

Kimmo Elo ja Olli Kleemola

Verkostoanalyysi historiallisten aineistojen eksploratiivisena analyysimenetelmänä: Esimerkkinä sotavalokuvat¹

Kimmo Elon ja Olli Kleemolan artikkeli havainnollistaa esimerkinomaisesti sosiaaliseen verkostoanalyysiin perustuvien tutkimusmenetelmien hyödyntämismahdollisuuksia historiallisten aineistojen eksploratiivisessa tutkimuksessa. Artikkelin primääriaineistona toimii otos kansallissosialistisen Saksan armeijan (*Wehrmacht*) propagandakomppanioiden Suomen alueella ottamista virallisista propagandakuvista. Analyysitulokset osoittavat, miten digitaalisten tutkimusmenetelmien avulla on mahdollista sekä ”louhia” aineistosta esiin uutta, relevanttia informaatiota että esittää tietoa uudella tavalla.

Viimeisten parinkymmenen vuoden aikana Internetin eksponentiaalinen kasvu ja digitaalisen maailman tunkeutuminen lähes kaikille inhimillisen toiminnan alueille on johtanut myös tutkimusaineistojen digitalisoitumiseen. Tällä hetkellä, hieman arvioista riippuen, 70–90 prosenttia uudesta informaatiosta syntyy Internetissä ja on siten lähtökohtaisesti digitaalista. Kun tähän di-

Dosentti, VTT Kimmo Elo on Turun yliopiston poliittisen historian ja Åbo Akademin saksan laitoksen yhteinen yliopistonlehtori. Sähköposti: kimmo.elo@utu.fi. VTM Olli Kleemola työskentelee Turun yliopiston poliittisen historian oppiaineen tohtorikoulutettavana. Sähköposti: owklee@utu.fi.

gitalisoitumiskehitykseen yhdistetään eri puolilla maailmaa käynnissä olevat tai jo valmistuneet historiallisten aineistojen (asiakirjat ja dokumentit, virallislähteet, kirjat, valokuvat ynnä muut) digitalisointihankkeet, digitaalisten ihmistieteiden menetelmät näyttäisivät olevan yhä relevantimpia myös niille tutkijoille, joiden käyttämät aineistot ovat syntyneet kauan ennen ensimmäistä tietokonetta. Samalla kehitys avaa tutkijoille uusia mahdollisuuksia, kun aiemmin kalliiden ja työläiden arkistovierailujen takana olleet aineistot ovat yhden tai kahden hiiren klikkauksen päässä.

Viime vuosien aikana ”digitaalisten ihmistieteiden” nimellä kulkeva tutkimussuuntaus on kasvattanut suosiotaan myös historiantutkijoiden parissa. Vaikka tutkimussuuntauksen määritelmä on edelleen osin kiistelty, digitaalisten ihmistieteiden voidaan laajasti ymmärrettynä katsoa pyrkivän sekä soveltamaan että kehittämään tietokoneavusteisia ja -pohjaisia analyysimenetelmiä ihmistieteille ominaisten kysymys- ja ongelmanasettelujen tutkimuksessa. Vaikka digitaaliset ihmistieteet käsitteenä onkin nuori ja vasta viimeisen vuosikymmenen aikana vakiintunut, sen juuret ovat tietokoneavusteisen tutkimuksen pitkässä traditiossa. Siinä missä historioitsijat ovat vasta vähitellen alkaneet löytää ja soveltaa digitaalisia tutkimusmenetelmiä, kielitieteet ja kulttuurintutkimus omaavat pitkän digitaalisen tradition.²

1. Kiitämme artikkelin käsikirjoituksen arvioitsijoita arvokkaasta palautteesta. Kiitämme myös Turun yliopiston poliittisen historian oppiaineen harjoittelijaa Saska Heinoa hänen avustaan artikkelissa käytetyn aineiston valmistelussa.

2. Ymmärrettävistä syistä digitaalisiin ihmistieteisiin tieteenalana liittyvät keskustelut keskittyvät vahvasti digitaalisiin foorumeille, mikä osaltaan vaikeuttaa kokonaiskuvan hahmottamista. Anna Haverinen ja Jaakko Suominen ovat tuoreessa katsausartikkelissaan koonneet ansiokkaasti yhteen kansainvälistä keskustelua digitaalisen humanismin ympärillä sekä pohtineet digitaalisen humanismin suhdetta ”perinteiseen” humanismiin. Ks. Anna Haverinen & Jaakko Suominen, Koodaamisen ja kirjoitta-

Tämän artikkelin kirjoittajat katsovat digitaalisten ihmistieteiden – jonka sateenvarjon alle sijoitamme humanistien ohella myös yhteiskuntatieteet – menetelmien avaavan historian tutkimukselle aiempaa parempia mahdollisuuksia eksploraatiivisen tutkimuksen tekemiseen. Eksploraatiivisella tutkimuksella tarkoitamme tutkimusta, jonka tavoitteena on kehittää uusia tapoja lähestyä olemassa olevia aineistoja uusien kysymysten, näkökulmien ja tutkimusalueiden löytämiseksi.³ Tämän ajatuksen mukaisesti artikkelimme tavoitteena on osoittaa esimerkinomaisesti, miten sosiaaliseen verkostoonalyysiin (engl. *social network analysis*, SNA) perustuvia tutkimusmenetelmiä voidaan hyödyntää historiallisten aineistojen eksploraatiivisessa tutkimuksessa. Artikkelin empiirisenä aineistona käytämme kansallissosialistisen Saksan armeijan *Wehrmachtin* propagandakomppanioiden Suomen alueella ottamia virallisia propagandakuvia (N=215), jotka on arkistoitu osana toisen maailmansodan aikaisen Scherl-kuvatoimiston kokoelmaa.

Artikkelissa oleva aineisto on lukumääräisesti vähäinen, sillä se on valittu ensisijaisesti menetelmäteknisistä syistä. Artikkelin ensisijaisena tavoitteena on havainnollistaa digitaalisten tutkimusmenetelmien soveltuvuutta myös sellaisten historiallisten aineistojen tutkimiseen, joiden valmistelu analyysiä varten edellyttää ainakin kohtuullisessa määrin tutkijan omaa käsityötä ennen kuin analyysissä voidaan hyödyntää digitaalisia tutkimusmenetelmiä. Anna Haverisen ja Jaakko Suomisen katsauksessaan hahmotteleman nelikentän mukaisesti katsomme artikkelimme suhteen digitaalisuuteen olevan yhtäältä tekemistä ja kehittämistä painottava, toisaalta välineellinen⁴. Ensimmäinen siksi, että artikkelimme keskiössä ovat historiallisista lähteistä muodostettavien määrällisten digitaalisten aineistojen analyysitavat ja näiden soveltuvuuden arviointi. Jälkimmäinen puolestaan siksi, että sovellamme digitaalisia analyysivälineitä sekä löytääksemme uusia näkökulmia aineistoon

(eksploraatiivisuus) että rakentaaksemme menetelmiä hyödyntämällä tietynlaista sotahistoriallista tulkintaa aineistomme pohjalta. Artikkelimme punaisena lankana on siten ajatus, jonka mukaan tutkimusmenetelmillä on olennainen vaikutus siihen, millaista tietoa aineistosta on ylipäätään mahdollista saada esille ja millaisia tulkintoja aineistosta voidaan esittää. Toisin sanoen se, *miten* katsomme vaikuttaa olennaisesti siihen, *mitä* näemme. Digitaaliset ihmistieteet tutkimusalan kehittää ja tarjoaa uusia välineitä katsoa materiaaleja ja siten vaikuttaa olennaisesti siihen, miten aineistojamme katsomme – ja siten luonnollisesti vaikuttaa myös siihen, mitä me aineistoissamme ja niiden avulla näemme.

Artikkelin aluksi esittelemme tarkemmin käytettävän empiirisen aineiston sekä tutkimusmenetelmänä sovellettavan historiallisen verkostoonalyysin. Kolmas alaluku koostuu itse analyysistä, jonka puitteissa tarkastelemme sotakuvakorpuksen temaattista rakennetta sekä havainnollistamme, miten metatietoihin⁵ liittyviä tietoja voidaan hyödyntää verkostoonalyysiin tukeutuvassa sisällöntutkimuksessa. Artikkelin lopuksi pohdimme saatuja tuloksia ja niiden merkitystä laajemmin digitaalisten ihmistieteiden kontekstissa.

misen vuoropuhelu? Mitä on digitaalinen humanistinen tutkimus. *Ennen ja Nyt* 19.5.2015, <http://www.ennenjanynt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/> (01.06.2015).

3. Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner & Jeffrey Schnapp, *Digital Humanities*. The MIT Press 2012.

4. Ks. tark. Haverinen & Suominen 2015.

5. Metatiedolla, jonka synonyymeinä käytetään usein termejä metadata, liitännäistieto tai kuvailutieto, tarkoitetaan kuvailevaa ja määrittävää lisätietoa. Tämä ”tieto tiedosta” pyrkii standardisoimaan eri tietoyksiköiden käsittelyä luomalla kaikille samaan tietokokoelmaan kuuluville tiedoille yhtenäiset lisätiedot. Esimerkkejä metatiedosta ovat esimerkiksi CD-levyihin liittyvät kuvailutiedot (levyn nimi, esittäjä, julkaisuvuosi) tai kirjastotietokantonjen teostietoihin sisältyvät (tekijä, julkaisija, painos, asia-sanat).



