



Heidi Kurvinen

Mediahistorian näkökulma tarpeen digitaalisten lehtiarkistojen käytössä

Digitaalisten tutkimusmenetelmien käyttö on lisääntynyt kaikissa humanistisissa tieteissä ja yhteiskuntatieteissä 2010- ja 2020-luvulla, ja valtaosa analogisia aineistoja hyödyntävistä tutkijoistakin käyttää erilaisia digitaalisiin arkistoihin liittyviä työkaluja, kuten hakukoneita ja aineistoluetteloita.¹ Andreas Fickers, Juliane Tatarinov ja Tim van der Heijden korostavat nykytutkimuksen hybridiluonnetta,² mikä tulee lähelle Mats Fridlundin käsitettä digitaalinen historia 1.5 (*digital history 1.5*). Sen tekemiseen ei tarvita ohjelmointitaitoja, mutta tutkijalta edellytetään digitaalisten menetelmien

ja aineistojen hyödyntämistä ja tietoisuutta niiden erityispiirteistä.³ Kriittinen katse kohdistuu tällöin niin digitaalisiin infrastruktuureihin ja käyttöliittymiin kuin aineistoihin ja niiden käyttötapoihinkin.⁴

Tässä katsauksessa avaan ”digiajassa tehtävää historian tutkimusta”⁵ keskittymällä digitoituihin sanoma- ja aikakauslehtiarkistoihin.⁶ Niitä on viime vuosina käsitelty lukuisissa alan hankkeissa ja yksittäisissä julkaisuissa,⁷ mutta tutkijat ovat harvoin pohtineet digitaalisten lehtiaineistojen käyttöä suhteessa mediahistorian kysymyksenasetteluihin.⁸ Sen sijaan digitaalisia lehti-

1. Ks. esim. Tim Hitchcock, *Confronting the Digital. Or How Academic History Writing Lost the Plot*. *Cultural and Social History* 10:1 (2013), 9–11; Lara Putnam, *The Transnational and the Text-Searchable. Digitized Sources and the Shadows They Cast*. *The American Historical Review* 121:2 (2016), 377–380; Kati Katajisto, *Digitalisaation haasteet historian tutkimuksessa*. *Historiallinen Aikakauskirja* 118:1 (2020), 70–71; Mats Fridlund, *Digital History 1.5. A Middle Way between Normal and Paradigmatic Digital Historical Research*. Teoksessa Mats Fridlund, Mila Oiva & Petri Paju (toim.) *Digital Histories. Emergent Approaches within the New Digital History*. Helsinki University Press 2020, 73–75, <https://doi.org/10.33134/HUP-5-4>; Adam Crymble, *Technology and the Historian. Transformations in the Digital Age*. University of Illinois Press 2021, 1–4; Gerben Zaagsma, *Digital History and the Politics of Digitization*. *Digital Scholarship in the Humanities* (2022), 1–3, <https://doi.org/10.1093/llc/fqac050>.
2. Andreas Fickers, Juliane Tatarinov & Tim van der Heijden, *Digital history and hermeneutics. Between Theory and Practice. An Introduction*. Teoksessa Andreas Fickers & Juliane Tatarinov (toim.) *Digital History and Hermeneutics. Between Theory and Practice*. De Gruyter Oldenbourg 2022, 7, <https://doi.org/10.1515/9783110723991-001>.
3. Fridlund 2020, 79–82.
4. Fickers, Tatarinov & van der Heijden 2022, 6–11.
5. Crymble ehdottaa digitaalinen historia -käsitteen korvaajaksi ajatusta digiajassa tehtävästä historian tutkimuksesta (*doing history in the digital age*). Crymble 2021, 20.
6. Digitaaliset lehtiaineistot ovat usein saatavilla kansalliskirjastojen käyttöliittymien kautta. Tässä katsauksessa kirjoitan kuitenkin digitoiduista lehtiarkistoista korostaakseni lähderyhmän erityispiirteitä erotuksena muista kirjastojen digitoiduista aineistoista, kuten kirjoista ja käsikirjoituksista.
7. Esim. Maud Ehrmann, Estelle Bunout & Marten Düring, *Historical Newspaper User Interfaces. A Review*. *IFLA WLIC* 2019, <https://library.ifla.org/id/eprint/2578/1/085-ehrmann-en.pdf>; Eva Pfanzer et al., *Digital Interfaces of Historical Newspapers. Opportunities, Restrictions and Recommendations*. *Journal of Data Mining and Digital Humanities* (2021), <https://doi.org/10.46298/jdmdh.6121>; Estelle Bunout, Maud Ehrmann & Frédéric Clavert (toim.) *Digitised Newspapers. A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. De Gruyter 2023.
8. Mediahistorian näkökulmasta digitaalisia aineistoja on käsitelty joissakin artikkeleissa. Ks. esim. Nicole Maurantonio, *Archiving the Visual*. *Media History* 20:1 (2014), 88–102; Huub Wijffes, *Digital Humanities and Media History. A Challenge*

arkistoja on lähestytty esimerkiksi kulttuurisen perinnön säilyttämisen alustoina, menneisyyden yhteiskuntien peleinä ja erilaisten julkisessa keskustelussa esiin nousseiden teemojen ja yllärajan tiedonkierron tarkastelun materiaaleina.⁹ Samanaikaisesti etenkin sanomalehtien lähdekäyttö on kasvanut eksponentiaalisesti,¹⁰ ja moni mediahistoriaan perehtymätön tutkija operoi nykyisin lehtiaineistoilla. Tuoreen artikkelikokoelman otsikossa kysytäänkin, ovatko digitoidut sanomalehdet historian tutkijoiden uusi Eldorado, nykytutkimuksen ehtymätön lähdeaineistovarasto.¹¹

Katsaukseni kiinnittyy samoihin käsitteellisiin pohdintoihin, joita edellä mainitun artikkelikokoelman kirjoittajat tarkastelevat. Päätaavoitteena on kuitenkin osoittaa, miten digihistoriaan liittyvää käsitteellistä tietoa voi soveltaa käytännön tutkimustyössä¹². Ajattelen Lara Putnamin tavoin, että historian tutkijoiden enemmistöä koskettava digitalisaatio ansaitsee oman menetelmällisen tarkastelunsa.¹³ Siksi nostan esiin sellaisia lähdekriittisiä kysymyksiä, joita digitaalisten lehtitekstien laadullinen sisällönanalyysi

edellyttää.¹⁴ Valotan erityisesti sitä, missä määrin lehtiarkistoja käyttävän tutkijan pitäisi olla tietoinen lähdeaineiston mediahistoriallisesta kontekstista. Lisäksi konkretisoin aineistoesimerkeillä niitä käytännön tutkimustyöhön liittyviä kysymyksiä, jotka ovat seurausta lehtiaineistojen 2000-luvulla tapahtuneesta massadigitoinnista.¹⁵ Toivon, että mallintamalla omaa oppimisprosessiani pystyn rohkaisemaan digitaalisia lehtiaineistoja käyttäviä tutkijoita ja opiskelijoita yhdistämään tulkinnoissaan sekä digitaalisen historian että mediahistorian näkökulmia.

Lähdekritiikkiin uusia tasoja

Tietokoneavusteinen historian tutkimus ei ole nykypäivän ilmiö, sillä tutkijat kiinnostuivat tietokoneen käyttömahdollisuuksista jo 1960-luvulla.¹⁶ Esimerkiksi Viljo Rasila korosti tuolloin tietokoneen hyötyjä talous- ja sosiaalishistorian tutkimuksessa, mutta toi esiin myös tietokoneavusteiseen tutkimukseen liittyviä haasteita: määrällinen tarkastelu ei soveltunut samalla tavoin tieteenalalle ominaisten uniikkien ilmiöiden tutkimiseen.¹⁷ Hieman ohimennen Rasila pohti

for Historical Newspaper Research. *TMG Journal for Media History* 20:1 (2017), 4–24; Henrik Grue Bastiansen, Nettarkivernes digitale objekter som intellektuell utfordring. *Mediehistorie 2.0 og behovet for en ny filologi. Mediehistorisk Tidsskrift* 1 (2020), 33–46; Zoé Kergomard, A Source Like Any Other? Including Digitised Newspapers in a “Hybrid” Research Project. Teoksessa Estelle Bunout, Maud Ehrmann & Frédéric Clavert (toim.) *Digitised Newspapers. A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. De Gruyter 2023, 359–377.

