

MAIJA PAAVOLAINEN

Digitaaliset ihmistieteet huipulta huipulle?

HELDIG Summitissa

Digitaalisia ihmistieteitä luonnehtii kaksi lähestymistapaa. Toinen tarkastelee digitalisaatiota, kuten verkkoa ja viestintää, kulttuurintutkimuksen menetelmillä ja toinen tutkii perinteisiä humanististen tieteiden kysymyksiä ja aineistoja digitalisaation mahdollistamilla metodeilla. Kulttuurintutkimuksen menetelmät ovat viime vuosina korostaneet kontekstia, valtarakenteita ja marginaaleja, siinä missä digitaalista lähestymistapaa on leimannut keskeisten, tunnettujen aineistojen saattaminen digimuotoon ja uuden tarkastelun alle. Helsingin yliopiston pienessä juhlasalissa lokakuun puolivälissä järjestetty HELDIG Summit -tapahtuma esitteli digitaalisten ihmistieteiden tutkimuksen tilaa Suomessa.

Yhteen päivään oli mahduttu yli kahdeksankymmentä viiden minuutin esitystä posterit mukaan lukien. Tuntui siltä, että kaikki halusivat olla mukana digitaalisten ihmistieteiden maailmanvalloituksessa. Tutkimusryhmien lisäksi esillä oli kirjastoja, arkistoja, museoita ja julkishallinnon edustajia. Tapahtuma järjestetään jatkossa vuosittain, ja ensimmäisellä kerralla tavoitteena oli varmasti myös tutustuminen ja verkostoituminen.

Vuonna 2016 perustettu HELDIG-keskus kuuluu Helsingin yliopiston strategisiin kärkihankkeisiin, joita on tuettu Suomen Akatemian profiloitirahoituksella. Vaikka keskuksella ei vielä ole erikseen suunniteltuja tiloja, vuoden 2017 aikana on valittu yhteensä seitsemän *tenure track* -professoria eri tiedekuntiin.¹ Helsingin yliopistoon on perustettu digitaalisten aineistojen professori jo vuonna 2014, ja virkaa on vuodesta 2015 hoitanut Mikko Tolonen. Hän on käynnistänyt digitaalisten ihmistieteiden sivuaineopetuksen ja koonnut Helsingin yliopistoon pienen mutta aktiivisen ja kansainvälisen yhteisön digitaalisten ihmistieteiden ympärille.

HELDIG-keskuksen johtajaksi valittiin vuonna 2016 professori Eero Hyvönen Aalto-yliopistosta. Hyvönen on eri yhteyksissä korostanut verkostomaista toimintatapaa ja roolissaan SeCo-tutkimusryhmän vetäjänä profiloitunut monenlaisten aineistojen hyödyntäjänä yleisten kirjastojen luetteloista lakitekstiin.² Tapahtumapäivä kuvasti tätä pyrkimystä hyvin.

Perinteet, aineistot ja avoimuus

Päivän ohjelmassa olivat vahvasti edustettuina Helsingin yliopiston Nykykielten laitoksen tutkimusryhmät kor-

puslingvistiikkaan ja kieliteknologiaan liittyvine hankkeineen sekä Aalto-yliopiston linkitetyn datan ja semanttisen verkon tutkimukseen vuosia keskittynyt SeCoryhmä. Kieliteknologialla tarkoitetaan kirjoituksen tai puheen jäsentämistä koneella käsiteltävään muotoon ja kielikorpuksilla kieltä koskevien hypoteesien testaamista varten koottuja teksti- tai puheaineistoja. Semanttisilla verkoilla puolestaan kuvaillaan metatasolla verkkosäilytystä siten, että sisältöjä voi tulkita osin automaattisesti. Tämä lähestymistapa mahdollistaa eri sisältöjen yhdistelyn ja automaattisen rikastamisen isommiksi kokonaisuuksiksi, niin sanotuksi linkitettyä dataksi. Digitaalisten ihmistieteiden tutkimusmenetelmien kehittäminen Suomessa nojaakin vahvasti näihin kahteen suuntaan: yhtäältä tekstiaineistoja tutkitaan kielentutkimuksen menetelmillä, ja toisaalta aineistoja yhdistellään linkitetyn datan periaatteiden mukaisesti.

Digitaalisten aineistojen käyttäminen kielentutkimuksessa ei ole uutta. Esimerkiksi suomalainen FIN CLARIN -konsortio ja sen lähinnä suomenkieliset kielikorpuseineistot Kielipankki-palvelussa ovat olleet tutkijoiden käytössä jo kymmenisen vuotta. Samoin linkitettyä dataa on Suomessa tutkittu jo yli viisitoista vuotta. Kumpikin ryhmä tarjoaa vakiintuneena toimijana menetelmiään aineistojen käsittelyyn ja uusiin hankkeisiin.