Kuva 1. Esimerkki aineistoon sisältyvästä kuvamateriaalista. Kuvassa skannaamalla digitalisoidun valokuvan etupuoli. Lähde: Bundesarchiv, Bild 183, B13382.

Sotavalokuvat aineistona

Artikkelin primaariaineisto koostuu saksalaisesta, Suomeen liittyvästä sotapropagandakuva-aineistosta. Analysoitavat Saksan sodanaikaisten propagandayksiköiden (*Propagandakompagnie*, lyh. PK) ottamat kuvat ovat peräisin Scherl-kuvapalvelun kuva-arkistosta, jota säilytetään Saksan liittoarkistossa (*Bundesarchiv*) arkistotunnuksella ”Bild 183”. Scherl-kuvapalvelun perusti vuonna 1883 saman niminen kustantamo, ja se ehti vuoteen 1945 mennessä koota noin 1,2 miljoonan kuvan arkiston. *Bundesarchivin* Scherl-kokoelmaan sisältyy alkuperäisvedokset kaikista noin 300 000 PK-kuvasta, joista noin 15 000–20 000 on otettu itärintamalla. Vuonna 1945 arkisto, joka Berliinin pommitusten vuoksi oli hajasijoitettuna, joutui osin amerikkalaisten ja osin neuvostoliittolaisten sotasaaliiksi. Neuvostoliiton sotasaaliiksi saama suurempi osa arkistoa

luovutettiin myöhemmin DDR:lle, missä se muodosti tärkeän osan *Zentralbild*-kuvapalvelun arkistoa. Viimeksi mainittu oli vuodesta 1956 lähtien osa *Allgemeiner Deutscher Nachrichtenagentur* (ADN) -uutistoimistoa. Kuvapalvelu ajautui Saksojen yhdistymisen jälkeen konkurssiin, jolloin kuva-aineisto siirrettiin *Bundesarchivin* haltuun.⁶ Amerikkalaisten sotasaaliiksi saama osa puolestaan on nykyään *Süddeutscher Verlag* -kustantamon hallussa, eikä kyseistä aineistoa luovuteta tutkimuskäyttöön kuin poikkeustapauksissa.

Tässä artikkelissa analysoitavat Bild 183 -kokoelman kuvat on luokiteltu aihepiireittäin. Artikkelissa analysoitavat kuvat (N=215) olemme

6. Rainer Hofmann, *Das Bildarchiv der früheren DDR-Nachrichtenagentur ADN. Mitteilungen aus dem Bundesarchiv* 1 (1993), 23–25.



Kuva 2. Esimerkki aineistoon sisältyvästä kuvamateriaalista. Kuvassa skannaamalla digitalisoidun valokuvan taustapuoli, jossa näkyy kuvavedoksen taustalle liimattu "saateliуска" (Bildbegleitzettel), josta käyvät ilmi kuvan ottoajankohta, kuvaajan nimi, kuvateksti ja sensuurimerkinnt. Lähde: Bundesarchiv, Bild 183, B13382.

valinneet aineistosta laajan temaattisen otannan perusteella käymällä läpi ne tema-alueet, joiden oletimme kytkeytyvän Suomeen. Otannan ensimmäisessä vaiheessa kävimme läpi ne kuvat, jotka *Bundesarchivin* aineistoluokittelun mukaan on otettu itärintamalla. Näiden joukosta poimimme otannan toisessa vaiheessa sellaiset kuvat, jotka alkuperäisten kuvatekstien sisältämien paikanmäärittysten perusteella sijoittuivat Suomen alueelle. Käytimme valinnassa paikanmääreitä, jotka viittasivat itsestään selvästi vuoden 1939 rajojen sisällä sijainneisiin paikkakuntiin (Helsinki, Petsamo, Salla, Viipuri), mutta myös viittauksia laajempiin alueisiin kuten Pohjois-Suomeen. Esimerkki tutkimusaineistoon sisältyvästä alkuperäisestä kuvamateriaalista on esitetty oheisissa kuvissa (Kuva 1 ja 2).

Historiallinen verkostoaalyysi tutkimusmenetelmänä

Historiantutkimuksessa kiinnostus kohdistuu useimmiten asioiden, tapahtumien, ilmiöiden ja ihmisten välisiin vuorovaikutussuhteisiin, näistä muodostuviin kokonaisuuksiin sekä vuorovaikutusrakenteessa ajan mukana tapahtuviin muutoksiin. Erityisesti viime vuosien aikana sosiaaliiseen verkostoaalyysiin perustuva historiallinen verkostoaalyysi on lisännyt suosiotaan myös historiantutkijoiden keskuudessa ja siten osataan vahvistanut digitaalisten tutkimusmenetelmien asemaa. Historiallista verkostoaalyysiä on sovellettu lähinnä välineellisistä lähtökohdista, kun menetelmää on hyödynnetty erilaisten historiallisten vuorovaikutusverkostojen analysoinnissa ja graafisessa esittämisessä (visuali-

sointi).⁷ Historiallisen verkostoaalyysin voi- daankin nähdä tuoneen historiantutkimukseen uuden tavan hahmottaa ja analysoida toimijoi- den (ihmiset, asiat, valtiot ja niin edelleen) väli- siä suhteita ja vuorovaikutusta, näiden seurauk- sia sekä vuorovaikutussuhteiden merkitystä laa- jemmalle tapahtumahistorialle.

Verkostoaalyysin formaalina perustana on graafiteoria, jonka mukaan verkostolla tarkoite- taan pisteistä (*nodes, vertices*) ja niitä yhdistävi- tä viivoista (*edges*) muodostuvaa kuviota eli graafia, jolla on tietty rakenne eli topologia.⁸ Sekä pisteisiin että viivoihin voidaan liittää niiden ominaisuuksia tarkentavia määreitä (attribuu- tit), joita voidaan edelleen hyödyntää varsinais- ssa analyysissä. Verkoston eri osia tällä tavoin rikastamalla verkostoa ja sen rakennetta on mahdollista tarkentaa vastaamaan mahdolli- simman hyvin kohteena olevaa ilmiötä.

Vaikka verkostoaajattelun juuret ulottuvat aina 1800-luvun loppupuolelle⁹, vasta tietokonei- den yleistyminen ja erityisesti niiden laskentate- hon kasvu sekä verkostojen graafiseen esittämi- seen eli visualisointiin tarkoitettujen helppo- käyttöisten ohjelmistojen tarjonnan laajenemi- nen on lisännyt humanististen ja yhteiskuntatie- teiden kiinnostusta verkostoaalyysiä kohtaan. Kehitystä ovat tukeneet ne tietoteknisesti orien- toituneet digi-ihmistieteilijät, jotka ovat pyrki- neet räätälöimään perinteisen verkostoaalyy- sin menetelmiä soveltumaan paremmin ihmis- tieteille relevanttien kysymysten tutkimiseen.¹⁰ Tieteenalan erityispiirteet huomioiden ei liene kovinkaan yllättävää, että suurin kiinnostus on kohdistunut tekstiaineistojen kannalta relevant- tien tutkimusmenetelmien – esimerkiksi teks- tinlouhinta tai semanttinen analyysi – edelleen kehittämiseen.

Tässä yhteydessä on syytä korostaa, että vaika verkostoaalyysi menetelmänä analysoi, mal- lintaa ja visualisoi tutkimusaineistoon sisältyviä vuorovaikutussuhteita, tutkimuskohteen itses- sään ei tarvitse (mutta se toki voi) olla verkosto. Monissa tutkimuksissa verkostoaalyysiä sovel-

letaan samalla eksploratiivisella perusajatuksel- la kuin tässä artikkelissa eli sovelletaan verkos- toanalyysimenetelmiä aineistossa piilevien yh- teyksien löytämiseksi, analysoimiseksi ja graafi- seksi esittämiseksi. Vaikka tämä toki edellyttää aineiston valmistelua muuttamalla aineisto ver- kostaalyysin edellyttämään määrälliseen muotoon sekä mahdollisesti aineiston rikasta- mista muista lähteistä saaduilla aineistoilla, ai- neistoon sisältyvien suhteiden analysoinnin tuli- si tapahtua soveltamalla verkostoaalyysimene- telmiä.

Artikkelin aineisto valmisteltiin analyysiä varten erottamalla kuvien metatiedoista kaksi kokonaisuutta. Ensinnäkin alkuperäiset saksan- kieliset kuvatestit käännettiin ja niistä muodos- tettiin oma taulukko. Tämän jälkeen taulukon teksti prosessoitiin tarkoitusta varten laaditun Python-ohjelman avulla¹¹, joka erotti kuvateks- teistä yksittäiset sanat ja palautti ne perusmuo- toonsa (*stemming*). Saatu sanalista käytiin lo-

7. Viime vuosien aikana historiallista verkostoaalyysiä on käytetty mm. natsihallinnon aikana maanpakoon läh- teneiden taiteilijoiden välisten verkostojen, talouseliitin verkostoitumisen, keskiaikaisten kauppiassukujen, kyl- män sodan aikaisten tiedusteluverkostojen tai historial- listen tekstien sisäisten tai kirjoitusten välisten yhteyk- sien tutkimuksessa. Kattava ja ajantasainen bibliografia historiallista verkostoaalyysiä soveltavista tutkimuksis- ta on luettavissa kansainvälisen *Historical Network Rese- arch* -tutkimusverkoston sivuilta, <http://historicalnet- workresearch.org/bibliography/> (02.06.2015).