9. Ks. esim. NewsEye-projektin kuvaus, <https://www.newseye.eu/about/> (28.10.2022); Impresso: Media Monitoring the Past -projektin tavoitteet, <https://impresso-project.ch/project/objectives/> (28.10.2022); Oceanic Exchanges: Tracing Global Information Networks In Historical Newspaper Repositories, 1840–1914 -hankkeen kuvaus, <https://oceanicexchanges.org/> (28.10.2022).
10. Adrian Bingham, The Digitization of Newspaper Archives. Opportunities and Challenges for Historians. *Twentieth Century British History* 21:2 (2010), 225–231, <https://doi.org/10.1093/tcbh/hwq007>; Bob Nicholson, The Digital Turn. *Media History* 19:1 (2013), 59–73; Petri Paju, Heli Rantala & Hannu Salmi, Tietokannoista tulkintoihin. Digitaalisen historian tutkimuksen käytäntöjä. *Ennen ja nyt. Historian tietosanomat* 2 (2019), <https://journal.fi/ennenjanyt/article/view/108925/63917> (28.2.2023); Fridlund 2020, 75–76.
11. Bunout, Ehrmann & Clavert (toim.) 2023.
12. Tiedon muodoista ja oppimisesta esim. John Biggs & Catherine Tang (toim.) *Teaching for Quality Learning at University*. McGraw-Hill Education 2011, 81–94.
13. Putnam 2016, 379.
14. Keskityn sellaisiin aineistokokonaisuuksiin eli korpusten tarkasteluun, joita tutkijan on mahdollista tulkita lähiluvun keinoin, kun taas tietokoneavusteiset etälukemisen menetelmät rajautuvat tarkastelun ulkopuolelle. Korpuksen rakentamisesta sekä lukemisen tavoista ks. esim. Heidi Hakkarainen et al., Digitoidut sanoma- ja aikakauslehdet lähdeaineistona. Teoksessa Rami Mähkä et al. (toim.) *Kulttuurihistorian tutkimus. Lähteistä menetelmiin ja tulkintaan*. Kulttuurihistorian seura 2022, 113–120.
15. Käytännön opastusta annetaan nykyisin myös opetuksessa. Lisäksi muutamissa tuoreissa artikkeleissa on pyritty samaan. Mediahistoriallinen keskustelu on kuitenkin rajautunut niissä pääsääntöisesti käsittelyn ulkopuolelle. Hakkarainen et al. 2022, 105–123; Kaisa Kyläkoski, Historiantutkija verkon ääressä. Hakee, hakee ja soisi löytävänsä. *Ennen ja nyt* 22:4 (2022), 33–47, <https://doi.org/10.37449/ennenjanyt.122103>.
16. Varhaisena esimerkkinä on Edward Shorterin tutkimusenteko-opas *The Historian and the Computer. A Practical Guide* (1971). Ks. Crymble 2021, 2; Shawn Graham, Ian Milligan & Scott Weingart, *Exploring Big Historical Data. The Historian's Macroscope*. 2 painos. Imperial College Press 2022, 17–20; Wijffes 2017, 5.
17. Viljo Rasila, Tietokone historian tutkimuksessa. *Historiallinen Aikakauskirja* 65:2 (1967), 140–146. Rasilasta ks. myös Mikko Tolonen & Leo Lahti, Digitaaliset ihmistieteet ja historian tutkimus. Teoksessa Matti Hannikainen, Mirrka Danielsbacka & Tuomas Tepora (toim.) *Menneisyyden rakentajat. Teoriat historian tutkimuksessa*. Gaudeamus 2018, 235.



historiantutkijan valmiuksia uuden tutkimusmenetelmän käyttöönottoon:

[T]ietokone avaa aivan uusia näkymiä myös historiantutkimukselle. Ennen kuin mahdollisuuksia kuitenkin voidaan käyttää kyllin tehokkaasti tutkimuksen hyväksi, on tutkijain tiedettävä: 1) mitä ja minkälaisia tehtäviä tietokone voi suorittaa, 2) minkälaisia tehtäviä tietokoneella kannattaa suorittaa, 3) mitä on tiedettävä ohjelmoinnista sekä 4) mitä on tiedettävä aineiston syöttötekniikasta.¹⁸

Rasilan peräänkuuluttama ymmärrys tietokoneen toimintamekanismeista esiintyy myös nykyisissä keskusteluissa digitaalisesta lähdekritiikistä sekä digitaalisten työkalujen kritiikistä.¹⁹ Osa tutki-

■ Kuva 1. Ennen digitaalisia arkistoja leikekokoelmat autoivat tutkijoita sanomalehtikeskustelujen äärelle. Leikkeitä arkistoiitiin myös lehtien toimituksissa. Kuvassa *Hufvudstadsbladetin* arkisto ja kirjasto vuodelta 1934. Lähde: Journalistinen kuva-arkisto, JOKAHBL3B_E01:20. CC BY 4.0.

joista on puhunut jopa historiantutkimuksen vallankumouksesta,²⁰ kun taas toiset ovat olleet varovaisempia ja korostaneet digitaalisissa tutkimusympäristöissä vaadittavia uudenlaisia taitoja.²¹ Petri Paju, Heli Rantala ja Hannu Salmi kirjoittavat verkkomaailman konehuoneista, joissa ”tehdyt päätökset ohjaavat aineiston käyttöä ja käyttäjiä”.²² Tähän keskusteluun kiinnittyy myös Helle Strandgaard Jensenin digitaalisen arkistolukutaidon (*digital archival literacy*) käsite. Sen

18. Rasila 1967, 146.

19. Esim. Kimmo Elo, Digitaalisen historiantutkimuksen kenttää louhimassa. Teoksessa Kimmo Elo (toim.) *Digitaalinen humanismi ja historiatieteet*. Turun historiallinen yhdistys 2016, 24; Marijn Koolen, Jasmijn van Gorp & Jacco van Ossenberg, Toward a Model for Digital Tool Criticism. Reflection as Integrative Practice. *Digital Scholarship in the Humanities* 34:2 (2019), 368–385; Estelle Bunout, Contextualising Queries. Guidance for Research using Current Collections of Digitised Newspapers. Teoksessa Estelle Bunout, Maud Ehrmann & Frédéric Clavert (toim.) *Digitised Newspapers. A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. De Gruyter 2023, 279, 291–292.

20. Julia Flanders, The Productive Unease of 21st-century Digital Scholarship. *Digital Humanities Quarterly* 3:3 (2009), <http://www.digitalhumanities.org/dhq/vol/3/3/000055/000055.html> (28.2.2023); Anders Frickers, Towards a New Digital Historicism? Doing History in the Age of Abundance. *View* 1:1 (2012), 19–26; Wijfjes 2017, 5–6.

