

Laskennalliset koneoppimisen menetelmät politiikan tutkimuksen kannalta: Institutionaalinen tarkastelu ja tieteenfilosofisen ja teoreettisen syventämisen tarve

PERTTI AHONEN

Lähestymistapa ja vastattavat kysymykset

Olen tarkastellut laskennallisia menetelmiä politiikan tutkimuksessa aikaisemmin (Ahonen 2015; 2018; Ahonen ja Wiberg 2018). Tässä esityksessäni paneudun kuitenkin vain sellaisiin laskennallisiin menetelmiin, joita luonnehtii tekoälyn tavallisin soveltamistapa, koneoppiminen. Sen sijaan minunkin toisessa yhteydessä käyttämäni nimike ”digitaaliset menetelmät” on niin yleinen, ettei se kelpaa tähän yhteyteen (ks. Ahonen 2018). Laskennallisen koneoppimisen menetelmien tekniset peruspiirteet ovat haarukoitavissa erikoiskirjallisuudesta (EMC 2015; DMI 2018; Ingersoll ym. 2014; O’Neill ja Schutt 2014; Wilkerson ja Casas 2016).

Aikaisemmista aihepiiriä koskevista esityksistäni poiketen tuen tätä esitystäni sillä teoreettisella juonteella, että harjoitan nyky-yhteiskunnan tiettyjen instituutioiden eräänlaista sosiaaliantropologiaa, enkä suinkaan ensimmäisenä tutkijana (Becher ja Trowler 2001; Meyer ja Rowan 1978; Trowler ym. 2012). Tieteenfilosofisesti paikannan esitykseni lähinnä ymmärtävän filosofisen fenomenologisen hermeneutiikan piiriin. Tarkastelen tieteen institutionaalisen ytimen suojelua, kun ytimen vartijat pyrkivät suojaamaan ydintä uhkilta rationalisoiduin myytein (rationalized myths) (Meyer ja Rowan 1977; 1978). Niissä on kysymys legitimiuden nimissä omaksutuista instituutioelementeistä kuten rakennemalleista ja toimintaprosessin kaavoista. Niihin viittaamalla instituutio ja sen ehdoin toimivat tähdentävät tekevänsä sitä mitä pitää pyrkien samalla retorisisilla tai muilla käytävissä olevilla keinoilla epäämään rationalisoidut myytit kiistävien haastajien väitteiden legitimitettiin. Tarkasteluuni tulee samalla instituutioiden taipumus torjua ja eristää muutosimpulsseja suojellakseen institutionaalista ydintään – ”kaikkein pyhintään”.

Muutosimpulssit ja -haasteet muodostuvat tässä tarkastelussani laskennallisista koneoppimisen menetelmistä ja niiden sovelluksista.

Instituutioiden yleiseen tapaan institutionalisointunein politiikan tutkimus tarkastelee kohdettaan, politiikkaa, institutionalisoitumisensa ehtojen mukaisen identiteettinsä mukaisesti toiseuttaen (*othering*) sitä, mikä poikkeaa tuosta identiteetistä (Barth 1982; Zerubavel 1995). Poliitiikan tutkimuksen identiteetin perusaineksina korostuvat ne sen osa-alueet, jotka on institutionaalisesti legitimoiden sinetöity tutkimus- ja opetusaluenimikkein sekä haarniskoitu vakanssirakenteiden pätevyyssuoritusmäärittäjin. Lisäksi tuota identiteettiä tosiasiallisesti määrittelevät toteutuneet rekrytoinnit etenkin toistaiseksi voimassa oleviin työsuhteisiin.

En ole liikkeellä provokatorisissa merkeissä argumentoiden institutionalisointuneimman politiikan tutkimuksen enkä sen haastajien puolesta tai kumpaakaan niistä vastaan. Hyväksyn sen ajatuksen, että kutakin tutkimuskohdetta koskeva institutionalisointunein tutkimus tarjoaa vakautta, jatkuvuutta ja ennustettavuutta. Haluan kuitenkin ottaa huomioon myös sen mahdollisuuden, että tuota tutkimusta saattavat vaivata kaavamaisuus, suvaitsemattomuus sekä kyyvyttömyys uudistua. Institutionalisointuneimman tieteen kohteisiin sen ulkopuolelta käsin kohdistuvan tutkimuksen mahdollisiin etuihin puolestaan saattavat kuulua esimerkiksi joustavuus, innovatiivisuus sekä tieteenharjoituksen repertuaarin monipuolistaminen ja uhkiin epävakaus, epäjatkuvuus ja vaikiutumattomuus.

Kaksi ensimmäistä tutkimuskysymystäni ovat seuraavat:

1. Miten laskennalliset koneoppimisen menetelmät asemoituvat institutionalisointuneimmissa kansainvälisessä politiikan tutkimuksessa?