8. Hyvänä johdatuksen graafiteoriaan ks. Keijo Ruohon- en, *Graafiteoria*. Tampereen teknillinen yliopisto 2013, <http://math.tut.fi/~ruohonen/GT.pdf> (1.10.2015). Verkos- toaajattelusta ja verkostoaalyysista laajemmin ks. esim. Christina Prell, *Social Network Analysis. History, theory and methodology*. SAGE 2012; Kai Eriksson (toim.) *Verkostot yhteiskuntatutkimuksessa*. Gaudeamus 2015.

9. Ks. Eriksson (toim.) 2015.

10. Hyvänä katsauksena tähän kehitykseen ks. Barbara Schultz-Jones, Examining information behavior through social networks. An interdisciplinary review. *Journal of Documentation* 66:4 (2009), 592–631.

11. Python-ohjelmointikieli on laajalle levinnyt, moni- puolinen ja melko helposti opittavissa oleva ohjelmointi- kieli. Hyvä johdatus Python-kieleen ks. esim. <https://www.cs.tut.fi/~jkorpe/python/> (16.9.2015).

puksi manuaalisesti läpi ja lopulliseen analyysiin poimittiin ne asiasanat, jotka arviomme mukaan olivat keskeisimmät kuvien sisällön tematiikan kannalta. Tämä oli välttämätöntä asiasanajoukon koon rajaamiseksi sekä irrallisten asiasanojen analyysiin aiheuttaman taustakohinan minimoimiseksi. Kuvateksteihin perustuvan asiasanaluettelon lisäksi laadimme toisen asiasanaluettelon, joka perustui kuvien manuaaliseen tarkasteluun ja sen pohjalta toteutettuun kuvien asiasanoittamiseen. Tämän menettelyn tavoitteena oli luoda edellytykset arvioida alkuperäisiin kuvateksteihin sisältyvien asiasanojen luotettavuutta tietokoneavusteisissa analyysissä. Toiseksi kuvien metatiedoista kerättiin tiedot kuvien ottopaikoista niiden kuvien osalta, joiden kohdalla kuvauspaikka oli yksilöity. Nämä paikatiedot rikastettiin kuvauspaikkojen sijaintikoordinaateilla aineiston maantieteellisen projisoinnin mahdollistamiseksi.¹²

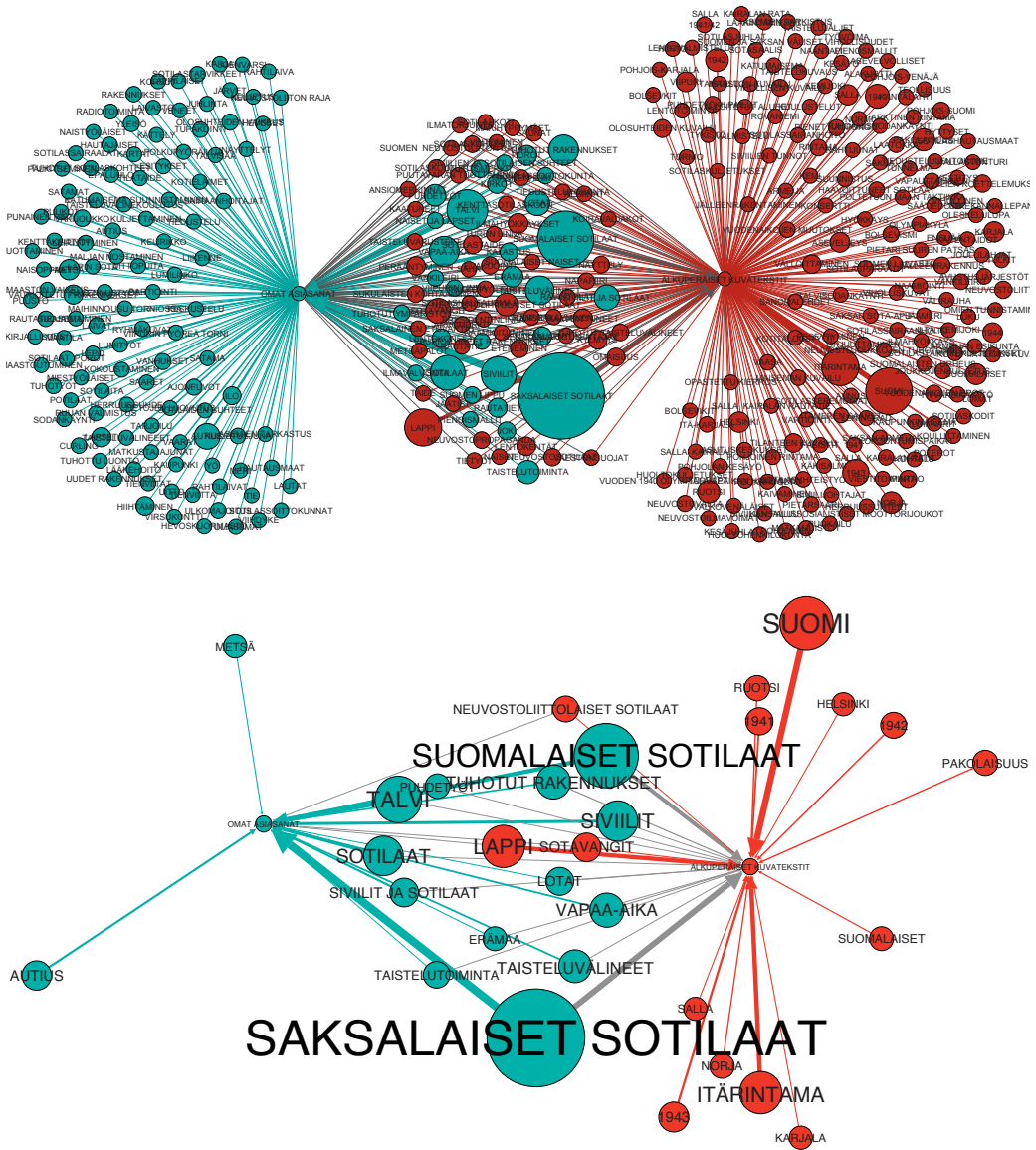
Sotavalokuvien verkostanalyysi

Seuraavassa analysoimme sotakuvakorpukselta kerättyä metadataa kolmesta eri näkökulmasta. Analyysin tavoitteena on osoittaa, millä tavoin historiallista verkostanalyysiä voidaan hyödyntää eräänlaisena laadullisen sisällönanalyysin laajennoksena, jonka avulla on mahdollista havainnoida, analysoida ja visualisoida aineistoon sisältyviä piileviä rakenteita. Analyysimme punaisena lankana ovat aineistoa eri näkökulmista kuvaavat visualisoinnit, jotka samalla tiivistävät ja esittävät graafisesti aineistosta esiin louhittua tietoa. Verkostanalyysissä juuri verkoston graafinen esittäminen vaikuttaa hyvin vahvasti siihen, millä tavoin tutkija ylipäätään voi havainnoida aineistoon sisältyviä piirteitä. Kaikki visualisoinnit perustuvat erityisten visualisointimallien (*layout*) käyttöön, joista kaikki painottavat hieman eri tekijöitä verkoston graafiseksi esittämiseksi ja siten tuottavat *samasta* datasta *erilaisia* verkostoja. Verkostanalyysit ja -visualisoinnit on toteutettu Gephi-verkostanalyysi-ohjelmistolla.¹³

Aloitamme tarkastelun vertailemalla alkupe-
räisistä kuvateksteistä poimittuja asiasanoja manuaalisesti laadittuun asiasanoitukseen sen arvioimiseksi, kuinka luotettavina tiedonlähteinä kuvatekstejä voidaan pitää. Alkuperäisistä kuvateiedoista poimittiin yhteensä 257 asiasanaa, kun taas manuaalisessa luokittelussa käytettiin yhteensä 203 asiasanaa eli lukumääräisesti arviotuna asiasanajoukot ovat melko samankaltaiset. Kun asiasanat yhdistetään kuvakorpukseseen, asiasanoitusten välillä vallitsevat erot tulevat selkeästi näkyviin (Kuva 3, graafi (a)). Oma graafimme rakentuu kahden poolin – *Alkuperäiset kuvatekstit* ja *Omat asiasanat* – ympärille, joiden väliin jäävälle alueelle sijoittuvat ne asiasanat, jotka esiintyvät sekä alkuperäisissä kuvateksteissä että manuaalisesti laaditussa asiasanoituksessa. Graafissa asiasanoja esittävän solmun koko on suoraan suhteessa kyseisen asiasanan esiintymismäärään eli mitä suurempi solmu on

12. Paikannimien sijaintitietojen keräämisessä hyödynnettiin Internetin hakukoneiden ohella maanmittauslaitoksen avoin paikkatieto -palvelua, <http://www.paikkatietoikkuna.fi/web/fi/avoin-paikkatieto> (16.9.2015) sekä nimistön kyselypalvelua, <http://www.maanmittauslaitos.fi/aineistot-palvelut/rajapintapalvelut/nimiston-kyselypalvelu-wfs> (15.9.2015).

