



Big data ja Porthan

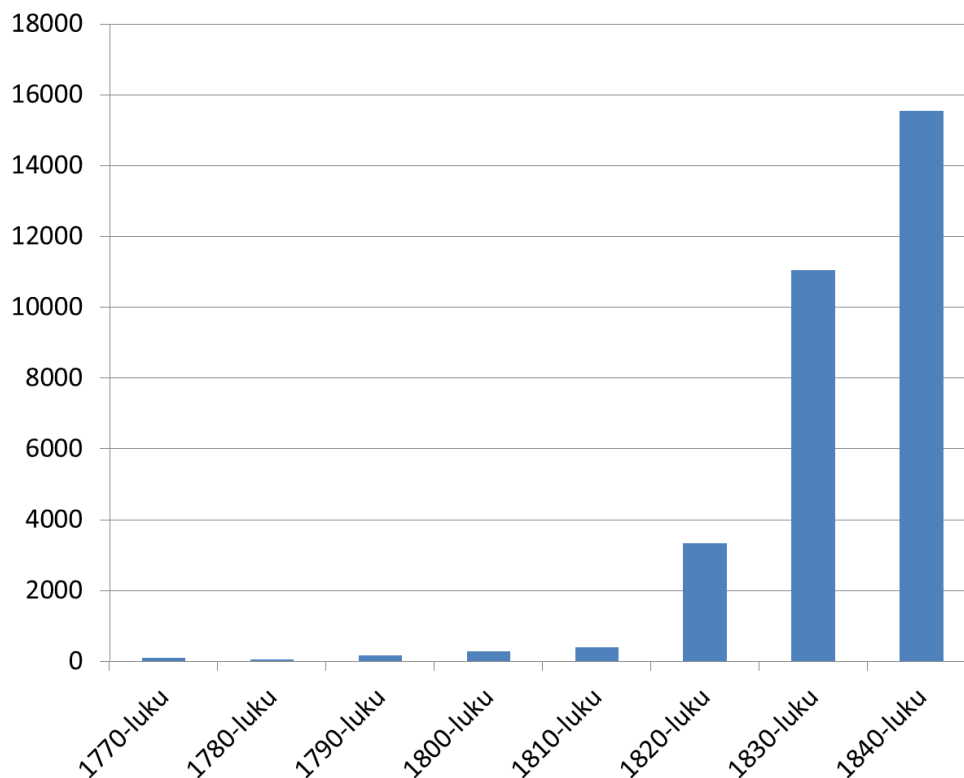
Hannu Salmi

Aloitimme vuoden 2016 alussa tutkimushankkeen *Computational History and Transformation of Public Discourse in Finland, 1640–1910*, joka perustuu Helsingin yliopiston, Kansalliskirjaston ja Turun yliopiston tutkijoiden yhteistyöhön. Tavoitteena on hahmottaa suomalaisen julkaisu- ja aikakauslehdistön historiaa kokonaisuutena, joka ulottuu kirjatutannon alkuvaiheista sanoma- ja aikakauslehdistön kukoistukseen 1800- ja 1900-lukujen vaihteessa. Turun yliopiston osuus perustuu informaatioteknologian tutkijoiden ja kulttuurihistorioitsijoiden yhteistyöhön: ryhmä analysoi vuosien 1771–1910 sanoma- ja aikakauslehdistöä, joka on kattavasti digitoitu. Kokoelma sisältää kaksi miljoonaa sivua sanomalehtiaineistoa, josta puolet ruotsin, puolet suomen kielellä. Juuri kaksikielisyys tekee Suomen tilanteesta ainutlaatuisen: autonomian ajan aineistoa voi käyttää laboratoriona sellaisten menetelmien tutkimiseen, joita voidaan hyödyntää myöhemmin Euroopan kielellisesti fragmentoituneen kulttuurin tutkimuksessa.