21. Esim. Nils Brügger, *The Archived Web. Doing History in the Digital Age*. The MIT Press 2018, 14.

22. Paju, Rantala & Salmi 2019.

keskiössä on ymmärrys niistä taloudellisista, kulttuurisista ja poliittisista reunaehdoista, jotka määrittelevät digitaalisten arkistojen muotoutumista ja niiden käyttöä.²³

Jensenin mukaan digitaalinen arkistolukutaito auttaa meitä ymmärtämään erilaisia kokoelmien syntyyn vaikuttaneita taustatekijöitä. Lisäksi se haastaa pohtimaan, mitä meidän on mahdollista löytää, kun toteutamme käyttöliittymässä erilaisia hakuja tai käytämme tietokoneavusteisia analyysimenetelmiä.²⁴ Käsite korostaa käyttöliittymään ja digitoitujen dokumenttien syntymekanismiin kohdistuvaa kriittisyyttä, jota Estelle Bunout on omassa mallissaan täydentänyt digiaineistoja käyttävän tutkijan itsereflektiolla. Niin sanotun kysymyskritiikin (*query criticism*) eli tekemiensä hakujen kriittisen tarkastelun ja mahdollisen uudelleen suuntaamisen lisäksi nykytutkijoilta vaaditaan ymmärrystä siitä, minkälaisia kysymyksiä digitaalisille aineistoille on ylipäätään mahdollista esittää.²⁵ Myös digitaalisen hermeneutiikan käsite yhdistää historian tutkimuksen käytänteiden kriittisen tarkastelun digitaalisten tutkimusympäristöjen lukutaitoon.²⁶

Sinällään tutkimusympäristön voimakas muutos ei ole ennen kokematon. Aineistojen digitalisaatio on verrattavissa 1800-luvun lopulla tapahtuneeseen kansallisarkistojen perustamiseen. Myös tuolloin tehtiin valintaa sen suhteen, mitä säilytetään ja asetetaan käyttäjien saataville.²⁷ Samalla tavoin toimitaan nyt. Tutkijoina olemme kuitenkin tottuneita tiedostamaan arkistoissa

säilyneiden aineistojen valinnaisuuden ja sen vaikutuksen tulkintoihin, kun taas tietokoneen ruudulle ilmestyvien aineistojen kohdalla samat kysymykset jäävät helposti pohtimatta.²⁸

Koska digitalisaatio on aiheuttanut tutkimuksellisen käänteeseen, joka tavalla tai toisella koskettaa kaikkia historian tutkijoita, näiden kysymysten pohtiminen on ensiarvoisen tärkeää. Onhan jo yksinomaan optisen tekstintunnistuksen (*optical character recognition*, OCR) kehittyminen vaikuttanut merkittävästi niihin tapoihin, joille rajaamme ja läpikäymme aineistoja. Uudet työskentelytavat edellyttävät tutkijalta myös kykyä toimia aineistopaljouden keskellä sekä tarkkuutta hakuosumien kontekstoinnissa,²⁹ mitä käsitellen seuraavaksi.

Erilaisten kontekstien ymmärtäminen tärkeää

Yksi keskeisimmistä digitalisaatioon liittyvistä hyödyistä on aineistojen helppo saatavuus. Osin tästä johtuen uusimman ajan tutkijat käyttävät sanoma- ja aikakauslehtiaineistoja yhä useammin tutkimuksensa täydentävänä lähteenä. Lehtiaineistojen lukeminen eri puolilta maailmaa on myös entistä helpompaa, ja uusien lehtisivujen digitointi tuo koko ajan laajenevan lähdejoukon tutkijoiden ruuduille. Esimerkiksi Iso-Britannian sanomalehtiarkistossa (*The British Newspaper Archive*) digitoidaan viikottain satoja tuhansia sivuja, ja käyttöliittymä avaa jo nyt pääsyn yli 70 miljoonaan lehtisivuun noin 2000 sanomaleh-

23. Helle Strandgaard Jensen, Digital Archival Literacy for (all) Historians. *Media History* 27:2 (2021), 251–254, doi:10.1080/13688804.2020.1779047. Jensenin kokoamia ajatuksia löytyy myös muiden tutkijoiden teksteistä. Ks. esim. Ehrmann, Bunout & Düring 2019, 2; Henrik Grue Bastiansen, Når mediehistorien blir søkbar – og grafisk. Introduksjon til Nasjonalbibliotekets første IT-verktøy. *Mediehistorisk Tidsskrift* 2 (2019), 133; Bastiansen 2020 42–43; Nicholson 2013, 64–67; Wijffes 2017, 10; Bunout 2023, 285–292.

24. Jensen 2021, 259–260.

25. Bunout 2023, 279–280.

26. Fickers, Tatarinov & van der Heijden 2022, 6–11. Ks. myös Tolonen & Lahti 2018, 236; Sarah Oberbichler et al., Integrated Interdisciplinary Workflows for Research on Historical Newspapers. Perspectives from Humanities Scholars, Computer Scientists, and Librarians. *Journal of the Association for Information Science and Technology* 73:2 (2022), 226.

27. Jani Marjanen, How Digitalization Changes History (and maybe also Historical Research). NewsEye-projektin blogi 12.3.2018, <https://www.newseye.eu/blog/news/how-digitalization-changes-history-and-maybe-also-historical-research/> (28.10.2022); Johan Jarlbrink & Pelle Snickars, Cultural Heritage as Digital Noise. Nineteenth Century Newspapers in the Digital Archive. *Journal of Documentation* 73:6 (2017), 1229.

28. Zaagsma 2022, 1–3, 8–15.

29. Käytännön tutkimustyön muutoksia käsitellään seikkaperäisesti lukuisissa alan julkaisuissa. Samanaikaisesti monet elementit tutkijan työssä ovat kuitenkin säilyneet digitaalisissa ympäristöissä ennallaan, mistä johtuen muutoksia ei tulisi korostaa liikaa. Ks. esim. Nicholson 2013, 61–64; Paul Gooding, *Historic Newspapers in the Digital Age. "Search all about it!"*. Routledge 2017, 1–3; Maud Ehrmann, Estelle Bunout & Frédéric Clavert. Digitised Historical Newspapers. A Changing Research Landscape. Introduction. Teoksessa Estelle Bunout, Maud Ehrmann & Frédéric Clavert (toim.) *Digitised Newspapers. A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. De Gruyter 2023, 1–5; Bunout 2023, 277–285.

destä.³⁰ Aineistopaljouden keskellä hakuosumia arvioidaan kuitenkin helposti painoarvoaan merkittävämmiksi tai niitä tulkitaan virheellisesti, jos tutkija ei perehdy riittävästi osumiensa yhteiskunnalliseen ja mediahistorialliseen kontekstiin.³¹ Näin tapahtui minulle, kun kirjoitin englanninkielistä ensyklopedia-artikkelia feminismistä eurooppalaisissa sanomalehdissä.

Keräsin käsittelyäni varten lehtitekstejä muun muassa Iso-Britannian sanomalehtiarkiston kokoelmista ja valikoin tekstiin sitaatin, joka kuvasi hyvin argumenttiani.³² Kontekstualisoin sitaatin asianmukaisesti omana aikanaan kierätettyihin feminismiin stereotyyppioihin sekä toin esiin sen, kuinka toimittaja tarkasteli siinä brittiläistä naisasialiikettä suhteessa yhdysvaltalaiseen feminismiin. Sen sijaan en pohtinut tarkemmin artikkelin julkaisukontekstia vaan havahduin vasta jälkepäin siihen, että kyse oli Aberdeenin kaupungissa Skotlannissa julkaistusta paikallislehdestä. Vaikka tulkintani oli riittävä lyhyelle ensyklopediatekstillemme, olisin voinut rikastuttaa sitä kiinnittämällä huomiota paikallislehden rooliin uusien ajattelutapojen leviämässä. Mediahistoriallinen konteksti häivytti analyysissäni kuitenkin taka-alalle, kun tein tulkintoja ilman kunnollista kulttuurista tuntemusta, mistä Putnam on varoittanut.³³

Laajempi konteksti katoaa herkästi, kun tutkija innostuu yksittäisestä hakuosumasta. Esimerkkinä voidaan tarkastella *Karjala*-lehden vuonna 1917 julkaisemaa mainintaa naiseksi pukeutuneesta miehestä, joka joutui venäläisten

sotilaiden pahoinpitelemäksi.³⁴ Jos sukupuolen moninaisuuden historiasta kiinnostunut tutkija tulkitsee uutista perehtymättä tarkemmin julkaisujankohdan muuhun uutisointiin, häneltä jää huomaamatta, että sotilaiden aiheuttamat järjestyshäiriöt olivat *Karjalassa* toistuvan uutisoinnin aiheena.³⁵ Näin ollen normeja rikkonut sukupuolen esittäminen ei näyttäydä keskeisenä syynä uutiselle tai edes sitä edeltäneelle tapahtumaketjulle, vaan kyse oli ajan poliittisessa ilmapiirissä tapahtuneesta kiehuunasta, joka herätti Viipurissa julkista keskustelua Suomen itsenäistymistä edeltäneinä kuukausina.