13. Gephi on avoimeen lähdekoodiin perustuva ilmainen, käyttöjärjestelmästä riippumaton ohjelmisto, joka tarjoaa kohtuullisen hyvät välineet erityisesti verkostojen graafiseksi esittämiseksi. Gephi on saavuttanut laajan käyttäjäkunnan etenkin ihmistieteiden parissa, mihin pääsyyinä lienee ohjelman helppokäyttöisyys sekä kohtuullisen hyvä laajennettavuus nk. plugin-sovellusten avulla. Ohjelma on ladattavissa ilmaiseksi projektin kotisivuilta, <http://gephi.github.io/> (9.10.2015). Ks. myös Mathieu Bastian, Sebastian Heymann, & Mathieu Jacomy, *Gephi. An open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media 2009, <http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154> (3.6.2015). Sosiaaliseen verkostanalyysiin perustuvalle tutkimukselle on kuitenkin tarjolla myös lukuisia muita sovelluksia, joten menetelmästä kiinnostuneen lukijan kannattaa perehtyä myös muihin tarjolla oleviin vaihtoehtoihin. Kattava lista sovelluksista ks. esim. <http://www.gmw.rug.nl/~huisman/sna/software.html> (3.6.2015).



Kuva 3. Kuvatekstien asiasanat vs. manuaalisesti laaditut asiasanat (modularisoitu, layout: Yifan Hu¹⁴): (a) kaikki ja (b) vähintään kymmenen kertaa esiintyvät asiasanat.

kooltaan, sitä useammin se esiintyy asiasanakorpuksessa. Vastaavasti asiasanaa ja kuvakorpusta edustavien poolien välinen viiva on sitä paksumpi, mitä useammin asiasanaa on käytetty joko kuvateksteissä tai manuaalisessa asiasanoituksessa. Graafissa käytetty väritys puolestaan pyrkii havainnollistamaan asiasanojen ensisijaista

käyttöympäristöä. Punaisella on esitetty ne asiasanat, jotka liittyvät vahvemmin alkuperäisiin kuvateksteihin, turkoosilla puolestaan on esitetty ne asiasanat, jotka liittyvät vahvemmin manuaaliseen asiasanoitukseen. Esimerkkeinä voidaan nostaa esille molemmissa ryhmissä esiintyvät asiasanat ”lappi” ja ”saksalaiset sotilaat”,

joista ensimmäinen kytkeytyy vahvemmin alkuperäisiin kuvateksteihin, jälkimmäinen manuaaliseen asiansanoitukseen.

Hyödyntämällä verkostometriikkaa eli joko verkostoon itseensä tai sen osiin liittyviä tunnuslukuja, graafista voidaan suodattaa esiin tai pois analyysin kannalta relevantteja elementtejä. Esimerkinomaisesti olemme suodattaneet graafista pois ne asiansanat, jotka esiintyvät vähemmän kuin kymmenen (10) kertaa (Kuva 3, graafi (b)). Kuten voidaan havaita, useimmat jäljelle jääneistä asiansanoista ovat yhteisiä alkuperäisille kuvateksteille ja manuaaliselle asiansanoitukselle, mutta samalla voidaan havaita tiettyjen asiansanojen toistuvan ainoastaan alkuperäisissä kuvateksteissä. Suodatettu graafi antaa siis viitteitä siitä, että alkuperäisissä kuvateksteissä käytettiin myös asiansanoja, jotka eivät näyttäisi ainkaan suoraan liittyvän itse kuvaan tai sen aihepiiriin. Mikäli alkuperäiset kuvatestit olisivat pääosin vastanneet kuvien sisältöjä, eniten käytettyjen asiansanojen tulisi olla sellaisia, jotka esiintyvät sekä alkuperäisissä kuvateksteissä että manuaalisessa asiansanoituksessa.

Alkuperäisten kuvatekstien yksipuolisuutta selittää osaltaan se, että kuvilla on syntyajankohdanaan ollut nimenomaan sotapropagandistinen funktio. Siten kuvaa katselevan henkilön huomio on kuvatekstin avulla haluttu kiinnittää tiettyyn, propagandan kannalta tärkeään asiaan, ja siten saada katsoja näkemään esitetty asia propagandistisesti oikeassa valossa. Kuvatekstillä onkin ollut kuvan katselijan huomion ohjaamisessa keskeinen rooli¹⁵. Siksi kuvatekstejä saatettiin jopa tietoisesti väärentää: jos kuvan esittäjä todellisuus ei sopinut yhteen saksalaisen propagandan kanssa, kuvan tulkintaa ohjattiin toisille urille uuden kuvatekstin avulla¹⁶. Tyypillistä propagandakuvien kuvateksteille oli myös tietty jargon, joka osaltaan selittää kuvatekstien yksipuolisuutta.

Toinen kuvatekstien perusteella tehdyn asiansanoituksen ja myöhemmän, kuvien tarkastelemaan perustuvan asiansanoituksen eroja selittävä

tekijä on kuvaajien tapa ottaa yksittäiskuvien sijaan sarjoja¹⁷. Näihin sarjoihin liitettiin monesti vain yksi kuvateksti, mistä syystä kuvateksti ei välttämättä korreloinut kaikkien sarjaan kuuluneiden kuvien kanssa. Hyvä esimerkki tästä sarjojen tekstitystavasta on puolustusvoimien Internet-kuva-arkistosta¹⁸ hakulauseella ”Mannerheim AND Saksa” löydettävissä oleva kuvasarja vuodelta 1942, jonka kuvatekstinä on ”Mannerheim seurueineen matkalla Saksassa, tapaa Hitlerin ym.”. Tämän sarjan kuviin kuuluu muun muassa Göringin Karinhalle-metsästysmajaa esittäviä otoksia (esim. SA-kuva nro. 93886), joihin kaikkiin em. kuvateksti siis liittyy.

Vaikka edellinen graafi toikin esille eri tavoin laadittujen asiansanoitusten eroja ja yhtäläisyyksiä, varsinaisen sisällönanalyysin kannalta kiinnostavampaa on analysoida kuviin liitettyjen asiansanojen välillä vallitsevia suhteita. Näiden niin kutsuttujen yhteisesiintymien (*co-occurrence*, suomeksi myös kollokaatiot) kautta on mahdollista päästä kiinni koko kuvakorpuksen kannalta keskeisiin asiansanoihin sekä analysoida kuvakorpuksen sisällä mahdollisesti olevia te-

14. Yifan Hu -layout perustuu pisteiden välisen ”vetovoiman” analysointiin ja pyrki tiivistämään graafin vetovoimaisimpien pisteiden ympärille. Layout soveltuu erityisen hyvin suurten verkostojen visualisointiin, mutta sen idea tulee hyvin esille myös Kuvassa 3. Ks. tark. Yifan Hu, *Efficient and High Quality Force-Directed Graph*. 2007, http://yifanhu.net/PUB/graph_draw_small.pdf (3.6.2015).

15. Kuvatekstin merkityksestä lehdistökuvan tulkinnassa ks. Elina Heikkilä, *Kuvan ja tekstin välissä. Kuvateksti uutiskuvan ja lehtijutun elementtinä*. SKS 2006.

16. Susan Sontag, *Über Fotografie*. Fischer 1977; Jörn Glasenapp, *Vergangenheit und Zukunft. Wie fotografischer sinn entsteht*. *Fotogeschichte* 124 (2012), 5–12.

17. Organisaatioitaan ja toimintatavaltaan saksalaisia vastineitaan muistuttaville suomalaisille tiedotuskompanioille annettiin useaan otteeseen ohjeita kuvata tapahtumia nimenomaan jatkuvina sarjoina yksittäiskuvien sijaan. Ks. esim. Kansallisarkisto, Sönnäisten toimipiste, kansio T20680/13, Ohjekäskyt tiedotuskomppanioille, käsky n:o 38.

18. www.sa-kuva.fi (3.6.2015).

maattisia samankaltaisuuksia. Aiemmissä tutkimuksissa yhteisesiintymäverkostoanalyysiä on sovellettu etenkin aineistojen temaattisten rakenteiden ja niiden taustalla vaikuttavien tai niihin yhteydessä olevien tekijöiden löytämiseksi.¹⁹ Taustalla on siten oletus, että yhteisesiintymät muodostavat sosiaalisten verkostojen kaltaisia vuorovaikutussuhteita, jotka auttavat havaitsemaan kuvien välillä vallitsevia temaattisia polkuja ja kuvakorpuksen temaattisia ”pakopisteitä”.