2. Miten laskennalliset koneoppimisen menetelmät näkyvät Suomessa institutionalisoituneimmassa ja muussa politiikan tutkimuksessa?

Jätän tarkastelustani pois aineksen, jonka olen jo analysoinut aikaisemmissa teksteissäni (Ahonen 2018; Ahonen ja Wiberg 2018). Argumentointiketjujeni jatkuvuuden säilyttääkseni en kuitenkaan ole minimoinut sellaisia sisällöltään varsinaisesti avaamatta jättämiäni lähdeviitteitä, jotka saattavat informatiivisuudessaan johtaa lukijat muillekin jäljille kuin niille, joita olen itse tässä yhteydessä seurannut.

En kannata menetelmien tarkastelua erossa tutkimuksen filosofisista, teoreettisista ja muista yleisemmistä kysymyksistä. Niiden ottamiseksi huomioon, joskin lähinnä vain alustavasti, asetan kolmannen tutkimuskysymyksen vastatakseni siihen nimenomaan esitykseni lopulla:

3. Millaisille lähtökohdille olisi perusteltua pohjautua sellaisen politiikan tutkimuksen, jossa sovelletaan laskennallisia koneoppimisen menetelmiä?

Miten laskennalliset koneoppimisen menetelmät asemoituvat institutionalisoituneimmassa kansainvälisessä politiikan tutkimuksessa?

Politiikan tutkimuksessa on sovellettu monia samoja laskennallisia koneoppimisen menetelmiä kuin muidenkin alojen tutkimuksessa (Wilkerson ja Casas 2017). Olen kuitenkin itse halunnut tarkastella erityisesti laskennallisia koneoppimisen menetelmiä, joita on syntynyt politiikan tutkimuksen piirissä, joka on tältä osin arvioni mukaan suorastaan kunnostautunut.

Muuna kuin kontingenttina tapahtumasarjana tuskin voidaan pitää sitä, että erääksi ensimmäisistä laskennallisten koneoppimisen menetelmien käyttökentistä muodostui institutionalisoituneimmassa politiikan tutkimuksessa puolueohjelmatutkimus. Sen piirissä ei vain sovellettuja vaan myös suorastaan syntyneitä ohjattuja (*supervised*) ja ohjaamattomia (*unsupervised*) laskennallisen koneoppimisen menetelmiä olen tarkastellut muissa yhteyksissä (Ahonen 2018; Ahonen ja Wiberg 2018). Toiston välttämiseksi esitän tässä yhteydessä vain lähdeviitteitä sekä laskennallisen koneoppimisen soveltajien kritikoimasta että harjoittamasta puolueohjelmatutkimuksesta sekä valtavirtatutkijoiden ja laskennallisen koneoppimisen menetelmiä soveltaneiden ja kehitelleiden tutkijoiden debattien alalta (Ahonen 2015; Benoit

ym. 2016; Budge ja Laver 1993; Budge ja Pennings 2007; Klingemann ym. 2006; Laver ym. 2002; Laver ja Garry 2000; Lo ym., 2016; Lowe 2018; Proksch ja Slapin 2008).

Aikaisemmissa teksteissäni en vielä ottanut huomioon sitä tähdellistä seikkaa, että politiikan tutkimuksen piirissä on puntaroitu sitä, mitä laskennalliset koneoppimisen menetelmät ylipäättään voisivat tarjota politiikan tutkimukselle. Alkujaan laskennallisen koneoppimisen vakiomenetelmiä soveltanut Justin Grimmer (2010) on pohtinut, millä tavoin vakiintuneita tai muita laskennallisen koneoppimisen menetelmiä soveltaen voitaisiin tukea kausaalialanalyttistä tutkimusta sekä kokeellisessa että ei-kokeellisessa mielessä (Grimmer 2015). Grimmer itse on tutkinut laskennallisia koneoppimisen menetelmiä ja massadataa käyttäen seikkoja, jotka selittävät Yhdysvaltain kongressin jäsenten tyyliä kansanedustajina heidän poliittisissa julkilausumissaan (Grimmer 2013). Hän on osallistunut myös uusien menetelmien kehittämiseen (King ja Grimmer 2016). Olen aiemmin puntaroinut eräiden sosiologiain suurta innostusta laskennallisten koneoppimisen menetelmien johdosta (DiMaggio ym. 2013; Ruppert ym. 2013). Institutionalisoituneimman politiikan tutkimuksen edustajat ovat yleensä olleet konservatiivisempia. Grimmerin (2015) odotukset laskennallisten menetelmien avaamista mahdollisuuksista uudentyyppeeseen kausaalialanalyysiin politiikan tutkimuksessa tekstiaineistoja analysoitaessa kuuluvat sikäli poikkeuksiin.