Tutkijan on syytä kiinnittää huomiota myös uutisen alkuperään sekä sijoittumiseen lehdessä. *Karjalasta* tuttu ilmaisu 'naiseksi pukeutunut mies' toistuu esimerkiksi *Keskisuomalaisen* uutisessa vuodelta 1932³⁶. Tällä kertaa sukupuolien rikkominen on nostettu uutisen kärkeen, sillä otsikko tiesi kertoa "'Daami' olikin – mies!'"³⁷ Suomen tietotoimiston (STT) välittämä uutinen kertoi dramaattisin sanakäntein tukholmalaisessa yökerhossa tapahtuneesta välikohtauksesta sekä sitä seuranneesta poliisin takaa-ajosta. Uutinen toi eittämättä lukijoiden ulottuville tiedon henkilöstä, joka kyseenalaisti käytöksellään sukupuolen kaksinaapaista jakoa, mutta normin rikkomus tapahtui turvallisesti Suomen rajojen ulkopuolella. Lisäksi uutisen journalistista merkitystä vähensi humoristinen kerronta sekä tekstin sijoittaminen viihteellisten pikku-uutisten joukkoon. Kyse oli täyteuutisesta, jonka tarkoituksena oli viihdyttää lukijoita.

30. Rose Staveley-Wadham, We've Reached 50 Million Pages. Celebrating 50 Fantastic Things About The Archive. The British Newspaper Archive -sivuston blogi 5.4.2022, <https://blog.britishnewspaperarchive.co.uk/2022/04/05/archive-reaches-50-million-pages/> (28.2.2023); The British Newspaper Archive -sivuston etusivu, <https://www.britishnewspaperarchive.co.uk/> (29.8.2023).

31. Bingham 2010, 227–228; Hitchcock 2013, 11–18; Putnam 2016, 383–389; Bunout 2023, 276–280; Kergomard 2023, 360–366. Niin sanottu rusinat pullasta -lähestymistapa (*cherry-picking*) ei ole ollut vieras analogisia lehtiaineistoja käyttäville tutkijoillekaan, mutta siihen on entistä helpompaa hairahtua digitaalisissa tutkimusympäristöissä. Ks. Penny Tinkler, Fragmentation and Inclusivity Methods for Working with Girls' and Women's Magazines. Teoksessa Rachel Ritchie et al. (toim.) *Women in Magazines. Research, Representation, Production and Consumption*. Routledge 2016, 26, 37–38.

32. Heidi Kurvinen, Feminism in the News. Teoksessa Karen Ross et al. (toim.) *The International Encyclopedia of Gender, Media and Communication*. Wiley-Blackwell 2020, 4, <https://doi.org/10.1002/9781119429128.iegmc202>.

33. Putnam 2016, 394–497.

34. Toissayöllinen katumeteli. *Karjala* 30.6.1917, 3, <https://digi.kansalliskirjasto.fi/sanomalehti/binding/1244056?page=3> Kansalliskirjaston digitaaliset aineistot (KDA).

35. Esim. Sotilasmielivalta Viipurissa ja tšekäläinen poliisilaitos. *Karjala* 1.7.1917, 3, <https://digi.kansalliskirjasto.fi/sanomalehti/binding/1244367?page=3>, KDA; Sotilasmielivalta. *Karjala* 30.8.1917, 3, <https://digi.kansalliskirjasto.fi/sanomalehti/binding/1244370?page=3>, KDA.

36. Sama uutinen julkaistiin myös *Savon Sanomissa*, *Kannaksen Lehdessä*, *Käkisalmen Sanomissa*, *Ylä-Vuoksessa*, *Länsi-Savossa* ja *Etelä-Suomen Sanomissa*.

37. "Daami" olikin – mies! *Keskisuomalainen* 3.5.1932, 3, <https://digi.kansalliskirjasto.fi/sanomalehtibinding/1864532?page=3>, KDA.

Hakuosumien ja niiden julkaisukontekstin välisen yhteyden katoamista edesauttaa se, että lehtiaineistot ovat digitoituna etäännyneet toimituksellisesta kontekstistaan, mikä historian-tutkijoilla on muutoinkin ollut tapana häivyttää näkyvistä³⁸. Mediahistoriallisen kontekstin vaikutus tulkintojen kestävytyteen vaihtelee toki lähestymistavasta ja tutkimuskysymyksestä riippuen. Jos tutkija esimerkiksi käyttää isoja aineistomasoja selvittääkseen lehtien välisiä tiedonkierron tapoja,³⁹ yksittäisten lehtien journalistisen kontekstin avaaminen ei ole välttämättä olennaista tutkimuskysymyksen ratkaisemisessa. Sen sijaan uudessa julkaisukontekstissa teksteille annettujen merkitysten analysoiminen edellyttää lähilukemista.⁴⁰ Tällöin lehtitekstien mediahistoriallinen kontekstointi on välttämätöntä, jotta voimme tehdä niistä perusteltuja tulkintoja. Esimerkiksi ymmärrys tekstien toimituksellisista syntyprosesseista sekä eri lehtityypeille⁴¹ ominaisista piirteistä syventävät olennaisesti lähilukemalla tehtyä analyysiä.⁴² On myös hyvä muistaa, ettei journalistinen konteksti ole irrallaan muista tutkimuksellisista konteksteista, kuten tutkimusajan laajemmasta yhteiskunnallisesta keskustelusta, vaan asettuu tulkintoja tehtäessä vuoropuheluun niiden kanssa.⁴³ Lisäksi toimitustyön luonteessa eri aikoina tapahtuneet muutokset

vaikuttavat tulkintoihin. Esimerkiksi suorat käännöstekstit ja niin sanottu leikkaa-liimaa-journalismi sekä tekstien kierrättäminen olivat leimalista 1800-luvun journalismille, mutta samaisen vuosisadan jälkipuoliskolta lähtien toimittajien aktiivinen rooli uutistuotannossa on koko ajan kasvanut.⁴⁴

Lehtitekstin kontekstoiminen oman aikansa mediaympäristöön ja siihen julkaisuun, josta teksti on löydetty, haastaa tutkijaa esittämään jatkokysymyksen: mitä lähteitä digitaalinen haku ei ole tavoittanut ja miten mahdolliset katveet vaikuttavat tutkimustuloksiin?⁴⁵ Tällöin lähdekriittinen katse kohdistuu digitaalisen lehtiarkiston institutionaaliseen ja teknologiseen kontekstiin, kuten Bunout kirjoittaa.⁴⁶ Käytännössä kaikkea analogisten arkistojen aineistoista ei ole vielä digitoitu eikä ehkä koskaan tullakaan digitoimaan. Samanaikaisesti aineistojen näennäisen helppo saatavuus kannustaa tutkijoita tietynlaisen tutkimuskysymysten äärelle ja vinouttaa käsitystä siitä, mikä on tutkimisen arvoista.⁴⁷ Esimerkiksi Suomessa naistenlehdet ovat olleet keskeisiä yhteiskunnallisen keskustelun paikkoja,⁴⁸ mutta Kansalliskirjaston digitoiduista kokoelmista niihin ei pääse käsiksi kuin 1940-lukua edeltävältä ajalta. Kotimaisista sanomalehdistä on katsauksen julkaisuhetkellä löydettävissä muutama