Tämän artikkelin empiirisenä aineistona käytettävän sotavalokuvakorpuksen muodostavien kuvien oletetaan siten muodostavan oman temaattisen tilansa, jonka pääsisältö on kiteytetyssä muodossa löydettävissä korpukseen liitetyistä asiasanoista. Tässä kohdin teemme oletuksen, että samoja asiasanoja on käytetty kuvaamaan samoja – tai ainakin samantyyppisiä – sisältöjä vastaavalla tavalla kuin esimerkiksi Internet-aineistoissa käytetään tageja²⁰. Looginen seuraus tästä on, että mikäli kullekin kuvalle konstruoidaan yhteisesiintymämatriisi, tematiikaltaan samantyyppisten kuvien yhteisesiintymämatriisien voidaan olettaa olevan samankaltaisia. Koska kuvien yhteydessä käytettyjen asiasanojen joukko on rajallinen, oletamme verkostoanalyysin löytävän materiaalista temaattisia polkuja eli keskeisiä avainsanoja, jotka toimivat linkeinä korpukseen kuuluvien kuvien välillä ja siten paljastavat tematiikalta samankaltaisia kuvajoukkoja.

Yhteisesiintymäverkostojen konstruoinnin toteutimme sen ajatuksen mukaisesti, että kukin kuvakorpuksen kuvista muodostaa oman temaattisen tilansa. Näin ollen kuvan metatietoihin liittyvien asiasanojen voidaan olettaa kuvaavan kuvaan yhdistettyä temaattista tilaa, jolloin kytkemällä kuvaan liittyvät asiasanat toisiinsa pareittain konstruoidaan kuvan sisältöä kuvaavia käsittepareja. Loimme avainsanaparit kaikkien avainsanojen kesken, kuitenkin siten, ettei kaksi avainsanaa voi saman kuvan kohdalla esiintyä kuin kerran yhdessä. Käytännössä tämä

tarkoittaa, että mikäli kuvaan X liittyy avainsanat A, B ja C, näiden perusteella voidaan muodostaa yhteisesiintymät A-B, A-C ja B-C. Verkostotermein ilmaistuna avainsanaparit ovat suuntaamattomia eli pari A-B tulkitaan samaksi kuin B-A. Tämän menettelyn analyttinen etu on, että avainsanaparin A-B lukumäärä aineistossa kertoo suoraan niiden kuvien lukumäärän, jotka jakavat tämän avainsanaparin. Tulkinallisesti tämä tarkoittaa, että yhteisesiintymien esiintymistiheyden perusteella voidaan identifioida tematiikaltaan samankaltaiset kuvat. Yhteisesiintymien konstruoinnissa käytimme ainoastaan manuaalisesti laadittua asiasanoitusta. Perustellemme tätä ratkaisua sillä, että alkuperäisten kuvatekstien kohdalla asiasanoja näytetään – kuten edellä osoitimme – käytetyn melko yleisluonteisesti, mistä syystä alkuperäisiin kuvateksteihin perustuvat asiasanaverkostot eivät välttämättä anna luotettavaa kuvaa kuvakorpuksen sisältävistä temaattisista rakenteista. Koska ma-

19. Keith Stuart & Ana Botella, *Corpus Linguistics, Network Analysis and Co-occurrence Matrices*. *International Journal of English Studies* S1 (2009), 1–20; Alan Brier & Bruno Hopp, *Computer assisted text analysis in the social sciences*. *Quality & Quantity* 45:1 (2011), 103–128; Sangno Lee, Jaeki Song & Yongjin Kim, *An Empirical Comparison of Four Text Mining Methods*. *Journal of Computer Information Systems* 51:1 (2010), 1–10; Jana Diesner, Kathleen M. Carley & Laurent Tambayong, *Extracting socio-cultural networks of the sudan from open-source, large-scale text data*. *Computational and Mathematical Organization Theory* 12:3 (2012), 328–339; Yi-Yu Hsu & Hung-Yu Kao, *Coin. A network analysis for document triage*. *Database* (2013), 1–11; Ma Feicheng & Li Yating, *Utilising Social Network Analysis to Study the Characteristics and Functions of the Co-occurrence Network of Online Tags*. *Online Information Review* 38:2 (2014), 232–247; Kimmo Elo, *A Spider Spinning its Web. East German HUMINT Networks on Nordic Affairs*. Teoksessa Thomas Wegener Friis & Władysław Bulhak (toim.) *Need to Know. Eastern and Western Perspectives, Studies in Intelligence and Security Series*. University Press of Southern Denmark 2014, 61–79. Kimmo Elo, *The Content Structure of Intelligence Reports*. *Connections* 35:1 (2015), 20–38.

20. Tagien käytöstä verkostoanalyysissä ks. Feicheng & Yating 2014.

Asiasanoja symboloivien ympyröiden väritys on suhteutettu kunkin asiasanan kyseisen asiasanan kytkösten määrään (*degree*). Väriskaala liukuu sinisestä (vähän kytköksiä) keltaisen kautta punaiseen (paljon kytköksiä). Värityksen perusteella on siis mahdollista identifoida ne asiasanat, jotka esiintyvät monien muiden asiasanojen kanssa. Asiasanoja yhdistävien viivojen paksuus puolestaan kuvastaa yhteisesiintymän yleisyyttä kuvakorpuksessa eli on riippuvainen yhteyden painoarvosta (*weight*). Mitä useammin kaksi asiasanaa esiintyy yhdessä, sitä painavampi niiden suhde ja sitä paksumpi niitä yhdistävä viiva. Näin useimmin yhdessä esiintyvät asiasanat, niin kutsutut ydintermit, on helppo tunnistaa graafista. Asiasanoja symboloivien solmujen koko ja niihin liittyvä tekstin koko puolestaan suhteutuu kyseisen asiasanan asemaa ja sijaintia koko verkostossa – verkostotermein ilmaistuna keskeisyyttä (*centrality*) – ilmaisevaan *betweenness centrality* -arvoon²². Näin voidaan löytää ne asiasanat, jotka toimivat sotakuvakorpuksen sisältäviä asiasanaklustereita toisiinsa yhdistävinä linkkeinä ja siten indikoivat korpuksen sisältäviä temaattisia polkuja.²³

Verkostometriikan perusteella kolme tärkeintä asiasanaa ovat ”saksalaiset sotilaat” (*betweenness centrality* -arvo: 7897.2), ”talvi” (4103.2) ja ”siviilit” (3432.6), mikä vastaa varsin hyvin graafin esittämää verkstorakennetta. ”Saksalaiset sotilaat” esiintyy yhteensä 85 kertaa 69 kuvan yhteydessä, ”talvi” puolestaan 51 kertaa 43 kuvan yhteydessä ja ”siviilit” kaikkiaan 41 kertaa 26 kuvan yhteydessä. Vahvimmat temaattiset linkit muodostuvat puolestaan asiasanaparien ”saksalaiset sotilaat suomalaiset sotilaat” (20 kuvaa), ”saksalaiset sotilaat talvi” (11 kuvaa) sekä ”sotilaat talvi” (10 kuvaa) välille. Näiden tulosten valossa analysoidun sotakuvakorpuksen vahvin temaattinen polku linkittäisi toisiinsa suomalaisia ja saksalaisia sotilaita talvisissa olosuhteissa esittävät kuvat.

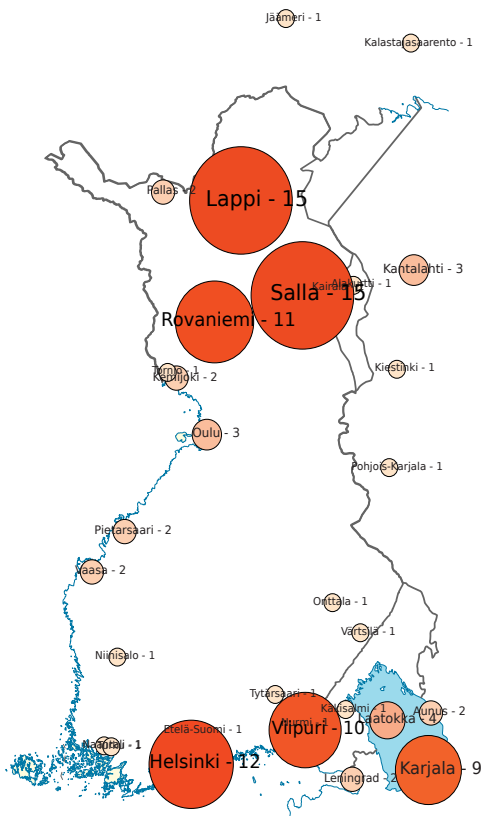
Kolmantena esimerkkinä verkostoaalyysin mahdollisuuksista haluamme nostaa esille geo-

spatiaalisen metadatan hyödyntämisen. Valmistellussa aineistossa oli yhteensä 109 kuvaa, joiden metatiedot mahdollistivat kuvien paikallistamisen myös maantieteellisesti. Paikkatietomatriisien perusteella laadimme kaksi verkostomatriisia. Ensimmäiseen lisäsimme pelkästään kunkin kuvauspaikan kokonaiskuvamäärän, minkä perusteella laitimamme visualisointi sijoittaa kuvauspaikat kuvamäärätietoineen kartalle siten, että paikkakuntaa symboloivan ympyrän koko ja väritys ovat positiivisessa suhteessa paikkakunnalla otettujen kuvien määrään (Kuva 5, graafi (a)). Kuten graafista voidaan nähdä kaksi maantieteellistä aluetta – eteläinen (Helsinki, Viipuri ja Karjala) sekä pohjoinen (Salla, Rovaniemi, Lappi) – dominoivat kuvauspaikkakorpuista.