Yleensä ei voida väittää laskennallisten koneoppimisen menetelmien lyöneen itseään läpi kansainvälisessä politiikan tutkimuksessa saati muodostuneen osaksi sen valtavirtaa. Mainittakoon, että myös yleiskäyttöisempien laskennallisen koneoppimisen menetelmännovaatioiden taustalta paljastuu politiikan tutkijoita, kuten laskennallisen aihehallintamisen (*structural topic modeling*) kehittäjien joukosta Margaret E. Roberts, Brandon M. Stewart ja Dustin Tingley (Roberts ym. 2014). He ovat julkaisseet myös yhdessä edellä mainitun Justin Grimmerin kanssa (Chuang ym. 2015; ks. myös Roberts ym. 2016).

Tunnen huolta siitä, että taidoiltaan tietojenkäsittelytieteilijöiden ja tilastotieteilijöiden uusille risteilyille datatieteilijöille vertoja vetävät politiikan tutkimuksen laskennallisten koneoppimisen menetelmien kehittäjät – kuten Sven-Oliver Proksch ja Jonathan B. Slapin, edellä mainitut Roberts, Stewart ja Tingley, joskaan eivät monialaisemmat Grimmer

ja Kenneth Benoit – jäisivät soljumaan tutkimuksen valtavirtaan nähden rinnakkaisessa mutta valtavirrasta erillisessä menetelmäkehittämisen kanavassa. Viittaamiani menetelmänkehittäjiä esiintyy huomattavasti sellaisissa yliopistoissa kuin Harvard, Stanford, Chicago, Princeton tai London School of Economics. Yliopistorankingien suhteellisuudesta huolimatta viestinsä on sillä, että noiden viiden yliopiston joukkoon sijoittuu Shanghai-rankingissa neljä seitsemästä politiikan tutkimuksen alalla korkeimmalle luokitetusta yliopistosta (Shanghai 2017). Niissä tuskin puuttuu varoja pitää yllä valtavirtatutkimuksen rinnakkaiskanavia toisin kuin tieteellisen toimintakyvyn alarajoilla kituuttavissa institutionaalisen ohuuden ja tutkija-opettajien pakotetun monitoimisuuden rasittamissa massayliopistoissa, paitsi Suomen kaltaisissa pienissä maissa myös muualla.

Miten laskennalliset koneoppimisen menetelmät näkyvät Suomessa institutionalisoituneimmassa ja muussa politiikan tutkimuksessa?

Olen jo aikaisemmin kiinnittänyt huomiota siihen kontingenttiin asiantilaan, että Suomen institutionalisoituneimman politiikan tutkimuksen piirissä laskennallisen koneoppimisen menetelmien soveltamisessa ehätti ensimmäiseksi Turun yliopiston valtio-opin professori Matti Wiberg ja toiseksi ilmeisesti minä (Ahonen 2015; Karlsson ja Wiberg 2010a; 2010b; Winter ja Wiberg 2016a; 2016b).

Tämän tekstin laadintaa varten nimenomaan tarkistin tietyn asian. Suomessa toimivien tutkijoiden toteuttamia laskennallisten koneoppimisen menetelmien sovelluksia puolueohjelmatutkimukseen ei ilmeisesti vielä ollut ilmestynyt ainakaan vuoden 2018 touko kuuhun mennessä.

En ole löytänyt institutionalisoituneimman politiikan tutkimuksen piirissä Suomessa toimivien tutkijoiden julkaistujen töiden joukosta suuremmalti laskennallisen koneoppimisen menetelmien sovelluksia. Tätä aikaisemmissa esityksissäni vaille kypsyttelyä jäänyttä kysymystä koskee ilmeisesti se, että noita menetelmiä käyttävät tutkijat pääasiassa ovat joko pitkällä väitöskirjatöissään mutta eivät vielä väitelleitä kuten Laura Sibinescu (Sibinescu 2016) ja Mari Marttila (yhtenä kirjoittajista artikkelissa Laaksonen ym. 2017) Helsingin yliopistossa tai väitöskirjaprosessinsa varhaisemmassa vaiheessa kuten pro gradu -tutkielmassaan Wordfish latent trait scaling

-menetelmää soveltanut Juha Koljonen samassa yliopistossa (Koljonen 2016).

Sitoudun aikaisempiin päätelmiin siitä, että laskennallisten menetelmien sovellukset politiikan tutkimukselle ominaisiin aiheisiin ovat Suomessa toistaiseksi selvästi monilukuisimmat institutionalisoituneimman politiikan tutkimuksen ulko- kuin sisäpuolella (Ahonen 2018; Ahonen ja Wiberg 2018). Tätä käsitystä vahvisti myös tätä esitystä varten toteuttamani aikaisempaa laajempi kirjallisuuden haku (Kekkonen 2018; Laaksonen ym. 2017; Nelimarkka ym. 2018; Purhonen ja Toikka 2016; Sormanen ja Dutton 2015; Ylä-Anttila 2017; 2018).