38. Aiheesta esim. Heidi Kurvinen, Toimittajat ja toimitusprosessit mediatekstien takana. *Historiallinen Aikakauskirja* 116:3 (2018), 310–322; Kergomard 2023, 362.
39. Esim. Hannu Salmi et al., Återanvändningen av text i den finska tidningspressen 1771–1853. *Historisk Tidskrift för Finland* 103:1 (2018), 46–76, <https://journal.fi/htf/article/view/80405>; Sofia Paasikivi et al., Infectious Media. Cholera and the Circulation of Texts in the Finnish Press, 1860–1920. *Media History* (2022), doi:10.1080/13688804.2022.2054408.
40. Hannu Salmi et al., Shared and Parallel Histories of Swedish-language Newspapers. An introduction. Teoksessa Patrik Lundell et al. (toim.) *Information Flows across the Baltic Sea. Towards a Computational Approach to Media History*. Föreningen Mediehistoriskt arkiv 2023, 16; Patrik Lundell, Recontextualisations and Transformations in the Information Flows. A worm’s-eye view through the temperance question. Teoksessa Patrik Lundell et al. (toim.) *Information Flows across the Baltic Sea. Towards a Computational Approach to Media History*. Föreningen Mediehistoriskt arkiv 2023, 81–82.
41. Historiantutkijoiden yleisesti käyttämistä lehdistä voidaan erottaa ainakin sanoma- ja aikakauslehdet, pilalehdet, järjestölehdet, ammattilehdet sekä organisaatioiden sisäiseen ja ulkoiseen viestintään tarkoitetut lehdet. Lehtityyppi vaikuttaa osaltaan siihen, millä tavoin niissä käsitellään tutkijaa kiinnostavia kysymyksiä.
42. Bingham 2010, 230; Nicholson 2013, 71; Wijffes 2017, 12–13, 21; Kergomard 2023, 362–366. Ks. myös Bunout 2023, 292–296; Putnam 2016, 392. Kontekstoinnin merkitys koskee yhtä lailla analogisia lehtiaineistoja käytäviä tutkijoita. Ks. esim. Tinkler 2016, 25–26.
43. Jukka Kortti, *Mediahistoria. Viestinnän merkityksiä ja muodonmuutoksia puheesta bitteihin*. SKS 2016, 33–34; Tinkler 2016, 31–32.
44. Pirjo Munck, *Valistajista ammattimieheksi. Toimittajien ammattilaistumisen pitkä tie 1771–1921*. Helsingin yliopisto 2016, 62–74; Hannu Salmi et al., The Reuse of Texts in Finnish Newspapers and Journals, 1771–192. A Digital Humanities Perspective. *Historical Methods. A Journal of Quantitative and Interdisciplinary History* 54:1 (2021), 14, 26; Lundell 2023, 78–79.
45. Jensen 2021, 253–255.
46. Bunout 2023, 278–279, 285–292.
47. Jensen 2021, 253–255; Putnam 2016, 389–392; Wijffes 2017, 16.
48. Esim. Laura Saarenmaa & Iris Ruoho, Women’s Magazines in the Nordic Style. Politics, Politicians and the Welfare State. *European Journal of Communication* 29:3 (2014), 289–303.

kokonaisuudessaan digitoitu lehti, joita ovat *Maaseudun Tulevaisuus*, *Länsi-Savo*, *Etelä-Suomen Sanomat* ja *Uusi Suomi*.⁴⁹ Lisäksi esimerkiksi *Apu*, *Hufvudstadsbladet*, *Ny Tid*, *Tekniikan Maaailma* ja *Suomen Kuvalehti* ovat suurelta osin luettavissa digitaalisessa muodossa myös viime vuosikymmeniltä. Varsin sattumanvaraiselta vaikuttava valikoima selittyy sillä, että kyseisten lehtien vanhoja vuosikertoja on digitoitu ulkopuolisella rahoituksella niiden vuosikymmenten osalta, joihin Kansalliskirjaston massadigitointi ei ole vielä ulottunut.⁵⁰ Näin ollen tutkimusaineistoissa on ollut selkeä vinouma, mikä on viime vuosina jossain määrin korostanut keskustalais-kokoomuslaisia poliittisia näkökantoja⁵¹. Lähi-tulevaisuudessa tosin suomenruotsalaisen lehdistön digitoitujen versiot avautuvat kokonaisuudessaan Kansalliskirjaston palvelun kautta tutkijoiden käyttöön, ja ulkopuolisella rahoituksella digitoidaan parhaillaan myös esimerkiksi *Helsingin Sanomia* ja *Ilta-Sanomia*.⁵²

Tällä hetkellä Kansalliskirjaston digipalvelussa on yli 17 miljoonaa sanomalehtisivua ja noin seitsemän miljoonaa aikakauslehtisivua. Avoimesti käytettävissä ovat suomenkieliset aineistot, jotka on julkaistu ennen vuotta 1939 sekä ruotsinkieliset aineistot vuoteen 1949 asti.⁵³ Käyttöliittymän ominaisuuksia myös parannetaan koko ajan. Vuosien 1771–1914 sanomalehtiaineistot on esimerkiksi viime vuosina tekstintunnistettu uudelleen, mikä on parantanut hakuosumien laatua keskimäärin 17 prosenttiyksikköä.⁵⁴ Digitaalisille arkistoille ominainen jatkuva muutos vaikuttaa näin ollen varsin konkreettisesti hakuhetkellä saatavien

osumien listaukseen ja tulosten luotettavuuteen. Tästä johtuen hakuhetkeä määrittäneet reunaehdot on syytä kirjoittaa läpinäkyvästi auki valmiiseen tutkimukseen. Tutkijan on lisäksi kyettävä arvioimaan, minkälaisia tulkintoja aineistosta on kussakin hetkessä mahdollista tehdä.⁵⁵

Omassa tutkimuksessani digiarkistojen jatkuvasti täydentyvä luonne on korostunut käyttäessäni Kansalliskirjaston digipalvelun aineistoja lähivuosisikymmeniltä. Koska eri sanomalehtien digitaalisia vuosikertoja on ilmestynyt käyttöliittymään pitkin tutkimusprosessia, olen joutunut muuttamaan ymmärrystäni tutkimusaineiston koostumuksesta. Zoé Kergomardin tavoin olen siirtynyt suljetun korpuksen ihanteesta kohti aukinaista lähdeaineistoa, joka voi saada uusia muotoja tutkimusprosessin kuluessa.⁵⁶ Samanaikaisesti olen hyödyntänyt aineistonkeruussa erilaisia lähteidenlukutapoja. Olen yhdistänyt yksittäisten tekstien läheltä lukemisen niiden katsomiseen etäältä ja havainnut siten kehityskulkuja ja yhteyksiä, jotka ovat ohjanneet uusien aineistojen äärelle.⁵⁷ Se on avannut mahdollisuuden löytää jotain sellaista, jota en alun perin osannut edes etsiä.⁵⁸ Seuraavaksi käsittelen tarkemmin digitaalisten arkistojen käyttöliittymien erityispiirteitä ja niiden vaikutuksia tutkimusprosessissa.

Arkistojen käyttö lisää käsitteellistä ymmärrystä

Digitoisissa ja digitaalisten aineistojen esillepanossa on aina kyse valinnoista, jotka tiivistyvät käyttäjille tarjotussa käyttöliittymässä. Sen toi-

49. Usein kysytyt kysymykset, <https://digi.kansalliskirjasto.fi/faq> (28.8.2023).

50. Hakkarainen et al. 2022, 109. Vastaavanlainen tilanne on esimerkiksi Hollannissa. Ks. Wijffes 2017, 15–17.

51. Vuotta 1939 uudemmat aineistot ovat katsauksen kirjoittamishetkellä tutkijoiden käytössä haka-kirjautumisen kautta.

52. Digitoituista ruotsinkielisistä sanomalehdistä tuli yleisömenestys. Kansalliskirjaston digipalvelu 14.11.2022, <https://www.kansalliskirjasto.fi/fi/uutiset/digitoituista-ruotsinkielisista-sanomalehdista-tuli-yleisömenestys> (28.2.2023); Johanna Lilja & Tuula Pääkkönen, Kansalliskirjasto digitoi. Tutkijoille ja kansalaisille. *Ennen ja nyt* 22:4 (2022), 11–12, <https://doi.org/10.37449/ennenjanyt.122833>.