Tällä hetkellä olemme tutkineet tekstin uudelleenkäyttöä. Ajatuksena on etsiä laskennallisin menetelmin isosta tekstimassasta toistettuja tekstejä tai tekstien katkelmia. Kansainvälinen tekijänoikeussopimus solmittiin vasta 1880-luvulla, ja sitä ennen lehdet saattoivat vapaasti kopioida tekstejä toisiltaan. Niin sanomalehdistöön muodostui toistojen ketjuja tai rihmastoja, jotka levisivät yhtä kontrolloimattomasti ja epälineaarisesti kuin meemit nykypäivän digitaalisissa verkoissa. Kansainvälisesti sanomalehdistön ekspansio oli erityisen voimakasta 1820-luvulta eteenpäin, ja saman voi havaita Suomessakin. Turun palon jälkeen lehtiä painettiin yhä enemmän ja myös useilla paikkakunnilla. Vuosisadan puolivälissä sanomalehtiä tuotettiin Turun ja Helsingin ohella myös Oulussa, Porvoossa, Vaasassa ja Viipurissa. Olemme tällä hetkellä jo prosessoineet koko sanomalehtiaineiston¹, ja tuloksena on miljoonia tekstitietoja. Tämä on jo itsessään tulos, joka saa pohtimaan näkemystämme historiasta. Olemmeko liiankin kiinnostuneita alkuperäisistä ja originaalista – ja jätämme huomiotta toiston ja kopioinnin, joka on menneisyyden ’pimeää ainetta’? Kopiointi tuntuu mekaaniselta, mutta sen kautta voi ymmärtää, miten ilmiöt voimistuivat ja saivat painoarvoa ja miten tieto menneisyydessä liikkui.

Käymme parhaillaan läpi ajanjaksoa vuodesta 1771, suomalaisen sanomalehdistön synnystä, 1840-luvun loppuun. Tältä aikaväliltä olemme löytäneet saman tekstin toistoketjuja, klustereita, yli 30 000 kappaletta, joista suurimmissa on satoja toistoja (kuva 1). Valtaosa on pieniä klustereita, joissa on vain 2–4 esiintymää. Tähän on vielä lisättävä, että olemme poistaneet saman lehden sisällä tapahtuneen kopioinnin. Tärkeää on myös, että toistot eivät välttämättä ole tapahtuneet lyhyellä aikavälillä vaan saattavat ulottua koko tutkitun periodin läpi, vuoteen 1910 asti.

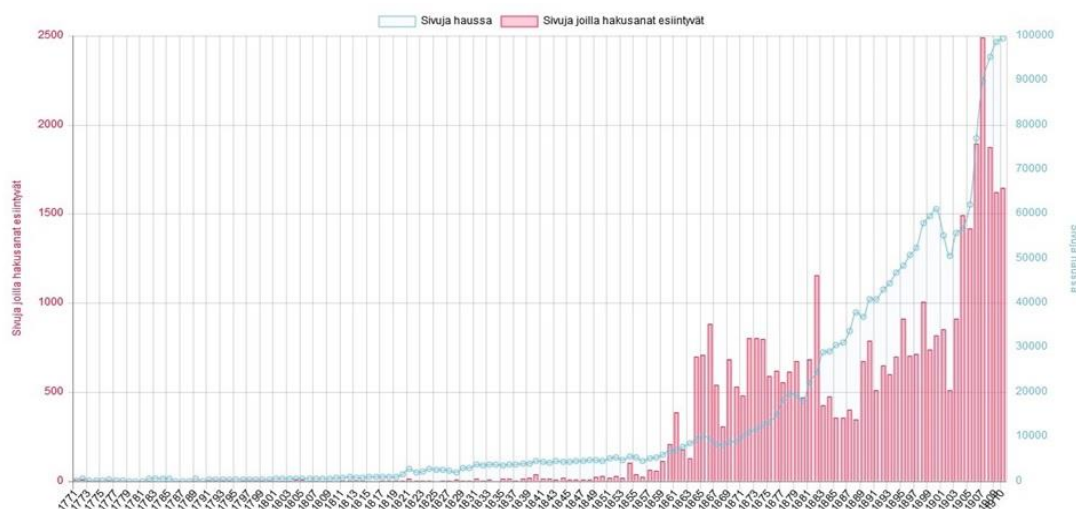
¹ Kiitän tutkimusryhmäämme, johon kuuluvat Filip Ginter, Asko Nivala, Heli Rantala, Tapio Salakoski ja Aleks Vesanto.



Kuva 1. Toistettujen tekstien klusterit suomalaisessa sanomalehtiaineistossa vuosikymmenittäin vuoteen 1850 asti. Klusteri on ajoitettu ensimmäisen esiintymän mukaan. Lähde: *Computational History and Transformation of Public Discourse in Finland, 1640–1910* -hanke.