Suomen institutionalisoituneimman politiikan tutkimuksen piirissä laskennalliset koneoppimisen menetelmät ovat jääneet toistaiseksi vain harvojen senioritutkija-opettajien soveltamiksi. Menetelmien valtavirtaistumisesta on toistaiseksi aivan liian aikasta puhua. Nuoremmissa tutkijapolvessa on kuitenkin nousemassa lisää soveltajia. Kenties viimeistään osan heistä vakiintuessa postdoc-vaiheensa jälkeen tutkijanurille tapahtuu laskennallisten koneoppimisen menetelmien ainakin alustava valtavirtaistuminen Suomessa arviolta 2020-luvun puoleen väliin mennessä. Jossakin määrin on ilmestynyt yhteisjulkaisuja, joiden kirjoittajiin on kuulunut sekä institutionalisoituneimman politiikan tutkimuksen että muiden tutkimusalojen edustajia. Varsinaisesta tutkimusalojen sulautumisesta tuskin kuitenkaan on kysymys, vaikka etenkin politiikan tutkimuksen ja viestinnän tutkimuksen lähtymistä on pidetty jopa yleismaailmallisena trendinä (ks. Ahonen 2018).

Institutionalisoituneimmassa politiikan tutkimuksessa eriteltyjen aiheiden analysointi muilla tutkimusaloilla ei esiinny vakavana ensin mainitun tutkimusalan kilpailijana laskennallisia koneoppimisen menetelmiä sovellettaessa. Noita menetelmiä soveltavat institutionalisoituneimman politiikan tutkimuksen harjoittajat tuskin Suomessa kilpailevat samaisia menetelmiä soveltavien viestinnän tutkijoiden tai muiden tutkijoiden kanssa esimerkiksi sosiaalisen median poliittisten sisältöjen analyysissa. Muiden alojen tutkijoiden olisi myös liioiteltua väittää tunkevan tutkimuksen piiristä pois esimerkiksi poliittisia puheita tutkivia tai populistisia ilmiöitä eritteleviä institutionalisoituneimman politiikan tutkimuksen edustajia. Kummankin aiheen tutkimuksessa on tilaa hyvinkin monen alan tutkijoille.

Olisi väärä todistus väittää meistä institutionalisoituneimman politiikan tutkimuksen edustajista,

että olisimme muka laiskoja tai osaamattomia ajatellen laskennallisen koneoppimisen menetelmien soveltamista. Näkisin, että kysymys on tutkijoiden pienestä määrästä institutionalisoituneimman politiikan tutkimuksen piirissä, alan jakautumisesta kuuheen eri yliopistoon ja suurimpienkin alan yksiköiden jäädessä sittenkin kooltaan varsin pieniksi mutta silti hyvinkin jakautuneiksi sisäisesti (Ahonen 2018). Siksi on tuskin mikään ihme, että politiikan tutkimukselle ominaisia aiheita tutkitaan Suomessa myös alan nimenomaisimman kentän ulkopuolella ja julkaisuvalvolymin kannalta nimenomaan siellä.

Päätelmiä ja puntarointia

Päätelmiä tulosten perusteella

Kaksi ensimmäistä kolmesta tutkimuskysymyksestäni vaativat empiiristä tarkastelua. Ne olivat seuraavat:

1. Miten laskennalliset koneoppimisen menetelmät asemoituvat institutionalisoituneimmissa kansainvälisessä politiikan tutkimuksessa?
2. Miten laskennalliset koneoppimisen menetelmät näkyvät Suomessa institutionalisoituneimmissa ja muussa politiikan tutkimuksessa?

Institutionalisoituneimmissa kansainvälisessä politiikan tutkimuksessa on yletty itsenäiseen laskennallisten koneoppimisen menetelmien kehittämiseen. Menetelmät ovat kuitenkin edelleen pikemminkin kehittelyn kuin laajamittaisen soveltamisen saati valtavirtaistumisen vaiheessa.

Suomessa laskennallisten koneoppimisen menetelmien soveltaminen institutionalisoituneimmissa politiikan tutkimuksessa eli valtio-opin tutkimusalalla on toistaiseksi jäänyt rajoitetuksi. Soveltajina ovat pikemminkin esiintyneet politiikkaa eritelleet sosiologian ja sen lähialojen edustajat. Suomessa sen enempää institutionalisoituneimman politiikan tutkimuksen edustajat kuin heidän kollegansa muilla yhteiskuntatieteellisillä aloilla eivät toistaiseksi ole ylittäneet laskennallisten koneoppimisen menetelmien kehittäjiksi.