Erityisessä asemassa digitaalisten ihmistieteiden tutkimuksessa ovat muistiorganisaatiot eli kirjastot, arkistot ja museot. Nämä organisaatiot ovat aina tarjonneet humanisteille tutkimusaineistoja, kuten painettuja kirjakokoelmia, käsikirjoituksia ja esineitä. Kokoelmien jatkuvasti etenevä digitointi muuttaa niiden käyttöä. Arkistoon tai kirjastoon ei tarvitse välttämättä fyysisesti mennä, vaan kokoelmia voi tarkastella omalta

päätelaitteelta dokumentti kerrallaan. Seuraava askel otetaan, jos kokonaisia kokoelmia lähdetään analysoimaan niiden rakenteisten piirteiden kuten metadatan tai tekstin osien avulla. Käytännössä tämä tarkoittaa esimerkiksi isojen tekstimäärien tutkimista tilastollisin menetelmin.

Samaan aikaan tutkimusrahoituksen haku muuttuu yhtenäisemmäksi eri tieteenaloilla, ja rahoittajat edellyttävät aineistonhallintasuunnitelmia myös humanisteilta. Jos aiemmin humanistit ovat voineet kuitata aineistoa koskevan suunnitelmansa sanalla ”kirjallisuus”, digitaalisten ihmistieteiden kaltainen tutkimus tuo tulevaisuudessa lisää humanisteja pohtimaan aineistojen käsittelyä, käyttöä ja jakamista. Opetusministeriön Avoin tiede ja tutkimus -hanke sekä Tieteen tietotekniikan keskus CSC toivat esiin, että humanistien kannattaa kuvailla, avata ja jakaa digitaalisia tutkimusaineistojaan.

Aineiston jakamisesta yleiseen käyttöön voi tulla uudenlainen tutkimuskontribuutio. Digitaalisten aineistojen saatavuuden parantuminen tukee osaltaan avoimen datan ja avointen menetelmien eli tutkimusta tukevien avoimen lähdekoodin tietokoneohjelmistojen käyttöä myös humanistisessa tutkimuksessa.³

Saatavuus ja menetelmät

Kiinnostavinta digitaalisten ihmistieteiden tutkimus on, kun uusien aineistojen ja menetelmien avulla tarkastellaan aiheita, jotka eivät muuten tulisi esiin. Valokuvaaja Tyyne Savia (1887–1959) kuvasi muotokuvia useille eri studioille vuosisadan vaihteesta kuolemaansa saakka, mutta hänen elämäntyötään ei ole tallennettu yhtenäisenä kokoelmana mihinkään museoon. Tunnetumpien kollegoiden töitä sen sijaan on säilytetty systemaattisesti negatiiveja myöten ja kuvailtu luetteloihin⁴. Helsingin kaupunginmuseon tutkija Satu Savia esitteli projektiaan, jonka tarkoituksena on koota Tyyne Savian kuvat yhteen eri puolilta maata ja tarkastella niiden valokuvallista ilmaisua. Tyyne Savia oli muotokuvavalaisemisen pioneeri ja ensimmäisiä sähkövalon käyttäjiä. Savian kohdalla kuvien paikasta riippumaton käyttö mahdollistaa uuden tutkimusaineiston kokoamisen. Vaikka digitaalinen versio toistaa alkuperäisen kuvan olennaiset piirteet, esimerkiksi jälkikäsitteilyn tarkasteluun kelpaavat vain alkuperäiset kuvat ja negatiivit. Voidaan ajatella, että materiaalisesta ja digitaalisesta kuvasta tutkitaan eri asioita, eikä digitointi siksi poista tarvetta säilyttää alkuperäisiä dokumentteja vaan ainoastaan avaa niihin toisen näkökulman.

Suomalaisia sanomalehtiä on ollut saatavana avoimesti verkossa jo jonkin aikaa Kansalliskirjaston tarjoamana⁵. Viime kevään Digital Humanities Hackathonissa syntynyt, opiskelijoista ja tutkijoista koostuva ryhmä selvitti, voiko algoritmin opettaa tunnistamaan sanomalehdissä ilmestynyttä kaunokirjallisuutta, joka on pääosin tähän asti jäänyt dokumentoimatta⁶. Voiko runoja erottaa muusta tekstistä kielen piirteiden avulla

koneellisesti? Mitä sanomalehtirunot käsittelevät? Ryhmä keräsi ensin lehdistä runoja käsin ja vertasi sen jälkeen tätä testidataa koko sanomalehtikorpuksen. Algoritmi näytti oppivan erottamaan runoja suhteellisen hyvin. Tätä suurten tekstimassojen tutkimista koneellisesti kutsutaan digitaalisissa ihmistieteissä kaukolukemiseksi, *distant reading*. Kaukolukemisen ja perinteisen lähilukemisen vuorottelu antaa löytyneille runoilta kontekstin ja auttaa niiden tulkinnassa.