Kun aineisto on käsillä, väistämättä tulee pohtineeksi, millaisiin tutkimuskysymyksiin sitä voisi hyödyntää. On selvää, että 1800-luvulla suomalainen sanomalehdistö oli merkittävä yhteiskunnan ja kulttuurin muovaaja. Se ei ollut vain passiivinen heijastuspinta vaan aktiivisesti tuotti sitä Suomea, joka lopulta syntyi. Aineistoa seuloessa olen miettinyt, millainen merkitys sanomalehdistöllä oli historiakulttuurin kannalta ja miten esimerkiksi suomalaiset suurmiehet vakiinnuttivat asemaansa painetun sanan kautta. Olen aiemmin käsitellyt *Auraicassa* Henrik Gabriel Porthanin muistoa 1800-luvun Suomessa.² On selvää, että Porthanin suurmieheyttä rakennettiin erityisesti 1850-luvulta eteenpäin. Jos Kansalliskirjaston historiallisen sanomalehtiarkiston aineistosta etsii Porthan-mainintoja, näyttää, että absoluuttinen määrä kasvoi voimakkaasti vuosisadan loppua kohti (kuva 2).

² Hannu Salmi, Porthanin arkku ja historian perintö. *Auraica* Vol. 6, 2015: 21–32.



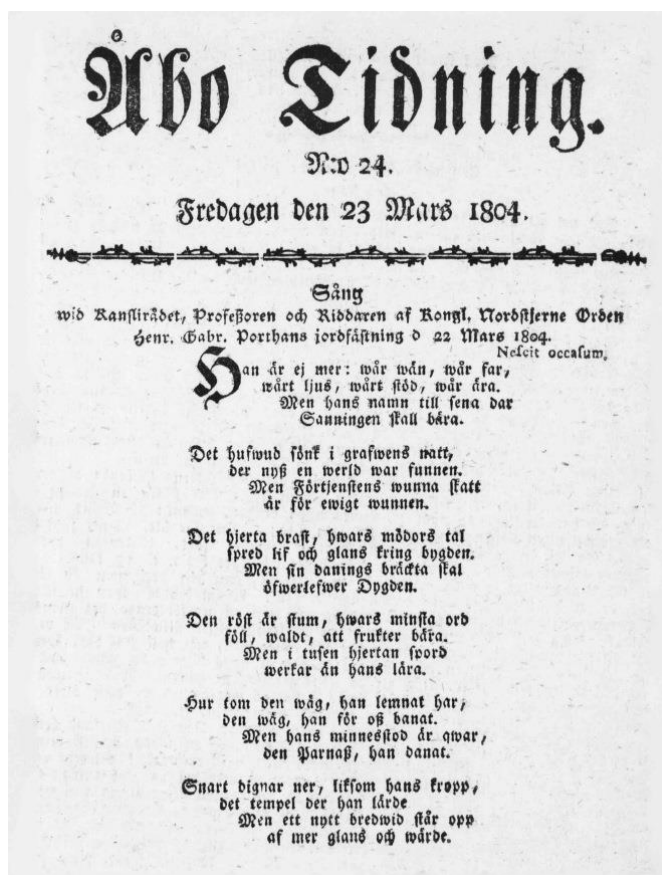
Kuva 2. Porthan-viitteet historiallisessa sanomalehtiarkistossa. Lähde: Kansalliskirjasto.

Porthan-viitteet sisältävät tässä kaiken mahdollisen, historiallisen henkilön käsittelystä höyrylaiva Porthanin aikatauluihin, mutta viittaaminen Porthaniin on selvästi ollut nousujohteista. Kaiken kaikkiaan viitteitä on yli 40 000, ja suhteellinen osuus näyttää erityisen korkealta tilanteessa, jossa suurmiehelle puuhattiin patsasta Turkuun 1860-luvulla, mutta myös fennomanian kiihtyessä ja vielä vuosisadan vaihteen sortokausien keskellä.

Kuva 2 osoittaa, ettei Porthaniin liittyvä tekstin uusiokäyttö voinut olla kovin vilkasta ennen 1850-lukua, sillä vasta tuossa vaiheessa suomalaisen sanomalehdistön volyymi lähti räjähdysmäiseen kasvuun. Koska projektimme on vielä kesken, on vaikea arvioida, miten suuren kierrätyksen kohteena juuri Porthan oli. Tosin etukäteistuntuma on, että 1800-luvun lopulta löytyy varmasti kymmenien toistojen Porthan-klustereita.