Toisen verkostomatriisin laadimme kytke-mällä yhteen kuvien asiasanat ja kuvauspaikat. Näin saamamme verkostomatriisi mahdollistaa kuvakorpuksen sisältävien asiasanojen maantieteellisen jakauman tarkastelun (Kuva 5, graafi (b)). Verkostometriikan perusteella asiasanojen ja maantieteellisten paikkojen välillä vallitsevista yhteisesiintymistä kolme vahvinta ovat ”saksalaiset sotilaat lappi/Salla/Karjala” (yhteensä 19 esiintymää), ”tuhotut rakennukset Helsinki/Viipuri” (7) sekä ”siviilit/Helsinki” (5). Siten asiasanojen geospatiaalinen jako näyttäisi noudattavan tiettyä pohjoinen–etelä-jakoa, jossa saksalaisiin joukkoihin sekä varsinaiseen taistelutoimintaan liittyvät kuvat painottuvat Pohjois-Suomeen, kun taas sodan tuhoista ja etenkin siviilien kärsimyksistä kertovat kuvat painottuvat

22. *Betweenness centrality* -arvo lasketaan sen perusteella, kuinka usein solmu sijoittuu kahta muuta solmua yhdistävälle polulle siten, että tämä polku on samalla lyhin näitä kahta muuta solmua yhdistävä polku. Ks. myös Prell 2012, 103–104; Hsu & Kao 2013.

23. Dmitry Paranyushkin, *Identifying the Pathways for Meaning Circulation using Text Network Analysis*. Research report. Nodus Labs 2012, 13–14, <http://noduslabs.com/publications/Pathways-Meaning-Text-Network-Analysis.pdf> (3.6.2015). Ks. myös Feicheng & Yating 2014.

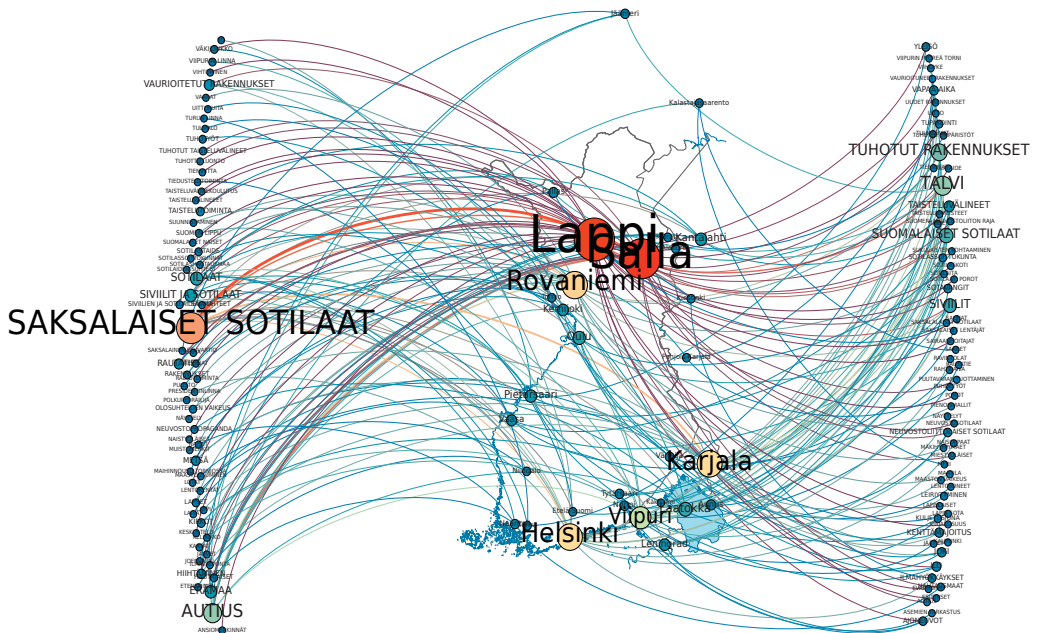


eteläiseen Suomeen sekä Karjalan kannaksen alueella otettuihin kuviin.

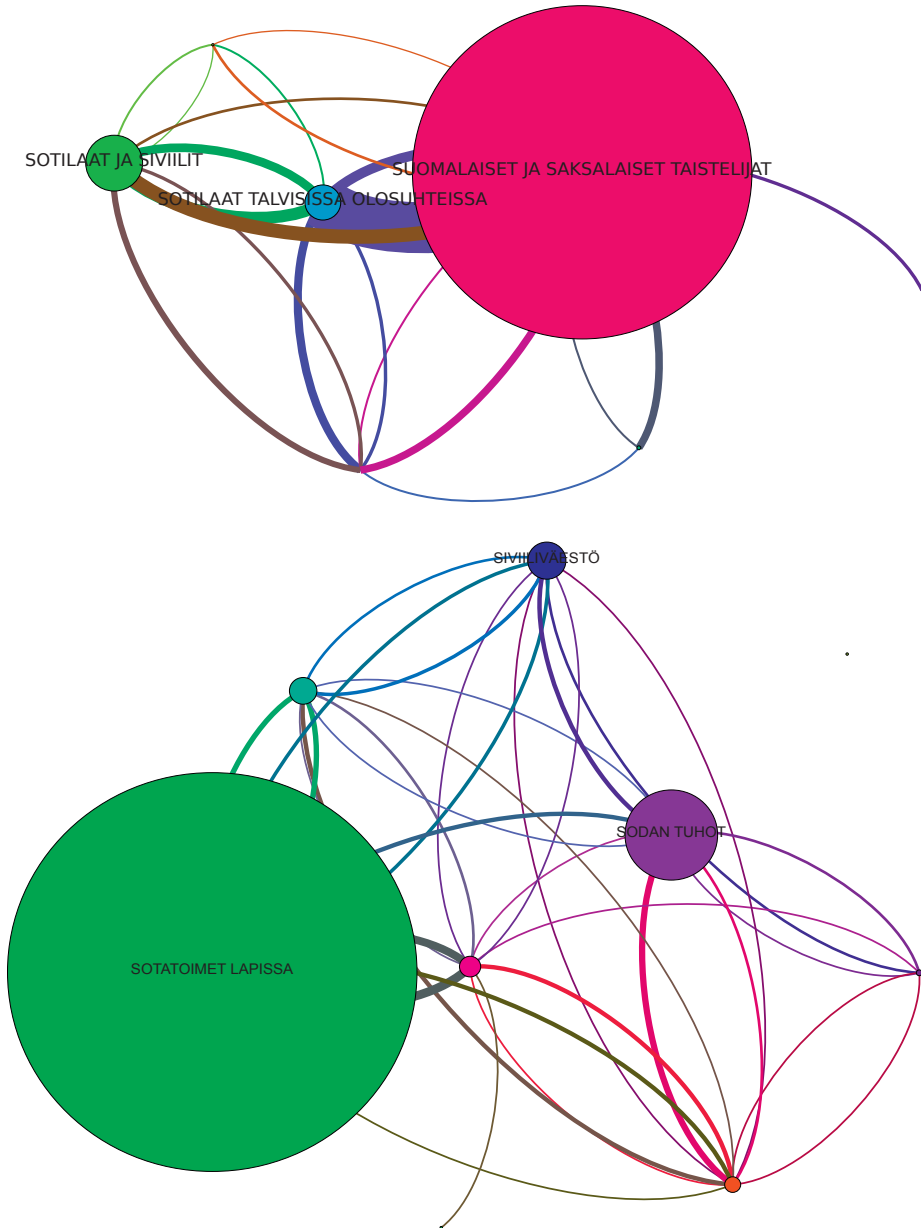
Jo edeltävät analyysivaiheet ovat antaneet viitteitä siitä, että käyttämämme empiirinen aineisto saattaisi koostua temattisesti erilaisista ryhmistä. Verkostoanalyysin näkökulmasta kyse on siitä, voidaanko verkostosta löytää toisistaan erottuvia yhteisöjä, klustereita tai klikkejä. Digitalisaation tuottama aineistojen kasvu, niin kutsuttu *Big Data*, on entisestään lisännyt tarvetta tehokkaiden ja luotettavien klusterointimenetelmien kehittämiseksi.²⁵

24. GeoLayout -asettelu rakentaa verkoston verkoston pisteiden maantieteellisten koordinaattien perusteella karttaprojektiona, jolloin verkostograafi esittää verkoston rakennetta suhteessa tiettyyn maantieteelliseen alueeseen. Ks. myös <https://forum.gephi.org/viewtopic.php?t=224> (3.6.2015).

25. Lisää klusterointiin liittyvästä problematiikasta ks. esim. Santo Fortunato, Community detection in graphs. *Physics Reports* 486 (2010), 75–174; Zhiang Wu, Jie Cao, Junjie Wu, Youquan Wang & Chunyang Liu, Detecting Ge-



Kuva 5. Kuvakorpuksen geospaatialinen ulottuvuus: (a) kuvauspaikat ja (b) asiasanat kuvauspaikoittain. Layout: Geo Layout²⁴.



Kuva 6: Verkoston klusterointi: (a) asiasanojen yhteisesiintymät ja (b) asiasanat kuvauspaikoittain (Louvain-menetelmä²⁶).