Omaksuin tämän tarkasteluni perspektiiviksi nyky-yhteiskunnan instituutioiden sosiaaliantropologian (Barth 1982; Becher ja Trowler 2001; Meyer ja Rowan 1977; 1978; Trowler ym. 2012; Zerubavel 1995). Tuosta perspektiivistä tarkastellen laskennal-

lisiä menetelmiä soveltavan tutkimuksen tulosten julkaisun esteeksi referoiduissa tieteellisissä lehdissä saati tuollaisen tutkimuksen valtavirtaistamisen tielle asettuu institutionalisoituneimmissa politiikan tutkimuksessa esteitä. Ne aiheutuvat valtavirtatutkimuksen institutionaalisen ytimen vakiintuneiden käytäntöjen mukaisista puolustustoimista kuten laskennallisia koneoppimisen menetelmiä soveltaneiden artikkelikäsitkirjoitusten hylkäämisestä käyttäen perusteluargumentteina myös seikkoja, joihin rakentuu rationalisoituja institutionaalisia myyttejä. Puolueohjelma-analyysin aihepiirissä käydyt debatit perinteisen kvantitatiivisen sisällönanalyysin tietyin muunnelman edustajien ja toisaalta laskennallisten koneoppimisen menetelmien kannattajien kesken ovat kuvaavia. Jälkimmäisten edellisille asettamat haasteet ovat ilmenneet suoranaisena pyhäinhäväistyksenä (Budge ja Pennings 2007), ja edes yhtä kelpoisten tulosten saavuttaminen uusin kuin perinteisin menetelmin on asetettu määrätietoisesti kyseenalaiseksi. Perinteisten menetelmien kiistaton kelpoisuus muodostaa tässä tapauksessa rationalisoidun myytin.

Kokemuksen mukaan laskennallisten menetelmien soveltaminen etenkin pienehköihin aineistoihin saattaa herättää sen torjuntareaktion, että väitetään – jopa tarkemmin yksilöimättömillä – laadullisilla menetelmillä helposti saavutettavan vähintään yhtä päteviä ja luotettavia tuloksia. Rationalisoidun myytin ytimenä ovat tällöin laadulliset eivätkä perinteiset kvantitatiiviset menetelmät. Vaihtoehtoisesti kenties vaaditaan, että rinnan laskennallisen koneoppimisen menetelmän kanssa tulisi soveltaa joko tiettyjä tai jopa yksilöimättömiä kvantitatiivisia menetelmiä, jolloin rationalisoitu myytti koostuu noista ideaaleiksi asetetuiksi menetelmistä. Niinkin saattaa käydä, että laskennallisia menetelmiä vaaditaan sovellettavan tiukasti tietyllä tavalla haluamatta sallia soveltamistavan joustoa esimerkiksi tutkimustarkoitusten tai sovellettujen teoreettisten tai empiiristen tutkimuslähtökohtien mukaan. Tällöin rationalisoitu myytti koostuu siitä käsityksestä, että kutakin metodia tulisi sallia käytettävän vain yhdellä tietyllä metodisen kaanonin mukaisella tavalla.

Laskennallisten menetelmien kehittämisen painotus institutionalisoituneimmissa kansainvälisessä politiikan tutkimuksessa paljastaa sivukanavan, joka on muodostunut vauraimmissa yliopistoissa tutkimuksen valtavirtojen rinnalle mutta niihin yhtymättä. Sekä sivukanava että tarkasteleman rationalisoidut myytit viittaavat siihen, että ainakin osa valtavirta-

tutkimuksen portinvartijoista kokee laskennallisten koneoppimisen menetelmien soveltamisen uhkana. Mikäpä muu selittäisi esiintyneitä jyrkkiä tuomioita, pyrkimyksiä poikkeuksellisen ankaran menetelmiä koskevan tilivelvollisuuden soveltamiseen sekä menetelmänkehittämisen sivukanavien muodostumista valtavrannalle mutta siitä näköjään turvallisesti eroon omiin uomiinsa.

Puntarointia tieteenfilosofiaa kaihtamatta

Ilmoitin kirjoitukseni aluksi, että vastustan menetelmien tarkastelua erossa tutkimuksen filosofisista, teoreettisista ja muista yleisemmistä kysymyksistä. Niiden tarkastelemiseksi asetin kolmannen tutkimuskysymyksen vastatakseni siihen esitykseni lopulla:

3. Millaisille lähtökohdille olisi perusteltua pohjautua sellaisen politiikan tutkimuksen, jossa sovelletaan laskennallisia koneoppimisen menetelmiä?