Suomi24-keskustelupalstan tekstejä analysoinut Turun yliopiston tutkimusryhmä esitteli aihehallinnusmenetelmää eli *topic modelingia* ja niitä yhteyksiä, joissa sana ”sosiaalipummi” palstalla esiintyy⁷. Aihehallinnusta käytetään ryhmittelemään sanoja tietyssä tekstikokonaisuudessa ja nostamaan esiin yleisimpiä, tekstin sisällön kannalta keskeisiä ilmaisuja. Kielteisten tunneilmaisuksen lisäksi ”sosiaalipummiin” liittyvistä aiheista erottui kuluttamiseen ja säästämiseen liittyviä teemoja. Samoja keskustelupalsta- ja muita sosiaalisen median aineistoja tutkiva Helsingin yliopiston Citizen Mindscapes -projekti on laatinut erillisen datajulkaisun näiden aineistojen käytön helpottamiseksi. Julkaisu kuvaa tarkemmin aineistojen syntytapaa ja tallennusmuotoja sekä auttaa tutkijoita arvioimaan mahdollisia uusia tutkimussuuntia.⁸

Digitaaliset ihmistieteet muuttavat humanistisen tutkimuksen suuntaa, mikäli tutkimuskäytännöt alkavat osalla tutkijoista muistuttaa enemmän muiden tieteenalojen käytäntöjä. Jos humanistinen tutkimus on perinteisesti ollut yksittäisten tutkijoiden pitkäjänteistä tulkintatyötä kirjallisuuden tai arkistoaineiston parissa ja sen tuloksena on vuosien päästä syntynyt monografia, joka esittää jostain ilmiöstä kokonaistulkinnan, ryhmässä tutkiminen, aineistojen käsittely yhdessä ja artikkelijulkaisut ovat uusia työtapoja. Monitieteiset tutkimusryhmät, digitaaliset kvantitatiiviset aineistot eli data ja ohjelmistojen käyttäminen aineistojen analyysiin ovat arkipäivää luonnontieteissä ja mahdollistavat tulevaisuudessa uusia työtapoja kielentutkijoiden lisäksi myös historian ja taiteentutkijoille. Toisaalta digitaalisten ihmistieteiden on pidettävä kiinni humanistiselle tutkimukselle keskeisistä huolellisesta kontekstualisoinnista ja muutokselle avoimista tulkinnoista pysyäkseen osana ihmistieteiden traditiota.

Lyhyistä esityksistä koostuva pitkä päivä ilman keskustelulle varattua aikaa ei ole tyypillisin tapa esitellä humanistista tutkimusta. Esitysten muoto muistuttikin enemmän yrity maailmasta tuttuja *pitchejä* eli esittelypuheita. Toisaalta tapahtuman verkkosivuja ja sinne vietyjä abstrakteja voi todella katsoa hakemistona siitä, minkälaisia aloitteita digitaalisten ihmistieteiden piirissä Suomessa on käynnissä, ja etsiä yhteistyökumppaneita⁹. Lisää kansainvälisiä esimerkkejä digitaalisista ihmistieteistä saadaan maaliskuussa 2018, kun vuosittainen pohjoismaainen konferenssi Digital Humanities in the Nordic Countries rantautuu Helsinkiin yli sadan osallistujan voimin.¹⁰

Viitteet & Kirjallisuus

- 1 Eero Hyvönen, Digitaalisten ihmistieteiden keskus HELDIG käynnisti toimintansa (Digital Humanities Centre HELDIG Started Operations). *Tieteessä tapahtuu* 2/2017. Verkossa: journal.fi/tt/article/view/62348
- 2 Tutkimusryhmän kotisivu: seco.cs.aalto.fi/
- 3 Verkossa: portti.avointiede.fi/tutkimusmenetelmat/sosiaalinen-koodailu-ja-avoimen-tieteen-vallankumous
- 4 Suomalaisten museoiden digitoituihin valokuviin voi tutustua kulttuuriportaali Finna.fi:ssa.
- 5 Verkossa: digi.kansalliskirjasto.fi/
- 6 Tutkimusryhmän jäsenet: Mikko Koho, Agata Dominowska, Elsi Hyttinen, Péter Ivanics, Elizabeth Oakes, Ilona Pikkanen, Leena Tulkki & Risto J. Turunen; ks. myös helsinki.fi/en/helsinki-centre-for-digital-humanities/helsinki-digital-humanities-hackathon-2017-dhh17
- 7 Veronika Laippala, Lotta Lehti ym. Myös Krista Lagus kirjoittaa oivaltavasti sosiaalipummi-käsitteestä ja siitä, miten sitä on käytetty 2010-luvulla tietynlaisen moraalipaniikin lietsomiseen, ks. *Sosiaaliturvariippuvuus. Sosiaalipummit oleskeluyhteiskunnassa?* Toim. Juho Saari. Tampere University Press, Tampere 2017.
- 8 Krista Lagus, Mika Pantzar, Minna Ruckenstein & Marjoriikka Ylisiurua, Suomi24. Muodonantoa aineistolle. Kuluttajatutkimuskeskus 2016:10. Helsingin yliopisto, Helsinki 2016. Verkossa: blogs.helsinki.fi/citizenmindscapes/files/2016/05/257383_HY_VALT_suomi24_muodonantoa_aineistolle.pdf
- 9 Ohjelma ja abstraktit osoitteessa: helsinki.fi/en/helsinki-centre-for-digital-humanities/heldig-digital-humanities-summit-2017
- 10 Verkossa: helsinki.fi/en/helsinki-centre-for-digital-humanities/dhn-2018