Suoritimme klusteroinnit sekä yhteisesiintymäverkoston että asiasanojen maantieteellisiin yhteyksiin liittyvän verkoston osalta. Kummankin verkoston kohdalla soveltamamme klusterointimenettely löysi aineistosta selkeitä klustereita (Kuva 6). Asiasanojen yhteisesiintymiin perustuvan temaattisen verkoston kohdalla kolme

nuine Communities from Large-Scale Social Networks. A Pattern-Based Method. *The Computer Journal* (2013), 1–15.

26. Tässä artikkelissa klusterointi toteutetaan Gephi-ohjelmistoon sisäänrakennettua Louvain-menetelmää hyödyntämällä. Menetelmän perusteella samaan klusteriin luokitellaan ne verkoston solmut, jotka ovat keskenään tiiviimmässä vuorovaikutuksessa kuin verkoston muiden

suurinta klusteria sulkevat sisäänsä lähes 70 prosenttia aineiston avainsanoista. Suurimpaan klusteriin sisältyy 24,3 prosenttia aineiston avainsanoista ja sen merkittävimpinä asiasanoina ovat ”suomalaiset sotilaat”, ”saksalaiset sotilaat” ja ”taisteluvälineet”. Asiasanojen painoarvoa klusterin sisällä analysoituumme päädyimme siihen johtopäätökseen, että tämän klusterin teemana ovat *Suomalaiset ja saksalaiset taistelijat*. Toinen pääklusteri on vain hieman pienempi sisältäen 22,8 prosenttia aineiston avainsanoista. Klusterin sisällä tärkeimmiksi asiasanoiksi nousivat ”siviilit”, ”siviilit ja sotilaat” ja ”lapset”, minkä perusteella katsomme klusterin teemana olevan *Sotilaat ja siviilit*. Kolmas pääklusteri sisältää 22,3 prosenttia avainsanoista ja sen tärkeimpiin avainsanoihin lukeutuivat ”talvi”, ”sotilaat” ja ”taistelutoiminta”. Tämän perusteella päädyimme siihen, että kolmannen klusterin teemana ovat *Sotilaat talvisissa olosuhteissa*.

Asiasanojen maantieteelliseen jakaumaan perustuvan verkoston kohdalla klusterointimenetelmä onnistui mielestämme jopa yllättävän hyvin tavoittamaan kuvakorpuksen maantieteellisen jakauman. Kaksi suurinta klusteria ovat yhtä suuret, molempiin sisältyy 19,5 prosenttia avainsanoista ja paikannimistä, mutta klusterin sisältöjen analyysi osoitti niiden olevan teemoiltaan hyvin erilaiset, mikä osaltaan vahvistaa klusteroinnin luotettavuutta. Ensimmäisen klusterin teema on selkeästi *Sotatoimet Lapissa*, mikä ilmenee sekä klusteriin sisältyvien asiasanojen että etenkin maantieteellisten paikkojen kautta. Klusterin sisältöä hallitseviin asiasanoihin kuuluvat ”saksalaiset sotilaat” ja ”taistelutoiminta”, keskeisiin maantieteellisiin paikkoihin ”Lappi” ja ”Oulu”. Toista klusteria dominoivat sellaiset asiasanat kuten ”autius”, ”hautajaiset” ja ”tuhotut taisteluvälineet/ympäristöt”, joiden maantieteellinen jakauma ei osoita selkeää painopistettä. Tästä johtuen päädyimme siihen, että tämän klusterin teemana ovat eri puolilla maata havaitut *Sodan tuhot*. Kolmas klusteri sisältää 15,1 prosenttia avainsanoista ja paikanni-

mistä, mutta myöskään tämän klusterin osalta maantieteellinen jakauma ei osoita selkeää painopistettä. Klusterin sisällön kannalta keskeisimmät asiasanat liittyvät siviiliväestön asemaan (”siviilit ja sotilaat” sekä ”lapset”), mistä syystä katsomme tämän klusterin teemana olevan *Siviiliväestö*.

Klusterien melko selkeästä rakenteesta huolimatta on syytä kiinnittää huomiota siihen, että myös klusterien välillä on vuorovaikutussuhteita. Erityisen selkeästi tämä on havaittavissa asiasanojen yhteiseiintymäverkostossa, jossa etenkin 1. ja 3. klusterin yhdistävä paksu viiva indikoi vahvaa vuorovaikutussuhdetta näiden kahden klusterin välillä (Kuva 6, graafi (a)). Kun huomioidaan klustereihin tyyppikuvauksina tiivistettyjen teemojen kohtuullinen päällekkäisyys, tätä vahvaa vuorovaikutussuhdetta voidaan kuitenkin pitää varsin luonnollisena.

Klusteroinnin kohdalla havaitut temaattiset päällekkäisyydet herättävät kysymyksen klusteroinnin luotettavuudesta johtopäätösten pohjana. Klusteroinnin luotettavuutta arvioidaan niin kutsutun modulaarisuusarvon (Q) perusteella, joka voi vaihdella välillä $[0,1]$. Aiempien tutkimusten perusteella luotettaviksi katsotaan klusteroinnit, joiden $Q > 0,4$.²⁷ Tämän artikkelin osalta avainsanojen yhteiseiintymäverkoston klusteroinnin $Q = 0,488$ ja asiasanojen ja kuvauspaikkojen välisiä yhteyksiä kuvaavan verkoston $Q = 0,561$ eli identifioituja klustereita voidaan pitää luotettavina. Asiasanojen pienempi Q -arvo selittyy pitkälti pääkorpusten toisiaan lähellä olevalla tematiikalla, mistä syystä asiasanarakenteeltaan samankaltaisia kuvia ei täysin luo-

solmujen kanssa. Mentelmästä laajemmin ks. Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, & Etienne Lefebvre, Fast unfolding of communities in large networks. *Journal of Statistical Mechanics. Theory and Experiment* P10008 (2008).

27. Esim. Mark E. J. Newman, Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103:23 (2006), 8577–8582; Paranyushkin 2011, 18.

tettavasti voida erotella toisistaan. Klusteroinnin luotettavuus on kuitenkin osoitus siitä, että aineiston tietokoneavusteinen luokittelu voi tarjota perinteiselle lähiluentaiselle sisällönanalyysille tietokoneavusteisen vaihtoehdon, jonka avulla isoista aineistosta on mahdollista nopeastikin löytää yhdistäviä tekijöitä ja siten päästä helpommin kiinni aineistoon sisältyviin teemoihin.

Sotahistoriallisen tutkimuksen näkökulmasta arvioituna edellä saadut tulokset korreloivat suhteellisen hyvin sen kanssa, mitä saksalaisten toiminnasta Suomessa tiedetään. Geospatiaalisen analyysin keskeinen havainto, jonka mukaan ylivoimaisesti suurin osa kuvista on otettu Lapisissa, on varsin looginen, sillä Saksan joukkoja oli määrällisesti eniten nimenomaan Pohjois-Suomessa. Suomalaiset, jotka ulkopoliittisista syistä pyrkivät korostamaan käyvänsä Saksan sodasta erillistä, omaa taisteluaan, pyrkivät myös kaikin tavoin korostamaan tätä seikkaa tiedotustoiminnassaan. Tätä kuvastaa hyvin suomalaisille TK-kuvaajille 2.7.1941, siis aivan jatkosodan alussa, annettu ohjekäskey²⁸, jossa kuvaajia kehoitettiin ryhtymään ”tervehenkiseen ja toverilliseen kilpailuun” saksalaisten PK-miesten kanssa, jotta ”saksalaiset liittolaisemme” joutuisivat mahdollisimman paljon lainaamaan suomalaista aineistoa. Tällä tavoin pyrittiin välttämään se kiusallinen tilanne, että Suomessa olevien saksalaisten joukkojen tiedotustoiminta olisi tullut rinnastetuksi suomalaiseen tiedotustoimintaan.²⁹

Entä sitten Helsingistä ja Karjalasta otetut kuvat? Kuten aineistoa esitellessämme totesimme, tämän artikkelin primääriaineiston muodostava Scherl-aineisto on peräisin saksalaisen kuvatoimiston arkistosta. Tähän kuvatoimistoon toimitettiin kuitenkin myös suomalaisia TK-kuvia, mikä käy ilmi Suomen puolustusvoimien sodanaikaisista kuvaleivitysraporteista³⁰. Ainakin osa asiansanan ”Suomi” alta löydettävistä kuvista on varmuudella tunnistettavissa suomalaisiksi TK-kuviksi, mitä seikkaa myös klusteroinnin tu-

lokset osaltaan tukevat: ulkomaille levitetyissä, suomalaisissa kuvissa suosittuja aiheita olivat niin Viipuri, Karjala kuin sodan aiheuttamat tuhotkin – siis asiat, jotka myös analyysissämme nousivat esille asiansanojen joukosta.