53. <https://digi.kansalliskirjasto.fi/stats> (22.8.2023); Digi.kansalliskirjasto.fi. 25 miljoonaa sivua historiaa. Kansalliskirjaston digipalvelu 12.12.2022, <https://www.kansalliskirjasto.fi/fi/uutiset/digikansalliskirjastofi-25-miljoonaa-sivua-historiaa> (28.2.2023).

54. Erno Liukkonen, Tuula Pääkkönen & Minna Kaukonen, ”Turho hölytyt” on nyt ”turha hälytyt”. Digitoidun sanomalehtiaineiston sisältöhaun parannukset käytössä. *Tietolinja* 2 (2021), <https://urn.fi/URN:NBN:fi-fe2021121661233>; Lilja & Pääkkönen 2022, 11–12.

55. Wijffes 2017, 16; Jensen 2021, 260; Martina Hjertman, *Afloat and Aflame. Deconstructing the Long 19th Century Port City Gothenburg through Newspaper Archaeology*. Universus Academic Press 2022, 69–70.

56. Kergomard 2023, 367–368.

57. Tim Hitchcock, Big Data, Small Data and Meaning. *Historyonics* 9.11.2014, http://historyonics.blogspot.com/2014/11/big-data-small-data-and-meaning_9.html (28.2.2023); Putnam 2016, 386–387; Julia Laite, The Emmet’s Inch. Small History in a Digital Age. *Journal of Social History* 53:4 (2020), 975.

58. Putnam 2016, 377, 386–387; Bunout 2023, 278; Kergomard 2023, 371–373.

mintamekanismien ymmärtäminen on tutkijalle välttämätöntä. Vaikka lehtiarkistojen käyttöliittymät muistuttavat monelta osin toisiaan, niissä on lukuisia hakujen tekemiseen ja tulosten esittämiseen liittyviä eroavaisuuksia.⁵⁹ Esimerkiksi Kansalliskirjaston digipalvelussa haku on mahdollista kohdentaa tekstien ohella lehtikuviin, kun taas Ruotsin Kuninkaallisen kirjaston sanomalehtiarkistossa kuvahakumahdollisuutta ei ole olemassa. Norjan Kansalliskirjastossa on käytössä N-gram-työkalu,⁶⁰ joka muistuttaa Googlen palvelua ja antaa mahdollisuuden nähdä selkeitä trendejä erilaisten hakusanojen kehityksessä. Norjalainen käyttöliittymä näyttää osumat suhteellisina frekvensseinä, kun taas Kansalliskirjaston hakuosumista katsottavissa oleva kaavio esittää absoluuttiset frekvenssit. Kansalliskirjaston digipalvelussa osumat on kuitenkin helppo suhteuttaa haussa mukana olleiden lehtisivujen määrään käyttöliittymästä saatavan Excel-tiedoston avulla. Kaikissa lehtiarkistoissa suhteuttaminen ei onnistu puuttuvien metatietojen vuoksi, ja ne saattavat myös käyttää erilaisia suhteuttamisen mittareita. Lehtisivujen määrän ohella osuman frekvenssi voidaan laskea esimerkiksi miljoonaa sanaa kohden.

Tietoisuus arkistokohtaisista eroista on edellytys sille, että hakutuloksista voidaan tehdä luotettavia tulkintoja.⁶¹ Käytännössä uuteen lehtiarkistoon tutustuminen tapahtuu usein yrityksen ja erehdyksen kautta. Se voi kuulostaa epätehtäiselältä, mutta digiaineistojen kohdalla käsillä tekemisen on itse asiassa osoitettu kehittävän käsitteellistä ymmärrystä, jos se yhdistetään kriittiseen ajatteluun (*thinking*).⁶² Etenkin aluksi digitaalisten lehtiarkistojen rohkea testailu on jopa välttämätöntä, jotta tutkijan ymmärrys niiden tutkimukselle avaamista mahdollisuuksista kehittyi. Minä esimerkiksi ehdin tehdä Tanskan Kuninkaallisen kirjaston sanomalehtiarkistossa useampia hakuja ennen kuin ymmärsin, miten digitoitujen lehtien sisällä oli mahdollista liikkua ilman, että lehteä oli selattava sivu sivulta.

En myöskään aluksi huomannut, että Ruotsin Kuninkaallisen kirjaston sanomalehtiarkistossa oli mahdollista tallentaa suosikkisivuja, vaan keräsin aineistoa talteen yksinomaan kuvamalla. Lopulta aineistonkeruutapojen yhdistelmä osoittautui tässä tapauksessa toimivimmaksi, sillä omalle listalle tallennettuja leikkeitä ei ole mahdollista lukea Suomessa. Ruotsalaiset ja tanskalaiset aineistot ovatkin kokonaisuudessaan tavoitettavissa vain kirjastojen omilla hakupäätteillä, kun taas Islannin, Färsearten ja Grönlandin lehdistöä kokoava Timarit.is sekä osa Norjan Kansalliskirjaston digitoimista sanoma- ja aikauslehtiaineistoista on luettavissa kotikoneelta Suomessa.

Pohjoismaisten lehtiarkistojen käyttöliittymien eroavaisuudet ilmentävät hyvin Maud Ehrmannin, Estelle Bunoutin ja Marten Düringin ajatusta digitoitujen sanomalehtiarkistojen neljästä sukupolvesta: niiden edetessä arkistot kehittyvät vuorovaikutteisemmiksi ja tutkijoiden tarpeita paremmin palveleviksi. Kehittyneemmissä käyttöliittymissä tutkija voi esimerkiksi rakentaa omia leikekokoelmia. Aineiston käyttömahdollisuuksia on myös voitu rikastaa tarjoamalla vaikkapa mahdollisuus nimentunnistukseen (*Named-entity recognition*, NER). Tällöin käyttöliittymä laskee esiintymisfrekvenssejä esimerkiksi henkilöiden, paikkojen ja tapahtumien nimille.⁶³

Useimmille tutkijoille keskeisin suurten aineistomassojen läpikäyntiä helpottanut ominaisuus on kuitenkin lehtiarkistoille standardiksi muodostunut sanahaku.⁶⁴ Myös kaikki pohjoismaiset digiarkistot tarjoavat käyttäjälleen ensimmäisenä avoimen sanahakukentän sekä erilaisia hakuja rajoittavia vaihtoehtoja, kuten ilmestymisajankohdan ja julkaisun nimen. Yksinkertaista sanahakua on usein mahdollista täydentää sumealla haulilla, jolloin osumiin sisältyy haetun sanan kaltaisia sanoja. Käyttäjää ohjeistetaan lisäksi erilaisten tuloksia parantavien hakumuotojen, kuten kirjastojen käyttöliittymistä tuttujen Boolean operaattoreiden, käyttöön.⁶⁵

59. Nicholson 2013, 64–66; Koolen, van Gorp & van Ossenbruggen 2019, 369–370, 379–382; Jensen 2021, 251–252; Bunout 2023, 285–288.