Laskennallisia koneoppimisen menetelmiä soveltavat institutionalisoituneimman politiikan tutkimuksen harjoittajat eivät juuri kunnostaudu tieteenfilosofisten lähtökohtiensa avaamisessa. Sama päätelmä soveltuu myös useimpiin politiikkaa tutkiviin viestinnän tai muiden sosiaalitieteiden tutkijoihin. Christian Fuchs (2017) kutsuu tavallisia tutkijoiden suhtautumistapoja kaikkea muuta kuin mairittelevasti digitaaliseksi positivismiksi sekä hallinnolliseksi massadata-analytiikaksi. Lisäksi se seikka, että laskennallisten koneoppimisen menetelmien soveltajien piirissä esiintyy vahvoja eksploratiivisen tutkimuksen painotuksia selittävän tai muun teorialähtöisen tutkimuksen asemesta (Pentland 2014), on omiaan saattamaan eksploraatiota pidemmälle tähtäävien tutkijoiden pyrkimykset tuhoavan kritiikin alaisiksi niiden pyrkimysten altistuessa väitteille siitä, että nuo menetelmät kelpaisivat muka yksinomaan vain aineiston eksploraatioon. Yksistään filosofisten metodologioiden runsaus (ks. esim. Moses ja Knutsen 2012) vaatii kuitenkin paneutumista myös peruskysymyksiin.

Eräänä näkyvänä vaihtoehtona laskennallisen koneoppimisen menetelmien ja niiden sovellusten tarkasteluun on esitetty vahvan konstruktionismin mukaisen ja jopa post-foundationalistisen eli jälkiperustahakuisen tieteenfilosofian mukaisia kantoja (boyd ja Crawford 2012; Ruppert 2014). Ongelmaksi jää tällöin kuitenkin se, että tietyt jo ennestään sovelletut kannat siirretään uuden kohteen tarkasteluun ilman, että laskennallisen koneoppimisen erityisyys saisi

osakseen sanottavaa huomiota. Vahva konstruktionismi ja jälkiperustahakuisuus löytänevät perustelluimmat soveltamiskohteensa siellä, missä noihin kantoihin sitoutuneisuus voi tukeutua vahvaan laskennallisen koneoppimisen menetelmien osaamiseen sen sijaan, että syntyisi vaikutelma niiden tieteenfilosofisten kantojen pelkästä päälleliimaamisesta tarkastelun kuorrutteeksi.

Marc Keuschniggin, Niclas Lövsjön ja Peter Hedströmin mukaan laskennallisten koneoppimisen menetelmien soveltajien keskuudessa ei ole ollut harvinaista esittää suurempaa filosofista ja teoreettista syvällisyyttä kuin olisi perusteltua (Keuschnigg ym. 2018). Kirjoittajien mukaan piilevien rakenteiden etsiminen ja löytäminen tekstiaineistosta ei sellaisenaan oikeuta väittämään, että eriteltävien tekstien taustalle sijoittuvat oleelliset poliittiset tai sosiaaliset syvärakenteet olisi onnistuttu tuomaan esiin, saati että tekstien sisältö olisi selitetty viittaamalla tuollaisiin rakenteisiin. He toivovat oleellisesti tähänastista paneutuneempaa perehtymistä filosofisiin ja teoreettisiin perusteisiin tutkimuksessa, jossa sovelletaan laskennallisia koneoppimisen menetelmiä. Mekanismpohjaisen tieteellisen selittämisen tutkijoina he puntaroivat nimenomaan tuon selitystavan soveltamisen mahdollisuuksia tutkimuksessa, jossa käytetään laskennallisen koneoppimisen menetelmiä. Omassa mielessäni mainitsemieni kirjoittajien pohdinnat herättävät sen jatkoajatuksen, että eikö laskennallisten koneoppimisen menetelmien soveltamisen tarjoamia mahdollisuuksia ja asettamia haasteita voitaisi tarkastella myös esimerkiksi filosofisen tieteellisen realismin tai kriittisen realismin katsantokantoja soveltaen (Porpora 2017). On kiinnostavaa seurata, laajenevatko mekanismpohjaisen selittämisen tutkijoiden kaavailemat tarkastelut ja syntykö luonnehtimiani tieteellisen tai kriittisen realismin viitoittamia pohdiskelujä.

LÄHTEET

- Ahonen, Pertti. 2015. Institutionalizing big data methods in social and political research. *Big Data & Society* 2:1, 1–12.
- Ahonen, Pertti. 2018. The situation of research and teaching in the discipline of political science – Finland: Facing the digital revolution. Artikkelikäsitkirjoitus IPSA:n (International Political Science Association) piirissä tekeillä olevaa toimitettua teosta varten. Helsinki: Helsingin yliopisto.
- Ahonen, Pertti ja Wiberg, Matti. 2018. Laskennallisten menetelmien mahdollisuuksia politiikan tutkimuksessa. *Politiikka* 60:1, 38–46.