Entä miten saksalaisten, pääosin Lapin alueelle sijoittuvien PK-kuvien aiheet korreloivat saksalaisen propagandan tavoitteiden kanssa? Saksalaisen propagandan, kuten propagandan yleensäkin, tavoite oli kertoa omien joukkojen menestyksekkäästä taistelusta. Tämä selittää sotatoimiin Lapisissa liittyvien asiansanojen vahvan roolin asiansanoituksessa. Suomen Lapisissa olleiden saksalaisten joukkojen ensisijainen tavoite oli Murmanskin valtaaminen, koska siten olisi saatu keskeytettyä Murmanskin kautta Neuvostoliittoon virranneet liittoutuneiden avustustoimitukset. Kun hyökkäys ei edennyt toivotulla tavalla eikä vastustajan ylivoimaisuutta saksalaisen propaganda-ajattelun mukaisesti voitu myöntää, syytä huonoon sotamenestykseen alettiin etsiä ilmasto- ja maasto-oloista. Lapin maasto, joka esti saksalaisia hyödyntämästä motorisoidun armeijansa liikkuvuutta, sekä pohjoisen talvien ankaruus nousevat hallitseviksi tekijöiksi jopa saksalaisten sotilaiden henkilökohtaisissa muistelmissa, joten ei liene kovin yllättävää, että samat teemat korostuvat myös valokuvapropagandassa.³¹

Suomalaisten sotilaiden läsnäolo kuvissa puolestaan selittyy sillä, että siinä missä suoma-

28. Kansallisarkisto, Sörnäisten toimipiste, kansio T20680/13, Ohjekäskyt tiedotuskomppaniolle, käsky n:o 6.

29. Ks. tark. Olli Kleemola, Kamera aseena. Teoksessa Helena Pilke & Olli Kleemola (toim.) *Suomi taisteli. Kuvat kertovat*. Readme.fi 2013, 8–25.

30. Kansallisarkisto, kansio T10602/6, Puolustusvoimien kuvien levitysraportit.

31. Maaston roolista saksalaisessa valokuvapropagandassa sekä Lapisissa taistelleiden saksalaisotilaiden muistelmista ks. Olli Kleemola, Vihollinen vai liittolainen? Metsä suomalaisten ja saksalaisten itärintaman sotapropagandakuvien elementtinä 1941–1944. *Vuosilusto* 10 (2014), 29–42.

laiset pyrkivät korostamaan Lapissa käytävän sodan erillisluonnetta, Saksa pyrki korostamaan suomalaisten taistelevan rinnallaan. Talvisodan myötä maailmalla demokraattisena maana tunnetuksi tullut Suomi toimi eräänlaisena ”näyteikkunamaana” maailmalle ja siten esimerkkinä siitä, miten demokratia ja kansallissosialistinen valtio voivat yhdessä taistella bolsevistista Neuvostoliittoa vastaan. Suomalaisten siviilien läsnäolo kuvissa puolestaan korostuu erityisesti Lapin sodan (1944–1945) myötä. Saksan armeija näyttäytyy omista kuvissaan siviiliväestön suojeelijana, joka pelastaa suomalaissiviilit Lappiin vöyryvän bolseviarmeijan tieltä Suomen hallituksen jätettyä omat kansalaisensa bolsevistien armoille. Tämä trendi ei ole mitenkään poikkeuksellinen saksalaisessa sotapropagandassa, samantyyllisiä kuvia julkaistiin myös muilta itärintaman osilta.

Kohti digitaalista historiantutkimusta?

Tämän artikkelin tavoitteena on havainnollistaa, millä tavoin historiallista verkostoanalyysiä voidaan soveltaa laadulliseen tutkimusmetodiikkaan lukeutuvan sisällönanalyysin tietokoneavusteisena laajennoksena. Tutkimusprosessina tietokoneavusteinen analyysi ei juurikaan eroa perinteisestä sisällönanalyysistä aineiston koodauksineen, ryhmittelyineen ja tyyppittelyineen.

Tässä artikkelissa käytetty aineisto oli, ainakin suhteessa *Big Data* -projekteihin, varsin pieni. Kriittinen lukija saattaa kysyä, miksi analysoida tällaista aineisto tietokoneavusteisesti, kun se olisi vielä manuaalisestikin hallittavissa. Aineiston koko oli harkittu ratkaisu artikkelin tavoitteiden kannalta, pyrimmehän edellä osoittamaan nimenomaisesti tietokoneavusteisten menetelmien luotettavuuden myös historiallisten aineistojen analysoinnille. Koko analyysin perustana olleet asiasanalistat laadittiin sekä automaattisesti kuvateksteistä poimimalla että manuaalisesti kuvia tarkastelemalla. Näin konstruoitujen asiasanalistojen vertailu osoitti, että pelkistä kuvateksteistä pöimitut asiasanat kor-

reloivat varsin hyvin manuaalisesti (ilman kuvatekstejä) laaditun asiasanalistan kanssa ja siten ne muodostivat luotettavan pohjan aineiston analyysille. Tämän vertailun perusteella voidaan olettaa, että myös määrällisesti suuremman sotakuvakorpuksen kuvatekstien automaattinen asiasanoittaminen tuottaa luotettavan pohjan kuvien sisältöjen analysoimiseksi. Ja mitä suuremmaksi tutkimusaineisto kasvaa, sitä mahdollisempaa sitä on manuaalisesti hallita.

Sekä omien tulostemme että aiempien tutkimusten valossa katsomme digitaalisten ihmistieteiden avaavan uusia mahdollisuuksia juuri eksploraatiiviselle historiantutkimukselle. Suurta potentiaalia näemme sisällönanalyysin ohella myös visuaalisessa historiantutkimuksessa, jossa tietokoneavusteisten kuvantunnistustelmien nopea kehitys tulee jo lähitulevaisuudessa avaamaan uusia mahdollisuuksia tutkimukselle.³² Samalla pidämme tärkeänä muistuttaa, että tietokoneavusteisten menetelmien tehokkuus perustuu puhtaaseen algoritmipohjaiseen laskentatehoon, jonka hintana on täydellinen kontekstittomuus. Dataa murskaavaa tietokonea ei kiinnosta, mistä aineistossa on kyse – se luokittelee ja lajittelee, klusteroi ja strukturoi pelkästään ennakkoon määrätyn algoritmin mukaisesti. Tutkija ei siis voi mennä analyysiohjelmiston taakse piiloon ja sysätä vääriä tuloksia koneen syyksi. Digitaalinen tutkimus edellyttää tutkijalta jopa aiempaa laajempaa aihepiirinsä tuntemusta, kun digitaalinen analyysi voi nostaa esille yllättäviäkin yhteyksiä ja tekijöitä, joiden merkityksen arviointi edellyttää aihepiirin syvällistä tuntemus-

32. *Visual history* -suuntauksesta, ks. <http://docupedia.de/zg/VisualHistoryVersion3.0GerhardPaul> (3.6.2015). Yhtenä esimerkkinä tämän tyyppisistä uusista tutkimusorientaatioista haluamme tässä nostaa esille Euroopan integraation historiaan liittyvän HistoGraph-projektin, joka pyrkii rakentamaan verkostoanalyttistä näkökulmaa integraation henkilöhistoriaan, <http://www.cvce.eu/en/dhlab/projects> (3.6.2015). – *Nyt julkaistu Kimmo Elon ja Olli Kleemolan artikkeli on käynyt läpi tieteellisen vertaisarvioinnin.*

ta. Samalla tulosten arviointi edellyttää kuitenkin myös sovellettujen tietokoneavusteisten menetelmien tuntemusta: ilman ymmärrystä siitä, miten tulokset syntyvät, on niiden luotettavuuden arviointi mahdotonta. Vaikka digi-ihmistieteilijän ei ehkä tarvitse jatkossakaan osata ohjelmoida, hänen tulee kuitenkin ymmärtää niitä tietojenkäsittelyn peruseräitä, joihin digimenetelmät perustuvat.

Digitutkimuksen ja ”perinteisen” historian tutkimuksen suhdetta ei kuitenkaan ole syytä tarkastella nollasummapelinä tai joko-tai-asetelmana. Vaikka digitalisaatio väistämättä tulee sähköistämään myös historian tutkimuksen aiheistot, on tulevaisuudessa tarvetta sekä digitutkimukselle että perinteisille, humanistisille tieteille ominaisille tavoille tehdä tutkimusta. Mutta erityisesti on tarvetta näiden kahden tutkimussuuntauksen väliselle vuoropuhelulle, koska vain sitä kautta historian tutkijat(kin) voivat osaltaan varmistaa, että historian tutkimukselle ominaiset kysymys- ja ongelmanasettelut ovat mukana digitalisaatioon ja sen vaikutuksiin liittyvissä keskusteluissa.

Abstract: Network analysis and exploratory research of historical sources: German war photography in Finland during the Second World War

In the recent years, ‘digital humanities’ as a research paradigm has gained in popularity among historians. This article seeks to exemplify how historical network analysis could be used as an exploratory research method for historical sources. The primary material used consists of war photography from the Nazi German *Wehrmacht* taken by the *Wehrmacht*'s propaganda troops in Finland during the World War II. We exemplify how computational tools can be applied in order to explore new perspectives on the material on the one hand, and to construct a certain historical interpretation based on these photographs on the other. The analytical results reliably prove how computational research techniques allow us both to mine historical sources for new, relevant information and to visualize knowledge in a wholly new way. We consider the latter to be especially important and fruitful, since visualization techniques allow us to gain completely new insights into phenomena being subjected to research.

Keywords: digital humanities, historical network analysis, war photography, Germany, Finland, Second World War, 20th century