60. Bastiansen 2019, 134–139.

61. Pfanzelter et al. 2021, 13–16.

62. Esim. Fickers, Tatarinov & van der Heijden 2022, 7–8.

63. Ehrmann, Bunout & Düring 2019, 8–18.

64. Esim. Hitchcock 2013, 14–15; Hakkarainen et al. 2022, 110–113.

65. Jensen 2021, 257–258; Oberbichler et al. 2022, 232–233; Hjertman 2022, 72–73.

Samanaikaisesti onnistunut aineistonkeruu edellyttää tutkijalta ymmärrystä tutkittavasta ilmiöstä ja sen laajemmasta kontekstista: saamme käyttöliittymästä ulos sen, mitä osaamme siltä kysyä.⁶⁶ Yrittäessäni paikantaa feminististä liikettä käsitellyttä kirjoittelua 1960- ja 1970-luvun vaihteen suomalaisista lehdistä en esimerkiksi aluksi huomannut käyttäviä englanninkielisen ilmaisun käännoästä 'naisten vapautusliike', koska ajattelin, ettei se suorana käännoäslainana olisi voinut olla vakiintuneessa käytössä. Sen sijaan suoritin hakuja asiasanoilla 'naisasia', 'naiskysymys', 'sukupuolirooli', 'feminismi' ja 'feministi' sekä edistyneempien hakujen yhdistelmillä, kuten eri taivutusmuodot hakuihin sisällyttävällä katkaisumerkillä *⁶⁷. Vasta törmätyäni suomenkielisessä lehtiartikkelissa englanninkieliseen ilmaukseen 'women's lib' ymmärsin lisätä hakusanoihini naisten vapautusliikkeen eri variaatioineen ja kolmella kielellä. Tämä oli tutkimukselleni ratkaisevaa, sillä tulokset avasivat laajemman näkökulman siihen keskusteluun, jota Suomessa tuohon aikaan käytiin.

Esimerkkini on hyvä osoitus siitä, että digitaalisissa ympäristöissä korostuu tutkijan kyky käyttää oikeanlaisia hakusanoja. Siksi käytettyjen hakusanojen kriittinen tarkastelu ja niiden aukikirjoittaminen on välttämätöntä.⁶⁸ Impresso-projektin kaltaisissa kehittyneemmissä käyttöliittymissä on tosin jo nykyisin tutkijan hakusanastoa kehitäviä ominaisuuksia, kuten mahdollisuus nähdä haetun sanan OCR-muunnelmia eli optisen tekstintunnistuksen väärintulkintamaisia sanoja.⁶⁹ Lisäksi erilaisten hakusanojen yhdistely ja saatujen tulosten lukeminen ristiin lisäävät kaikissa käyttöliittymissä tulosten luotettavuutta. Tutki-

jalta vaaditaan myös ymmärrystä historiallisesta kielenkäytöstä, siinä tapahtuneista muutoksista ja sanojen mahdollisista piilomerkityksistä.⁷⁰ Omaa sanavarastoa on mahdollista kartuttaa perehtymällä esimerkiksi tutkimusajankohdan lehtikieleen ja siinä käytettyihin ilmaisiin. Sukupuoli- ja seksuaalivähemmistöjen historian osalta sanomalehtien käyttämää suomenkielistä terminologiaa on kartoitettu ansiokkaasti Sateenkaarihistorian hakusanakirjassa.⁷¹

Käyttöliittymä on keskeisessä roolissa myös hakuosumien esittämisessä, mikä aiheuttaa esimerkiksi hakutulosten näennäistä tasa-arvoisuutta.⁷² Toisin sanoen hakutuloksia selatessa kaikki tulokset näyttävät yhtä tärkeitä, vaikka hakukone tarjoaisikin mahdollisuuden järjestää osumat osuvuuden perusteella, kuten Kansalliskirjaston digipalvelussa on mahdollista tehdä. Niinpä esimerkiksi hakukomentoni "Betty Friedan" AND "Naisellisuuden harhat" tuotti osumiksi muun muassa kirja-arvioita Friedanin vuonna 1967 suomennetusta kirjasta *Naisellisuuden harhat* ja artikkeleita hänen samana vuonna tapahtuneesta Suomen-vierailustaan, mutta myös yleisiä artikkeleita feminismistä⁷³. Käytännössä haun mahdollistanut ohjelmisto laski jokaiselle osumalle arvon, jossa hakusanaparin sivukohtaista toistuvuutta laskettiin suhteessa sivun sanojen kokonaismäärään sekä sanan yleiseen toistuvuuteen.⁷⁴

Yllättäen kustantamon mainokset Friedanin kirjasta eivät olleet osumien joukossa, vaikka tekemäni haku sisälsi suomennoksen nimen. Tässä tapauksessa käyttöliittymälle aiheutti ongelmia kirjan otsikon ensimmäisen sanan tavutus (naisellisuuden) sekä sen rivitys kolmelle riville. Tieto-

66. David Deacon, Yesterday's Papers and Today's Technology. Digital Newspaper Archives and 'Push Button' Content Analysis. *European Journal of Communication* 22:1 (2007), 8, <https://doi.org/10.1177/0267323107073743>; Bastiansen 2019, 136–137.

67. Esim. Hakkarainen et al. 2022, 111–112.

68. Bunout 2023, 281–283; Wijffes 2017, 16–17.

69. Ks. Design, <https://impresso-project.ch/project/design/>; Computational Linguistics, <https://impresso-project.ch/project/linguistics/>; FAQ: What are 'word embeddings', <https://impresso-project.ch/app/faq> (28.2.2023).

70. Esim. Bingham 2010, 229–230; Hakkarainen et al. 2022, 113.

71. <https://sateenkaarisanakirja.herokuapp.com/> (17.11.2022).

72. Maarten van den Bos & Hermione Giffard, The Grapevine. Measuring the Influence of Dutch Newspapers on Delpher. *TS. Tijdschrift voor tijdschriftstudies* (2015), 29.

73. Olin hakua tehdessäni kiinnostunut siitä, minkälaisen painoarvon Friedanin teoksen suomennos ja hänen samalle vuodelle ajoittunut Suomen-vierailunsa on saanut kotimaisessa uutisoinnissa lähivuosisikymmeninä. Tein haun 1.8.2020 ja toteutin sen vuosille 1967–2018. Kohdistin haun sanoma- ja aikakauslehtiaineistoihin ja edellytin osumilta kaikkia hakusanoja. Haku tuotti kaikkiansa 22 osumaa, joista osa oli valtavirtalehdistön sijasta aikakausjulkaisuista, kuten Naistutkimusmonisteista.

74. Usein kysyttyä: Haku ja metadata, <https://digi.kansalliskirjasto.fi/faq> (22.2.2023).



■ Kuva 2: Sanomalehtien digitointi on tehty useimmiten mikrofilmiin pohjalta, mikä on paikoin vaikuttanut tekstintunnistuksen laatuun. Kuvassa mikrofilminlukulaite vuodelta 1971. Kuvaaja: Yves Tessier. Lähde: Lavalin yliopiston arkistohallintaosasto, 334-2-001. Wikimedia Commons.

koneella saattaa olla hankaluuksia tunnistaa hakusanoja myös optisen tekstintunnistuksen puutteista ja palstojen segmentoinnissa esiintyneistä ongelmista johtuen⁷⁵. Alkuperäistä paperilehteä ja sen digitoitua versiota onkin syytä tarkastella editioina, joilla on omat ominaisuutensa. Digitoitu lehti esimerkiksi sisältää aina sellaista tietoa, joka puuttuu sen paperisesta versiosta. Tämä digitoinnissa syntyvä kohina (*noise*) saattaa tuoda tekstintunnistuksen keskelle sellaista tekstiä, jota alkuperäisessä lehdessä ei ole ollut, kuten Johan Jarlbrink ja Pelle Snickars ovat kuvanneet.⁷⁶