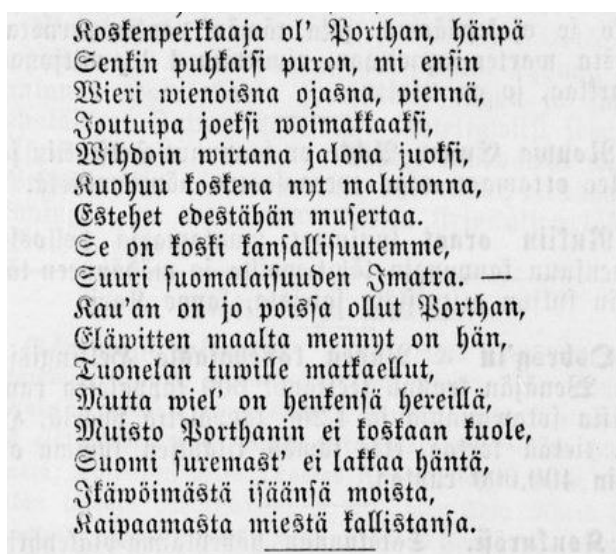
Nyt käsiteltävänä oleva aineisto 1770-luvulta 1840-luvulle sisältää 18 Porthan-aiheista klusteria. Aiheena ovat muun muassa Suomen historia ja Turun Akatemian kirjasto, ja mukana on myös monia muistikirjoituksia. Klustereissa on vain 2–4 toistoa, joten ne ovat vielä hyvin pieniä siihen nähden, mitä sanomalehdistössä oli sittemmin tulossa. Vaikka lehdistön määrä kasvoikin nopeasti 1830- ja 1840-luvuilla – kuten myös klustereiden määrä kuvan 1 osoittamalla tavalla – lehdistö ei vielä ollut julkisuuden määrittelijä. Joka tapauksessa 1800-luvun alun aineisto sisältää kiinnostavia esimerkkejä Porthanin muistosta. Kun Porthan oli menehtynyt 16. maaliskuuta 1804, *Åbo Tidningin* julkaisi muistorunon 23. maaliskuuta (kuva 3).

Tästä runosta tuli teksti, joka liikkui ajassa vuosikymmenien mittaan. Se nostettiin uudelleen esiin, kun Porthanin kuolemasta oli kulunut 50 vuotta ja julkaistiin *Suometaressa* 9. syyskuuta 1864. Runo löydettiin myös sadan vuoden kuluttua, sillä se esiteltiin lukijoille vielä *Wiipurissa* 18. maaliskuuta 1904.



Kuva 2. Porthanin muistoruno *Åbo Tidningarissa* 23.3.1804.
Lähde: Kansalliskirjasto.

Kun tutkimushanke *Computational History and Transformation of Public Discourse in Finland, 1640–1910* etenee, saamme varmasti lisää tietoa ja näkemystä siitä, miten Porthanin varsinainen kulttiasema 1800-luvun lopussa vakiintui ja miten hänestä tehtiin fennomanian edelläkävijä. Jo nyt historiallisesta sanomalehtiarkistosta voi löytää voimakkaan tunteellisia kuvauksia. *Päijänne*-lehti julkaisi 12. marraskuuta 1878 runon, jossa Porthanin sivistystehtävää kuvattiin vertaamalla häntä koskenperkaajaan (kuva 4).



Kuva 4. Porthan koskenperkaajana. *Päijänne* 21.11.1878.
Lähde: Kansalliskirjasto.

Runon mukaan koskenperkaaja oli avannut esteet kansallisuuden ”koskelta”, joka saattoi nyt kuohua ”maltitonna”. Yhtä kuohuvaa ja esteetöntä oli lopulta suomalainen sanomalehdistö, josta 1800-luvun kuluessa kasvoi moniaineksinen, moneen suuntaan haarautuva rihmasto, jonka silmukoiden ja risteyskohtien, umpikujien ja jatkumoiden tutkimus tempaa mukaansa.

Hannu Salmi

Kulttuurihistorian professori, Turun yliopisto
*Computational History and Transformation of Public
Discourse in Finland, 1640–1910* -konsortion johtaja
hansalmi (apud) utu.fi