- Barth, Fredrik. 1982. Introduction. Teoksessa Fredrik Barth (toim.), *Ethnic groups and boundaries: The social organization of cultural difference*. 2. uusintap., 1. p. 1969. Oslo: Universitetsforlaget, 9–38.
- Becher, Tony ja Trowler, Paul. 2001. *Academic tribes and territories: Intellectual enquiry and the culture of disciplines*. 2 p. London: Open University Press.
- Benoit, Kenneth, Conway, Drew, Lauderdale, Benjamin E. ja Laver, Michael. 2016. Crowd-sourced text analysis: Reproducible and agile production of political data. *American Political Science Review* 110:2, 278–295.
- Boyd, Danah ja Crawford, Kate. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15:5, 662–679.
- Budge, Ian ja Laver, Michael. 1993. The policy basis of government coalitions: A comparative investigation. *British Journal of Political Science* 23:4, 499–519.
- Budge, Ian ja Pennings, Paul. 2007. Do they work? Validating computerised word frequency estimates against policy series. *Electoral Studies* 26:1, 121–129.
- Chuang, Jason, Roberts, Margaret E., Stewart, Brandon M., Weiss, Rebecca, Tingley, Dustin, Grimmer, Justin ja Heer, Jeffrey. 2016. TopicCheck: Interactive alignment for assessing topic model stability. Julkaisussa *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*. Denver, CO: Association for Computational Linguistics, 175–184.
- DiMaggio, Paul, Nag, Manish ja Blei, David. 2013. Exploiting affinities between topic modeling and the sociological perspective on culture. *Poetics* 41:6, 570–606.
- DMI. 2018. Digital Methods Initiative. <https://wiki.digitalmethods.net/Dmi/DmiAbout>.
- EMC. 2015. *Data science and big data analytics: Discovering, visualizing and presenting data*. Indianapolis, IN: EMC Education Services.
- Fuchs, Christian. 2017. From digital positivism and administrative big data analytics towards critical digital and social media research! *European Journal of Communication* 32:1, 37–49.
- Grimmer, Justin. 2010. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis* 18:1, 1–35.
- Grimmer, Justin. 2013. *Representational style in Congress: What legislators say and what it matters*. Cambridge: Cambridge University Press.
- Grimmer, Justin. 2015. We are all social scientists now: How big data, machine learning, and causal inference work together. *PS* 48: 1, 80–83.
- Ingersoll, Grant, S., Morton, Thomas S. ja Farris, Andrew L. 2013. *Taming text: How to find, organize and manipulate it*. Shelter Island, NY: Manning.
- Karlsson, Fred ja Wiberg, Matti. 2010a. Puolueohjelmien kieliopillinen kompleksisuus. *Sananjalka* 52, 89–103.
- Karlsson, Fred ja Wiberg, Matti. 2010b. Puolueilla on eroja! Periaateohjelma-analyysi. *Politiikka* 52:1, 54–66.
- Kekkonen, Arto. 2018. Koulutuslupaus – koulutuspoliittinen kampanjointi ja keskustelu aihetunnisteella. Teoksessa Pekka Isotalus, Jari Jussila ja Janne Matikainen (toim.), *Twitter viestintänä: Ilmiöt ja verkostot*. Tampere: Vastapaino 2018, 111–133.
- Keuschnigg, Marc, Lövsjö, Niclas ja Hedström, Peter. 2018. Analytical sociology and computational social science. *Journal of Computational Social Science* 1:1, 3–14.
- King, Gary ja Grimmer, Justin. 2016. Method and apparatus for selecting clusterings to classify a data set. U.S. Patent 9519705, 13.12.2016. <https://gking.harvard.edu/publications/method-and-apparatus-selecting-clusterings-classify-predetermined-data-set>
- Klingemann, Hans-Dieter, Volkens, Andrea, Bara, Judith, Budge, Ian ja Macdonald, Michael. 2006. *Mapping policy preferences II: Estimates for parties, electors and governments in Central and Eastern Europe, European Union and OECD 1990–2003*. Oxford: Oxford University Press.
- Koljonen, Juha. 2016. Finnish industrial relations interest group policy positions between 2003 and 2015: Computer assisted text analysis for comparative politics in Finnish language. Yleisen valtio-opin pro gradu -tutkielma. Helsinki: Helsingin yliopisto.
- Laaksomen, Salla-Maaria, Nelimarkka, Matti, Tuokko, Mari, Marttila, Mari, Kekkonen, Arto ja Villi, Mikko. 2017. Working the fields of big data: Using big data-augmented online ethnography to study candidate-candidate interaction at election time. *Journal of Information Technology & Politics*, 14:2, 1–22.
- Laver, Michael, Benoit, Kenneth K. ja Garry, John. 2002. Extracting policy positions from political texts using words as data. *American Political Science Review* 97:2, 311–332.
- Laver, Michael ja Garry, John. 2000. Estimating policy positions from political texts. *American Journal of Political Science* 44:3, 619–634.
- Lo, James, Proksch, Sven-Oliver ja Slapin, Jonathan B. 2016. Ideological clarity in multiparty competition: A new measure and test using election manifestos. *British Journal of Political Science* 46:3, 591–610.
- Lowe, Will. 2018. Austin. Laskennallisten tekstianalyysimenetelmien tietokoneohjelmistovusto. <http://conjugateprior.org/software/austin/>
- Meyer, John W. ja Rowan, Brian. 1977. Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology* 83:2, 340–56.
- Meyer, John W. ja Rowan, Brian. 1978. The structure of educational organizations. Teoksessa Marshall Meyer (toim.), *Environments and organizations*. San Francisco, CA: Jossey-Bass, 78–109.
- Moses, Jonathan ja Knutsen, Torbjörn L. 2012. *Ways of knowing: Competing methodologies in social and political research*. 2.p. Houndmills: Palgrave.
- Nelimarkka, Matti, Laaksomen, Salla-Maaria, Marttila, Mari, Kekkonen, Arto, Tuokko, Mari ja Villi, Mikko. 2018. Influencing the news through social media: Online agenda building and normalization during a pre-electoral campaign period. To be presented at the 66th ICA Annual Conference.
- O'Neill, Cathy ja Schutt, Rachel. 2014. *Doing data science: Straight talk from the frontline*. Sebastopol, CA: O'Reilly.
- Pentland, Alex. 2014. *Social physics: How ideas turn into actions*. New York: Penguin Press.
- Porpora, Douglas V. 2017. Dehumanization in theory: Anti-humanism, non-humanism, post-humanism, and trans-humanism. *Journal of Critical Realism* 16:4, 353–367.
- Proksch, Sven-Oliver ja Slapin, Jonathan B. 2008. A scaling model for estimating time-series party positions from texts. *American Journal of Political Science* 52:3, 705–722.
- Purhonen, Semi ja Toikka, Arho. 2016. ”Big datan” haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät: esimerkitapauksena aiheallianaalyysi tasavallan presidenttien uuden vuodenpuheista 1935–2015. *Sociologia* 53:1, 6–27.