Friedan-esimerkissä alkuperäisten osumien lista täydentyi avoimemman haun (hakukomento ”Betty Friedan”) ansiosta muutamilla vuoteen 1967 liittyneillä maininnoilla,⁷⁷ mikä vahvisti tulkintaa yhdysvaltalaisfeministin ja hänen kirjansa merkityksestä 1960-luvun suomalaisessa keskustelussa. Kustantamon mainonnassa kirja saikin suhteellisen suuren painoarvon, kun taas kirja-arvioiden kirjoittajat nostivat paikoin esiin teoksen keskeisten argumenttien soveltumattomuuden suomalaiseen kontekstiin. Analysoidessani Betty Friedania käsitteleviä osumia niiden tulkitseminen kokonaisuutena ja suhteessa toisiinsa nousi näin ollen keskeiseksi. Käytännössä kontekstualisoin uudelleen sen, mitä tietokoneen ruudulta oli luettavissa yksittäisinä artikkeleina ja mainoslauseina⁷⁸. Tämä edellytti ymmärrystä niin lehtitekstien historiallisista konteksteista kuin käyttöliittymän toteutuksessa tehdyistä valinnoista sekä kyvyistäni käyttöliittymän käyttäjänä. Tekemiäni tulkintoja ei voikaan irrottaa käyttämistäni hakukomennoista tai siitä institutionaalista ja teknologisesti kontekstista (Kansalliskirjasto ja sen tarjoama käyttöliittymä), jossa suoritin hakuja ja selasin saamiani osumia.⁷⁹

Käytännössä erilaiset kontekstit asettuvat digitaalisissa arkistoissa usein vuoropuheluun keskenään. Lehtitekstien kontekstoinnissa on nimittäin mahdollista hyödyntää metatietoja, jotka löytyvät digitoitujen lehtiarkistojen käyttöliittymistä varsin helposti. Esimerkiksi Kansalliskirjaston digipalvelussa metatietoihin on kirjattu muun muassa lehden julkaisupaikka ja siinä mahdollisesti tapahtuneet muutokset, lehden nimi eri variaatioineen sekä julkaisijan mahdollinen poliittinen sidonnaisuus. Metatietojen merkitys ei ole kuitenkaan yksinomaan informaatioarvossa, vaan ne auttavat tutkijaa tarkastelemaan digitoitujen aineistojen materiaalisuutta – tai oikeammin digitointiprosessissa muuttunutta materiaalisuutta. Henrik Grue Bastiansen muistuttaa paperilehdistä digitoitujen sanoma- ja aikakauslehtien koostuvan ykkösistä ja nolista. Ne tuottavat tietokoneen ruudulle kuvan alkupe-

75. Yleistajuisesti OCR-tunnistuksesta ja siihen liittyvistä haasteista esim. Hakkarainen et al. 2022, 111; Lilja & Pääkkönen 2022, 6; Wijfjes 2017, 18–19.

76. Jarlbrink & Snickars 2017, 1228–1243, 1229. Ks. myös Oberbichler et al. 2022, 231–232; Hakkarainen et al. 2022, 107, 111.

77. Vuodelle 1967 osumista sijoittui seitsemän, joista kaksi oli kirjan mainoksia. Kaikkien osumien määrä vuosina 1967–2018 oli 100. Haun rajaukset noudattivat edellisen haun vastaavia.

78. Thomas Smits, Looking for The Illustrated London News in Australian Digital Newspapers. *Media History* 23:1 (2017), 84.

79. Bunout 2023, passim.

räislähteestä, jota emme pysty koskettamaan.⁸⁰ Näin ollen esimerkiksi tutkitun sanoma- tai aikakauslehden fyysinen koko ja sivumäärä häipyvät tietokoneen ruudulta helposti näkyvistä. Sen sijaan alkuperäistä paperilehteä selaava tutkija näkee lähteensä materiaalsen muodon konkreettisesti edessään ja pystyy suhteuttamaan tulkinsansa osaksi tätä ymmärrystä.⁸¹ Digitoituja aineistoja tarkasteleva tutkija voi rakentaa vastaavaa ymmärrystä numeerisista tiedoista koostuvien metatietojen avulla, mutta lehden kuvittelu materiaalisena objektina jää tutkijan mielikuvituksen varaan. Tähänkin on tosin kehitetty ratkaisuja. Esimerkkinä voidaan mainita Helsingin yliopiston COMHIS-tutkimusryhmässä kehitetty *materiality explorer* -työkalu, jonka avulla kotimaisten sanomalehtien koon ja ulkoasun muutoksia on pystytty lähestymään tietokoneavusteisesti⁸².

Lopuksi

Digitaaliset aineistot antavat meille mahdollisuuden esittää kokonaan uudenlaisia tutkimuskysymyksiä sekä löytää ennestään tutkimattomia näkökulmia menneisyyteen. Laajat sanomalehtiaineistot tarjoavat esimerkiksi mahdollisuuksia analysoida julkisessa keskustelussa käytyjä debatteja sekä hahmottaa erilaisten käsitteiden ja ilmiöiden historiallista kehitystä. Aineistojen käyttö edellyttää kuitenkin tietoisuutta digitaalisten työkalujen rajoitteista sekä niiden käytön vaikutuksista tutkimustuloksiin. Koska digitaalinen sivu on aina alkuperäisen lehtisivun editio, tekstien tulkitseminen edellyttää myös ymmärrystä kahdesta erilaisesta mediaympäristöstä. Viime kädessä digitaalisuuden synnyttämässä murroksessa onkin kyse historian tutkijan perusmenetel-

mien hallitsemisesta sekä niiden soveltamisesta uudennlaisissa tutkimusympäristöissä.

Tässä katsausartikkelissa olen pyrkinyt avaamaan digitaalisten lehtiarkistojen erityispiirteitä omien käytännön esimerkkieni avulla, koska keskustelu digitalisaatioon liittyvistä peruskysymyksistä on nähdäkseni edelleen tärkeää. Vaikka tällä hetkellä opiskeleva sukupolvi on kasvanut digitaalisten ympäristöjen kyllästävässä maailmassa, heidänkään diginatiiviuutensa ei tarkoita automaattista kykyä tarkastella digitaalisessa muodossa olevia aineistoja kriittisesti. Tästä johtuen meidän on tutkijoina ja opettajina oltava tietoisia digitoitujen aineistojen käyttöön liittyvistä menetelmällisistä kysymyksistä. Samanaikaisesti meidän on tärkeä tarkastella digitaalisia lehtiaineistoja lähdeyyppinä, jonka pohjalta tehtävät tulkinnat edellyttävät ymmärrystä yksittäisten tekstien laajemmasta mediahistoriallisesta kontekstista.

Haluan kiittää lämpimästi vertaisarvioitsijoita heidän esittämistään kriittisistä kommentteista, jotka auttoivat kirkastamaan katsauksen argumenttia. Lisäksi kiitän Turun yliopiston kulttuurihistorian professori Hannu Salmea sekä tutkijatohtori Heidi Hakkarasta, dosentti Petri Pajua ja väitöskirjatutkija Satu Sorvalia, joilta olen viime vuosina oppinut paljon digitaalisesta historiasta.

Dosentti Heidi Kurvinen työskentelee kollegiumtutkijana Turun yliopiston ihmistieteiden tutkijakollegiumissa (TIAS). **Sähköposti:** heidi.kurvinen@utu.fi

80. Bastiansen 2020, 39–40. Ks. myös Nicholson 2013, 64; Wijffes 2017, 18.

81. Hieke Huistra & Bram Mellink, Phrasing History. Selecting Sources in Digital Repositories. *Historical Methods* 49:4 (2016), 221–222; Wijffes 2017, 20; Tinkler 2016, 26.

82. Eetu Mäkelä et al., Interdisciplinary Collaboration in Studying Newspaper Materiality. Teoksessa S. Krauwer & D. Fiaer (toim.) *Twin Talks Workshop at DHN 2019. Proceedings of the Twin Talks Workshop at DHN 2019*. CEUR-WS.org 2019, 55–66. Ks. myös <https://newspaper-materiality-comhis.rahtiapp.fi/> (17.11.2022).