin tekemiseen teos ei kuitenkaan ole, vaan se oppi pitää hakea muualta.

Data Theory on paitsi mainio johdatus laskennallisten menetelmien ja yhteiskunnallisten teorioiden yhdistämiseen, myös voimakas puheenvuoro monimenetelmällisen yhteiskuntatieteellisen tutkimuksen puolesta. Lindgren esittää varsin vakuuttavasti, että suuret aineistomassat ja laskennallisten menetelmien kehitys eivät suinkaan ole tehneet yhteiskuntatieteistä tarpeettomia, vaan itse asiassa entistä ajankohtaisempia. Hienoinkin laskennallinen malli kun vaatii tulkintaa, käsitteitä ja teoreettista viitekehystä tarjotakseen vastauksia. Jotta teoria kuitenkin pystyy kiinni muuttuvassa maailmassa, sitä tulee kuitenkin toisinaan ravistella ja venyttää rohkeasti.

Kirjallisuus

- Anderson, Chris. 2008. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*, 23.6.2008. Luettu 14.2.2022. <https://www.wired.com/2008/06/pb-theory/>.
- Bourdieu, Pierre. 1977. *Outline of a Theory of Practice*. Cambridge: Cambridge University Press.
- Bourdieu, Pierre. 1986. The Forms of Capital. Teoksessa *Handbook of Theory and Research for the Sociology of Education*, toimittanut John G. Richardson, 241–258. Westport, CT: Greenwood Press.
- Deleuze, Gilles ja Felix Guattari. 1987. *A Thousand Plateaus: Capitalism and Schizophrenia*. Minneapolis: University of Minnesota Press.
- Haapoja, Jesse, Salla-Maaria Laaksonen ja Airi Lampinen. 2020. Gaming Algorithmic Hate-Speech Detection: Stakes, Parties, and Moves. *Social Media + Society*, April–June 2020: 1–10. <https://doi.org/10.1177/2056305120924778>.
- Laaksonen, Salla-Maaria, Jesse Haapoja, Teemu Kinnunen, Matti Nelimarkka ja Reeta Pöyhtäri. 2020. The Datafication of Hate: Expectations and Challenges in Automated Hate Speech Monitoring. *Frontiers in Big Data* (3) 3. <https://doi.org/10.3389/fdata.2020.00003>.
- Laclau, Ernest ja Chantal Mouffe. 1985. *Hegemony and Socialist Strategy: Towards a Radical Democratic Politics*. London: Verso.
- Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network Theory*. Oxford: Oxford University Press.
- Merton, Robert K. ja Elinor Barber. 1976. *Sociological Ambivalence and Other Essays*. New York: Free Press.
- Spinney, Laura. 2022. Are we witnessing the dawn of post-theory science? *The Guardian*, 9.1.2022. Luettu 14.2.2022. <https://www.theguardian.com/technology/2022/jan/09/are-we-witnessing-the-dawn-of-post-theory-science>.
- Toivanen, Pihla, Jukka Huhtamäki, Katja Valaskivi ja Minttu Tikka. 2020. Aihemallinnus hybridin mediatapahtuman ja merkitysten kierron tutkimuksessa. *Media & viestintä* 43 (1): 1–20. <https://doi.org/10.23983/mv.91078>.