- Roberts, Margaret E., Stewart, Brandon M., Tingley, Dustin, Lucas, Christopher, Leder-Luis, Jetson, Kushner Gadarian, Shana, Albertson, Bethany ja Rand, David G. 2014. Structural topic models for open-ended survey responses. *American Journal of Political Science* 58:4, 1064–1082.
- Roberts, Margaret E., Stewart, Brandon M. ja Tingley, Dustin. 2016. Navigating the local mode of big data: The case of topic models. Teoksessa R. Michael Alvarez (toim.), *Computational social science: Discovery and prediction*. Cambridge: Cambridge University Press, 51–97.
- Ruppert, Evelyn, Law, John ja Savage, Mike. 2013. Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30:4, 22–46.
- Shanghai 2017. Shanghai Ranking's Global Ranking of Academic Subjects 2017 – Political Sciences. 28.4.2018. <http://www.shanghai-ranking.com/Shanghai-ranking-Subject-Rankings/political-sciences.html>.
- Sibinescu, Laura. 2016. Protest, citizen engagement and the co-production of political communication: The case of Romania's 2014 presidential elections. *Medijske Studije – Media Studies* 7:14, 142–156.
- Sormanen, Niina ja Dutton, William H. 2015. The role of social media in societal change: Cases in Finland of fifth estate activity on Facebook. *Social Media + Society* 1:2, 1–16.
- Trowler, Paul ja Saunders, Veronica. 2012. *Tribes and territories in the 21st century: Rethinking the significance of disciplines in the 21th century*. Abingdon: Routledge.
- Wilkerson, John ja Casas, Andreu. 2017. Large-scale computerized text analysis in political science: Opportunities and challenges. *Annual Review of Political Science* 20, 529–244.
- Winter, Lasse ja Wiberg, Matti. 2016a. Lainsäädäntövolyymin kvantitatiivinen tekstintutkimus. *Edilex* 2016/34. Helsinki: Edita.
- Winter, Lasse ja Wiberg, Matti. 2016b. Presidentin uudenvuodenpuheet: kvantitatiivisen tekstianalyysin mahdollisuuksia. *Politiikka* 58:1, 80–88.
- Volkens, Andrea, Bara, Judith, Budge, Ian, McDonald, Michael D. ja Klingemann, Hans-Dieter (toim.). 2013. *Mapping policy preferences from texts: Statistics solutions for manifesto analysts*. Oxford: Oxford University Press.
- Ylä-Anttila, Tuukka. 2017. The populist toolkit. Artikkeliväitöskirjan johdantoteksti. Helsinki: Helsingin yliopisto. <https://helda.helsinki.fi/bitstream/handle/10138/224199/ThePopul.pdf?sequence=1>.
- Ylä-Anttila, Tuukka. 2018. Populist knowledge: 'Post-truth' repertoires of contesting epistemic authorities. *European Journal of Cultural and Political Sociology*, OnlineFirst, 9.1.2018. DOI: <https://doi.org/10.1080/23254823.2017.1414620>
- Zerubavel, Eviatar. 1996. Lumping and splitting: Notes on social classification. *Social Research* 11:3, 421–433.